

# Social emotion classification of short text via topic-level maximum entropy model



Yanghui Rao<sup>a</sup>, Haoran Xie<sup>b,\*</sup>, Jun Li<sup>a</sup>, Fengmei Jin<sup>a</sup>, Fu Lee Wang<sup>c</sup>, Qing Li<sup>d</sup>

<sup>a</sup> School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

<sup>b</sup> Department of Mathematics and Information Technology, The Hong Kong Institute of Education, Hong Kong

<sup>c</sup> Caritas Institute of Higher Education, Hong Kong

<sup>d</sup> Department of Computer Science, City University of Hong Kong, Hong Kong

## ARTICLE INFO

### Article history:

Received 14 July 2015

Received in revised form 15 January 2016

Accepted 10 April 2016

Available online 6 May 2016

### Keywords:

Topic-level maximum entropy model

Social emotion classification

Short-text analysis

Public opinion mining

## ABSTRACT

With the rapid proliferation of Web 2.0, the identification of emotions embedded in user-contributed comments at the social web is both valuable and essential. By exploiting large volumes of sentimental text, we can extract user preferences to enhance sales, develop marketing strategies, and optimize supply chain for electronic commerce. Pieces of information in the social web are usually short, such as tweets, questions, instant messages, messages, and news headlines. Short text differs from normal text because of its sparse word co-occurrence patterns, which hampers efforts to apply social emotion classification models. Most existing methods focus on either exploiting the social emotions of individual words or the association of social emotions with latent topics learned from normal documents. In this paper, we propose a topic-level maximum entropy (TME) model for social emotion classification over short text. TME generates topic-level features by modeling latent topics, multiple emotion labels, and valence scored by numerous readers jointly. The overfitting problem in the maximum entropy principle is also alleviated by mapping the features to the concept space. An experiment on real-world short documents validates the effectiveness of TME on social emotion classification over sparse words.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

The big data era has descended on many communities including e-commerce, governments, and health organizations [2]. While there are increasingly numerous forms of multimedia content, such as images, photos, and videos, text is rapidly becoming a major part of enterprise data, especially for e-commerce firms [3]. Understanding text and extracting knowledge from it can be valuable to the business world, allowing for consolidation and promotion of customer opinions of a product, brand, or organization [4]. Social web is an important source of big data and is quite suitable for text mining [2]. With the prevalence of mobile Internet, more and more users can conveniently post messages to express their feelings through Twitter, Flickr, YouTube, and other social web apps. These messages provide a voice for customers to praise or criticize a particular product or service, which is useful for helping e-commerce firms understand the opinions of individuals, or to measure aggregated emotions of the public [5]. However,

these data are not easily accessible to computers because they are not only primarily unstructured and specifically produced for human consumption [4], but also rapidly expanded and constantly evolved [6]. To analyze and better understand such large volumes of big data for e-commerce, it is indispensable to identify sentiments and opinions of users on them. A typical application of sentiment analysis is the stock price prediction, which extracts sentiment features from news articles to predict the directions of stock price movement [7,8]. In light of these considerations and applications for e-commerce, sentiment analysis is concerned with the automatic and accurate emotion classification of an opinion holder towards either a topic or the overall tone of a document [9], which can facilitate comparison shopping, product design, marketing strategies development, and supply chain optimization [10].

Sentiment analysis has been extensively studied for product and movie reviews, which probably form a small part of the social web [11]. The other abundant sentimental documents include messages concerning politics, news and sports, and daily discussions in social network sites. Several factors hinder the application of traditional sentiment analysis methods in the above documents. First, unlike standard reviews with many words for

\* The corresponding author.

E-mail address: [hrxie2@gmail.com](mailto:hrxie2@gmail.com) (H. Xie).

most samples, the document length in the social web is quite diverse [11]. For instance, many documents posted in Twitter, Tumblr and other micro-blogging services only consist of a few phrases or 1–2 sentences. Second, traditional sentiment analysis research focused mainly on classifying the attitudes of authors/writers who created the reviews. Typically, this is achieved by using reviews where the author rating was expressed either with stars or with some numerical value to train machine learning algorithms [12]. In the social web, however, the emotion labels of documents (e.g., news, blog posts, and tweets) are mainly provided by readers using social annotation services such as online crowdsourcing [13]. For this reason, the aggregation of emotional responses from the reader is termed *social emotion* [14,15]. Traditional sentiment analysis methods from the perspective of authors may not be applicable to readers because authors and readers do not always share the same emotion for the same text [16,17]. Thus, we aim to classify social emotions from the reader's perspective over short documents.

Work has been performed to exploit the social emotion of individual words because words play a central role in how we describe and understand emotions [18]. The SWAT algorithm [19] was one of the top-performing lexicon-based algorithms on the SemEval “affective text” task [20], which aimed to annotate news headlines according to the emotions they evoked in readers. In the SWAT algorithm, a word-emotion mapping lexicon was first constructed, in which, the emotions of each word was scored as the average of emotions of every headline contained this word. The emotion-term algorithm was created by improving the naïve Bayes classifier [14]. The algorithm differs from the traditional naïve Bayes classifier, by taking into account emotional ratings when calculating the probability of a category and the probability of a word given an emotion label. The limitation, however, of such word-level algorithms is that the words used to express certain emotions can be quite different across different contexts [21]. This difference is observed because words that have sentiment ambiguity and multiple emotions are difficult for word-level emotion lexicons to recognize [22].

As a concept-level approach [23], the topic model has emerged to measure the social emotions evoked by the online text. The emotion-topic model [14], the affective topic model [24], the multi-label supervised topic model, and the sentiment latent topic model [25] were built by introducing an additional layer of emotion modeling into latent Dirichlet allocation (LDA). Experimental results have shown that those topic-level models outperform word-level algorithms as well as support vector machines and several other methods in the classification of social emotions over normal documents. With the emergence of social media, short documents are prevalent on the Web. The existing social emotion classification methods based on conventional topic models, such as LDA, typically used statistical techniques to learn the topic proportions of each document and the topic-word distributions. Thus, those topic-level models may suffer from the severe data sparsity problem (i.e., the sparse word co-occurrence patterns in individual document), and their performance is limited on social emotion classification over short documents [26]. An effective solution to the data sparseness problem is to employ the maximum entropy (ME) method [12], whose flexible modeling capability has alleviated data sparseness more successfully than the other probabilistic models. However, the standard word-level ME method suffers from undesirable overfitting, especially for sparse words because the equality constraints on feature expectations are uncertain. Although inequality constraints [27] can be used to alleviate such problems, it introduces extra cost to determine the optimal constraint width, and again, the same word may evoke different social emotions in different topics.

Motivated by the above observation, this paper is concerned with the following research questions:

- Given a set of short documents with social emotion labels, how can one predict social emotions for the new short documents in an effective and stable way?
- How can one alleviate the overfitting problem caused by the ME method and the sparse problem caused by conventional topic models in a unified social emotion classification model for short documents?

To address the abovementioned research questions, we propose a topic-level maximum entropy (TME) model for social emotion classification over short documents. We bring the insight that the problem of overfitting when using the ME method on sparse words can be alleviated by latent topics. To mitigate the data sparsity problem of conventional topic models, and the overfitting problem of the standard ME method, we learn latent topics over short documents by modeling the word co-occurrence patterns explicitly, in addition to using the aggregated word co-occurrence patterns in both the training and testing set to smooth the equality constraints on feature expectations. A real-world corpus that contained 11,813 short documents, which were emotionally annotated by readers, was employed to evaluate the effectiveness of our model. Experimental results show that the proposed model is robust against overfitting on sparse words. The rest of this paper is organized as follows. A summary of related work on sentiment analysis, social emotion classification, and short-text topic modeling is given in Section 2. Our model and experiment analysis are presented in Section 3 and Section 4, respectively. Finally, the paper's conclusions are presented in Section 5.

## 2. Related work

In this section, we briefly summarize the related work from the following three perspectives: sentiment analysis, social emotion classification on normal documents, and topic modeling over short text.

### 2.1. Sentiment analysis

The first line of work on sentiment analysis applied many pre-developed sentiment lexicons, for example, Subjectivity Wordlist [28], WordNet-Affect [29], and SentiWordNet [30] to classify documents by emotions. The Subjectivity Wordlist is built by a manually selected seed set of subjective words, a small raw corpus, and an online dictionary, which is a subjectivity lexicon distinguishing subjective versus objective words. The WordNet-Affect is a linguistic resource in which the synsets representing emotional concepts are labeled. Synset is the synonym set that represents a sense or a concept in the WordNet lexicon. As many as 2874 synsets and 4787 words are annotated in the WordNet-Affect. The SentiWordNet is a lexical resource developed for supporting sentiment classification. This resource scores each synset in the WordNet along three emotional dimensions: positivity, negativity, and neutrality. The second method adopted either supervised [12] or unsupervised [31] learning algorithm for sentiment classification. However, these studies focus on extracting emotions from the perspective of writers primarily.

### 2.2. Social emotion classification

Social emotion classification annotates reader emotions towards certain contexts. The first line of research into social emotion classification is the “affective text” in SemEval-2007 tasks [19]. To annotate social emotions of unlabeled news headlines, the

proposed algorithms assumed that all words, even neutral ones, can effectively convey positive or negative emotions of readers. However, due to the limited information of news headlines or sentences, it is usually difficult to annotate the emotions consistently, even for a human [22]. Thus, the emotion-term [14] and the word-emotion [26] algorithms were proposed to make use of all words in the body of a news document. These algorithms were effective for associating different words with social emotions accurately, but they struggled to distinguish the sentiment ambiguity of the same word.

In order to recognize different emotional senses of the same word, topic models such as LDA were extended for social emotion classification, in which, a topic acted as an important component of an emotion, and informative and coherent topics were extracted and grouped under different emotions. In the emotion-topic model [14], a distribution over reader ratings is first generated from a multinomial distribution for each document. Then, for each word in the document, a single emotion label is sampled according to the above distribution. Finally, a latent topic is generated from a Dirichlet distribution conditioned to the emotion label, and a word is generated from the latent topic, which is modeled by another multinomial distribution over words. To have discriminative power between affective and background topics, the multi-label supervised topic model and the sentiment latent topic model [25] were proposed by representing the set of reader ratings as a bag of emotion labels. Another way to model all words, topics, and social emotions jointly is using the exponential distribution to generate reader ratings over each emotion label [24]. The limitation, however, of such topic-level models is that they were designed for normal text that contain sufficient words in an individual document.

### 2.3. Short-text topic modeling

Short text is prevalent on the Web, especially with the emergence of social network apps. The main feature of short text is the sparsity of content. Because most words only occur once in each short text, it is difficult to conduct topic modeling and other tasks accurately. To address this issue, a common solution is to aggregate short documents into long pseudo-documents before training a standard topic model, that is, to expand short text and use external knowledge for short-text topic modeling. Phan et al. [32] trained a topic model over massive external data collected from Wikipedia or MEDLINE, and then predicted the topic distribution of short text. Jin et al. [33] proposed a Dual-LDA model that learns topics over both the original short text and related long documents. External linguistic knowledge could alleviate the problem of sparsity and enhance the topic learning of short text; however, it usually leads to topic bias, and the effectiveness of such methods is highly dependent on the accuracy of the external information. Therefore, according to the empirical comparison among topic learning methods over current data of Twitter [34], it is necessary to design a specialized and universal topic model for short text. To address these issues, Cheng et al. [35] proposed a generative biterm topic model (BTM) to learn topics over short documents by directly modeling the generation of biterms (i.e., unordered word pairs co-occurring in a short context) in the whole corpus. Unfortunately, these models were designed to reveal the main topics in a collection, not directly applicable for social emotion classification in short text.

## 3. TME model

In this section, we propose the TME model for social emotion classification over short text. The extraction of topics from short text is first defined, and the framework of TME is subsequently

presented in detail. An ME model using words as features, that is, word-level maximum entropy (WME), is also briefly described for comparison. Finally, we describe the estimation and prediction of parameters for TME.

### 3.1. Topic extraction

The aim of topic extraction in the TME model is to identify the latent topics from short documents to tackle the data sparsity problem, which always occurs in word-level features. For example, two short documents “BMW X1 is a good SUV” and “The oil cost of Benz cars is really expensive” do not have any overlapping features (terms) in the word level, while they share the same topic “automobile” in topic level.

Without loss of generality, topics are represented as groups of correlated words in topic models. Conventional topic models, such as LDA, learn topics based on document-level word co-occurrence patterns, whose effectiveness will be highly influenced in short-text scenario where the word co-occurrence patterns become quite sparse in each document [32]. Mixture of unigrams [36] was designed for topic extraction over short text, which assumed that all the words in a document share a same topic. In other words, it models the whole corpus, rather than a document in LDA, as a mixture of topics. By leveraging the information of the whole corpus, it alleviates the sparsity problem in topic extraction of short text. However, the constraint that a document has a single topic is too strict to model fine topics in documents. Recently, the BTM was proposed to extract topics by breaking each document into biterms and learning a global topic distribution for short documents [35]. In the BTM, a biterm is constructed by any two distinct words co-occurring in a document.

BTM is an effective topic model over short documents. The generative process of BTM is as follows:

1. For the whole corpus, draw a topic distribution  $\theta \sim \text{Dir}(\alpha)$
2. For the  $k$ -th topic, draw a word distribution  $\phi_k \sim \text{Dir}(\beta)$
3. For each biterm  $(w_{i,1}, w_{i,2})$ 
  - (a) draw a topic assignment  $z_i \sim \text{Multi}(\theta)$
4. draw two words  $w_{i,1}, w_{i,2} \sim \text{Multi}(\phi_{z_i})$

In the above, the notation of “biterm” denotes an unordered word pair co-occurring in a short text, that is, a small, fixed-size window over a word sequence. In short documents with limited length, each document is often used as an individual context unit. For instance, given a window size equal to three, a document with four words will generate five biterms:  $(w_1, w_2, w_3, w_4) \rightarrow \{(w_1, w_2), (w_1, w_3), (w_2, w_3), (w_2, w_4), (w_3, w_4)\}$ . A larger size of the window will generate more biterms but weaken the probability of sharing the same topic for the two words. The biterm can also be extended to  $N$ -term (i.e., an unordered group of  $N$  words co-occurring in the same document), while the performance decreases gradually as  $N$  grows [35]. In particular, the model is equivalent to a mixture of unigrams when  $N$  is equal to or larger than the maximum length of documents in the corpus, which has a strict constraint of all the words in a document sharing the same topic.

BTM is appropriate for topic extraction of short text but is not designed for social emotion classification. In other words, BTM is an unsupervised topic model of learning topics from word co-occurrence patterns (i.e., biterms) in the whole corpus, while social emotion classification requires us to model words and multiple types of emotions voted by various readers jointly. A straightforward way to address the problem is to extend BTM by adding an extra distribution to generate reader ratings over each emotion, or using a two-layer generation process through representing the set of reader ratings as a bag of emotion labels. The challenge of these

solutions, however, is that a large window size is required to generate enough biterns for each short document. Because a larger size of the window will weaken the probability of sharing the same topic for the two words, we cannot rely on tuning the window size to achieve good performance on social emotion classification of short text.

### 3.2. Model description

In this section, we first define the issue of social emotion classification over short documents and the related notations, and then present our model that named TME .

A short-text collection consists of  $D$  documents  $\{d_1, d_2, \dots, d_D\}$  with word tokens and multiple emotion labels. We represent the set of all word tokens by  $\{w_1, w_2, \dots, w_V\}$ , where  $V$  is the number of unique word tokens. The  $k$ -th emotion label is represented as  $e_k \in [1, E]$ , where  $E$  is the amount of predefined emotion labels such as positive and negative. The numbers of training set with words and emotions jointly, and testing set with words only are  $D_{tr}$  and  $D_{te}$ , respectively. We formulate the task of social emotion classification over short text as learning an accurate classifier from the training set with sparse word co-occurrence patterns in individual documents to predict the evoked social emotions of unlabeled documents in the testing set. Following the convention of BTM, we denote the number of biterns and topics as  $N_B$  and  $K$ , respectively. The topic indicator of the  $i$ -th bitern  $(w_{i,1}, w_{i,2})$  is represented as  $z_i \in [1, K]$ . The prevalence of topics in the corpus is denoted by a multinomial distribution  $\theta$ . The word distribution for topics is represented as a  $K \times V$  matrix  $\phi$  where the  $k$ -th row is a  $V$ -dimensional multinomial distribution. For the multiple emotion labels and reader ratings, a matrix  $\mathbf{R}$  is used to represent the reader votes over each emotion label for all short documents in the training set. In the  $D \times M$  matrix  $\mathbf{R}$ , the element  $r_{jk}$  denotes the ratings over the  $k$ -th emotion among all users who have read the  $j$ -th short text. The reader ratings were normalized and summed to one for each short document.

To alleviate data sparseness, we associate word tokens and the generated topics with social emotions based on the ME principle. The ME principle states that from all the probability distributions, we should select the distribution that satisfies all prior conditions and constraints [37]. Different from the standard ME models using uni-grams or bi-grams as features, we generate features by the co-occurrences of  $w_i$  and  $e_k$ , in addition to  $z_i$  and  $e_k$ , that is,  $(z_i, e_k)$ . The set of all topic–emotion pairs is represented by  $\Phi$ . Table 1 summarizes the notations of frequently used variables.

The graphical representation of TME is shown in Fig. 1, in which shaded nodes are observed data, blank ones are latent (not observed), and arrows indicate dependence. The set of all emotions and features is denoted by  $\mathbf{e}$  and  $\mathbf{f}$ , respectively.

In our model, the prior conditions and constraints are the co-occurrences of word tokens, topic indicators, and emotion labels in the training set. The detail of associating word tokens with emotion labels can be found in Ref. [1], and we focus here on the

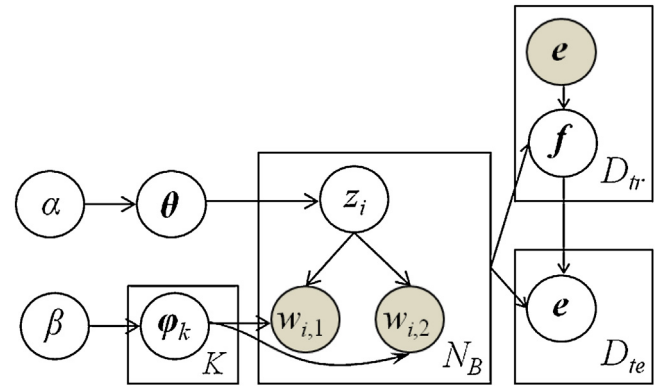


Fig. 1. Graphical representation of TME.

additional modeling of topics and emotions. Table 2 shows the samples of a training set with topic indicators, where  $D=3$ ,  $K=4$ ,  $E=3$ , and  $F=12$ .

To associate multiple emotion labels annotated by users with topics generated by unsupervised topic models in short training documents, we use the following feature function based on the Bernoulli model:

$$f_n(z_i, e_k) = \begin{cases} 1 & z_i \in [1, K], e_k \in [1, E], (z_i, e_k) \in \Phi \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where a binary indicator is assigned for each topic–emotion pair  $(z_i, e_k)$ . For example, the second and the third columns of Table 2 present the topic indicators and reader ratings over three emotion labels of each training document, respectively. The first topic indicator in  $d_1$  is associated with each of the three emotion labels  $e_1, e_2$ , and  $e_3$ , which generates three features, that is,  $f_1(1, e_1), f_2(1, e_2)$ , and  $f_3(1, e_3)$ . Among various feature representation methods, the Bernoulli model works best for modeling short documents [38]. Thus, the value of all features listed in the third column of Table 2 is one.

We then measure the strength of each topic–emotion pair by aggregating the reader ratings of each training document over  $e_k$ , as follows:

$$\bar{p}(z_i, e_k) \propto \left( \sum_{d_j \in T_i} r_{jk} \right) / |\Phi|, \quad (2)$$

where  $\bar{p}(z_i, e_k)$  is the empirical probability distribution of topic–emotion pair  $(z_i, e_k)$ ,  $\Phi$  is the set of all topic–emotion pairs in the training set, and  $T_i$  is the collection of training documents that contain  $z_i$ . In the above samples, we have  $|\Phi|=18$ , as shown in the last column of Table 2.

Given a set of features, the moment constraint is often used in the ME models [37]. It requires that the moment of the features as observed from the training set, that is,  $\bar{E}(f_n)$ , should be the same as that predicted from the model, that is,  $E(f_n)$ . The expected value of each feature  $f_n(z_i, e_k)$  with respect to the empirical distribution in the training set can be estimated by

$$\bar{E}(f_n) = \sum_{z_i, e_k} \bar{p}(z_i, e_k) f_n(z_i, e_k). \quad (3)$$

The expected value of  $f_n(z_i, e_k)$  with respect to the probability of emotion label  $e_k$  conditioned to topic indicator  $z_i$ , that is,  $p(e_k|z_i)$ , is derived as follows:

$$E(f_n) = \sum_{z_i, e_k} \bar{p}(z_i) p(e_k|z_i) f_n(z_i, e_k), \quad (4)$$

where  $\bar{p}(z_i)$  is the empirical distribution of  $z_i$  in the training set. Thus, the moment constraint of the TME is as follows:

$$E(f_n) = \bar{E}(f_n). \quad (5)$$

Table 1  
Notations of frequently used variables.

Symbol	Description
$D$	Number of short documents
$D_{tr}$	Number of training set
$D_{te}$	Number of testing set
$V$	Number of unique word tokens
$E$	Number of predefined emotion labels
$K$	Number of topics
$F$	Number of unique features
$\Phi$	Set of all topic–emotion pairs
$\mathbf{R}$	$[r_{jk}]$ : $D \times E$ matrix of normalized ratings for each emotion

**Table 2**  
Samples of a training set with topic indicators.

Short text	Topic indicators	Emotions and ratings	Features
$d_1$	{1, 2}	$e_1: r_{11}, e_2: r_{12}, e_3: r_{13}$	$f_1(1, e_1), f_2(1, e_2), f_3(1, e_3),$ $f_4(2, e_1), f_5(2, e_2), f_6(2, e_3).$
$d_2$	{3, 4}	$e_1: r_{21}, e_2: r_{22}, e_3: r_{23}$	$f_7(3, e_1), f_8(3, e_2), f_9(3, e_3),$ $f_{10}(4, e_1), f_{11}(4, e_2), f_{12}(4, e_3).$
$d_3$	{2, 3}	$e_1: r_{31}, e_2: r_{32}, e_3: r_{33}$	$f_4(2, e_1), f_5(2, e_2), f_6(2, e_3),$ $f_7(3, e_1), f_8(3, e_2), f_9(3, e_3).$

According to Eq. (3) and Eq. (4), we have

$$\sum_{z_i, e_k} \bar{p}(z_i) p(e_k|z_i) f_n(z_i, e_k) = \sum_{z_i, e_k} \bar{p}(z_i, e_k) f_n(z_i, e_k). \quad (6)$$

A mathematical measure of the uniformity of the conditional distribution  $p(e_k|z_i)$  is provided by the conditional entropy:

$$H(P) = -\sum_{z_i, e_k} \bar{p}(z_i) p(e_k|z_i) \log p(e_k|z_i). \quad (7)$$

Then, the model is formulated as the following optimization problem:

$$\begin{aligned} \min_{p(e|z)} & -H(P) = \sum_{z_i, e_k} \bar{p}(z_i) p(e_k|z_i) \log p(e_k|z_i) \\ \text{s.t.} & E(f_n) - \bar{E}(f_n) = 0 \quad 1 \leq n \leq F \\ & \sum_{e_k} p(e_k|z) - 1 = 0 \quad \text{for all } z. \end{aligned} \quad (8)$$

To estimate the values of  $p(e|z)$  that minimize  $H(P)$ , we resolve the above primal optimization problem to an unconstrained dual optimization problem by introducing the Lagrange parameters  $\lambda$ . The values of  $p(e|z)$  are estimated as follows:

$$p_\lambda(e|z) = \frac{1}{S_\lambda(z)} \exp\left(\sum_{n=1}^F \lambda_n f_n(z, e)\right), \quad (9)$$

$$S_\lambda(z) = \sum_e \exp\left(\sum_{n=1}^F \lambda_n f_n(z, e)\right). \quad (10)$$

Next, we develop an iterative algorithm to achieve optimal values of  $\lambda$ , that is, the weight of connections between topics generated by unsupervised topic models and multiple emotion labels annotated by users, with the purpose being to predict social emotions for the new unlabeled short documents.

### 3.3. Parameter estimation

To estimate the topic indicator of each biterm  $z$  and the optimal values of each parameter  $\lambda_n$  for TME, an iterative algorithm is proposed, as shown in Algorithm 1. The number of biterms assigned to topic  $z$  is denoted by  $c_z$ , and the number of times a word  $w$  is assigned to topic  $z$  is represented as  $c_w^z$ . The symbol  $-i$  means that the  $i$ -th biterm is excluded from the number. The model can also be simplified to a WME model when generating features  $f(w, e)$  for each word  $w$  in the 8-th step.

**Algorithm 1** Iterative algorithm for TME

---

**Input:**  
Topic number  $K$ , hyperparameters  $\alpha$  and  $\beta$ , window size  $s$

**Output:**  
Optimal values of each parameter  $\lambda_n$ , i.e., the weight of  $f_n(z, e)$

1. Extract a sequence of  $N$  words via a single scan over the documents
2. For  $b = 1$  to  $N$
3. For  $e = b + 1$  to  $b + s - 1$
4. Generate a biterm  $(w_{i,b}, w_{i,e})$  and  $i = i + 1$
5. Set  $c_z, c_{w_{i,1}}^z$ , and  $c_{w_{i,2}}^z$  to 0
6. Randomly initialize the topic assignments for all the biterms 7. For each biterm  $(w_{i,1}, w_{i,2})$ , repeat until convergence

Draw topic  $z$  from  $(c_{-i,z} + \alpha) \frac{(c_{w_{i,1}}^z + \beta)(c_{w_{i,2}}^z + \beta)}{(c_{-i} + V\beta + 1)(c_{-i}^z + V\beta)}$

(Continued)

**Algorithm 1** Iterative algorithm for TME

---

Update  $c_z, c_{w_{i,1}}^z$ , and  $c_{w_{i,2}}^z$

8. For each topic indicator  $z$  and emotion label  $e$  Generate the feature  $f_n(z, e)$  according to Eq. (1)
9. Set  $\lambda_n^{(0)}$  to 0
10. For each  $f_n$ , repeat until convergence

$$\lambda_n^{(t+1)} = \lambda_n^{(t)} + \frac{\log(\bar{E}(f_n)/E^{(t)}(f_n))}{M}$$

where  $(t)$  is the iteration index and the constant  $M$  is the length of the short text that contains the maximum number of topic indicators.

---

After estimating the optimal values of parameter  $\lambda$ , predicting the emotion label of unlabeled short text  $d$  is straightforward. We can first extract the topics of  $d$  using the same method of topic extraction used for the training set because the process is totally unsupervised. Next, the topic indicators of  $d$  and each emotion label can be used to generate features according to Eq. (1). For example, given the topic indicators of 1 and 4 for  $d$ , the emotion label of  $d$  can be predicted as follows:

$$p_\lambda(e_1|d) = \exp(\lambda_{1f_1} + \lambda_{10f_{10}}) / S_\lambda(d), \quad (11)$$

$$p_\lambda(e_2|d) = \exp(\lambda_{2f_2} + \lambda_{11f_{11}}) / S_\lambda(d), \quad (12)$$

$$p_\lambda(e_3|d) = \exp(\lambda_{3f_3} + \lambda_{12f_{12}}) / S_\lambda(d), \quad (13)$$

where

$$S_\lambda(d) = \exp(\lambda_{1f_1} + \lambda_{10f_{10}}) + \exp(\lambda_{2f_2} + \lambda_{11f_{11}}) + \exp(\lambda_{3f_3} + \lambda_{12f_{12}}). \quad (14)$$

The difference between TME and WME on the prediction of unlabeled short text is the scope of feature utilized. For word-level models such as WME, we cannot generate features from the words that appear in the unlabeled documents but do not occur in the training set. Thus, as number of iterations increases, the prediction of social emotions will be more likely to overfit the sentiment orientation of words that only appear in the training set. The above problem is alleviated in TME by mapping the original words in both the training set and unlabeled documents to a unified concept space because the topics can be extracted without any emotion label.

## 4. Experiments

In this section, we evaluate the performance of the proposed model for social emotion classification over short text. We designed the experiments to achieve the following two goals: (i) to analyze the influence of the number of iterations on the accuracy of TME and WME and (ii) to conduct comparative analysis with various baselines.

#### 4.1. Dataset

To test the adaptiveness, effectiveness, and robustness of our model on social emotion classification over short text, a real-world dataset was employed in the experiment.<sup>1</sup> The dataset contains 11,813 short documents including BBC Forum posts (BBC), Digg.com posts (Digg), MySpace comments (MySpace), Runners World forum posts (Runners World), Twitter posts (Twitter), and YouTube comments (YouTube). Each document was manually labeled by readers—who were allowed to use their own judgments rather than being trained to annotate in a predefined way—with the positive and negative sentiment strengths. The positive sentiment strength ranges from 1 (not positive) to 5 (extremely positive), and the negative sentiment strength ranges from –1 (not negative) to –5 (extremely negative). Table 3 summarizes the statistics of the dataset, where the second column presents the mean words of each document for the category.

We randomly selected 50% of short documents as the training set and used the remainder as the testing set.

#### 4.2. Experimental design

In order to evaluate the performance of the proposed TME model, as well as the ME model using words as features (WME), the following baselines were implemented for comparison:

1. SWAT algorithm, which was also designed for short text originally [19]. This algorithm scored the emotions of word  $w$  as the averaged emotional scores of every headline that contains  $w$ .
2. Emotion-term (ET) algorithm and emotion-topic model (ETM) [14]. The ET algorithm was proposed to improve the traditional naïve Bayes classifier, and the ETM was developed by introducing an additional topic layer into ET.
3. Multi-label supervised topic model (MSTM) and sentiment latent topic model (SLTM) [25]. The MSTM and SLTM designed a two-layer topic model to associate emotions with documents.
4. Affective topic model (ATM) [24]. The ATM used the exponential distribution to generate reader ratings over each emotion label.

To make an appropriate comparison between methods, all parameters in the topic-level models (i.e., TME, ETM, MSTM, SLTM and ATM) were set as follows: (i) the hyperparameters  $\alpha = 50/K$  and  $\beta = 0.01$ . The values of them were tuned via grid search, which always works well on the short-text collection [35]. (ii) The iteration number of Gibbs sampling is 300. Here we randomly select 50% of training documents as the validation set to determine the above value. (iii) The number of topics  $K$  ranges from 100 to 300, and the validation set as mentioned earlier is used to choose the optimal  $K$  value. Although the value of  $K$  can be also determined by estimating the probability of words conditioned to topics [39], it has high time complexity.

In terms of the biterm extraction, we set the window size  $s$  to 3 for TME, because a larger size of the window will weaken the probability of sharing the same topic for the two words [35]. This is consistent with the characteristics of our experimental corpus. There are 815 (6.9%) documents with less than four words, a model with  $s > 3$  will “force” two words in many different documents to be the same topic. For instance, given  $d_1 = \{w_1, w_2, w_3\}$  and  $d_2 = \{w_4\}$ , it will generate six biterns when  $s > 3$ :  $(w_1, w_2, w_3, w_4) \rightarrow \{(w_1, w_2), (w_1, w_3), (w_1, w_4), (w_2, w_3), (w_2, w_4), (w_3, w_4)\}$ . The words from different documents, that is,  $w_1$  and  $w_4$ , occurred in  $d_1$  and  $d_2$  are

**Table 3**  
Statistics of the dataset.

Category	Mean words	Documents
BBC	64.76	1000
Digg	33.63	1077
MySpace	19.76	1041
Runners World	64.25	1046
Twitter	16.81	4242
YouTube	17.38	3407
All six combined	27.00	11,813

combined to be a bitern, but they are unlikely to share the same topic. The value of  $s$  cannot be too small either, for example, when  $s = 2$ , the above sequence of four words will generate three biterns:  $(w_1, w_2, w_3, w_4) \rightarrow \{(w_1, w_2), (w_2, w_3), (w_3, w_4)\}$ . In that case, although the two words within two windows have a large probability of sharing the same topic over the whole corpus, the number of biterns is even less than the original length of the documents.

#### 4.3. Evaluation metrics

We employed three evaluation metrics as indicators of performance: accuracy at top 1 (Acc@1), mean of average precision (MAP), and averaged Pearson’s correlation (AP). Acc@1 is essentially the micro-averaged F1 measure that weights precision and recall equally [38]. This metric was also used to evaluate the performance of various social emotion classification baselines [14,24,25]. Given an unlabeled document  $d$ , the top-ranked predicted emotion  $e_p$ , and the truth emotion set  $E_{top@d}$ , which includes the top-ranked emotions, Acc <sub>$d$</sub> @1 is calculated as

$$\text{Acc}_{d@1} = \begin{cases} 1 & \text{if } e_p \in E_{top@d} \\ 0 & \text{else.} \end{cases} \quad (15)$$

The emotion distribution of  $d$  is predicted correctly if  $e_p \in E_{top@d}$ . Then, Acc@1 is estimated by averaging Acc <sub>$d$</sub> @1 of all testing documents. MAP is used to measure the quality of a ranked list when there are only two coarse-grained levels: relevant and irrelevant [40]. AP is conducted using the averaged Pearson measure of correlation between the gold standard scores and the predicted scores over all emotion labels [19]. The higher the values of Acc@1, MAP, and AP, the better the model performance.

#### 4.4. Influence of the number of iterations

To evaluate the influence of iteration number, we varied the number of iterations from 1 to 300. Fig. 2 presents the Acc@1 of WME and TME over all datasets (i.e., the combined collections from BBC, Digg, MySpace, Runners World, Twitter, and YouTube) when using different numbers of iterations.

According to Fig. 2, our first observation is that WME performed best in the first 50 iterations, while its performance was gradually decreased as the number of iterations increased. This is consistent with the ME principle that is employed by WME. Because the equality constraint was used for our models, as shown in Eq. (5), the weight of features (i.e., the co-occurrence of words and emotions in WME) overfitted to the training set with larger numbers of iterations. However, the drop rate of the performance was relatively slow over all datasets. The reason is that the number of generated features in all datasets is sufficient (the mean of words is 27). We also found that TME converged to its asymptote in approximately 100 iterations for 100, 200, and 300 topic numbers. This is because the topic indicator of each bitern can capture the same emotional orientation of different words. For example, in the category of electronic product reviews, the word “compact” is often used to express a positive sentiment. On the other hand, in

<sup>1</sup> The dataset is available in public at: <http://sentistrength.wlv.ac.uk/documentation/>.

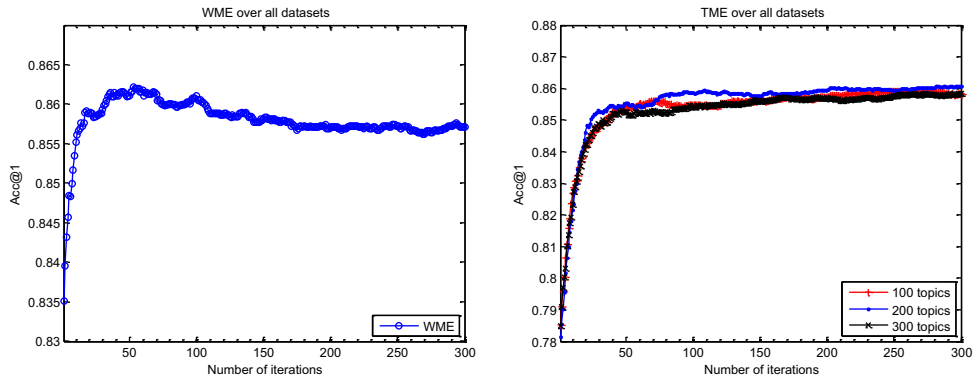


Fig. 2. Performance of WME and TME over all datasets.

the books category the word “exciting” expresses a primarily positive sentiment. The proposed TME is useful for modeling these words by combining them to biterms and mapping each biterm to the same topic indicator.

Fig. 3 illustrates the Acc@1 of WME and TME on Twitter subset. The results illustrate that the performance of WME was obviously decreased as the number of iterations increased. This is because the overfitting problem in WME becomes severe when the number of generated features is insufficient in the dataset (the mean of words is 16.81 in the Twitter subset). However, we can still find that TME has relatively flat curves for 100, 200, and 300 topic numbers over Twitter subset. The results in Fig. 2 and Fig. 3 indicate that the proposed TME can alleviate the problem of overfitting in ME and classify the social emotions accurately.

4.5. Comparison with baselines

In this section, we measure and compare the performance of different models on social emotion classification over all datasets comprehensively, as presented in Table 4.

The results indicate that both TME and WME outperformed others in terms of AP and yielded performance that is competitive with various baselines in terms of Acc@1 and MAP. First, the existing topic-level models, that is, ETM, ATM, SLTM, and MSTM performed poorly on this combined corpus from six categories, because they focused on associating emotions with topics specific to one context primarily. The generated biterm in TME acted as the “bridge” between different contexts, thus enhancing the social emotion classification performance in real-world environments.

Table 4

Performance of different models over all datasets.

Models	Acc@1 (%)	MAP	AP
TME	86.06	0.87	0.46
WME	85.67	0.86	0.48
ETM	84.66	0.86	0.01
ET	85.72	0.87	0.00
SWAT	81.88	0.86	0.16
ATM	62.62	0.64	0.23
SLTM	74.38	0.85	0.11
MSTM	75.63	0.86	0.17

Second, although ET and SWAT performed well in terms of MAP, that is, the quality of a ranked list when there are only two coarse-grained levels, their performance was not satisfactory in terms of the fine-grained evaluation metric AP. This is because they used the limited words as features only, while both TME and WME generated features by exploiting reader ratings over multiple emotion labels. Third, TME performed well in terms of Acc@1 and MAP, while slightly worse than WME in terms of AP. The reason is that we used a validation set to choose the optimal number of iterations for both TME and WME, which may introduce differences on the performance over the testing set. It follows that a better method of determining the optimal iteration number deserves further research.

Table 5 presents the performance of different models on Twitter subset. We found that most models performed well in terms of the coarse-grained evaluation metric MAP, especially for SWAT and ATM. However, most existing models (e.g., ETM, ET, SWAT, SLTM,

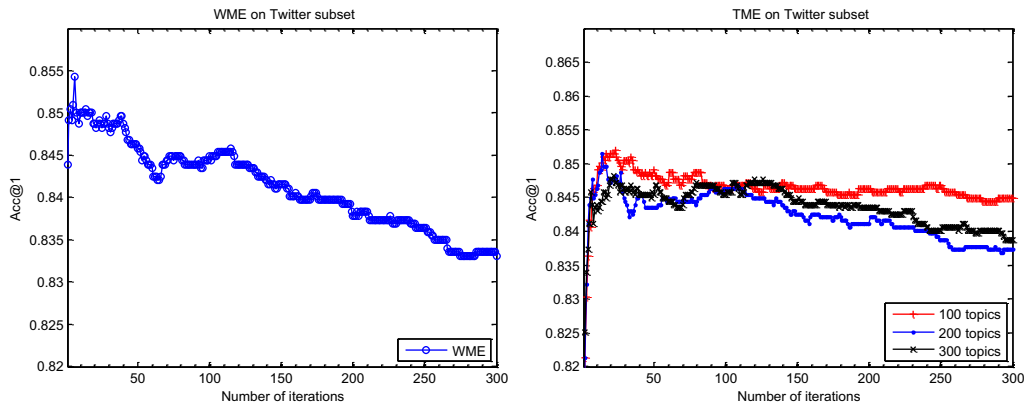


Fig. 3. Performance of WME and TME on Twitter subset.

**Table 5**

Performance of different models on a Twitter subset.

Models	Acc@1 (%)	MAP	AP
TME	85.01	0.84	0.31
WME	84.91	0.85	0.31
ETM	82.27	0.81	0.01
ET	85.20	0.83	0.00
SWAT	80.06	0.90	0.06
ATM	79.63	0.90	0.17
SLTM	78.88	0.86	0.05
MSTM	78.83	0.86	0.02

and MSTM) performed poorly in terms of AP. The experimental result shows that short documents lack enough content from which statistical conclusions can be drawn, thus the performance of these models was quite unstable. We exploited the reader ratings over multiple emotion labels to enrich features for both TME and WME, thus achieving better performance.

## 5. Conclusions

As increasing numbers of users share their experiences, ideas, and opinions on the Web [41,42], sentiment analysis has become a popular topic for those who wish to understand public opinion from online data [43]. For instance, electronic commerce websites such as Amazon ([www.amazon.com](http://www.amazon.com)) and Epinions ([www.epinions.com](http://www.epinions.com)) allow users to write reviews on products, which have generated large volumes of data for sentiment analysis. Measuring the opinions of the general public regarding company strategies, marketing campaigns, product preferences, and social events has also steadily attracted interest in business intelligence areas [44]. Although sentiment analysis has attracted numerous investigations, little research has been performed on social web data from the reader's perspective compared to product and movie reviews written by writers. Automatically classifying reader emotions can help us analyze the opinions/emotions embedded in user-contributed comments in the social web and further extract user preferences using big data analytics to enhance competitiveness for electronic commerce.

In this paper, we proposed a TME model to classify the social emotions evoked in readers. The topic indicators generated by unsupervised topic models were combined with word tokens to alleviate the overfitting problem in the ME principle. We conducted experiments to evaluate the effectiveness of the proposed models on real-world short documents. The results indicated that the performance of our approach is competitive when compared to various baselines. For future research, we plan to (i) explore the part-of-speech information for feature generation, (ii) integrate recent learning models [45–47] to enhance the efficiency of the classification, and (iii) extend our approach to other applications, such as the emotionally aware recommendation of events in social media enhanced systems [48,49] and multimedia retrieval systems by integrating with low-level features [50–52].

## Acknowledgments

The research described in this paper was substantially supported by the National Natural Science Foundation of China (No. 61502545), “the Fundamental Research Funds for the Central Universities” (No. 46000-31610009), and a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS11/E06/14). This work has also been supported, in part, by a Strategic Research Grant (Project no. 7004218) and an Applied Research Grant (Project no. 9667095), both of City

University of Hong Kong. This paper is an extended version of our previous conference paper [1].

## References

- [1] Y.H. Rao, J. Li, X.Y. Xiang, H.R. Xie, Intensive maximum entropy model for sentiment classification of short text, in Proceedings of the 20th International Conference on Database Systems for Advanced Applications (DASFAA) Workshops, (2015) pp. 42–51.
- [2] H. Chen, R.H.L. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, *MIS Q.* 36 (4) (2012) 1165–1188.
- [3] E.-P. Lim, H. Chen, G. Chen, Business intelligence and analytics: research directions, *ACM Trans. Manage. Inf. Syst.* 3 (4) (2013) 17–27.
- [4] E. Cambria, B. Schuller, B. Liu, H.X. Wang, C. Havasi, Knowledge-based approaches to concept-level sentiment analysis, *IEEE Intell. Syst.* 28 (2) (2013) 12–14.
- [5] X. Hu, L. Tang, J. Tang, H. Liu, Exploiting social relations for sentiment analysis in microblogging, Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM) (2013) 537–546.
- [6] X. Wu, X. Zhu, G.Q. Wu, W. Ding, Data mining with big data, *IEEE Trans. Knowl. Data Eng.* 26 (1) (2014) 97–107.
- [7] X. Li, H. Xie, L. Chen, J. Wang, X. Deng, News impact on stock price return via sentiment analysis, *Knowl.-Based Syst.* 69 (2014) 14–23.
- [8] X. Li, H. Xie, Y. Song, Q. Li, S. Zhu, F. Wang, Does summarization help stock prediction? News impact analysis via summarization, *IEEE Intell. Syst.* 30 (3) (2015) 26–34.
- [9] A. Gangemi, V. Presutti, D.R. Recupero, Frame-based detection of opinion holders and topics: a model and a tool, *IEEE Comput. Intell. Mag.* 9 (1) (2014) 20–30.
- [10] R.Y.K. Lau, C. Li, S.S.Y. Liao, Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis, *Decis. Support Syst.* 65 (2014) 80–94.
- [11] M. Thelwall, K. Buckley, G. Paltoglou, Sentiment strength detection for the social web, *J. Am. Soc. Inf. Sci. Technol.* 63 (1) (2012) 163–173.
- [12] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (2002) 79–86.
- [13] Y. Wang, A. Pal, Detecting emotions in social media: a constrained optimization approach, Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI) (2015) 996–1002.
- [14] S.H. Bao, S.L. Xu, L. Zhang, R. Yan, Z. Su, D.Y. Han, Y. Yu, Mining social emotions from affective text, *IEEE Trans. Knowl. Data Eng.* 24 (9) (2012) 1658–1670.
- [15] Y.H. Rao, Contextual sentiment topic model for adaptive social emotion classification, *IEEE Intell. Syst.* (2015), doi:http://dx.doi.org/10.1109/mis.2015.91.
- [16] Y.-J. Tang, H.-H. Chen, Mining sentiment words from microblogs for predicting writer-reader emotion transition, Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC) (2012) 1226–1229.
- [17] K. Lin, H. Chen, Ranking reader emotions using pairwise loss minimization and emotional distribution regression, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (2008) 136–144.
- [18] A. Kazemzadeh, S. Lee, S. Narayanan, Fuzzy logic models for the meaning of emotion words, *IEEE Comput. Intell. Mag.* 8 (2) (2013) 34–49.
- [19] C. Strapparave, R. Mihalcea, Semeval-2007 task 14: affective text, Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval) (2007) 70–74.
- [20] R. Snow, B.O. Connor, D. Jurafsky, A.Y. Ng, Cheap and fast-but is it good? Evaluation non-expert annotations for natural language tasks, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (2008) 254–263.
- [21] D. Bollegala, D. Weir, J. Carroll, Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL) (2011) 132–141.
- [22] C. Quan, F. Ren, An exploration of features for recognizing word emotion, Proceedings of the 23rd International Conference on Computational Linguistics (Coling) (2010) 922–930.
- [23] R.Y.K. Lau, Y. Xia, Y. Ye, A probabilistic generative model for mining cybercriminal networks from online social media, *IEEE Comput. Intell. Mag.* 9 (1) (2014) 31–43.
- [24] Y.H. Rao, Q. Li, L. Wenyin, Q.Y. Wu, X.J. Quan, Affective topic model for social emotion detection, *Neural Netw.* 58 (2014) 29–37.
- [25] Y.H. Rao, Q. Li, X.D. Mao, L. Wenyin, Sentiment topic models for social emotion mining, *Inf. Sci.* 266 (2014) 90–100.
- [26] Y.H. Rao, J.S. Lei, L. Wenyin, Q. Li, M.L. Chen, Building emotional dictionary for sentiment analysis of online news, *World Wide Web* 17 (2014) 723–742.
- [27] J. Kazama, J. Tsujii, Maximum entropy models with inequality constraints: a case study on text categorization, *Mach. Learn.* 60 (2005) 159–194.
- [28] C. Banea, R. Mihalcea, J. Wiebe, A bootstrapping method for building subjectivity lexicons for languages with scarce resources, Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC) (2008) 2764–2767.
- [29] C. Strapparava, A. Valitutti, Wordnet-affect: an affective extension of wordnet, Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC) (2004) 1083–1086.



- [30] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, *Proceedings of the 7th Conference on Language Resources and Evaluation (LREC) (2010)* 2200–2204.
- [31] P.D. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) (2002)* 417–424.
- [32] X.H. Phan, L.M. Nguyen, S. Horiguchi, Learning to classify short and sparse text & web with hidden topics from large-scale data collections, *Proceedings of the 17th International World Wide Web Conference (WWW) (2008)* 91–100.
- [33] O. Jin, N.N. Liu, K. Zhao, Y. Yu, Q. Yang, Transferring topical knowledge from auxiliary long texts for short text clustering, *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM) (2011)* 775–784.
- [34] L. Hong, B.D. Davison, Empirical study of topic modeling in Twitter, *Proceedings of the 1st Workshop on Social Media Analytics (SOMA) (2010)* 80–88.
- [35] X. Cheng, X. Yan, Y. Lan, J. Guo, BTM: Topic modeling over short texts, *IEEE Trans. Knowl. Data Eng.* 26 (12) (2014) 2928–2941.
- [36] K. Nigam, A. McCallum, S. Thrun, T. Mitchell, Text classification from labeled and unlabeled documents using EM, *Mach. Learn.* 39 (2) (2000) 103–134.
- [37] D. Yu, L. Deng, A. Acero, Using continuous features in the maximum entropy model, *Pattern Recognit. Lett.* 30 (2009) 1295–1300.
- [38] C.D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, University Press, Cambridge, 2008.
- [39] T.L. Griffiths, M. Steyvers, Finding scientific topics, *Proc. Natl. Acad. Sci. U. S. A.* 101 (Suppl. 1) (2004) 5228–5235.
- [40] Q. Wang, O. Wu, W. Hu, J. Yang, W. Li, Ranking social emotions by learning listwise preference, *Proceedings of the 1st Asian Conference on Pattern Recognition (ACPR) (2011)* 164–168.
- [41] H. Xie, Q. Li, X. Mao, X. Li, Y. Cai, Q. Zheng, Mining latent user community for tag-based and content-based search in social media, *Comput. J.* 57 (9) (2014) 1415–1430.
- [42] T. Ma, J. Zhou, M. Tang, Y. Tian, A. Al-Dhelaan, M. Al-Rodhaan, S. Lee, Social network and tag sources based augmenting collaborative recommender system, *IEICE Trans. Inf. Syst.* E98-D (4) (2015) 902–910.
- [43] A. Tsai, C. Wu, R. Tsai, J. Hsu, Building a concept-level sentiment dictionary based on commonsense knowledge, *IEEE Intell. Syst.* 28 (2) (2013) 22–30.
- [44] J.L. Zhao, S. Fan, D. Hu, Business challenges and research directions of management analytics in the big data era, *J. Manage. Anal.* 1 (3) (2014) 169–174.
- [45] B. Gu, V.S. Sheng, K.Y. Tay, W. Romano, S. Li, Incremental support vector learning for ordinal regression, *IEEE Trans. Neural Netw. Learn. Syst.* 26 (7) (2015) 1403–1416.
- [46] B. Gu, V.S. Sheng, Z.J. Wang, D. Ho, S. Osman, S. Li, Incremental learning for v-Support Vector Regression, *Neural Netw.* 67 (2015) 140–150.
- [47] X.Z. Wen, L. Shao, Y. Xue, W. Fang, A rapid learning algorithm for vehicle classification, *Inf. Sci.* 295 (1) (2015) 395–406.
- [48] H. Xie, X. Li, T. Wang, R.Y.K. Lau, T.-L. Wong, L. Chen, F.L. Wang, Q. Li, Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy, *Inf. Process. Manage.* 52 (1) (2016) 61–72.
- [49] H. Xie, D. Zou, R.Y.K. Lau, F.L. Wang, T.-L. Wong, Generating incidental word-learning tasks via topic-based and load-based profiles, *IEEE Multimedia* 23 (1) (2016) 60–70.
- [50] Z.Q. Pan, Y. Zhang, S. Kwong, Efficient motion and disparity estimation optimization for low complexity multiview video coding, *IEEE Trans. Broadcast.* 61 (2) (2015) 166–176.
- [51] Y.H. Zheng, B. Jeon, D.H. Xu, Q.M. Jonathan Wu, H. Zhang, Image segmentation by generalized hierarchical fuzzy C-means algorithm, *Jm Intell. Fuzzy Syst.* 28 (2) (2015) 961–973.
- [52] Z.H. Xia, X.H. Wang, X.M. Sun, Q.S. Liu, N.X. Xiong, Steganalysis of LSB matching using differences between nonadjacent pixels, *Multimedia Tools Appl.* 75 (4) (2016) 1947–1962.

**Yanghai Rao** is an assistant professor at the Sun Yat-sen University. He received his Ph.D. degree from the City University of Hong Kong in 2014, and his master's degree from the Graduate University of the Chinese Academy of Science in 2010. His research interests include social emotion classification, natural language processing, topic modeling, and event detection. Among these research areas, he has authored more than 20 papers for international journals and conferences. He has also served as the co-chair at the 1st International Workshop on User Modeling for Web-based Learning (IWUM) in conjunction with ICWL'15, and the PC member at the 2nd International Workshop on Semantic Computing and Personalization (SeCoP) in conjunction with DASFAA'15.

**Haoran Xie** is an assistant professor at The Hong Kong Institute of Education. He received his PhD and MSc from the City University of Hong Kong, and BEng from the Beijing University of Technology. His research interests include big data, social media, recommender systems, human computer interaction, and e-learning systems. He has published more than 60 research papers in many prestigious journals and conferences. He also served as a guest editor of *Neurocomputing* and *International Journal of Distance Education Technologies*, a co-chair/committee member of SETE, SECOP, IWUM, ICWL, U-MEDIA, WISE, and HKWS.

**Jun Li** is currently pursuing his M.E. degree at the Sun Yat-sen University. His research interests include social emotion detection, natural language processing, and short-text classification.

**Fengmei Jin** is a student at the Sun Yat-sen University. Her research interests include sentiment analysis, natural language processing, and topic modeling.

**Fu Lee Wang** is a professor and Vice-President (Research and Advancement) in Caritas Institute of Higher Education. He received BEng and MPhil from The University of Hong Kong, MSc from the Hong Kong University of Science and Technology, MBA from Imperial College London, and PhD from The Chinese University of Hong Kong. His research interests include electronic business, information retrieval, information systems, and e-learning. Before joining Caritas, he was a faculty member at the City University of Hong Kong. He is Past Chair of ACM Hong Kong Chapter and Chair of IEEE Hong Kong Section Computer Society Chapter. He served as program/conference chair of a number of international conferences.

**Qing Li** is a professor at the City University of Hong Kong. His research interests include object modeling, multimedia databases, social media, and recommender systems. He is a Fellow of IET, a senior member of IEEE, a member of ACM SIGMOD and IEEE Technical Committee on Data Engineering. He is the chairperson of the Hong Kong Web Society, and is a steering committee member of DASFAA, ICWL, and WISE Society.