Knowl Inf Syst
DOI 10.1007/s10115-015-0865-0

REGULAR PAPER

Tracking the evolution of social emotions with topic models

Chen Zhu[1] · Hengshu Zhu[2] · Yong Ge[3] · Enhong Chen[1] · Qi Liu[1] · Tong Xu[1] · Hui Xiong[4]

Received: 13 December 2014 / Revised: 7 April 2015 / Accepted: 9 July 2015

Abstract Many of today's online news Web sites have enabled users to specify different types of emotions (e.g., angry or shocked) they have after reading news. Compared with traditional user feedbacks such as comments and ratings, these specific emotion annotations are more accurate for expressing users' personal emotions. In this paper, we propose to exploit these users' emotion annotations for online news in order to track the evolution of emotions, which plays an important role in various online services. A critical challenge is how to model emotions with respect to time spans. To this end, we propose a time-aware topic modeling perspective for solving this problem. Specifically, we first develop two models named *emotion-Topic over Time* (eToT) and *mixed emotion-Topic over Time* (meToT), in which the topics of news are represented as a beta distribution over time and a multinomial distribution over emotions. While they can uncover the latent relationship among news, emotion and time directly, they cannot capture the evolution of topics. Therefore, we further develop another model named *emotion-based Dynamic Topic Model* (eDTM), where we explore the state space model for tracking the evolution of topics. In addition, we demonstrate that all of proposed models could enable several potential applications, such as emotion prediction, emotion-based news recommendations, and emotion anomaly detections. Finally, we validate the proposed models with extensive experiments with a real-world data set.

✉ Enhong Chen
cheneh@ustc.edu.cn

Chen Zhu
zc3930155@gmail.com

1 School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China

2 Big Data Lab, Baidu Research, Beijing 100085, China

3 Department of Computer Science, The University of North Carolina at Charlotte, University City Blvd, Charlotte, NC 28223-0001, USA

4 Management Science and Information Systems Department, Rutgers Business School, Rutgers University, Newark, NJ 07102, USA

Published online: 28 July 2015

Keywords Social emotions · Topic models · Sentiment analysis

1 Introduction

With the increasing prosperity of Web 2.0, people are encouraged to have various social interactions on the Web sites. A recent development trend of online news Web sites, such as Yahoo! and Sina, is to allow readers to specify different types of emotions (e.g., angry and shocked) after reading news. Compared with traditional users' feedback (e.g., reviews, tags), such specific emotion annotations are more accurate for expressing users' personal emotions. For example, Fig. 1a shows an example of users' aggregated emotions at Sina News,[1] which has six different kinds of emotions. Each user can choose one emotion, which most accurately reflects his impression after reading, to annotate a piece of news. If we collectively look at and analyze all user emotions from the news Web site, we may be able to get a good picture about the overall emotions of online social media users, namely *social emotions*. The social emotions often vary with respect to the topics of news and time and thus have intrinsic dynamics. For example, Fig. 1b shows the distribution of aggregated social emotions with respect to different time spans in our data set collected from Sina News. We can observe different distributions of emotions along the time, which indicates that the social emotions evolve over time. In fact, such evolution of social emotions is inherently driven by the dynamic topics of news at different time. Capturing such dynamic characteristics of social emotions is critically important for the successful development of various social services, such as social opinion monitoring and social event detections. In the literature, there are recent studies about social emotion-related problems. For example, some works focus on sentiment analysis [24,35], social emotion analysis [6,13,18] of online documents and user emotion modeling [3]. However, few of them have paid attention to the dynamic characteristics of social emotions.

To this end, in this paper, we propose to exploit the users' emotion annotations from online news to track the evolution of social emotions. A critical challenge is how to model emotions with respect to time spans. Along this line, we propose a time-aware topic modeling perspective for solving this problem. Specifically, we first develop a model named *emotion-Topic over Time* (eToT), where we represent each topic of news with a beta distribution over time and a multinomial distribution over emotions. In this model, the process of modeling topics of news is affected not only by the word co-occurrences but also the emotions and time. Particularly, with some experimental observations, we find that not all the topics have distinct relationships with time, such as the common topics or the topics about news terminology. Therefore, we further propose an extension of eToT model, named *mixed emotion-Topic over Time* (meToT), in which a word can be generated from two different topic sets, i.e., one is time-independent and the other is time-dependent. Indeed, similar to the popular Latent Dirichlet Allocation (LDA) [5], news topics in both eToT and meToT are static which do not change with respect to different time spans. However, while some researchers have revealed that the latent topics of documents may evolve as time unfolds [4], eToT and meToT cannot capture the dynamics of topics. Therefore, we propose another model, namely *emotion-based Dynamic Topic Model* (eDTM), to capture the dynamics of news topics. In eDTM, we first divide all news into different segmentations with respect to their timestamps and then implement topic modeling for each segmentation. Models learned from different segmentations are linked together by the Markov state space model. In fact, all eToT, meToT, and eDTM can model

[1] http://news.sina.com.cn/.

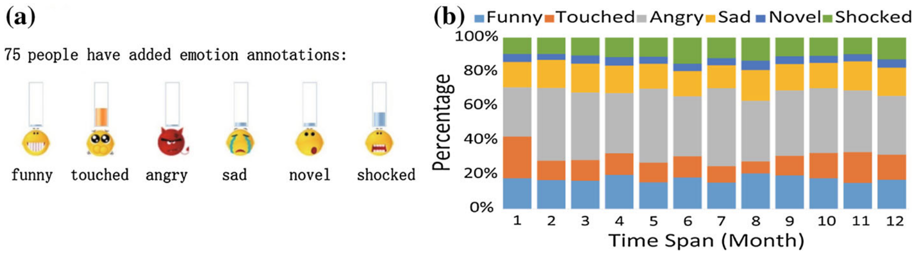


Fig. 1 **a** Examples of different user emotions. **b** The distribution of social emotions with respect to news in different time spans

the evolution of social emotion effectively and help us understand the semantic relationships between social emotion and news topics in different views. In addition, we demonstrate that all of them are general models and could enable several potential applications of social media, such as emotion prediction, emotion-based news recommendation and emotion anomaly detection. Specially, the contributions of this paper can be summarized as follows.

– First, we provide a comprehensive study to track the evolution of social emotions by exploiting the user emotion annotations from online news. Indeed, this study is vital for the successful development of various social services.
– Second, we propose three novel topic models, i.e., eToT, meToT, and eDTM, for solving the problem of tracking the evolution of social emotion. Particularly, the proposed models can effectively model the emotions, news and time from different views. Also, we introduce some novel emotion-based applications enabled by the proposed models.
– Third, we carry out extensive experiments on a real-world data set which was collected from Sina News to evaluate the proposed models. As shown in our experiments, the proposed models are all effective for modeling social emotions and other enabled applications.

The remainder of this paper is organized as follows. In Sect. 2, we briefly introduce some related works of this paper. Section 3 introduces the details of proposed topic models. In Sect. 4, we make some discussions about our models and introduce their potential applications. Section 5 reports the experimental results. Finally, we conclude this paper in Sect. 6.

2 Related work

The related works can be grouped into two categories: sentiment analysis and topic model. We will give brief discussions on these related works and the relationships between them and the proposed models in this paper.

2.1 Sentiment analysis

In the field of sentiment analysis, existing researches mainly focus on two aspects, namely sentiment classification and sentiment information extraction. Among them, since sentiment extraction aims to extract units that are related to emotion at sentence level or paragraph level, it is less relevant to our scenario. Therefore, in the following, we only discuss the works on sentiment classification [13,16,18,26,31,37].

Classification is a very fundamental problem in machine learning [26,39,40]. Traditionally, sentiment classification is often addressed by classic classification approaches directly.

Recently, researchers have proposed a lot of new approaches for sentiment classification and begun to study the readers' emotions. For instance, in [17], two ranking methods were presented to rank readers' emotions. Lin et al. [18] also studied the readers' emotions that are triggered by articles, and the authors mainly focus on feature selection. Kozareva et al. [13] designed an approach for classifying headline emotion based on the information collected from the World Wide Web. Besides, emoticon, as a kind of emotion label, has also been used for sentiment classification. Zhao et al. [38] built a system called MoodLens, in which 95 emoticons are mapped into four categories of sentiments for the sentiment classification of Chinese tweets in Weibo. Liu et al. [20] treated the emoticons as labeled data and integrated it and labeled data into a uniform sentiment classification framework.

We can find that most of existing works aim to study the sentiment directly from the documents (words). Since there is a consensus that a document is composed of topics, it is also reasonable to treat sentiment of a document as a combination of sentiment on these topics. Therefore, in this paper, we focus on analyzing sentiment from the perspective of the sentiment (emotion) of the topics. There have been some works about the emotion with respect to topics. For instance, Mei et al. [22] modeled texts through a mixture of topic model and sentiment model. Lin and He [16] also extended LDA to detect sentiment and topics. However, none of them considers the temporal effect.

2.2 Topic models

Topic model is a tool for analyzing texts or other types of discrete data. Many different kinds of topic models [4,5,21,33] have been proposed, among which Latent Dirichlet Allocation (LDA) [5] is very popular. Actually, the models proposed in this paper can be also traced back to LDA. The basic assumption of LDA is that each document can be treated as a mixture of topics and the generation of words in the document is attributable to one of those topics.

However, LDA is a static model that does not consider the influence of time, so it is inappropriate to be adopted in many applications, including social emotion analysis. To this end, several topic models have examined topics and their changes with time. For instance, Topics over Time (ToT) [33] model associates each topic with a beta distribution over time to directly capture the topic changes over time. The relative simplicity of ToT makes it easy to inject its ideas to other topic models. Different from ToT, Dynamic Topic Model (DTM) [4] uses state space model on the natural parameters to model the evolution of topic and the variational approximations depending on Kalman filters and wavelet regression. DTM requires that time must be discretized, and thus how to determine the length of time span is an important problem. Wang et al. [32] solved this problem by replacing state space model with Brownian motion which is actually an acontinuous generalization of state space model. Another way to solve this problem was proposed in [11], which considers multi-scale time span to track the topic changes. Topic Tracking Model (TTM) [9] is another Dynamic Topic Model for tracking consumer purchase behaviors. Compared with DTM, it simplifies the model by replacing Gaussian distribution with Dirichlet distribution so that the parameters could be estimated by Gibbs sampling which is simpler compared with variational methods. In summary, there are two main approaches to model the evolution of topics over time: treating time as an attribute of topics, e.g., ToT; using state space model or Brownian motion on natural parameters, e.g., DTM and TTM.

Meanwhile, some work extended LDA from another perspective. The assumption behind them, which is similar to that of meToT, is each word could come from several different topic sets. Beyond bag of words, Wang et al. [34] introduced phrases and proposed *topical n-grams*, in which a word may come from a topic-specific unigram distribution or a bigram distribution. Following similar idea, Pual et al. [27] tried to detect cultural differences from online blogs and forums. They assumed that an article from a special cultural group is a mixture of some general topics and culture-special topics. Besides, some researchers have modeled sentiment by this way. For instance, Titov and McDonald [31] proposed a Multi-Aspect Sentiment model consisting of an extension of LDA, Multi-Grain LDA, and sentiment predictors for sentiment summarization.

As introduced above, there have been some works aiming to analyze sentiment by topic model, but none of them focus on analyzing social emotions with respect to temporal information.

3 Modeling social emotion

In this section, firstly, some preliminaries and backgrounds will be introduced. Then, we explain the details of our time-aware topic modeling solutions for analyzing social emotion.

3.1 Preliminaries and backgrounds

Recently, many of todays online news Web sites have enabled users to specify their emotions after reading news. Therefore, for each piece of news, we can get not only the textual content and timestamp but also the emotion that was annotated by users. For example, Fig. 1a illustrates the emotion annotations from 75 users, and these emotions are classified into six categories, i.e., "*Funny*," "*Touched*," "*Angry*," "*Sad*," "*Novel*," and "*Shocked*." As shown in Fig. 1a, all emotion annotations of a piece of news from different users are usually aggregated together.

For social emotion modeling, intuitively, users' emotion about a document (e.g., news) is mainly decided by their emotion about the topics of the document, which could be extracted by topic model. Indeed, topic model is an unsupervised method to effectively extract topics from the texts based on the co-occurrence of words. Considering that different topics lead to different user emotions, e.g., some topics tend to depress people but some other topics may be very affecting, it is reasonable to associate each topic with a distribution with respect to emotion. In this way, we can model the evolution of social emotions by introducing an emotion and time generation layer into the topic models, in which the discovery of topics is influenced by social emotions, co-occurring of words, and also temporal information.

Formally, we assume that a document d is a bag of words $\mathbf{w}_d$, size of which is N_d, and $\mathbf{w}_d$ is from a vocabulary containing V words. Each document also has a timestamp t and an observed multinomial distribution over emotions $\mathbf{e}$. Thus, we can use a set of triples $\mathbf{D} = \{(\mathbf{w}_0, t_0, \mathbf{e}_0), \ldots, (\mathbf{w}_D, t_D, \mathbf{e}_D)\}$ to represent a collection of D documents. Next, we will explain our solutions to model the evolution of social emotion from the corpus $\mathbf{D}$.

3.2 Emotion-topic over time

In this subsection, we would propose a novel topic model named emotion-Topic over Time (eToT) to directly uncover the latent relationship among documents, emotion, and time. Compared with LDA, each topic discovered by eToT also has a distribution with respect to time and a distribution with respect to emotion. That is, besides understanding the content of a topic, we can know the people's impression of this topic (i.e., social emotion) and when this topic emerges.

We first introduce the way to connect social emotions with topics. It has been introduced briefly above that we assume the users' emotions about a specific document come from their emotions about the topics of this document and each topic should own a emotional tilt. Thus, it is reasonable to assume that each topic has a latent distribution with respect to emotion in eToT. Note that, the emotions are discrete and the number of each kind of emotions could be observed for any documents (news). Therefore, given a news article, we could easily get the distribution over emotions **e** with respect to this news. Here, **e** is observed, while the emotion distributions on topics η, which determine **e**, are latent. Specifically, we assume this latent distribution follows the Dirichlet distribution, i.e., each topic k has a Dirichlet distribution η_k with respect to emotion. Dirichlet distribution is a family of continuous multivariate probability distributions over simplex of which the summation is equal to one. Thus, it is the appropriate choice to model the distribution of emotions. Given a topic k, the probability of an observed emotion distribution **e** could be calculated as follows:

$$P(\mathbf{e}|\eta_k) = \frac{\Gamma\left(\sum_{l=1}^{E} \eta_{k,l}\right)}{\prod_{l=1}^{E} \Gamma(\eta_{k,l})} \cdot \prod_{l=1}^{E} e_l^{\eta_{k,l}-1}, \tag{1}$$

where η_k are parameters of Dirichlet distribution and the inference for them will be shown later, and **e** is the vector of proportion of emotions. E is the number of categories of emotions and is also the size of vector $\eta_k.\Gamma(\cdot)$ is gamma function.

Similarly, we also directly associate time with topics, i.e., each topic k has a latent distribution ψ_k with respect to time. Following the ideas in ToT, we select beta distribution as ψ_k, which could behave versatile shapes. Therefore, we need to normalize the timestamp into a range from 0 to 1 firstly. A point we should pay attention to is how to normalize the timestamps. Indeed, the timestamps can be normalized in different ways, which depends on different application requirements. To be specific, if they are normalized with respect to the entire time range, the model can capture the long-term relationships between times, emotion, and topics, which can be used for social opinion monitoring. In contrast, if we normalize them with respect to a specific time range (e.g., week or month), the model can capture the periodical effects of social emotions (e.g., the emotions during Christmas each year), which can be used for predicting social emotions in the future. Particularly, for simplicity, we just choose the entire time range for normalizing timestamps. Then, given a topic k, the probability of an observed timestamp t could be calculated as follows:

$$P(t|\psi_k) = \frac{\Gamma(\psi_{k,1}+\psi_{k,2})}{\Gamma(\psi_{k,1})\Gamma(\psi_{k,2})} \cdot (1-t)^{\psi_{k,1}-1} t^{\psi_{k,2}-1}, \tag{2}$$

where ψ_k are parameters of beta distribution and the inference for them will be shown later.

Note that, for simplifying the process of parameter estimation, we generate emotions and timestamp with respect to each word token. That is, we assume all word tokens in one document share the same emotion distribution **e** and the same timestamp t with the document. In summary, the corresponding graphical model of eToT is shown in Fig. 2, and its parameterizations are

$$\begin{aligned}
\theta_d|\alpha &\sim \text{Dirichlet}(\alpha),\\
\phi_k|\beta &\sim \text{Dirichlet}(\beta),\\
z_{d,i}|\theta_d &\sim \text{Multinomial}(\theta_d),\\
w_{d,i}|\phi_{z_{d,i}} &\sim \text{Multinomial}(\phi_{z_{d,i}}),\\
t_{d,i}|\psi_{z_{d,i}} &\sim \text{Beta}(\psi_{z_{d,i}}),\\
\mathbf{e}_{d,i}|\eta_{z_{d,i}} &\sim \text{Dirichlet}(\eta_{z_{d,i}}).
\end{aligned}$$

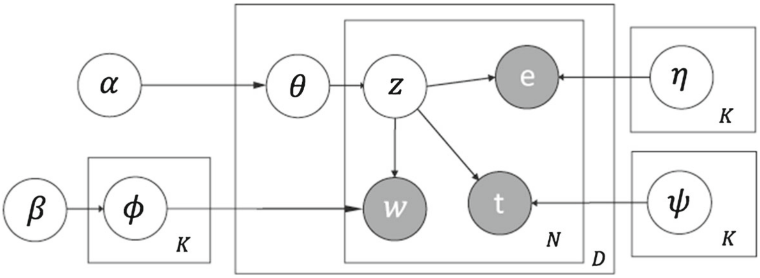


Fig. 2 The graphical representation of eToT

Thus, the generative process for a document d is as follows:

1. Draw a multinomial distribution θ_d over topics from a Dirichlet prior α;
2. Draw parameters of topic-word multinomial distributions ϕ_k from a Dirichlet prior β;
3. For a word token i in document d:
   (a) Draw a topic $z_{d,i}$ from Mult(θ_d);
   (b) Draw a word $w_{d,i}$ from Mult($\phi_{z_{d,i}}$);
   (c) Draw a timestamp $t_{d,i}$ from Beta($\psi_{z_{d,i}}$);
   (d) Draw a multinomial distribution $e_{d,i}$ over emotions from Dirichlet($\eta_{z_{d,i}}$);

We employ Gibbs sampling to estimate parameters, i.e., to "invert" the generative process and "generate" latent variables from given observations. For simplicity, we estimate ψ and η once per iteration of Gibbs sampling. The joint distribution of topic z, time t, word w, and emotion e is as follows:

$$P(\mathbf{w}, \mathbf{t}, \mathbf{z}, \mathbf{e}|\alpha, \beta, \eta, \psi) = \prod_{d=1}^{D}\prod_{i=1}^{N_d}(P(t_{d,i}|\psi_{z_{d,i}})P(\mathbf{e}_{d,i}|\eta_{z_{d,i}}))\cdot \prod_{z=1}^{K}\frac{\Delta(\mathbf{n}^z+\beta)}{\Delta(\beta)}\prod_{d=1}^{D}\frac{\Delta(\mathbf{n}^d+\alpha)}{\Delta(\alpha)}. \tag{3}$$

In Gibbs sampling procedure, what we need to calculate is the conditional distribution $P(z_{d,i} = k|\mathbf{w}, \mathbf{t}, \mathbf{z}_{\neg d,i}, \mathbf{e}, \alpha, \beta, \Psi, \eta)$ where $\mathbf{z}_{\neg d,i}$ means the topics for all of the word tokens except $w_{d,i}$. Because of conjugation, the conditional distribution in each iteration can be written as:

$$P(z_{d,i} = k|\mathbf{w}, \mathbf{t}, \mathbf{z}_{\neg d,i}, \mathbf{e}, \alpha, \beta, \Psi, \eta) \propto \frac{n^d_{k,\neg d,i}+\alpha_k}{\sum_{z=1}^{K}\left(n^d_{z,\neg d,i}+\alpha_z\right)}\cdot\frac{n^k_{w_{d,i},\neg d,i}+\beta_{w_{d,i}}}{\sum_{v=1}^{V}\left(n^k_{v,\neg d,i}+\beta_v\right)}\cdot \frac{(1-t_{d,i})^{\psi_{k,1}-1}(t_{d,i})^{\psi_{k,2}-1}}{B(\psi_{k,1},\psi_{k,2})}\cdot\left(\frac{\Gamma\left(\sum_{l=1}^{E}\eta_{k,l}\right)}{\prod_{l=1}^{E}\Gamma(\eta_{k,l})}\cdot\prod_{l=1}^{E}e_{d,i,l}^{\eta_{k,l}-1}\right). \tag{4}$$

The notation n^a_b refers to the number of times b has been assigned to a. For example, n^d_k is the number of word tokens in document d that are assigned to topic k, n^k_v is the number of word v assigned to topic k, and $e_{d,i,l}$ is the lth emotion proportion of the ith word token in document d. After each iteration of Gibbs sampling, we update ψ and η as

$$\hat{\psi_{k,1}} = \bar{t_k}\left(\frac{\bar{t_k}(1-\bar{t_k})}{{S^{(t)}}^2_k}-1\right), \tag{5}$$

$$\hat{\psi_{k,2}} = (1 - \bar{t_k})\left(\frac{\bar{t_k}(1 - \bar{t_k})}{{S^{(t)}}_k^2} - 1\right), \tag{6}$$

$$\hat{\eta_{k,l}} = \bar{e_{k,l}}\left(\frac{\bar{e_{k,1}}(1 - \bar{e_{k,1}})}{{S^{(e)}}_{k,1}^2} - 1\right), \tag{7}$$

where $\bar{t_k}$ and ${S^{(t)}}_k^2$ are the sample mean and the biased sample variance of the timestamps of word tokens belonging to topic k, respectively. Similarly, $\bar{e_{k,l}}$ and ${S^{(e)}}_{k,l}^2$ are the sample mean and biased sample variance of the ith emotion of word tokens belonging to topic k. After the process of Gibbs sampling, $\theta_{d,k}$ and $\phi_{k,i}$ can be evaluated as follows:

$$\hat{\theta_{d,k}} = \frac{n_k^d + \alpha_k}{\sum_{z=1}^{K}(n_z^d + \alpha_z)}, \tag{8}$$

$$\hat{\phi_{k,i}} = \frac{n_i^k + \beta_i}{\sum_{v=1}^{V}(n_v^k + \beta_v)}. \tag{9}$$

By eToT, we could extract topics and some additional information about them from a collection of texts. Though the topics are constant, we get their distributions with respect to time ψ and discrete distributions with respect to emotions η. By these distributions, we can understand the relationships among emotions, topics, and time. And the temporal information could help us to discover those emergencies effectively [33], which will also be shown in experiments well. At last, note that, the generation process of our eToT model is similar to the ToT model [33]. Indeed, in eToT, topic discovery is affected by all of the effects from word co-occurrences, emotion, and temporal information, while ToT only considers the influence from words and time. So ToT could not be used for sentiment analysis and modeling social emotions.

3.3 Mixed emotion-topic over time

In eToT, we associate all topics with time, which means all topics have distinct relationships with time. However, based on our preliminary experiments, we observe that the time distributions of some topics discovered by eToT are very smooth, such as the common topics or the topics about news terminology. Therefore, we argue that some topics are not sensitive to time. Following this idea, we propose an extension of eToT model, named *mixed emotion-Topic over Time*, in which there are two sets of topics, i.e., one is time-independent and the other is time-dependent.

In meToT, we employ the same methods of eToT to associate topics with time and emotion, while each document has two topic distributions, θ_1 and θ_2, corresponding to two sets of topics ϕ_1 and ϕ_2. Although both of them have word distributions and emotion distributions, only ϕ_1 owns beta distributions with respect to time. Thus, only ϕ_1 is time-dependent. The graphical representation of meToT is shown in Fig. 3. Similar to eToT, we also generate emotions and timestamps with respect to each word token. However, we employ an additional mixture component to distinguish where each word is generated from, time-dependent topics or time-independent topics, by a new set of discrete variables x. In meToT, each word token has two topic labels, z_1 and z_2. If the state of i-th word token, x_i, is 1, this word and emotion come from z_{1_i} and vice versa. Because the timestamp is only related to θ_1, if $x_i = 1$, the timestamp could be drawn from corresponding η but if $x_i = 0$, we draw the timestamp from a background time distribution, $Beta(\psi_0)$, which is a uniform distribution actually. Meanwhile x_i is sampled

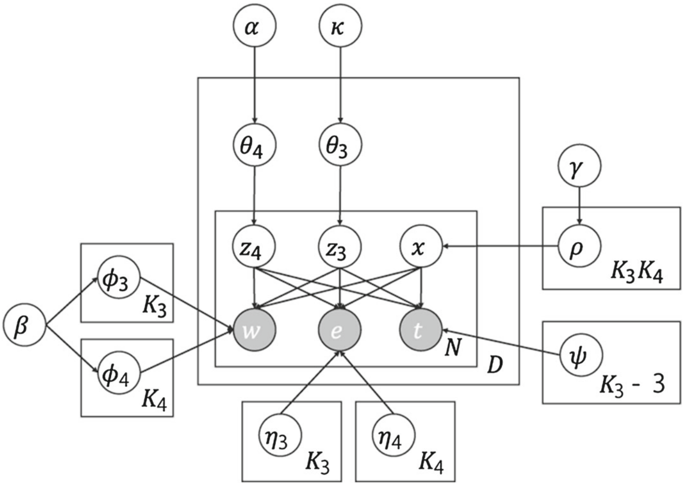


Fig. 3 The graphical representation of meToT

from the binomial distribution $\rho_{z1_i,z2_i}$. And γ is the prior of ρ. The meToT parameterizations are

$$\begin{aligned}
\theta_{1_d}|\alpha &\sim \text{Dirichlet}(\alpha),\\
\theta_{2_d}|\kappa &\sim \text{Dirichlet}(\kappa),\\
\phi_k|\beta &\sim \text{Dirichlet}(\beta),\\
\rho_l|\gamma &\sim \text{Beta}(\gamma),\\
z1_{d,i}|\theta_{1_d} &\sim \text{Multinomial}(\theta_{1_d}),\\
z2_{d,i}|\theta_{2_d} &\sim \text{Multinomial}(\theta_{2_d}),\\
x_{d,i}|\rho_{z1_{d,i},z2_{d,i}} &\sim \text{Bernoulli}(\rho_{z1_{d,i},z2_{d,i}}),\\
w_{d,i}|\phi_{z*_{d,i}} &\sim \text{Multinomial}(\phi_{z*_{d,i}}),\\
t_{d,i}|\psi_{z*_{d,i}} &\sim \text{Beta}(\psi_{z*_{d,i}}),\\
\mathbf{e}_{d,i}|\eta_{z*_{d,i}} &\sim \text{Dirichlet}(\eta_{z*_{d,i}}),
\end{aligned}$$

where $*$ in $z*_{d,i}$ is a label, which could be set as 0 or 1, indicating whether it is a time-dependent topic or time-independent one. The generative process for a document d is as follows:

1. Draw a multinomial distribution θ_{1_d} over topics from a Dirichlet prior α;
2. Draw a multinomial distribution θ_{2_d} over topics from a Dirichlet prior κ;
3. For a word token i in document d:
   (a) Draw a topic $z1_{d,i}$ from Mult(θ_{1_d});
   (b) Draw a topic $z2_{d,i}$ from Mult(θ_{2_d});
   (c) Draw $x_{d,i}$ from Bernoulli($\rho_{z1_{d,i},z2_{d,i}}$);
   (d) If $x_{d,i} = 1$:
      i Draw a word $w_{d,i}$ from Mult($\phi_{z1_{d,i}}$);
      ii Draw a timestamp $t_{d,i}$ from Beta($\psi_{z1_{d,i}}$);
      iii Draw a multinomial distribution $e_{d,i}$ over emotions from Dirichlet($\eta_{z1_{d,i}}$);

(e) if $x_{d,i} = 0$:
 i Draw a word $w_{d,i}$ from Mult($\phi_{z2_{d,i}}$);
 ii Draw a timestamp $t_{d,i}$ from Beta(ψ_0);
 iii Draw a multinomial distribution $e_{d,i}$ over emotions from Dirichlet($\eta_{z2_{d,i}}$);

We also infer the topic assignments by Gibbs sampling. The corresponding parameter inference is similar to eToT. The conditional distribution is as follows:

$$
\begin{aligned}
&P(z1_{d,i} = k_1, z2_{d,i} = k_2, x_{d,i} = s|\mathbf{w}, \mathbf{z}1_{\neg_{d,i}}, \mathbf{z}2_{\neg_{d,i}}, \mathbf{x}_{\neg_{d,i}}, \mathbf{t}, \mathbf{e}, \alpha, \beta, \psi, \eta) \\
&\propto \frac{n^d_{k_1,\neg_{d,i}} + \alpha_{k_1}}{\left(\sum_{z=1}^{K_1} n^d_{z,\neg_{d,i}} + \alpha_z\right) - 1} \frac{n^d_{k_2,\neg_{d,i}} + \gamma_{k_2}}{\left(\sum_{z=1}^{K_2} n^d_{z,\neg_{d,i}} + \gamma_z\right) - 1} \frac{n_s^{k_1,k_2} + \gamma_0}{n_{x=0}^{k_1,k_2} + n_{x=1}^{k_1,k_2} + \gamma_0 + \gamma_1} \\
&\cdot \begin{cases} \frac{n^{k_2}_{w_{d,i},\neg_{d,i}} + \beta_{w_{d,i}}}{\sum_{v=1}^{V}\left(n^{k_2}_{v,\neg_{d,i}}\right) + \beta_v} P(e_{d,i}|\eta_{k_2}) P(t_{d,i}|\psi_0), & \text{if } x_{d,i} = 0, \\ \frac{n^{k_1}_{w_{d,i},\neg_{d,i}} + \beta_{w_{d,i}}}{\sum_{v=1}^{V}\left(n^{k_1}_{v,\neg_{d,i}}\right) + \beta_v} P(e_{d,i}|\eta_{k_1}) P(t_{d,i}|\psi_{k_1}), & \text{if } x_{d,i} = 1. \end{cases}
\end{aligned} \tag{10}
$$

The approach to evaluate ϕ, η, ψ is similar to that of eToT. With emToT, we can better analyze the relationships among emotion, time, and topics. In experiments, we would show that meToT could distinguish stable topics and dynamic ones in terms of time.

3.4 emotion-based Dynamic Topic Model

Although eToT and meToT can uncover the latent relationship among documents, emotion, and time, the constant topics learned by eToT could not capture the dynamics of news topics. To this end, in this subsection, we propose another model, emotion-based Dynamic Topic Model (eDTM), to solve this problem.

Different from them, eDTM discovers topics by the content and emotion information in each time span, and the topics in different time spans are chained by state space model. In this way, eDTM could learn the distribution over words and the distribution with respect to emotion for each topic in different time spans. Thus, the evolution of news topics is demonstrated clearly and directly by eDTM. Here, we assume the distribution over words of a topic should evolve smoothly, while emotion may not. Actually, the study on social science has shown that the popular mind is weird [1,14]. Sometimes, it is bigoted, conservative, and imperious. At other times, it shows some totally different characteristics such as impulse, fickleness, and irritability.

Thus, we just chain topics depending on their distributions over words. Due to the state space model, we need firstly divide texts into different time spans by timestamps. Then eDTM models the documents in time span t with a topic model which has K topics, where these topics evolve from the topics associated with the time span $t - 1$. For simplicity, we use Dirichlet distribution to chain the neighboring topic models on the parameters of the multinomial distributions of topics over word tokens ϕ. We could also employ Gibbs sampling to evaluate these parameters. Specifically, the sequential structure is

$$
P(\phi_{t,k}|\phi_{t-1,k}, \lambda_{t,k}) \propto \prod_{v=1}^{V} \phi_{t,k,v}^{\lambda_{t,k}\phi_{t-1,k,v}-1}, \tag{11}
$$

where λ is a parameter to control the influence of the topic model in time span $t - 1$ on the topic model in time span t.

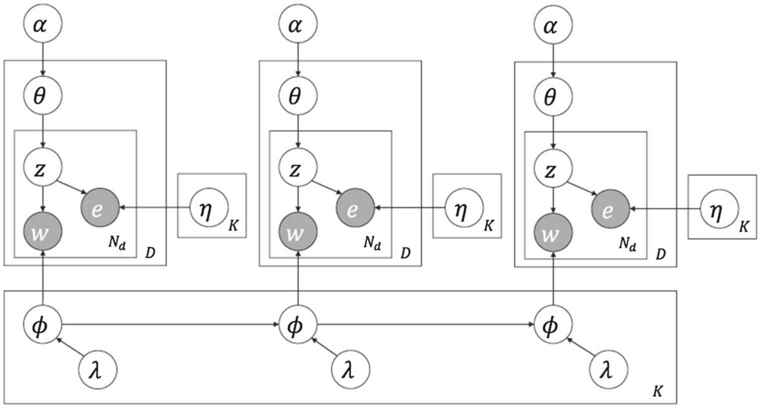


Fig. 4 The graphical representation of eDTM

Since the state space model requires the discretization of time, we need to modify some notations in this subsection. In time span t, the document collection is $\mathbf{D}_t$ and the size of it is D_t. Given a document d, the set of word tokens is $\{w_{t,d,i}\}$ and $|\{w_{t,d,i}\}|$ is $N_{t,d}$. We represent the emotion of document d (word tokens in d) by an observed multinomial distribution $\mathbf{e}_{t,d}.\theta_{t,d}$ is the topic distributions in document $d.\phi_{t,k}$ and $\eta_{t,k}$ are the parameters for Dirichlet distributions with respect to words and emotions for the topic k, respectively. E is the number of emotions. Finally, the graphical model of eDTM is shown in Fig. 4, and the eDTM parameterizations are

$$\begin{aligned}
\theta_{t,d}|\alpha_t &\sim \text{Dirichlet}(\alpha_t),\\
\phi_{t,k}|\phi_{t-1,k} &\sim \text{Dirichlet}(\lambda_{t,k}\cdot\phi_{t-1,k}),\\
z_{t,d,i}|\theta_{t,d} &\sim \text{Multinomial}(\theta_{t,d}),\\
w_{t,d,i}|\phi_{t,z_{t,d,i}} &\sim \text{Multinomial}(\phi_{t,z_{t,d,i}}),\\
\mathbf{e}_{t,d,i}|\eta_{t,z_{t,d,i}} &\sim \text{Dirichlet}(\eta_{t,z_{t,d,i}}).
\end{aligned}$$

The generative process for a document d in time span t is:

1. Draw a multinomial distribution $\theta_{t,d}$ over topics from a Dirichlet prior α_t;
2. Draw parameters of topic-word multinomial distributions $\phi_{t,k}$ from $Dirichlet(\lambda_{t,k}\cdot\phi_{t-1,k})$;
3. For a word token i in document d in time span t:
   (a) Draw a topic $z_{t,d,i}$ from Mult$(\theta_{t,d})$;
   (b) Draw a word $w_{t,d,i}$ from Mult$(\phi_{t,z_{t,d,i}})$;
   (c) Draw a multinomial distribution $e_{t,d,i}$ over emotions from Dirichlet$(\eta_{t,z_{t,d,i}})$;

We infer the topic assignments by Gibbs sampling. The joint distribution of topic z, word w, and emotion e is as follows:

$$\begin{aligned}
&P(\mathbf{w}_t,\mathbf{z}_t,\mathbf{e}_t|\alpha_t,\phi_{t-1},\lambda_t,\eta_t)\\
&\quad = P(\mathbf{w}_t|\mathbf{z}_t,\phi)p(\mathbf{e}_t|\eta,\mathbf{z}_t)P(\mathbf{z}_t|\alpha_t)P(\phi_t|\phi_{t-1},\lambda_t)\\
&\quad = \prod_{z=1}^{K}\frac{\Delta(\mathbf{n}^{t,z}+\phi_{t-1,z}\cdot\lambda_{t,z})}{\Delta(\phi_{t-1,z}\cdot\lambda_{t,z})}\prod_{d=1}^{D_t}\frac{\Delta(\mathbf{n}^{t,d}+\alpha_t)}{\Delta(\alpha_t)}
\end{aligned}$$

$$\cdot \prod_{d=1}^{D_t} \prod_{i=1}^{N_d} P(\mathbf{e}_{t,d,i} | \eta_{t,z_{t,d,i}}), \tag{12}$$

$$P(\mathbf{e}_{t,d,i} | \eta_{t,z_{t,d,i}}) = \frac{\Gamma\left(\sum_{l=1}^{E} \eta_{t,z_{t,d,i},l}\right)}{\prod_{l=1}^{E} \Gamma\left(\eta_{t,z_{t,d,i},l}\right)} \cdot \prod_{l=1}^{E} e_{t,d,i,l}^{\eta_{t,z_{t,d,i},l}-1}, \tag{13}$$

where $\mathbf{n}^{t,z}$ is a V-dimensional vector, where each entry represents the number of each word assigned to topic z in time span t, and $\mathbf{n}^{t,d}$ is a K-dimensional vector representing the number of word tokens assigned to each topic in text d of time span $t.K$ is the number of topics.

During Gibbs sampling, we need to calculate the full conditional distribution of assigning word token i on each topic, e.g., topic k, and this is

$$P(z_{t,d,i} = k | \mathbf{w}_t, \mathbf{z}_{t,\neg_{d,i}}, \mathbf{e}_t, \alpha_t, \phi_{t-1}, \lambda_t, \eta_t) = P(\mathbf{e}_{t,d,i} | \eta_{t,k}) \cdot \\ \frac{n_{t,w_{t,d,i},\neg_{d,i}}^{t,k} + \phi_{t-1,k,w_{t,d,i}} \lambda_{t,k}}{\sum_{v=1}^{V}\left(n_{t,v,\neg_{d,i}}^{t,k}\right) + \lambda_{t,k}} \frac{n_{t,k,\neg_{d,i}}^{t,d} + \alpha_{t,k}}{\sum_{z=1}^{K}\left(n_{t,z,\neg_{d,i}}^{t,d} + \alpha_{t,z}\right)}. \tag{14}$$

Similar to that in eToT, we update η after each iteration of Gibbs sampling by the following equation:

$$\hat{\eta_{t,k,l}} = \bar{e_{t,k,l}} \left(\frac{\bar{e_{t,k,1}}(1 - \bar{e_{t,k,1}})}{S^{(e)2}_{t,k,1}} - 1 \right), \tag{15}$$

where $\bar{e_{t,k,l}}$ and $S^{(e)2}_{t,k,l}$ are the sample mean and the biased sample variance of the lth emotion of word tokens belonging to the topic k in time span t, respectively. After enough iterations, $\theta_{t,d,k}$ and $\phi_{t,k,i}$ can be evaluated as follows:

$$\hat{\theta_{t,d,k}} = \frac{n_{t,k}^{t,d} + \alpha_{t,k}}{\sum_{z=1}^{K}(n_{t,z}^{t,d} + \alpha_{t,z})}, \tag{16}$$

$$\hat{\phi_{t,k,i}} = \frac{n_{t,i}^{t,k} + \lambda_{t,k}\phi_{t-1,k,i}}{\sum_{v=1}^{V}(n_{t,v}^{t,k} + \lambda_{t,k}\phi_{t-1,k,v})}, \tag{17}$$

the parameter λ controls the process of evolution, and it could be fixed or estimated through maximizing the joint distribution by the following equation [10,23]:

$$\lambda_{t,k} \leftarrow \lambda_{t,k} \cdot \frac{\sum_{v=1}^{V} \phi_{t-1,k,v} \cdot \Lambda_{t,k,v}}{\Psi(\sum_{v=1}^{V} n_{t,k,v} + \lambda_{t,k}) - \Psi(\lambda_{t,k})}, \tag{18}$$

$$\Lambda_{t,k,v} = \Psi(n_{t,k,v} + \lambda_{t,k} \cdot \phi_{t-1,k,v}) - \Psi(\lambda_{t,k} \cdot \phi_{t-1,k,v}), \tag{19}$$

where $\Psi(\cdot)$ is actually the derivative of the logarithm of the gamma function.

Different from eToT and meToT which output the constant topics, eDTM can get the distribution over words ϕ and distribution over emotions η for topics in any time span. Because of the state space model, for a given topic, eDTM could unveil the evolution of words and emotions directly. Compared with DTM, we choose the Dirichlet distribution as an alternative to model the evolution, which makes a relatively easy method of parameter estimation available.

4 Discussion and applications

In this section, we provide some discussions and potential applications about our novel models.

4.1 Model discussion

All eToT, meToT, and eDTM can model the evolution of social emotions effectively. Meanwhile, they could unveil the semantic relationships between social emotion and news topics from different views. To be specific, from the experiments, we have observed that some words, which represent people's names or addresses in some social events, have high generation probabilities in the topics learned by eToT. For example, a top-ranked word in one of the topics learned by eToT is the name of a doctor, who received unjustified treatment and attracted lots of public attention. Therefore, the two emotions with highest generation probabilities in this topic are "*Angry*" and "*Sad*," which reflect people's sympathy for the doctor. It implies that eToT has the ability to find those emergencies and could reflect the popular incline to them. Differently, eDTM can discover the hidden semantics from another dynamic view. For example, we find the top-ranked words in a topic learned by eDTM always include "Police," "Hospital," and "Investigation," which may indicate the semantics of crime. Therefore, at the beginning, the two emotions with highest generation probabilities in this topic are "*Angry*" and "*Shocked*." However, there are some interesting changes of the generation probability of emotion during the evolution of this topic. As time passes by, the emotion "*Sad*" became more and more important and finally exceeded "*Shocked*" in the topic, which may indicate people had changed their focus from the crime events to the victims.

Besides, compared with our work, blog data analysis is a very similar field, which has attracted a lot of attentions. In [36], authors seek to accurately identify those posts with high-volume response by a topic model, which aims to model the relationships between blogs and their comments. Authors of [29] aim to employ topic detection and tracking on blog data. In [25], authors present a large-scale study of weblog comments and their relation to the posts. Because there is no emotion label in blog data, we could not directly apply our models on it. But we could extract readers' emotion from the comments of blogs perviously or even add a new layer to employ comment.

4.2 Model applications

Indeed, many applications can be derived from our novel models, i.e., eToT, meToT, and eDTM. In the following, we demonstrate three motivating examples including the emotion prediction, the emotion-based news recommendation as well as the emotion anomaly detection.

4.2.1 Emotion prediction

Emotion prediction is a classic problem in sentiment analysis, which has great application value in both industry and academia [3,18]. Indeed, our models could be also leveraged for solving this problem. Specifically, given a document (e.g., a news) d with timestamp t, the goal is to predict the emotion e^* with the highest generation probability; that is,

$$e^* = \arg\max_e P(e|d) = \arg\max_e P(e|\mathbf{w}, t), \quad (\mathbf{w} \in d). \tag{20}$$

In eToT, the probability $P(e|d)$ can be computed by

$$P(e|d) \propto \sum_z \prod_{w \in d} P(w|z)P(t|z)P(e|z)P(z), \tag{21}$$

where $P(w|z)$, $P(t|z)$, $P(e|z)$, and $P(z)$ can be learned during the training of models. In meToT, the probability $P(e|d)$ is similar to that in eToT. In eDTM, the probability $P(e|d)$ can be computed by

$$P(e|d) \propto \sum_{z_t} \prod_{w \in d} P(w|z_t) P(e|z_t) P(z_t), \tag{22}$$

where $\{z_t\}$ is the set of topics learned in time span T, where $t \in T$.

In traditional emotion prediction methods, they only treat $P(e|d) = P(e|\mathbf{w})$, $(\mathbf{w} \in d)$, while the time information contained in documents is usually neglected. Therefore, compared with previous methods, our models can predict the emotions more accurately.

4.2.2 Emotion-based news recommendation

The existing news recommender systems usually recommend news according to the content similarity between news and user preferences [7,19], most of which neglect the impact of user emotions. However, users often have different emotion preferences during reading news. For example, some people may like news about funny stories that make them happy, while some people may like the shocked news for knowing the society. To this end, we first estimate the emotion preferences of a user u by calculating the probability $P(e|d_u)$, where d_u is the set of historical news read by u. Indeed, the probability $P(e|d_u)$ can be computed in the similar way of Equation 22. After that, the emotion preference $P(e|d_u)$ can be integrated into many state-of-the-art approaches [12,15] for news recommendation. For example, we can compute the user similarity by cosine similarity or KL divergence with the user preferences, and leverage user-based collaborative filtering approach to recommend news.

4.2.3 Emotion anomaly detection

Anomaly detection is an important task, and detecting the emotion anomaly has attracted lots of researchers' attentions [24,30]. Although some existing works can find the static anomaly of social emotions, few of them could detect the anomalous changes of social emotions with respect to the social events. To this end, with the eDTM model, we could directly capture the changes of social emotion during different time spans. To be specific, by calculating the cosine similarity or the KL divergence of emotion distributions, i.e., η in eDTM, between two neighboring time spans, we could measure the similarity of social emotions. After that, if the similarity is smaller (e.g., Cosine similarity) or larger (e.g., KL divergence) than a pre-defined threshold ϵ, we believe that there are some anomalies during the two time spans, e.g., some significant events happened.

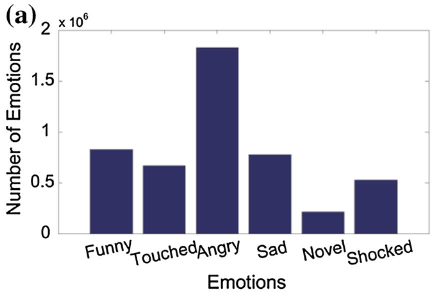
5 Experimental results

In this section, we evaluate our models, namely eToT, meToT, and eDTM, with extensive experiments on a real-world data set.

5.1 The experimental data

The experimental data were collected from the Society subsection of Sina News,[2] which is one of the biggest online news Web sites in China, from August 21, 2012, to November 11,

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



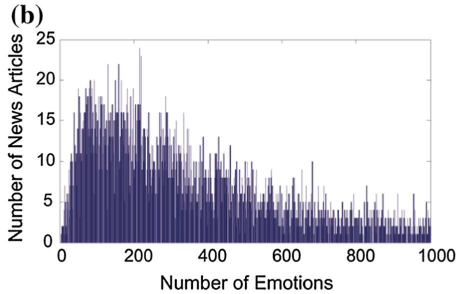


Fig. 5 The distribution of **a** the number of annotations on different emotions, and **b** the number of news articles with respect to different number of emotion annotations

2013. In this Web site, readers are allowed to choose one of the emotions, including "*Funny*," "*Touched*," "*Angry*," "*Sad*," "*Novel*," and "*Shocked*," to annotate the news after reading. In our data set, all the news articles were collected from the official top chart of each emotion every day, which guarantees there are enough emotion annotations for model training. Specifically, the data set contains 7504 news articles with 4,844,594 emotion annotations. In particular, we have made this data set publicly available for research purpose.[3] To guarantee the modeling performance, all stop words are removed. Note that, our models were trained with original Chinese news articles, and all experimental results were manually translated into English for facilitating demonstration.

Figure 5 shows some statistics of our data set. Specifically, Fig. 5a shows the distribution of the number of annotations on different emotions. We can observe that "*Angry*" is the most popular emotion, and other emotions have relatively even distribution, which may indicate people are more likely to show their emotions when they are angry. Moreover, Fig. 5b shows the distribution of the number of news articles with respect to different number of emotion annotations. We can find that most of the news articles have 50–600 annotations.

5.2 Performance evaluation of eToT

In this subsection, we show the overall performance of eToT. Specifically, the hyperparameters α and β were empirically set to $50/K$ and 0.01 according to [8], and the topic number K was set to 50 by the perplexity-based approach in [2]. The perplexity of eToT with different number of topics is shown in Fig. 6. Meanwhile, to guarantee the convergence of Gibbs sampling, all results were obtained after 500 iterations.

As we introduced above, eToT can discover the latent connections among time, emotions, and news topics. Therefore, we carefully studied six latent topics learned by eToT. Specifically, the distributions of different emotions in these topics, i.e., $P(e|z)$, are shown in Fig. 7. Moreover, the distributions with respect to different time spans in these topics, i.e., $P(t|z)$, are demonstrated in Fig. 8. From these results, we can observe that the topics #1, #2, #3, and #5 have similar distributions of emotions, where the generation probability of emotion "*Touched*" is the highest. However, these distributions over different time spans are totally different. For example, the distribution of topic #2 is relatively smooth. Therefore, we can argue that time has a significant impact on the topic generation of eToT. It also indicates the importance of time in the process of analyzing social emotions.

3 http://emotiondata.sinaapp.com/.

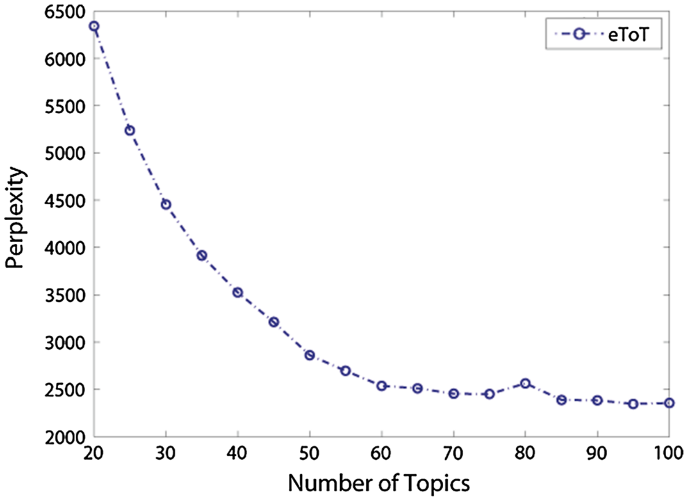


Fig. 6 The perplexity of eToT with different number of topics

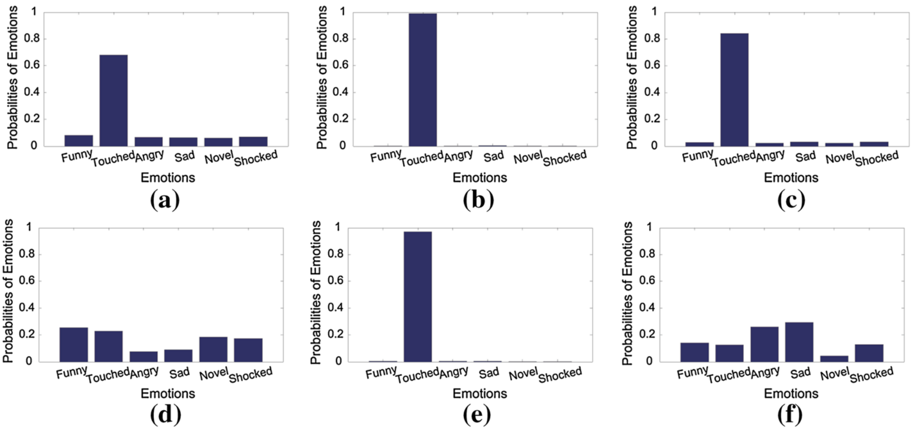


Fig. 7 The distributions of different emotions in six different news topics learned by eToT. **a** Topic #1, **b** topic #2, **c** topic #3, **d** topic #4, **e** topic #5, **f** topic #6

Furthermore, we inspect the topics in the view of words and social events. Specifically, Table 1 shows the top 10 ranked words in different learned topics. We can observe that there are some special words, which are associated with some social events closely, in topic #1, #3, and #5. For example, the "lottery-ticket" and "lost" in topic #1 can help us target a social event happened on April 2013, which is about *a woman returned a lottery with a huge bonus to its owner*. Similarly, in topic #3, "Xueying Hu" and "Suying" are names of two people who always help others. Therefore, the most important social emotion in both topic #1 and #3 is "*Touched*." Particularly, we find that words in topic #4 are insipid, and the corresponding distribution of emotions is even, which may indicate that this topic is about background.

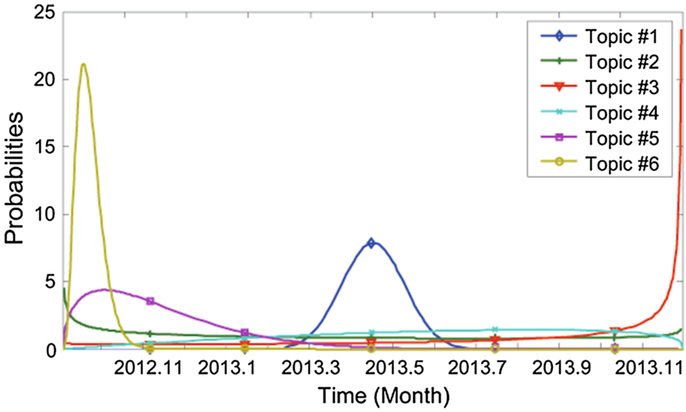


Fig. 8 The distributions of six topics learned by eToT with respect to different time spans

Table 1 The top 10 ranked words in six different topics learned by eToT

	Topic #1		Topic #2		Topic #3	
1	Lottery	0.00462	Shed	0.00797	Book	0.00768
2	Lost	0.00446	Fierce	0.00311	X. Hu[a]	0.00593
3	Inside	0.00413	Immensely	0.00305	Stall	0.00346
4	Immensely	0.00407	Money	0.00285	Rape	0.00345
5	Small	0.00382	The old	0.00278	Silence	0.00332
6	Hospital	0.00339	Leave	0.00264	Realm	0.00322
7	Yuan	0.00335	Vanish	0.00262	Rubbish	0.00286
8	Money	0.00331	Small	0.00257	Suying	0.00267
9	Make	0.00331	Vein	0.00247	Burglar	0.00267
10	Leave	0.00326	Lucky	0.00244	Heal	0.00257
	Topic #4		Topic #5		Topic #6	
1	Immensely	0.00393	H. Li[b]	0.00404	Go	0.00409
2	Able	0.00280	Z. Wang[c]	0.00321	Hospital	0.00396
3	Cat	0.00217	X. Tian[d]	0.00298	Intend	0.00327
4	Discover	0.00208	Leave	0.00288	Already	0.00302
5	Act	0.00206	Sudoku	0.00288	Yuan	0.00300
6	Friend	0.00205	Lonely	0.00247	Edit	0.00293
7	Nowadays	0.00202	Money	0.00245	Never	0.00291
8	Inside	0.00202	Sultan	0.00219	Make	0.00291
9	Dog	0.00197	Woman	0.00215	Immensely	0.00284
10	Leave	0.00187	Immensely	0.00214	Official	0.00270

[a] Xueying Hu, [b] Hunjun Li, [c] Zhanfu Wangm, [d] Xinbin Tian

5.3 Performance evaluation of meToT

In this subsection, we demonstrate the overall performance of meToT. Specifically, the hyper-parameters α and β were empirically set as same as them in ToT, and ρ was set to [10,10] as

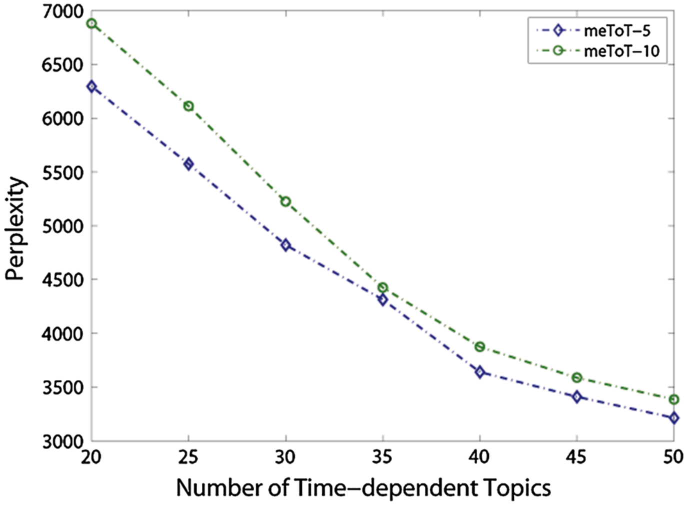


Fig. 9 The perplexity of meToT with different number of topics

Table 2 The top 10 ranked words in three different time-independent topics learned by meToT

	Topic #1		Topic #2		Topic #3	
1	Hospital	0.00314	Walk	0.00291	Police	0.00305
2	Discover	0.00277	Especially	0.00256	Hospital	0.00293
3	Yuan	0.00262	Man	0.00219	Discover	0.00291
4	Older	0.00236	Represent	0.00219	Man	0.00288
5	Current	0.00235	Daughter	0.00207	Situation	0.00256
6	Money	0.00221	Yuan	0.00205	Investigate	0.00247
7	Talk	0.00214	Small	0.001832	Currently	0.00231
8	Kindhearted	0.00203	Think	0.00175	Demand	0.00219
9	Work	0.00193	Title	0.00169	Court	0.00209
10	Women	0.00180	Student	0.001642	Introduce	0.00198

in [28]. The topic number K_1 and K_2 were set to 40 and 5 by the perplexity-based approach in [2], respectively. The perplexity of eToT with different number of topics is shown in Fig. 9, in which the "meToT-5/10" means that the number of time-independent topics is 5 and the number of time-dependent topics is 10. Meanwhile, all results were obtained after 500 iterations for guaranteeing the convergence of Gibbs sampling.

We randomly selected three time-dependent and time-independent topics, respectively. The top 10 ranked words of them are shown in Tables 2 and 3, and the emotion distributions are in Figs. 10 and 11. Figure 12 is the time distribution of time-dependent topics. We can find that, just as eToT, there are also some people names and addresses in the word list of time-dependent topics (i.e., Table 3). However, words in Table 2 are more common and time-independent. It indicates that meToT actually could distinguish time-dependent topics from time-independent ones, which is the assumption behind meToT. Furthermore, we inspect these time-independent topics in the views of emotions. In topic #1, the words appearing in

Table 3 The top 10 ranked words in three different time-dependent topics learned by meToT

	Topic #1		Topic #2		Topic #3	
1	Mother	0.00593	Great	0.00336	Netizen	0.00319
2	Aragonite	0.00497	C. Chen[b]	0.00299	Inside	0.00294
3	Older	0.00464	Inside	0.00299	Student	0.00267
4	Bother	0.00448	China	0.00286	Represent	0.00246
5	Q. Zou[a]	0.00336	Father	0.00274	Man	0.00243
6	Guilin	0.00304	Eat	0.00261	Yuan	0.00239
7	Talk	0.00304	Travel	0.00249	Small	0.00233
8	Student	0.00272	Work	0.00236	Tell	0.00221
9	Help	0.00224	Yuan	0.00236	Current	0.00217
10	Wish	0.00224	Watch	0.00224	Time	0.00215

[a] Qingdong Zou [b] Chaobo Chen

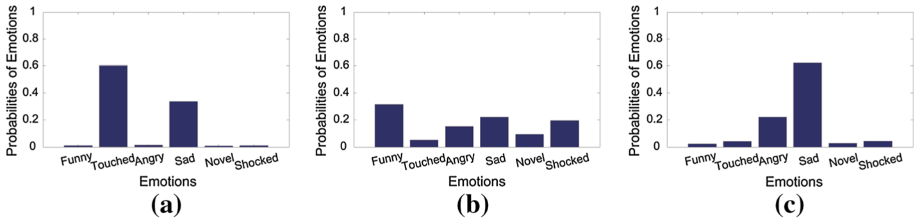


Fig. 10 The distributions of different emotions in three different time-independent topics learned by meToT. **a** Topic #1, **b** topic #2, **c** topic #3

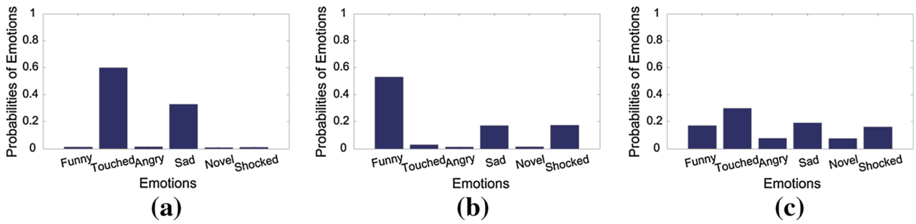


Fig. 11 The distributions of different emotions in three different time-dependent topics learned by meToT. Topic #1, **b** topic #2, **c** topic #3

top-ranked word list, such as "hospital," "older," and "kindhearted," probably indicate this topic is also about philanthropic act. Meanwhile, the emotion distribution of it has very high probability of "*Touched*." So we could treat these time-independent topics as background topics of emotion.

5.4 Performance evaluation of eDTM

In this subsection, we study the overall performance of eDTM. For training eDTM, we separated the data by month (one time span), and the parameter α was set to $50/K$ according to [8] firstly. With the perplexity-based approach in [2], the topic number K was set to 20. Figure 13 is the perplexity results of eDTM with different number of topics. In eDTM, we

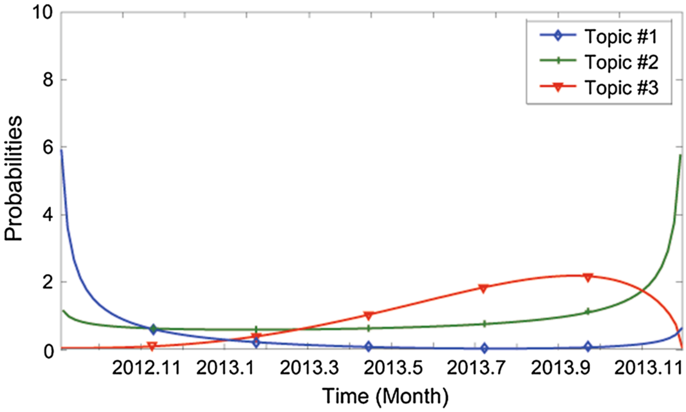


Fig. 12 The distributions of three time-dependent topics learned by meToT with respect to different time spans

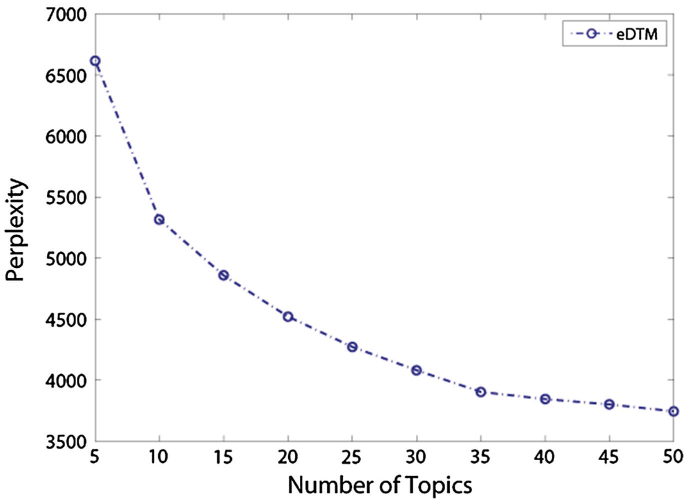


Fig. 13 The perplexity of eDTM with different number of topics

can treat the $\lambda_t \phi_{t-1}$ as the prior for ϕ_t. However, there is no prior for ϕ_1. For simplifying the process of parameter estimation, we empirically set $\phi_0 = 0.01$ and $\lambda_1 = 1$ in our experiments, and estimated the λ by maximizing the joint distribution. To guarantee convergence, we implemented 500 iterations in Gibbs sampling during model training.

Figures 14 and 15 show the distributions of social emotions in two randomly selected topics, which are learned by eDTM, with respect to different time spans. From these figures, we can observe that the social emotions consistently evolve with the evaluation of topics, which clearly validates the motivation of eDTM. Furthermore, Tables 4 and 5 show the top 10 ranked words in the two topics during different time spans. From these results, we find that the words in topics learned by eDTM are more common than those in topics learned by eToT. It may be because of the constraints of the state space model in eDTM; thus it cannot uncover special events effectively, while eDTM could unveil the evolution of words and emotions about special kinds of events directly. Particularly, we can observe that there

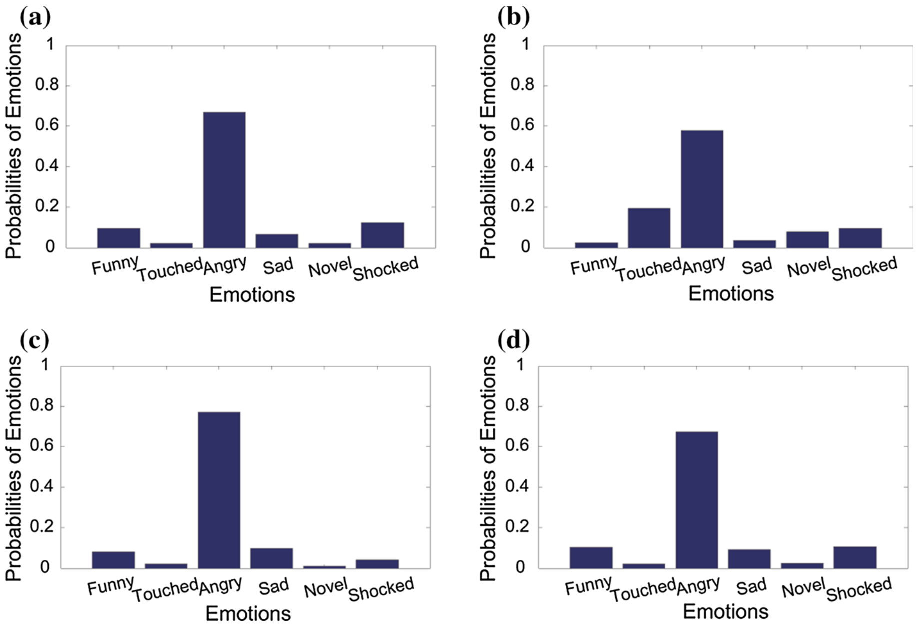


Fig. 14 The distribution of social emotions in topic #1 learned by eDTM with respect to different time spans. **a** December 2012. **b** January 2013. **c** February 2013. **d** March 2013

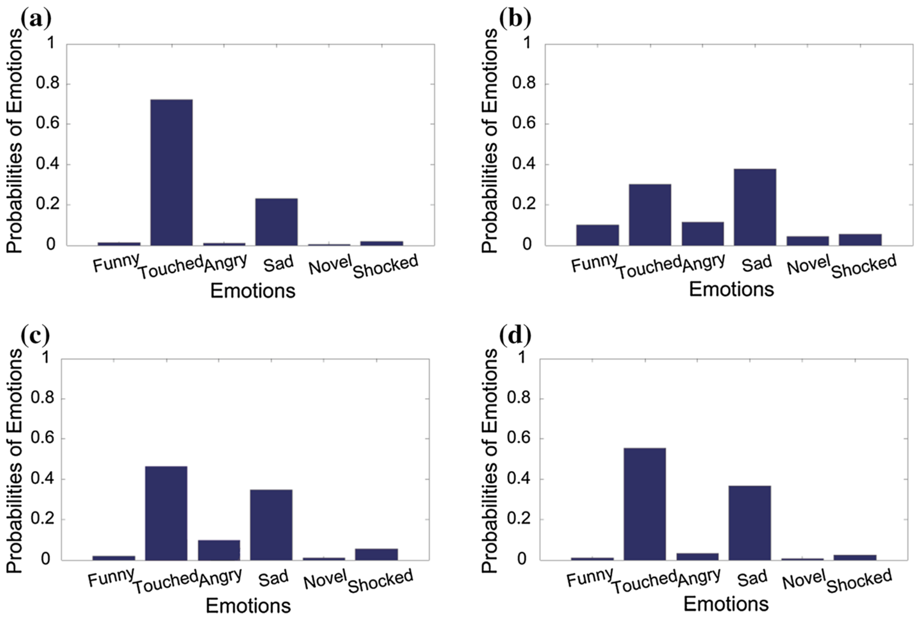


Fig. 15 The distribution of social emotions in topic #2 learned by eDTM with respect to different time spans. **a** March 2013. **b** April 2013. **c** May 2013. **d** June 2013

Table 4 The top 10 ranked words in topic #1 learned by eDTM during December 2012 to March 2013

	December 2012		January 2013		February 2013		March 2013	
1	**Thousand**	0.00392	**Yuan**	0.00324	**Yuan**	0.00374	**Yuan**	0.00386
2	**Yuan**	0.00363	**Thousand**	0.00314	Demonstrate	0.00321	**Thousand**	0.00342
3	**Police**	0.00343	**Called**	0.00314	**Man**	0.00319	**Called**	0.00316
4	Leave	0.00333	Judiciary	0.00308	**Called**	0.00309	Demonstrate	0.00282
5	Girl	0.00311	Police force	0.00301	Money	0.00282	Process	0.00269
6	Friend	0.00308	Demonstrate	0.00281	Friend	0.00274	**Man**	0.00267
7	**Called**	0.00293	Money	0.00277	Company	0.00272	Money	0.00266
8	Discover	0.00291	**Man**	0.00272	**Police**	0.00271	Discover	0.00264
9	Judiciary	0.00283	Leave	0.00265	**Thousand**	0.00267	Company	0.00261
10	**Man**	0.00272	**Police**	0.00257	Police force	0.00260	**Police**	0.00260

Table 5 The top-ranked 10 words in topic #2 learned by eDTM during March 2013 to June 2013

	April 2013		May 2013		June 2013		July 2013	
1	**Old man**	0.00571	**Leave**	0.00434	**Leave**	0.00433	**Son**	0.00461
2	**Son**	0.00471	**Son**	0.00424	**Son**	0.00423	**Leave**	0.00418
3	**Leave**	0.00431	**Old man**	0.00403	**Old man**	0.00402	**Old man**	0.00377
4	**Father**	0.00371	**Immensely**	0.00344	**Immensely**	0.00343	**Immensely**	0.00351
5	**Mother**	0.00335	**Father**	0.00333	**Father**	0.00332	Inside	0.00345
6	**Immensely**	0.00332	Inside	0.00322	job	0.00321	**Father**	0.00298
7	Inside	0.00321	**mother**	0.00304	**Mother**	0.00303	Mom	0.00294
8	Money	0.00320	Money	0.00297	Hospital	0.00296	Make	0.00287
9	Hospital	0.00303	Hospital	0.00292	Good	0.00292	Money	0.00270
10	Regularly	0.00282	Job	0.00289	Mom	0.00258	**Mother**	0.00268

are always words related to crime, such as "judiciary" and "police force," in topic #1 during different time spans. These results may imply that the topic is likely about crime. We can observe that "*Angry*" is always the most representative emotion in topic #1 during different time spans. These results may indicate that the social emotions evolve steadily with the evolution of the news topics.

Furthermore, we inspect another topic #2 learned by eDTM, of which the evolution is shown in Fig. 15 and Table 5. Although the emotion "*Touched*" always has the highest generation probability in this topic during different time spans, the probability of "*Sad*" and "*Angry*" increases abnormally after April 2013. To explain the reasons behind this evolution of social emotion, we manually checked some representative news of this topic from March 2013 to May 2013, which are shown in Table 6. Indeed, many representative news during this period are about good Samaritans; thus the social emotion has a trend of "*Touched*." Furthermore, a terrible earthquake occurred in Ya'an, China, on April 20, 2013, and many organizations began to call for donations. The news about the earthquake may result in the generation of emotion "*Sad*." Particularly, "Meimei Guo" in the first news on April 2013 is a key person related to the corruption of Red Cross China, which is the largest humanitarian organization in China. Although there was no evidence to prove that the corruption also

Table 6 The news about topic #2 between March 2013 and May 2013

News 1, March 2013
E-pal donated for a young ill girl voluntarily
A bus driver is loyal and devoted to the last
News 2, April 2013
The post office calls for donation for Ya'an with the slogan "here no MeiMei Guo"
A older man insisted on doing good for 14 years
News 3, May 2013
The girl hurt when combating the blaze passed away and the remaining donation has been return
The CEO of Alibaba are praised by people for his benefaction

existed in the disaster relief for Ya'an, the news about "Meimei Guo" may result in the social emotion "*Angry*."

5.5 Evaluation of model application

In this subsection, we evaluate the proposed models by emotion prediction, which is one of the potential applications introduced in Sect. 4. Besides our three novel models, we select one state-of-the-art model, namely *Emotion-Topic Model* (ETM), proposed in [3], and a classification model, *Maximum Entropy Model* (Maxent), as the baselines. Here we treat the emotion prediction as a multi-classification problem and use all of above models to calculate the posterior probability of each emotion given a news article. Specifically, each model can generate a ranked list of emotions by calculating the probability $P(e|d)$. Thus we can measure the ranking performance of each model by the popular metric Normalized Discounted Cumulative Gain (NDCG). Indeed, NDCG shows how well the rank order of a given ranked list returned by an approach with a cutoff rank N is. The larger value of NDCG indicates the better ranking performance. Specifically, the Discounted Cumulative Gain (DCG) of a ranked list for a given document can be calculated by $DCG@N = \sum_{i=1}^{N} \frac{2^{rel_i}-1}{log_2^{i+1}}$, where rel_i is the score of the ith emotion. Here we set $rel_i = C_i/C$, where C_i is the number of the ith emotion and C is the number of all of emotions. NDCG is the DCG normalized by IDCG, which is the DCG value of the ideal ranking list of the returned results and $NDCG@N = \frac{DCG@N}{IDCG@N}$.

In our experiments, we used the fivefold cross validation to evaluate each model. Table 7 illustrates the $NDCG@N$ performance of different models. From the results, we can observe

Table 7 The NDCG@N performance of different models

	eToT	meToT	eDTM month	eDTM season	ETM	Maxent
NDCG@1	0.75497	0.76057	0.58638	0.65808	0.56167	0.66334
NDCG@2	0.81362	0.82578	0.67006	0.74414	0.65794	0.74701
NDCG@3	0.83923	0.84182	0.73482	0.79291	0.73405	0.79331
NDCG@4	0.86471	0.86749	0.78007	0.82434	0.77139	0.82460
NDCG@5	0.88353	0.89279	0.81118	0.85236	0.80115	0.85188
NDCG@6	0.90073	0.91101	0.83085	0.87199	0.82229	0.86722

all of eToT, meToT, and eDTM (with different granularity for segmenting time spans) consistently outperform ETM with respect to different N, which clearly validates the importance of time for emotion modeling. Meanwhile, the outstanding performance of meToT implies the rationality of assumption behind it. However, the performance of eDTM is comparable with that of ETM. Moreover, it is interesting that Maxent has better prediction performance than eDTM, although it is still worse than eToT. By comparing the performance of eDTM with different granularity of time spans, we think it may be because the training data of eDTM are insufficient. To be more specific, since eDTM needs to separate training data into several time spans, the training data for each topic model are limited in our real-world data set. Moreover, since Maxent cannot capture the latent semantics between emotion and text, it has worse prediction performance than eToT and meToT. Besides, to further validate the experimental results, we conduct a series of t test of 0.95 confidence level which show that compared with other baselines, the improvements of our models on NDCG metrics are all statistically significant.

6 Conclusions

In this paper, we studied the problem of exploiting the user emotion annotations from online news to track the evolution of social emotions. We proposed three novel time-aware topic models, namely eToT, meToT, and eDTM, for building connections between news topics and social emotions. Specifically, in eToT, the news topics are associated with a beta distribution over time and a multinomial distribution over emotions. In meToT, we extend eToT by assuming that some topics are stable in terms of time. In eDTM, the state space model is leveraged for tracking the dynamics of news topics. Furthermore, we demonstrated some potential applications enabled by these two novel models, such as emotion prediction, emotion-based news recommendation and emotion anomaly detection. Finally, the extensive experiments on a real-world data set clearly demonstrate the effectiveness of our models.

In the future, we plan to evaluate the performance of our models with more potential applications, such as emotion-based news recommendation and social opinion monitoring.

Acknowledgments This research was partially supported by grants from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010), the National High Technology Research and Development Program of China (Grant No. 2014AA015203), the Fundamental Research Funds for the Central Universities of China (Grant No. WK2350000001), the Natural Science Foundation of China (Grant No. 61403358), NIH R21AA02395-01, and the Anhui Provincial Natural Science Foundation (Grant No. 1408085QF110).

References

1. Abrams D, Hogg MA (2006) Social identifications: a social psychology of intergroup relations and group processes. Routledge
2. Azzopardi L, Girolami M, van Risjbergen K, (2003) Investigating the relationship between language model perplexity and ir precision-recall measures. In: Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval. ACM, pp 369–370
3. Bao S, Zhang L, Yan R, Su Z, Han D, Yu Y (2009) Joint emotion-topic modeling for social affective text mining. In: ICDM'09, pp 699–704
4. Blei DM, Lafferty JD (2006) Dynamic topic models. In: Proceedings of the 23rd international conference on machine learning. ACM, pp 113–120
5. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

6. Chaumartin F-R (2007) Upar7: a knowledge-based system for headline sentiment tagging. In: Proceedings of the 4th international workshop on semantic evaluations. Association for Computational Linguistics, pp 422–425
7. Das AS, Datar M, Garg A, Rajaram S (2007) Google news personalization: scalable online collaborative filtering. In: Proceedings of the 16th international conference on World Wide Web. ACM, pp 271–280
8. Heinrich G (2005) Parameter estimation for text analysis. http://www.arbylon.net/publications/text-est.pdf
9. Iwata T, Watanabe S, Yamada T, Ueda N (2009a) Topic tracking model for analyzing consumer purchase behavior. In: IJCAI, vol 9, pp 1427–1432
10. Iwata T, Watanabe S, Yamada T, Ueda N (2009b) Topic tracking model for analyzing consumer purchase behavior. In: IJCAI, vol 9, pp 1427–1432
11. Iwata T, Yamada T, Sakurai Y, Ueda N (2012) Sequential modeling of topic dynamics with multiple timescales. ACM Trans Knowl Discov Data (TKDD) 5(4):19
12. Koren Y (2010) Collaborative filtering with temporal dynamics. Commun. ACM 53(4):89–97
13. Kozareva Z, Navarro B, Vázquez S, Montoyo A (2007) Ua-zbsa: a headline emotion classification through web information. In: Proceedings of the 4th international workshop on semantic evaluations. Association for Computational Linguistics, pp 334–337
14. Le Bon G (1897) The crowd: a study of the popular mind. Macmillan, London
15. Li B, Zhu X, Li R, Zhang C, Xue X, Wu X (2011) Cross-domain collaborative filtering over time. In: Proceedings of the twenty-second international joint conference on artificial intelligence-volume volume three. AAAI Press, pp 2293–2298
16. Lin C, He Y (2009) Joint sentiment/topic model for sentiment analysis. In: Proceedings of the 18th ACM conference on Information and knowledge management. ACM, pp 375–384
17. Lin KH-Y, Chen H-H (2008) Ranking reader emotions using pairwise loss minimization and emotional distribution regression. In: Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 136–144
18. Lin KH-Y, Yang C, Chen H-H (2007) What emotions do news articles trigger in their readers? In: Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval. ACM, pp 733–734
19. Liu J, Dolan P, Pedersen ER (2010) Personalized news recommendation based on click behavior. In: Proceedings of the 15th international conference on intelligent user interfaces. ACM, pp 31–40
20. Liu K-L, Li W-J, Guo M (2012) Emoticon smoothed language models for twitter sentiment analysis. In: AAAI
21. Liu Q, Ge Y, Li Z, Chen E, Xiong H (2011) Personalized travel package recommendation. In: Data mining (ICDM), 2011 IEEE 11th international conference on. IEEE, pp 407–416
22. Mei Q, Ling X, Wondra M, Su H, Zhai C (2007) Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th international conference on World Wide Web, ACM, pp 171–180
23. Minka T (2000) Estimating a dirichlet distribution, Technical report, MIT
24. Mishne G, de Rijke M (2006) Capturing global mood levels using blog posts. In: AAAI spring symposium: computational approaches to analyzing weblogs, pp 145–152
25. Mishne G, Glance N (2006) Leave a reply: an analysis of weblog comments. In: Third annual workshop on the Weblogging ecosystem. Edinburgh, Scotland
26. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on empirical methods in natural language processing, vol 10, pp 79–86
27. Paul M, Girju R (2009) Cross-cultural analysis of blogs and forums with mixed-collection topic models. In: Proceedings of the 2009 conference on empirical methods in natural language processing, vol 3. Association for Computational Linguistics, pp 1408–1417
28. Paul M, Girju R (2010) A two-dimensional topic-aspect model for discovering multi-faceted topics. Urbana 51:61801
29. Pinto et al JPGdS (2008) Detection methods for blog trends. Report of Dissertation Master in Informatics and Computing Engineering
30. Tang H, Tan S, Cheng X (2009) A survey on sentiment detection of reviews. Expert Syst Appl 36(7):10760–10773
31. Titov I, McDonald R (2008) A joint model of text and aspect ratings for sentiment summarization. Urbana 51:61801
32. Wang C, Blei D, Heckerman D (2012) Continuous time dynamic topic models. arXiv preprint. arXiv:1206.3298

33. Wang X, McCallum A (2006) Topics over time: a non-Markov continuous-time model of topical trends. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp 424–433
34. Wang X, McCallum A, Wei X (2007) Topical n-grams: phrase and topic discovery, with an application to information retrieval. In: Data mining, 2007. ICDM 2007. Seventh IEEE international conference on. IEEE, pp 697–702
35. Yang C, Lin KH-Y, Chen H-H (2007) Building emotion lexicon from weblog corpora. In: Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions. Association for Computational Linguistics, pp 133–136
36. Yano T, Smith NA (2010) What's worthy of comment? Content and comment volume in political blogs. In: ICWSM
37. Yu H, Hatzivassiloglou V (2003) Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: Proceedings of the 2003 conference on empirical methods in natural language processing, pp 129–136
38. Zhao J, Dong L, Wu J, Xu K (2012) Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1528–1531
39. Zhu H, Chen E, Xiong H, Cao H, Tian J (2014) Mobile app classification with enriched contextual information. IEEE Trans Mobile Comput 13(7):1550–1563
40. Zhu H, Xiong H, Ge Y, Chen E (2014) Mobile app recommendations with security and privacy awareness. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, KDD. ACM, New York, NY, USA, pp 951–960. doi:10.1145/2623330.2623705

Chen Zhu is an engineer in Baidu Inc. He received a master degree from University of Science and Technology of China in 2015 and a B.E. degree from Northwestern Polytechnical University in 2012. His research interests include social media analysis, sentimental analysis, and text mining. He has published a paper in ICDM.

Hengshu Zhu is a data scientist at Big Data Lab, Baidu Research, received the Ph.D. degree in 2014 and B.E. degree in 2009, both in Computer Science from USTC, China, was a visiting scholar at Rutgers, the State University of New Jersey, USA, from 2012 to 2013. His research interests include mobile data mining, business analytics, and computational social science. He has published prolifically in refereed journals and conference proceedings, including IEEE TKDE, IEEE TMC, IEEE TC, ACM TIST, ACM TKDD, SIGKDD, IJCAI, SDM, ICDM, and CIKM. He received the Best Student Paper Awards of KSEM-2011, WAIM-2013, and CCDM-2014. He received the Special Prize of President Scholarship for Postgraduate Students of CAS (2014), the National scholarship for Graduate Students of China (2013), and Scholarship Award for Excellent Doctoral Students of China (2012). He has served on program committees of many conferences including KDD, IJCAI, ICDM, and SDM. Also he has served as a reviewer for numerous journals, including IEEE TKDE, IEEE TC, ACM TIST, and KAIS.

Yong Ge received his Ph.D. in Information Technology from Rutgers, The State University of New Jersey in 2013, the M.S. degree in Signal and Information Processing from the University of Science and Technology of China (USTC) in 2008, and the B.E. degree in Information Engineering from Xi'an Jiao Tong University in 2005. He is currently an Assistant Professor at the University of North Carolina at Charlotte. His research interests include data mining and business analytics. He received the ICDM-2011 Best Research Paper Award, Excellence in Academic Research (one per school) at Rutgers Business School in 2013, and the Dissertation Fellowship at Rutgers University in 2012. He has published prolifically in refereed journals and conference proceedings, such as IEEE TKDE, ACM TOIS, ACM TKDD, ACM TIST, ACM SIGKDD, SIAM SDM, IEEE ICDM, and ACM RecSys. He has served as Program Committee members at the ACM SIGKDD 2013, the International Conference on Web-Age Information Management 2013, and IEEE ICDM 2013. Also he has served as a reviewer for numerous journals, including IEEE TKDE, ACM TIST, KAIS, Information Science, and TSMC-B.

Enhong Chen received the BS degree from Anhui University, master's degree from the Hefei University of Technology and the Ph.D. degree in computer science from USTC. He is currently a professor and the vice dean of the School of Computer Science, the vice director of the National Engineering Laboratory for Speech and Language Information Processing of USTC, winner of the National Science Fund for Distinguished Young Scholars of China. His research interests include data mining and machine learning, social network analysis and recommender systems. He has published lots of papers on refereed journals and conferences, including IEEE TKDE, TMC, KDD, ICDM, NIPS, and CIKM. He was on program committees of numerous conferences including KDD, ICDM, SDM. He received the Best Application Paper Award on KDD'08 and Best Research Paper Award on ICDM'11. He is a senior member of the IEEE.

Qi Liu is an Associate Researcher in University of Science and Technology of China (USTC). He received his Ph.D. in Computer Science from USTC. His general area of research is data mining and knowledge discovery. He has published prolifically in refereed journals and conference proceedings (10+ journal papers, and 20+ conference papers), e.g., TKDE, TOIS, TKDD, TIST, SIGKDD, IJCAI, ICDM, and CIKM. He has served regularly in the program committees of a number of conferences, and is a reviewer for the leading academic journals in his fields. He is a member of ACM and IEEE. Dr. Liu is the recipient of the IEEE ICDM'11 Best Research Paper Award, KSEM'2013 Best Paper Award, the Special Prize of President Scholarship for Postgraduate Students, Chinese Academy of Sciences (CAS) and the Distinguished Doctoral Dissertation Award of CAS.

Tong Xu is currently a Ph.D. student at the School of Computer Science, University of Science and Technology of China. He received a BS degree from Zhejiang University, China in 2010. His research interests include social network and social media analysis, recommender system, and mobile data mining. He has published several papers in refereed conference proceedings such as SDM, ICDM and PKDD.

Hui Xiong is currently a Full Professor and Vice Chair of the Management Science and Information Systems Department, and the Director of Rutgers Center for Information Assurance, at Rutgers University. He received the B.E. degree from the University of Science and Technology of China (USTC), China, the M.S. degree from the National University of Singapore (NUS), Singapore, and the Ph.D. degree from the University of Minnesota (UMN), USA. His general area of research is data and knowledge engineering with a focus on developing effective and efficient data analysis techniques for emerging data intensive applications. He is a co-Editor-in-Chief of Encyclopedia of GIS, an Associate Editor of IEEE Transactions on Data and Knowledge Engineering, IEEE Transactions on Big Data, the Knowledge and Information Systems journal, and ACM Transactions on Management Information Systems. He has served regularly on the organization and program committees of numerous conferences, including as a Program Co-Chair of the Industrial and Government Track for the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, a Program Co-Chair for ICDM-2013, and a General Co-Chair for ICDM-2015. He is an ACM Distinguished Scientist.