Journal of Web Engineering, Vol.17, No. 1&2 (2018) 073-092

SEMANTIC EMOTION-TOPIC MODEL IN SOCIAL MEDIA ENVIRONMENT

RUIRONG XUE

School of Computer Engineering and Science, Shanghai University, China
xueruirong@i.shu.edu.cn

SUBIN HUANG

School of Computer Engineering and Science, Shanghai University, China
Anhui Polytechnic University, China
huangsubin@shu.edu.cn

XIANGFENG LUO*

Shanghai Institute for Advanced Communication and Data Science, School of Computer Engineering and Science, Shanghai University, China
**Corresponding author: luoxf@shu.edu.cn*

DANDAN JIANG

School of Computer Engineering and Science, Shanghai University, China
emily921220@163.com

YIKE GUO

School of Computer Engineering and Science, Shanghai University, China
Department of Computing, Imperial College London, British
y.guo@imperial.ac.uk

YAN PENG

School of Mechatronic Engineering and Automation, Shanghai University, China
pengyan@shu.edu.cn

Received July 28, 2017
Revised October 15, 2017

With the booming of social media users, more and more short texts with emotion labels appear in social media environment, which contain users' rich emotions and opinions about social events or enterprise products. Social emotion mining on social media corpus can help government or enterprise make their decisions. Emotion mining models involve statistical-based and graph-based approaches. Among them, the former approaches are more popular, e.g. Latent Dirichlet Allocation (LDA)-based Emotion Topic Model. However, they are suffering from bad retrieval performance, such as the bad accuracy and the poor interpretability, due to them only considering the bag-of-words or the emotion labels in social media media environment. In this paper, we propose a LDA-based Semantic Emotion-Topic Model (SETM) combining emotion labels and inter-word relations to enhance the retrieval performance in social media media environment. The performance influence of four factors on SETM are considered, i.e., association relations, computing time, topic number and semantic interpretability. Experimental results show that the accuracy of our proposed model is 0.750, compared with 0.606, 0.663 and 0.680 of Emotion Topic Model (ETM), Multi-label Supervised Topic Model (MSTM) and Sentiment Latent Topic Model (SLTM) respectively. Besides, the computing time of our model is reduced by 87.81% through limiting word frequency, and its accuracy is 0.703, compared with 0.501, 0.648 and 0.642 of the above baseline methods. Thus, the proposed model has broad prospects in social media media environment.

Key words: Social emotion mining; Semantic discovery; Social emotion classification; Topic Model; Semantic Emotion Topic Model

Communicated by: M. Gaedke & Q. Li

1 Introduction

In recent years, with the rapid change of social media environment (e.g. Sina Microblog and Twitter), more and more users tend to share their opinions, experiences or emotions in the above social media environment. Users are increasingly interested in using more emotion labels as well as short texts to express their emotions and opinions. Thus, the mixtures of emotion labels and short texts carry users' rich emotions and opinions. Figure 1 shows an example of short texts with emotion labels. Many news websites, e.g., Sina Society Channel[a] have provided a news service for users to express their emotions and opinions after browsing news [1, 9]. In such websites, each article is shown with ratings by users who have read the article and voted over a set of predefined emotion labels/emoticons, as Figure 2 shows.

Figure 1 Short texts with emotion labels

Social emotion mining has been widely used, including opinion summarization [3] and sentiment retrieval [4], and has attracted lots of attention from researchers of natural language processing and machine learning [5, 6]. By mining social emotion, government can find out the emotions and opinions

---

[a] http://news.sina.com.cn/society/

of people towards specified social events. Enterprises can assess customers' satisfaction to help promote their products by analyzing emotions of comments.

Figure 2 An example of emotion labels and user ratings

Existing mainstream approaches to social emotion mining are based on statistical model. However, most of them are suffering from low accuracy and poor interpretability, since they only consider words and emotion labels in short texts. Besides, individual words' emotion is ambiguity [7], which may lead to a quite biased prediction of social affective texts. Thus, many researchers start to change Latent Dirichlet Allocation [8] involving emotion labels or emoticons [9-11] to correct words emotions. It has enhanced the accuracy of social emotion computing to some extent. But, all these LDA–based models are bag-of-words models, which carry less semantics in social media corpus. As shown in Table 1, the emotion that is represented by the distribution of words has bad semantic interpretability. It is hard to reveal the knowledge association and help researchers find out semantic context. Figure 3 shows the association relations we may extract from social media corpus. For example, negative words like "corrupt" or "arrested" without context may lead to misunderstanding of documents, while semantic context like "corrupt—arrested" express real positive emotion of readers.

Table 1 The word distribution of emotion "surprise" is shown as bag-of-words with weak semantics.

word	probability
rich	0.0246
children	0.0212
Samsung	0.0211
single	0.0113
witness	0.0018
hurt	0.0015
female anchors	0.0013
corruption	0.0011

In this paper, we propose a LDA-based Semantic Emotion-Topic Model (SETM) involving relations and emotion labels in social media environment. The model follows a several-step generation process for affective terms, which first generates a word's latent topic from a document-specific topical distribution and a word's emotion, label's emotion together with a relation's emotion from a document-specific emotional distribution, then generates a word from Multinomial distribution based on latent topics and emotions, an emotion label and a relation from Multinomial distribution based on their respective emotions.

According to psychology [12], we find it best to define six-dimension emotions(love、fear、joy、anger、sad、surprise) to describe human emotions. We evaluate the proposed model on an online collection collected from the Sina Society Channel. Since the website established a one-to-one relationship between each of the emotion labels and users' emotions shown in Figure 2, the emotion labels matrix and emotions matrix of the same document are fixed to be the same. Experimental results show that the performance of proposed model is affected by relations obviously. Mining results of the proposed model can also be interpreted more semantically by combining words and sentence environments.

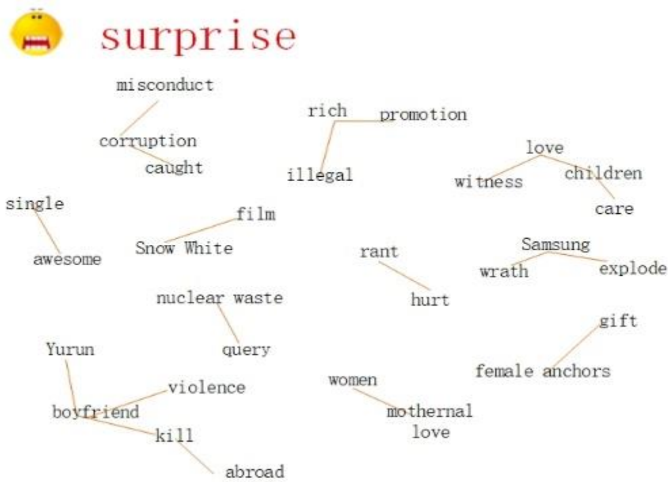

Figure 3 An example of "surprise" relations network extracted from social events, which has strong semantic context

The remainder of this paper is organized as follows. We describe related work in Section 2 and present our model for social emotion mining in Section 3. Data set, results, and discussion are illustrated in Section 4. Finally, conclusion is given in Section 5.

2 Related Work

In this section, we firstly review the related work on sentiment classification and analysis, and then introduce the related topic models used in the area of affective text mining.

Previous work on sentiment classification and analysis can be classified into three levels: document-level [13-17], sentence-level [18, 19] and word-level [20, 21].

Document-level sentiment computing can fall into two parts: supervised learning [13, 14] and unsupervised learning [15-17]. Document-level emotion computing is also a text classification problem, so all existed supervised learning methods can be used to solve it. Supervised learning features, like terms and their frequency, part of speech information, opinion words, negations, and syntactic dependence [13, 14], all have been applied in computing document sentiment. Since the supervised learning depends more on subjective factors, it costs more manpower and time to choose and evaluate the training corpus. Besides, only the categories defined in training samples can be recognized. So that the classification result may be influenced by some unknown categories. In unsupervised learning, Turney [15] extracted phrasal by relations and proposed an algorithm to calculate one phrase polarity according to the Pointwise Mutual Information (PMI) and Information Retrieval (IR) algorithm and search results of searching engine. Taboada [16] used the lexicon involving emotion words and phrases to compute each document's emotion score. Hu [17] computed emotion with emotion signals involved in social media. Li X [2] presented a Bayesian-based model named WMCM to learn document-level semantic features. Li X [31] leveraged unsupervised teaching models to incorporate semantic domain knowledge into the neural network to bootstrap its inference power and interpretability. But, reliable classification results can be available just after massive analysis, post-processing and labeled dataset.

Sentence-level sentiment computing can fall into two parts: First, subjectivity classification, which distinguishes objective sentences from subjective sentences. Second, sentiment classification on subjective sentences. If the sentence is judged as subjective, then we can identify its emotion orientation. All document-level method and lexicon-based method can be involved in sentence emotion computation. Besides, Yamamoto [18] determined a tweet sentiment based on the emoticon role. Tang [19] built an emotion classification framework on sentence level text. But, some objective sentences still contain opinion tendency. For example, production descended 0.3%, compared with last year. Just judging the sentence subjectivity and then classifying sentiment orientation may leak some objective sentences with opinions.

Word-level sentiment computing is the basis of sentiment computing in both sentence-level and document-level. Word-level computing methods can be applied in compiling sentiment lexicon. Kiritchenko [20] used the method proposed by Mohammad [21] that emotional words labeled by hashtag (#) in tweet implying the whole tweet express the same emotion, to construct a word-emotion association lexicon. But, word emotion computing relies on context so much that it is hard to judge words' emotion in different context.

In addition to the above methods, there are also numerous approaches for modeling, e.g., probabilistic Latent Semantic Indexing (PLSI) [22] and LDA [8]. Later researchers introduce other factors into these topic model. A model called emotion-topic model (ETM) [1] followed the Naive Bayes method by assuming words are independently generated from social emotion labels. It introduces an intermediate emotion layer into LDA and assumes each topic being an important component of an emotion. Rao [11] proposed another two topic models called Multi-label Supervised Topic Model (MSTM) and Sentiment Latent Topic Model(SLTM) to associate latent topics with evoked emotions of readers.

Most previous works only distinguish the polarity orientation (positive/ negative) of documents. Recent years also spring up many researches focus on multidimensional emotions. Yamamoto [18] described ten sentiment dimensions. Plutchik [23, 24] clustering eight basic emotions into four-dimensional sentiment vectors: "Acceptance – Disgust", "Anticipation – Surprise", "Joy – Sadness" and "Anger – Fear". Takaoka [25] proposed a method for extracting six-dimensional sentiment. Kumamoto [26] used another six sentiment dimensions to represent readers' emotions. However, our method use "love", "fear", "joy", "sad", "surprise" and "anger" as six emotions which considering the nature of human emotions[12].

Apart from the traditional features – words, many literatures [27-29] have included emoticons or emotion labels into sentiment computing process. Based on words and emotion labels, we find it more related with the language nature when we consider inter-word relations into the generation process since relations show more semantic relations between words. The details of extracting relations will be presented in section 3.

3 Proposed Model

In this section, we present an emotion topic model with more semantics in social media environment. We name the model as Semantic Emotion-Topic Model (SETM).In the following part, we will present the Semantic Emotion-Topic Model in detail.

3.1 Semantic Emotion-Topic Model

In this subsection, we will briefly introduce the Semantic Emotion-Topic Model. Figure 4 presents the graphical model of the proposed SETM in social media environment. SETM generates each word conditioning on topics and emotions simultaneously. But for an emotion label in a document, it is influenced by just the writers' emotion. As relations are composed by words and the process of generating words has already involved topics, so we don't consider the topics in generating relations to simplify our model.

As a complete generative model, SETM allows us to associate each emotion with word tokens and relation tokens jointly, and to predict the probabilities of emotions conditioned to unlabeled documents that contain word tokens and relation tokens (without emotion labels). Here, we define a social texts collection consists of D documents $\{d_1, d_2, \ldots, d_D\}$ with word tokens, relation tokens and user ratings. Word tokens are selected from a vocabulary containing V distinct terms. Relation tokens are selected from a relation list containing U distinct terms and a set of emotion user ratings are chosen from a predefined list of T emotion labels. The list of emotions is denoted by $e = \{e_1, e_2, \ldots, e_E\}$. In this paper, we define the instances of emotions as "love", "fear", "joy", "sad", "surprise" and "anger". Similarly, a document d consists of a sequence of N word tokens $\{w_{d,1}, w_{d,2}, \ldots, w_{d,N}\}$, a sequence of M emotion ratings over T emotion labels denoted by $\{l_{d,1}, l_{d,2}, \ldots, l_{d,M}\}$ and a sequence of Q relation tokens $\{\mathrm{r}_{d,1}, \mathrm{r}_{d,2}, \ldots, \mathrm{r}_{d,Q}\}$.In the d^{th} document, $w_{d,\mathrm{n}}$ represents the n^{th} word, $l_{d,m} \in e$ represents the t^{th} emotion label and $r_{d,q}$ represents the q^{th} relation. It's worth noting that emotion labels are different from the emotions. It means that one emotion label or emoticon may belong to several emotions in the shape of distribution.

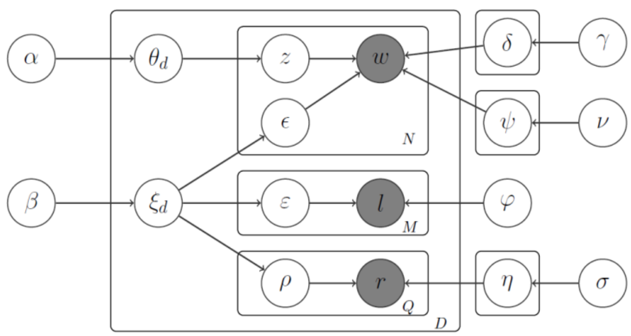

Figure 4 The graphical model of Semantic Emotion-Topic Model (SETM)

In our model, label token $l_{d,m}$ is generated form emotion-emotion distribution φ_e and φ is related with human true emotion matrix, thus we use ξ_d to represent the multinomial distribution of emotions specific to document d. w is generated from topic z and emotion ϵ . Since words can reflect the latent topics in documents, we use θ_d to represent the multinomial distribution of topics specific to document d.

Table 2 Notations of variables in our model

Symbol	Description	Symbol	Description
K	Number of topics	E	Number of emotions
D	Number of documents	U	Number of unique relation tokens
V	Number of unique word tokens	T	Number of unique predefined emotion labels
α	Dirichlet prior of θ	Γ	Dirichlet prior of δ
Β	Dirichlet prior of ξ	ν	Dirichlet prior of ψ
Σ	Dirichlet prior of η	N	Number of word token in each document
Q	Number of relation token in each document	M	Number of emotion labels in each document

e_t	The t^{th} emotion	$w_{d,n}$	The n^{th} word token in document d
$z_{d,n}$	The topic assigned to word token $w_{d,n}$	$\epsilon_{d,n}$	The emotion assigned to word token $w_{d,n}$
$l_{d,m}$	The m^{th} emotion label in document d	$\varepsilon_{d,m}$	The emotion assigned to emotion label $l_{d,m}$
$r_{d,q}$	The q^{th} relation token in document d	$\rho_{d,q}$	The emotion assigned to relation token $r_{d,q}$
θ_d	The multinomial distribution of topics specific to document d		
δ_k	The multinomial distribution of words specific to topic k		
ξ_d	The multinomial distribution of emotions specific to document d		
ψ_e	The multinomial distribution of words specific to emotion e		
φ_e	The multinomial distribution of emotion labels specific to emotion e		
η_e	The multinomial distribution of relation token specific to emotion e		

In our model, label token $l_{d,m}$ is generated form emotion-emotion distribution φ_e and φ is related with human true emotion matrix, thus we use ξ_d to represent the multinomial distribution of emotions specific to document d. w is generated from topic z and emotion ϵ . Since words can reflect the latent topics in documents, we use θ_d to represent the multinomial distribution of topics specific to document d.

Table 2 lists the notations of frequently used variables in this paper. In the graphical model as shown in Figure 4, shaded nodes are observed data, blank nodes are latent parameters, and arrows indicate dependence. The parameterization of the latent data in this model is shown as follows:

$$\theta_d|\alpha \sim Dir(\alpha)$$
$$\xi_d|\beta \sim Dir(\beta)$$
$$\delta_k|\gamma \sim Dir(\gamma)$$
$$\psi_e|\nu \sim Dir(\nu)$$
$$\eta_e|\sigma \sim Dir(\sigma)$$
$$z_{dn}|\theta_d \sim Mult(\theta_d)$$
$$\epsilon_{d,n}|\xi_d \sim Mult(\xi_d)$$
$$\varepsilon_{d,m}|\xi_d \sim Mult(\xi_d)$$
$$\rho_{d,q}|\xi_d \sim Mult(\xi_d)$$
$$w_{d,n}|\delta_{z,n},\psi_{\epsilon,n} \sim Mult((\delta_{z,n}+\psi_{\epsilon,n})|2)$$
$$l_{d,m}|\varphi_{\varepsilon,m} \sim Mult(\varphi_{\varepsilon,m})$$
$$r_{d,q}|\eta_{\rho,q} \sim Mult(\eta_{\rho,q})$$

The generation process of SETM can be described as:

1. Choose $\delta_k \sim Dir(\gamma)$, $\psi_e|\nu \sim Dir(\nu)$, $\eta_e|\sigma \sim Dir(\sigma)$
2. For each document d, the word tokens, emotion labels and relation tokens are generated as follows:
   1) Choose $\theta_d \sim Dir(\alpha)$, $\xi_d \sim Dir(\beta)$
   2) For each of the n^{th} word tokens $w_{d,n}$:
      (1) Choose a topic $z_{d,n} \sim Mult(\theta_d)$.
      (2) Choose a word emotion $\epsilon_{d,n} \sim Mult(\xi_d)$.
      (3) Choose a word token $w_{d,n} \sim Mult((\delta_{z,n}+\psi_{\epsilon,n})|2)$
   3) For each of the m^{th} emotion label $l_{d,m}$:
      (1) Choose a label emotion $\varepsilon_{d,m} \sim Mult(\xi_d)$.

(2) Choose an emotion label $l_{d,\mathrm{m}}$ from $p(l_{d,m}|\varepsilon_{d,m},\varphi)$.

4) For each of the q^{th} relation tokens $r_{d,\mathrm{q}}$:

(1) Choose a relation emotion $\rho_{d,q} \sim Mult(\xi_d)$.

(2) Choose an emotion label $r_{d,\mathrm{q}}$ from $p(r_{d,q}|\rho_{d,q}, \eta)$.

After generating D documents by the process above, the parameter ψ, η are used to predict the documents without emotion labels.

3.2 Extract Relations

Bag-of-words topic models can only get the distribution of words, which usually miss many semantic information [30]. To overcome the limitation, we add the inter-word relations into the topic model.

The association strength between keywords in the document can be defined as follows:

$$S(w_i, w_j)^{T_{d,n}} = \frac{Co(w_i, w_j)}{\sqrt{DF(w_i)DF(w_j)}} \tag{1}$$

where $T_{d,n}$ is the n^{th} transaction of document d, $Co(w_i, w_j)$ is the total co-occurrence number that word w_i and word w_j in the transaction of the document. $DF(w_i)$ is the word w_{i} occurrence number in the T_n。Besides, we also use parameter λ to adjust the importance of the relations selection. Algorithm 1 shows the process of extracting relations in documents set:

Algorithm 1 Extracting Relations

Input: documents set D, parameter λ

Output: relations set $\{r_{d,q}\}, d \le D, q \le Q$

```
Segment each document d_n in D by the slide window to generate transaction set T;
Extract word set W from d_n;
for each pair word <w_i, w_j>:
  calculate the S(w_i, w_j) through Eq. (1);
end for
for each document d in D:
  select the top λ percentage S(w_i, w_j) in each document d as relations r_{d,q} ;
end for
```

3.3 *Parameter Estimation and Prediction*

In SETM, the generative process of each word token, emotion label and relation token is similar to LDA. We use Gibbs sampling [31] to estimate the parameters.

First, we consider θ_d, which represents the topic distribution of the document d. It is a probability of each topics, where $\sum_{k=1}^{K} \theta_{dk} = 1$. θ_d is a Dirichlet distribution and its posterior conditional distribution on other variables can be observed as,

$$p\left(\theta_d \middle| \alpha, \{z_{d,n}\}\right) \propto \prod_{k=1}^{K} \theta_{d,k}^{\sum_{n=1}^{N} I_{z_{d,n}=k} + \alpha_k - 1} \tag{2}$$

$$\therefore \theta_d \sim Dir(\theta_d; \alpha + \sum_{n=1}^{N} I_K(z_{d,n})) \tag{3}$$

where $z_{d,n}$ is the candidate topic to which word token $w_{d,n}$ is assigned. $\sum_{n=1}^{N} I_{z_{d,n}=k}$ is the number of tokens with word $w_{d,n}$ that are assigned to topic k. θ_d is the document-topic distribution of d^{th} document generated from Dirichlet prior α .

Second, we consider the full conditional for the emotion distribution ξ_d of d^{th} document:

$$p(\xi_d|\beta,\{\epsilon_{d,n}\},\{\varepsilon_{d,m}\},\{\rho_{d,r}\}) \propto \prod_{e=1}^{E} \xi_{d,e}^{\sum_{n=1}^{N} I_{\epsilon_{d,n}=e}+\sum_{m=1}^{M} I_{\varepsilon_{d,m}=e}+\sum_{r=1}^{R} I_{\rho_{d,r}=e}+\beta_e-1} \tag{4}$$

$$\therefore \xi_d \sim Dir(\xi_d;\beta+\sum_{n=1}^{N} I_e(\epsilon_{d,n})+\sum_{m=1}^{M} I_e(\varepsilon_{d,m})+\sum_{r=1}^{R} I_e(\rho_{d,r})) \tag{5}$$

where $\sum_{n=1}^{N} I_{\epsilon_{d,n}=e}$ is the number of tokens with word $w_{d,n}$ that are assigned to emotion e, $\sum_{m=1}^{M} I_{\varepsilon_{d,m}=e}$ is the number of tokens with emotion label $l_{d,m}$ that are assigned to emotion e, and $\sum_{r=1}^{R} I_{\rho_{d,r}=e}$ is the number of tokens with relation $l_{d,m}$ that are assigned to emotion e.

Next, we sample the candidate topic $z_{d,n}$ and the candidate emotion $\epsilon_{d,n}$, $\varepsilon_{d,m}$ and $\rho_{d,q}$ in each document d.

$$p(z_{d,n}=k|\theta_d,w_{d,n},\{\delta_k\},\{\psi_e\},\epsilon_{d,n}) \propto \frac{\delta_{k,w_{d,n}}+\psi_{\epsilon_{d,n},w_{d,n}}}{2}\cdot\theta_{d,k} \tag{6}$$

$$\therefore p(z_{d,n}|\theta_d,w_{d,n},\{\delta_k\},\{\psi_e\},\epsilon_{d,n}) \propto \theta_d \circ \frac{\delta^{w_{d,n}}+\delta^{\psi_{d,n}}}{2} \tag{7}$$

where $z_{d,n}$ is the candidate topic and $\epsilon_{d,n}$ is the candidate emotion to which word token $w_{d,n}$ is assigned. k is a topic token. $z_{d,n}=k$ means the k^{th} topic is assigned to $w_{d,n}$.

$$p(\epsilon_{d,n}=e|\xi_d,w_{d,n},\{\delta\},\{\psi\},z_{d,n}) \propto \frac{\delta_{z_{d,n},w_{d,n}}+\psi_{e,w_{d,n}}}{2}\cdot\xi_{d,e} \tag{8}$$

$$\therefore p(\epsilon_{d,n}|\xi_d,w_{d,n},\{\delta\},\{\psi\},z_{d,n}) \propto \xi_d \circ \frac{\delta^{w_{d,n}}+\delta^{\psi_{d,n}}}{2} \tag{9}$$

Similarly, e is an emotion token. $\epsilon_{d,n}=e$ means that the e^{th} emotion is assigned to word woken $w_{d,n}$.

$$p(\varepsilon_{d,m}=e|\xi_d,l_{d,m},\varphi) \propto \varphi_{e,l_{d,m}}\cdot\xi_{d,e} \tag{10}$$

$$\therefore p(\varepsilon_{d,m}|\xi_d,l_{d,m},\varphi) \propto \xi_d \circ \varphi^{l_{d,m}} \tag{11}$$

where $\varepsilon_{d,m}$ is the candidate emotion to which emotion label token $l_{d,m}$ is assigned. $\varepsilon_{d,m}=e$ means that the e^{th} emotion is assigned to $l_{d,m}$.

$$p(\rho_{d,q}=e|\xi_d,r_{d,q},\eta) \propto \eta_{e,r_{d,q}}\cdot\xi_{d,e} \tag{12}$$

$$\therefore p(\rho_{d,q}|\xi_d,r_{d,q},\eta) \propto \xi_d \circ \eta^{r_{d,q}} \tag{13}$$

where $\rho_{d,q}$ is the candidate emotion to which relation token $\rho_{d,q}$ is assigned. $\varepsilon_{d,m}=e$ means that the e^{th} emotion is assigned to $\rho_{d,q}$.

For the unknown variable δ_k, it is topic k's word vocabulary distribution, the posterior distribution should be,

$$p(\delta_k|\gamma,z,w,\{\psi_e\},\epsilon) \propto (\prod_{\{(d,n):z_{d,n}=k\}}(\prod_{v=1}^{V}\frac{\delta_{k,v}^{I_{w_{d,n}=v}}+\psi_{\epsilon_{d,n},v}^{I_{w_{d,n}=v}}}{2}))\cdot(\prod_{v=1}^{V}\delta_{k,v}^{\gamma_v-1}) \tag{14}$$

$$\therefore \delta_k \sim Mult(w;(\delta_z+\psi_e)/2)\cdot Dir(\gamma) \tag{15}$$

For the unknown variable ψ_e, it is emotion e's emotion labels distribution, the posterior distribution should be,

$$p(\psi_e|\nu,\delta,z_n=k,w,\epsilon) \propto (\prod_{\{(d,n):\epsilon_{d,n}=e\}}(\prod_{v=1}^{V}\frac{\delta_{z_{d,n},v}^{I_{w_{d,n}=v}}+\psi_{e,v}^{I_{w_{d,n}=v}}}{2}))\cdot(\prod_{v=1}^{V}\mu_{e,v}^{\nu_v-1}) \tag{16}$$

$$\therefore \psi_e \sim Mult\left(w;\frac{\delta_z+\psi_e}{2}\right)\cdot Dir(\gamma) \tag{17}$$

For the unknown variable η_e, it is emotion e's relations distribution, the posterior distribution should be,

$$p(\eta_e|\sigma,r,\rho) \propto \prod_{u=1}^{U}\eta_{e,u}^{\sum_{i\rho_i=e} I_{r_i=u}+\sigma_u-1} \tag{18}$$

$$\therefore \eta_e \sim Dir(\eta_{e,u}; \sigma + \sum_{i\rho_i=e} I_u(r_i)) \quad (19)$$

where $\sum_{i\rho_i=e} I_u(r_i)$ is the number of the item r which assigned to emotion e.

We listed the detailed information as how to derive the parameters used in the model in Appendix 1.

Then we iteratively sample θ_d, ξ_d, $z_{d,n}$, $\epsilon_{d,n}$, $\varepsilon_{d,m}$, $\rho_{d,q}$, δ_k, ψ_e and η_e. The sampling process is summarized in Algorithm 2.

Algorithm 2 Gibbs sampling for Semantic Emotion Topic Model

Input: The number of topic K, emotion number 6, word matrix W, emotion label matrix l, relation matrix r, label-emotion matrix φ, hyper parameters $\alpha, \beta, \gamma, \nu, \sigma$
Output: multinomial distribution $\theta, \xi, \delta, \psi, \eta, z, \epsilon, \varepsilon, \rho$

```
Randomly initialize θ, ξ, δ, ψ, η, z, ϵ, ε, ρ
for iter = 1 to N_iter do
    for each document d ∈ D,d = 1…D do:
        draw θ_d through Eq. (3) ;
        draw ξ_d through Eq. (5) ;
        for word token n<= word number in each document d, n=1…N do:
            draw z_{d,n} through Eq. (7) ;
            draw ϵ_{d,n} through Eq. (9) ;
        end for
        for emotion label token m<= emotion label number in document d, m=1…M do:
            draw ε_{d,m} through Eq. (11) ;
        end for
        for relation token q<= relation number in document d, q=1…Q do:
            draw ρ_{d,q} through Eq. (13) ;
        end for
    end for
    for each topic k<= total topic number, k = 1…K do:
        draw δ_k through Eq. (15) ;
    end for
    for each emotion e<= total emotion number, e = 1…E do:
        draw ψ_e through Eq. (17) ;
        draw η_e through Eq. (19) ;
    end for
end for
```

4 Examples

In this section, we firstly analyze the influence of relations to find the right amount of relations to improve accuracy. Secondly, we experiment on different word frequency to balance the accuracy and computing time. Thirdly, we analyze the effects of topic numbers to see the performance of SETM on different number of latent aspects. Lastly, we analyze relations to show the semantic information and interpretability of SETM.

4.1 Experiments Design

To prove the efficiency of the proposed model, we have crawled 43768 articles about "female drivers" from the Sina Society Channel. The attributes of each article include user comments and user ratings over six emotion labels: "love", "fear", "joy", "sad", "surprise" and "anger". Before the modeling process, all news articles are preprocessed and cleaned using the following steps:

1) Extract the title and user comments of each news article. Merge all comments under the same news article into a new document.
2) Segment all the words for each document with Ansj[b]. Ansj is an integrated Chinese lexical analysis system based on HMM, n-Gram and CRF. We enjoy its faster segment speed and higher accuracy than ICTCLAS[c] (another Chinese lexical analysis system).

We split the dataset into training set (26942 comments) and testing set (16826 comments). The previous work Emotion-Topic Model (ETM)[1], Multi-label Supervised Topic Model(MSTM) and Sentiment Latent Topic Model(SLTM)[11] are used as comparison. We use an important evaluation metric: the accuracy at top $\mathrm{N}(\mathrm{Accu@N}, \mathrm{N} = 1,2,3)$. For an unlabeled document, we already has its truth emotion labels set $E_{topk@d}$ by users rating and predicted emotion labels e_p by our model. $\mathrm{E}_{\mathrm{topk@d}}$ include the top k emotion labels. Document d is considered predicted correctly if $e_p \in E_{topk@d}$.

4.2 The Influence of Relations on Accuracy

As we know, excessive relations can cause noise and too few relations will result in missing information. To analyze the influence of relations on accuracy and find an appropriate value of relations, we experiment on different threshold of relations.

In SETM, to prove it is really help to improve accuracy by adding relations into our model, we use another line "*SETM with no relations*", in which r of SETM is a zero matrix. Top λ relations are selected from 10% to 100%. Also, $wdtf$ determines the lower bound of word frequency we obtained. For example, $wdtf = 0$ means we obtain all words, and $wdtf = 1$ means we obtain word frequency bigger than 1. We set the number of topics K=10 for our models. The recommended value of α is 0.5, β = 0.5 , γ = 1 , υ = 1 , $\sigma = 0.4$.

Figure 5 and Figure 6 shows the accuracy curves of SETM, MSTM, SLTM and ETM, with $wdtf = 0$ and $wdtf = 1$ respectively. Note that the performance of the baseline MSTM, SLTM and ETM does not change since they do not consider any relations. As the curves of SETM $wdtf = 0$ and SETM $wdtf = 1$ shows the relationship between λ and accuracy is not linear. The peak value of $wdtf = 1$ and $wdtf = 0$ are 0.703 and 0.750.

In the curve of *SETM with* $wdtf = 0$, it is clear that the accuracy are reduced sharply for involving all relations, which resulted from adding too much noise. $\lambda = 0.3$ and $\lambda = 0.8$ show fine accuracy respectively. While in SETM with $wdtf = 1$, the curve tend to be more gentle.

---

[b] https://github.com/NLPchina/ansj_seg
[c] http://ictclas.nlpir.org/

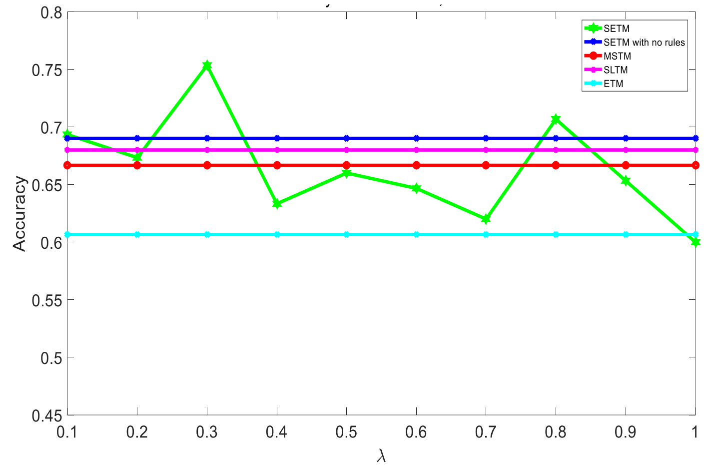

Figure 5 Accuracy on different λ with wdtf=0

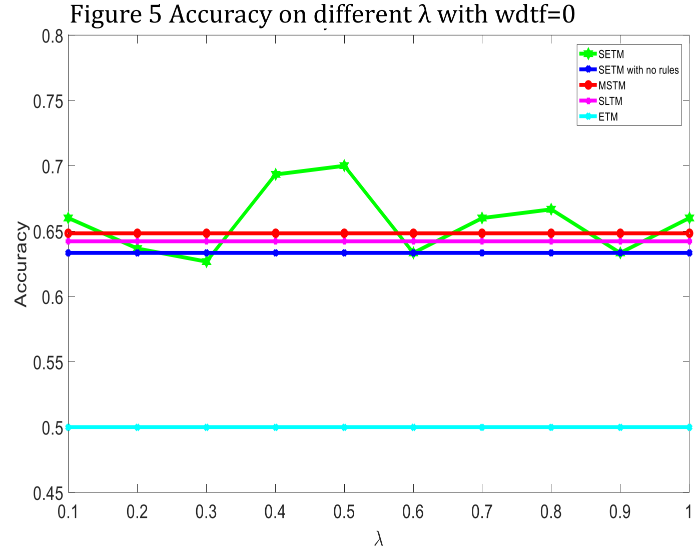

Figure 6 Accuracy on different λ with wdtf=1

Table 3 Statistics of SETM

$wdtf$	Mean(%)	Variance
0	66.40	0.0020
1	65.70	0.0006

Table 3 shows the mean and variance of accuracy for *SETM* $wdtf = 0$ and $wdtf = 1$.Compared with $wdtf = 1$, the averaged accuracy of $wdtf = 0$ is reduced by 1.1%. The variance of the accuracy for $wdtf = 1$ is much smaller than $wdtf = 0$, meaning that the performance of $wdtf = 1$ is more stable than $wdtf = 0$ with respect to different λ.

According to the results, it does not mean that the more relationship, the better average precision. Too few relations may result in missing information, while too many relations may lead to too much noise. The change of parameters cannot lead to linear or nonlinear relations between densities and precisions. Therefore, it needs some experience to select appropriate parameters to get better performance.

4.3 The Influence of Computing Time on Accuracy

Sampling process of LDA-based model is time consuming, especially our model has added more elements than traditional LDA. To balance the time consumption and accuracy of SETM, we experiment on different word frequency.

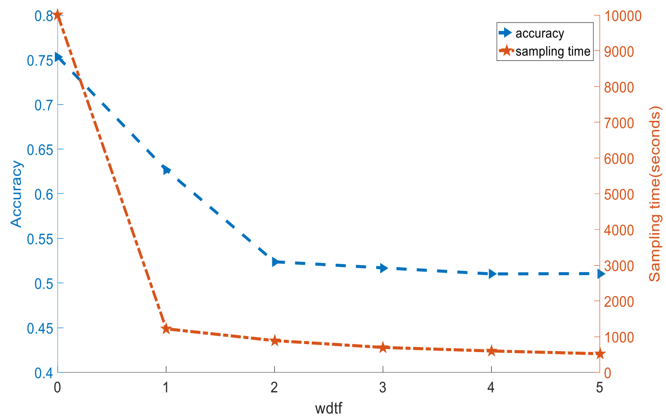

Figure 7 Accuracy and computing time of SETM with different wdtf

Figure 7 shows the accuracy and computing time of Semantic Emotion Topic Model with different $wdtf$ where $\lambda = 0.3$. The blue line represents the accuracy, while the orange line represents computing time. Here we show the time of sampling 1000 times by Matlab, while the real sampling number is much more than that. The green line highlight the accuracy and computing time at $wdtf = 1$. As $wdtf$ grows to 1, both the computing time and the accuracy drop. The computing time fall sharply from 10000 seconds to 1219 seconds, while the accuracy fall from 0.75 to 0.62. This can translated into a relative time reduction of 87.81 percent and a relative accuracy reduction of 17 percent. But the two curves tend to be gentle while $wdtf$ continue grows. Besides, when we set $\lambda = 0.5$, the accuracy reduction of our model is only 6.2 percent according to Figure 5 and Figure 6.

The experimental results show it best to set $wdtf = 1$to best maintain the semantic information of data and sharply reduce computing time. Besides, the mean accuracy of $wdtf = 1$ is more stable than $wdtf = 0$. Additionally, the reduced accuracy can be partly increased by adjusting $\lambda = 0.5$ according to Figure 5 and Figure 6, which is very worthwhile compared with the time consumption.

4.4 The Influence of Topic Number on Accuracy

The number of topics K indicates how many latent aspects of articles can be derived, which may influence the performance of the baseline ETM, MSTM, SLTM and the proposed SETM. Considering the performance of SETM under different threshold of relations may differ, we choose top accuracy λ= 0.5, according to the "$SETM = 1$ " in Figure 7. To evaluate the influence of K, we vary it from 2 to 30. The hyper parameters α=0.5, $\beta = 0.5$, $\gamma = 1$, $\upsilon = 1$, $\sigma = 0.4$. $wdtf = 1$.

The experimental results show that the performance of ETM, SLTM and MSTM. The "SETM λ=0.5" out performs the baseline MSTM when K is larger than 4, while performing worse than MSTM with K = 2 and K=4. For K lager than 7, baseline ETM and SLTM yield competitive performance. Both SETM and the baseline SLTM tend to fall with K larger than 20.

As a whole, the performance of Semantic Emotion Topic Model (SETM) is better than the baseline ETM, MSTM and SLTM. Our model has better performance with more than 4 latent aspects of articles.

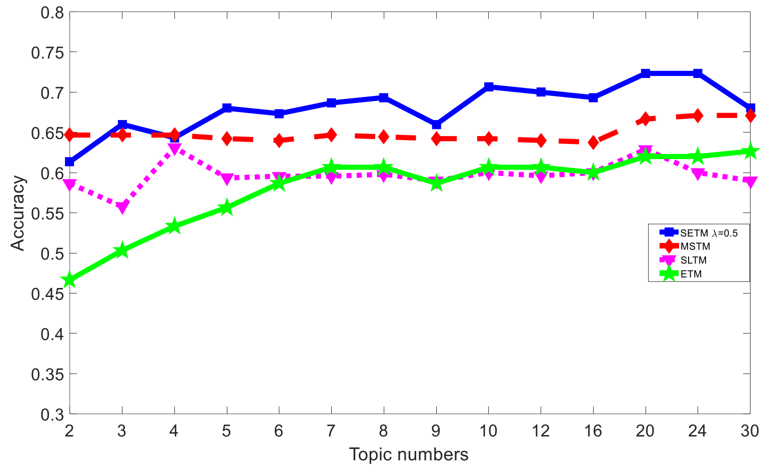

Figure 8 The performance with different topic number

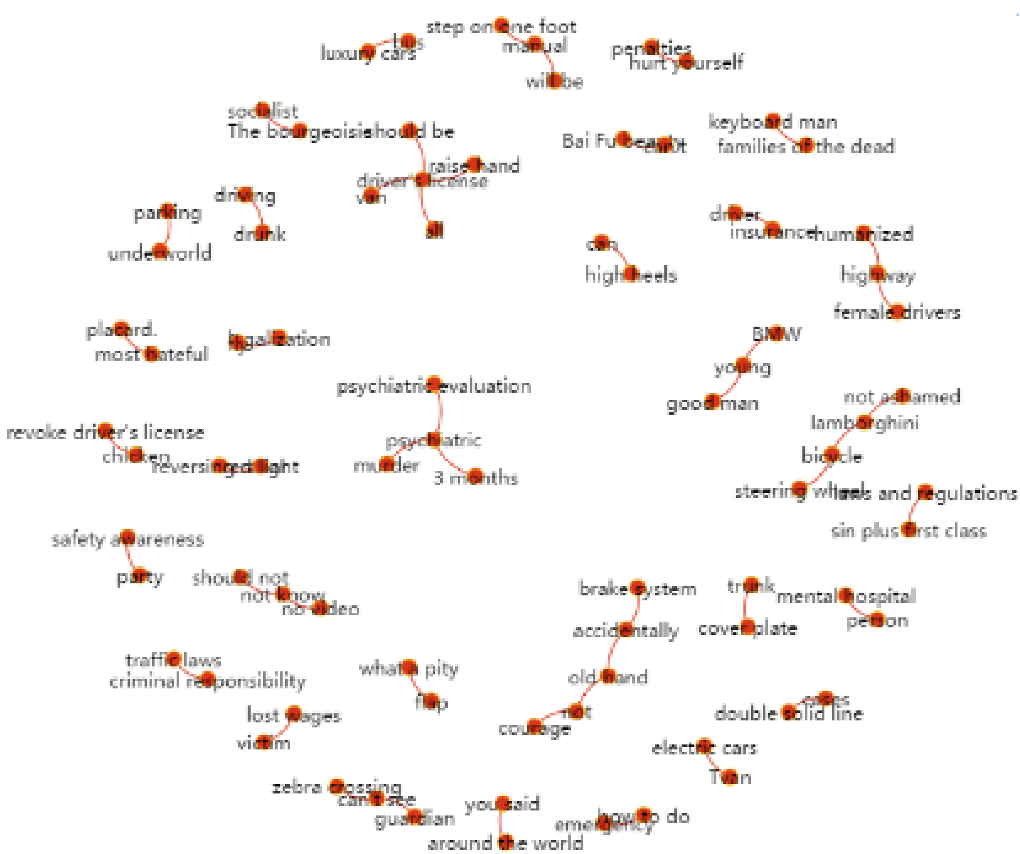

Figure 9 Semantic emotion network of emotion "angry"

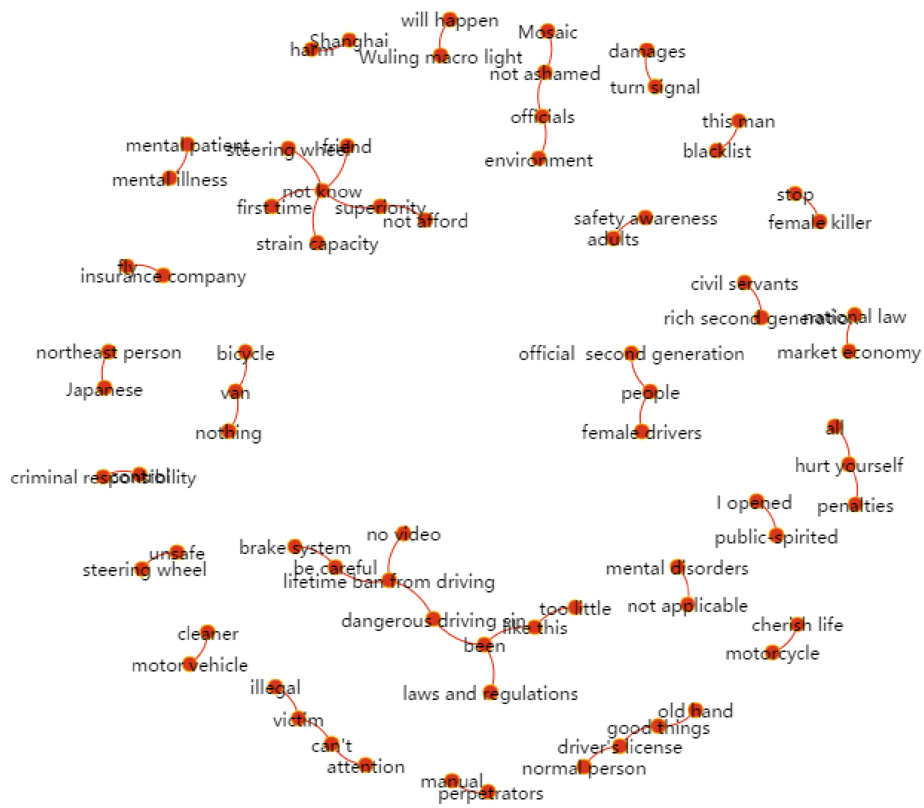

Figure 10 Semantic emotion network of emotion "sad"

Table 4 Top 10 words and relations with their probability of emotion "Angry" and "Sad"

Top 10 words and relations with their probability of emotion "Angry"					
word	probability	relations		probability	
female drivers	0.2156162	not know	no video	0.000999	
not know	0.050256	penalties	hurt yourself	0.000994	
revoke driver's license	0.024627	step on foot	manual	0.000926	
must	0.021925	accidentally	brake system	0.000874	
people	0.020934	Lamborghini	bicycle	0.000854	
life forbidden	0.020536	what a pity	flap	0.000694	
can	0.015925	placard.	most hateful	0.00069	
penalty	0.015431	old hand	not	0.00069	
public safety	0.014477	will be	manual	0.000688	
driver	0.014383	bus	luxury cars	0.000688	
Top 10 words and relations with their probability of emotion " Sad "					
word	probability	relations		probability	
safety belt	0.01608	people	female drivers	0.000736	
situation	0.014829	be careful	brake system	0.000708	
safety	0.013745	not know	strain capacity	0.000703	

toy	0.010991	safety awareness	adults	0.000702	
motor	0.007463	not know	friend	0.0007	
rearview mirror	0.005928	mental illness	mental patient	0.000699	
driver	0.005383	control	criminal responsibility	0.000695	
female driver	0.005374	can't	victim	0.000695	
impossible	0.004595	not afford	superiority	0.000695	
pity	0.004526	motorcycle	cherish life	0.000695	

4.5 The Interpretability of Mining Results in SETM

As we have noted, our mining relations has stronger interpretability than traditional bag-of-words model. We additionally list part of our mining relations to analyse the semantic information and interpret the mining results.

Table 4 lists the top 10 words and relations with their probability of emotion "*Angry*" and "*Sad*". Neutral words, like "female drivers", have carried emotion strongly due to many social news and horrible accidents. For each emotion, we combine top 50 relations with words to generate semantic emotion net.

Figure 9 and Figure 10 show the network of emotion "*Angry*" and "*Sad*" respectively. In emotion "*Angry*", Single word like "*psychiatric*" do not carry any evident emotion, but relations like "*psychiatric-murder*" and "*psychiatric evaluation - psychiatric*" arouse social users strongly angry. Relations like "*Lamborghini - bicycle*", and "*bus - luxury cars*", which connect two contrast words, reflect angry emotion. In emotion "*Sad*", word like "*strain capacity*" do not carry emotion, but "*not know- strain capacity*", which resulting accidents, reflects social users sad feeling about the accidents. Apart from evident emotion words like "*penalty*", "*hateful*", our model mines more relations closely related to current news. These relations can be regarded as emotion patterns containing rich semantic information.

All the above results prove the Semantic Emotion Topic Model contains more semantic contexts compared to traditional LDA-based method. These semantic contexts help interpret the emotion mining results from computer, which prove the mining result of SETM is reasonable.

5 Conclusion

As social media users increasing, more and more mixtures of abundant emotion labels and short texts appear, which contain rich emotions of social media users about social events or enterprise products in social media environment. Social emotion mining has been applied into many projects. By emotion mining, governments understand online users' emotions towards their policies and enterprises know consumers' opinions towards their products. So that, mining emotions from online social texts can help governments make their policies more satisfactory and help enterprises change products to cater to people. In this article, we have proposed a LDA-based Semantic Emotion-Topic Model involving the words, emotion labels and relations. Unlike the mainstream statistical-based approaches, our model additionally considered inter-words relations, which containing more semantic information than traditional bag-of-words model. We compared our model with three baseline methods (Emotion Topic Model, Multi-label Supervised Topic Model and Sentiment Latent Topic Model). Experiment results show that our Semantic Emotion Topic Model can not only improve the accuracy, but also semantically interpret the mining results.

The main conclusions are summarized as follows:

1) For improving the social emotion mining accuracy, we add inter-word relations into LDA-based model, then, we proposed a Semantic Emotion Topic Model. Experiment results show

that the accuracy of our proposed model is 0.750, compared with 0.606, 0.663 and 0.680 of above baselines respectively.

2) In order to balance the accuracy of social emotion mining result and the computing time of our model, we removed unimportant words with low word frequency because traditional LDA uses all words as input, which is very time-consuming. Experiment results show that the computing time of our model has been reduced by 87.81% while its accuracy is 0.703, compared with 0.501, 0.648 and 0.642 of the above baselines.

For the future work, we aim to apply our model to online social events emotion mining, for which we use the comments as our dataset instead of news contents. In social media like Sina Microblog, the social emotion of an event are closed with the emotion patterns extracted from inter-word rules. In the future, the relations closed to social news will be used to create emotion pattern for generalized using. Besides, we can further study the inter-word relations and evolution of relations on different emotions.

Acknowledgements

The research work reported in this paper was supported in part by the National Science Foundation of China under grant no. 91746203, 61471232 and 61403245. This work was jointly supported by the Shanghai Science International Cooperation Project under grant no. 16550720400.

References

1. Bao S, Xu S, Zhang L, et al. Mining Social Emotions from Affective Text[J]. IEEE Transactions on Knowledge & Data Engineering, 2012, 24(9):1658-1670.
2. Li X, Rao Y, Xie H, et al. Social emotion classification based on noise-aware training[J]. Data & Knowledge Engineering, 2017.
3. Ku L W, Liang Y T, Chen H H. Opinion extraction, summarization and tracking in news and blog corpora[C]//Proceedings of AAAI. 2006: 100-107.
4. Eguchi K, Lavrenko V. Sentiment retrieval using generative models[C]//Proceedings of the 2006 conference on empirical methods in natural language processing. Association for Computational Linguistics, 2006: 345-354.
5. Cambria E, White B. Jumping NLP curves: A review of natural language processing research[J]. IEEE Computational intelligence magazine, 2014, 9(2): 48-57.
6. Wang Q, Wu O, Hu W, et al. Ranking social emotions by learning listwise preference[C]//Pattern Recognition (ACPR), 2011 First Asian Conference on. IEEE, 2011: 164-168.
7. Quan C, Ren F. An exploration of features for recognizing word emotion[C]//Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010: 922-930.
8. Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3(Jan): 993-1022.
9. Bao S, Xu S, Zhang L, et al. Joint emotion-topic modeling for social affective text mining[C]//Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on. IEEE, 2009: 699-704.
10. Lin C, He Y. Joint sentiment/topic model for sentiment analysis[C]//Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009: 375-384.
11. Rao Y, Li Q, Mao X, et al. Sentiment topic models for social emotion mining[J]. Information Sciences, 2014, 266: 90-100.
12. Parrott W G. Emotions in social psychology: Essential readings[M]. Psychology Press, 2001.
13. Vinodhini G, Chandrasekaran R M. Sentiment analysis and opinion mining: a survey[J]. International Journal, 2012, 2(6): 282-292.

14. Da Silva N F F, Hruschka E R, Hruschka E R. Tweet sentiment analysis with classifier ensembles[J]. Decision Support Systems, 2014, 66: 170-179.
15. Turney P D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002: 417-424.
16. Taboada M, Brooke J, Tofiloski M, et al. Lexicon-based methods for sentiment analysis[J]. Computational linguistics, 2011, 37(2): 267-307.
17. Hu X, Tang J, Gao H, et al. Unsupervised sentiment analysis with emotional signals[C]//Proceedings of the 22nd international conference on World Wide Web. ACM, 2013: 607-618.
18. Yamamoto Y, Kumamoto T, Nadamoto A. Multidimensional sentiment calculation method for Twitter based on emoticons[J]. International Journal of Pervasive Computing and Communications, 2015, 11(2): 212-232.
19. Tang D, Qin B, Wei F, et al. A joint segmentation and classification framework for sentence level sentiment classification[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2015, 23(11): 1750-1761.
20. Kiritchenko S, Zhu X, Mohammad S M. Sentiment analysis of short informal texts[J]. Journal of Artificial Intelligence Research, 2014, 50: 723-762.
21. Mohammad S M. # Emotional tweets[C]//Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation. Association for Computational Linguistics, 2012: 246-255.
22. Hofmann T. Probabilistic latent semantic indexing[C]//Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1999: 50-57.
23. Plutchik R. The multifactor-analytic theory of emotion[J]. the Journal of Psychology, 1960, 50(1): 153-171.
24. Plutchik R. The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice[J]. American scientist, 2001, 89(4): 344-350.
25. Takaoka K, Nadamoto A. Words-of-wisdom search based on multi-dimensional sentiment vector[C]//Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services. ACM, 2011: 66-73.
26. Kumamoto T. Design of impression scales for assessing impressions of news articles[C]//Database Systems for Advanced Applications. Springer Berlin/Heidelberg, 2010: 285-295.
27. Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining[C]//LREc. 2010, 10(2010).
28. Davidov D, Tsur O, Rappoport A. Enhanced sentiment learning using twitter hashtags and smileys[C]//Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, 2010: 241-249.
29. Dayalani G G, Patil B K. Emoticon-based unsupervised sentiment classifier for polarity analysis in tweets[J]. International Journal of Engineering Research and General Science, 2014, 2(6).
30. Wen Ma, Xiangfeng Luo, Junyu Xuan, Ruirong Xue: Discover Semantic Topics in Patents within a Specific Domain. J. Web Eng. 16(7&8): 653-675 (2017)
31. Korsos L F, Taddy M. Gibbs Sampling for n-Gram Latent Dirichlet Allocation[J]. 2011.
32. Li X, Rao Y, Xie H, et al. Bootstrapping Social Emotion Classification with Semantically Rich Hybrid Neural Networks[J]. IEEE Transactions on Affective Computing, 2017, PP(99):1-1. F

Appendix: Detailed Derivation of Parameters

1. $p(\theta_d|\alpha,\{z_{d,n}\}) \propto p(\{z_{d,n}\},\theta_d|\alpha) = p(z_{d,n}|\theta_d)p(\theta_d|\alpha) = \prod_{n=1}^{N} p(z_{d,n}|\theta_d)\cdot p(\theta_d|\alpha) \propto \prod_{n=1}^{N}\prod_{k=1}^{K}\theta_{d,k}^{I_{z_n=k}} \cdot \prod_{k=1}^{K}\theta_{d,k}^{\alpha_k-1} = \prod_{k=1}^{K}\theta_{d,k}^{\sum_{n=1}^{N} I_{z_{d,n}=k}} \cdot \prod_{k=1}^{K}\theta_{d,k}^{\alpha_k-1} = \prod_{k=1}^{K}\theta_{d,k}^{\sum_{n=1}^{N} I_{z_{d,n}=k}+\alpha_k-1}$
$$\therefore \theta_d \sim Dir(\theta_d;\alpha+\sum_{n=1}^{N} I_K(z_{d,n}))$$
2. $p(\xi_d|\beta,\{\epsilon_{d,n}\},\{\varepsilon_{d,m}\},\{\rho_{d,r}\}) \propto p(\{\epsilon_{d,n}\},\{\varepsilon_{d,m}\},\{\rho_{d,r}\},\xi_d|\beta) = \prod_{n=1}^{N} p(\epsilon_{d,n}|\xi_d)\cdot \prod_{m=1}^{M} p(\varepsilon_{d,m}|\xi_d)\cdot\prod_{r=1}^{R} p(\rho_{d,r}|\xi_d)\cdot p(\xi_d|\beta) \propto \prod_{n=1}^{N}\prod_{e=1}^{E}\xi_{d,e}^{I_{\epsilon_{d,n}=e}}\cdot\prod_{m=1}^{M}\prod_{e=1}^{E}\xi_{d,e}^{I_{\varepsilon_{d,m}=e}}\cdot \prod_{r=1}^{R}\prod_{e=1}^{E}\xi_{d,e}^{I_{\rho_{d,r}=e}}\cdot\prod_{e=1}^{E}\xi_{d,e}^{\beta_e-1} = \prod_{e=1}^{E}\xi_{d,e}^{\sum_{n=1}^{N} I_{\epsilon_{d,n}=e}}\cdot\prod_{e=1}^{E}\xi_{d,e}^{\sum_{m=1}^{M} I_{\varepsilon_{d,m}=e}}\cdot\prod_{e=1}^{E}\xi_{d,e}^{\sum_{r=1}^{R} I_{\rho_{d,r}=e}}\cdot \prod_{e=1}^{E}\xi_{d,e}^{\beta_e-1} = \prod_{e=1}^{E}\xi_{d,e}^{\sum_{n=1}^{N} I_{\epsilon_{d,n}=e}+\sum_{m=1}^{M} I_{\varepsilon_{d,m}=e}+\sum_{r=1}^{R} I_{\rho_{d,r}=e}+\beta_e-1}$
$$\therefore \xi_d \sim Dir(\xi_d;\beta+\sum_{n=1}^{N} I_e(\epsilon_{d,n})+\sum_{m=1}^{M} I_e(\varepsilon_{d,m})+\sum_{r=1}^{R} I_e(\rho_{d,r}))$$
3. $p(z_{d,n}=k|\theta_d,w_{d,n},\{\delta_k\},\{\psi_e\},\epsilon_{d,n}) \propto p(w_{d,n},z_{d,n}=k|\theta_d,\{\delta_k\},\{\psi_e\},\epsilon_{d,n}) = p(w_{d,n}|\{\delta_k\},z_{d,n}=k,\{\psi_e\},\epsilon_{d,n})\cdot p(z_{d,n}=k|\theta_d) = \frac{\delta_{k,w_{d,n}}+\psi_{\epsilon_{d,n},w_{d,n}}}{2}\cdot\theta_{d,k}$
$$\therefore p(z_{d,n}|\theta_d,w_{d,n},\{\delta_k\},\{\psi_e\},\epsilon_{d,n}) \propto \theta_d \circ \frac{\delta^{w_{d,n}}+\delta^{\psi_{d,n}}}{2}$$
4. $p(\epsilon_{d,n}=e|\xi_d,w_{d,n},\{\delta_k\},\{\psi_e\},z_{d,n}) \propto p(w_{d,n},\epsilon_{d,n}=e|\xi_d,\{\delta_k\},z_{d,n}=k,\{\psi_e\}) = p(w_{d,n}|\epsilon_{d,n}=e,\{\delta_k\},z_{d,n},\{\psi_e\})\cdot p(\epsilon_{d,n}=e|\xi_d) = \frac{\delta_{z_{d,n},w_{d,n}}+\psi_{e,w_{d,n}}}{2}\cdot\xi_{d,e}$
$$\therefore p(\epsilon_{d,n}|\xi_d,w_{d,n},\{\delta_k\},\{\psi_e\},z_{d,n}) \propto \xi_d \circ \frac{\delta^{w_{d,n}}+\delta^{\psi_{d,n}}}{2}$$
5. $p(\varepsilon_{d,m}=e|\xi_d,l_{d,m},\{\varphi_e\}) \propto p(l_{d,m},\varepsilon_{d,m}=e|\xi_d,\{\varphi_e\}) = p(l_{d,m}|\{\varphi_e\},\varepsilon_{d,m}=e)\cdot p(\varepsilon_{d,m}=e|\xi_d) = \varphi_{e,l_{d,m}}\cdot\xi_{d,e}$
$$\therefore p(\varepsilon_{d,m}|\xi_d,l_{d,m},\{\varphi_e\}) \propto \xi_d \circ \varphi^{l_{d,m}}$$
6. $p(\rho_{d,q}=e|\xi_d,r_{d,q},\eta) \propto p(r_{d,q},\rho_{d,q}=e|\xi_d,\eta) = p(r_{d,q}|\eta,\rho_{d,q}=e)\cdot p(\rho_{d,q}=e|\xi_d) = \eta_{e,r_{d,q}}\cdot\xi_{d,e}$
$$\therefore p(\rho_{d,q}|\xi_d,r_{d,q},\eta) \propto \xi_d \circ \eta^{r_{d,q}}$$
7. $p(\delta_k|\gamma,z,w,\{\psi_e\},\epsilon) \propto p(\delta_k,w|\gamma,z,\{\psi_e\},\epsilon) = \prod_{\{(d,n):z_{d,n}=k\}} p(\delta_k,w|\gamma,z,\{\psi_e\},\epsilon)\cdot p(\delta_k|\gamma) \propto (\prod_{\{(d,n):z_{d,n}=k\}}(\prod_{v=1}^{V}\frac{\delta_{k,v}^{I_{w_{d,n}=v}}+\psi_{\epsilon_{d,n},v}^{I_{w_{d,n}=v}}}{2}))\cdot(\prod_{v=1}^{V}\delta_{k,v}^{\gamma_v-1})$
$$\therefore \delta_k \sim Mult(w;(\delta_z+\psi_e)/2)\cdot Dir(\gamma)$$
8. $p(\psi_e|\nu,\{\delta_k\},z_n=k,w,\epsilon) \propto p(w,|\nu,\{\delta_k\},z_n=k,\epsilon) = \prod_{\{(d,n):\epsilon_{d,n}=e\}} p(w_{d,n}|\{\delta_k\},\psi_e,z_{d,n})\cdot p(\psi_e|\nu) \propto (\prod_{\{(d,n):\epsilon_{d,n}=e\}}(\prod_{v=1}^{V}\frac{\delta_{z_{d,n},v}^{I_{w_{d,n}=v}}+\psi_{e,v}^{I_{w_{d,n}=v}}}{2}))\cdot(\prod_{v=1}^{V}\mu_{e,v}^{\nu_v-1})$
$$\therefore \psi_e \sim Mult(w;(\delta_z+\psi_e)/2)\cdot Dir(\gamma)$$
9. $p(\eta_e|\sigma,r,\rho) \propto p(r,\eta_e|\sigma,\rho) = p(r|\eta_e,\rho)\cdot p(\eta_e|\sigma) \propto \left(\prod_{i:\rho_i=e}\prod_{u=1}^{U}\eta_{e,u}^{I_{r_i=u}}\right)\cdot$

$$\left(\prod_{u=1}^{U}\eta_{e,u}^{\sigma_u-1}\right)=\prod_{i:\rho_i=e}\prod_{u=1}^{U}\eta_{e,u}^{I_{r_i=u}+\sigma_u-1}=\prod_{u=1}^{U}\eta_{e,u}^{\sum_{i\rho_i=e}I_{r_i=u}+\sigma_u-1}$$

$$\therefore\ \eta_e\sim Dir(\eta_{e,u};\sigma+\sum_{i\rho_i=e}I_u(r_i))$$