Latent Discriminative Models for Social Emotion Detection with Emotional Dependency

XIAOJUN QUAN, Institute for Infocomm Research, A*STAR
QIFAN WANG, Purdue University YING
ZHANG, Nankai University
LUO SI, Alibaba Group Inc / Purdue University
LIU WENYIN, City University of Hong Kong

Sentimental analysis of such opinionated online texts as reviews and comments has received increasingly close attention, yet most of the work is intended to deal with the detection of authors' emotion. In contrast, this article presents our study of the social emotion detection problem, whose objective is to identify the evoked emotions of readers by online documents such as news articles. A novel latent discriminative model (LDM) is proposed for this task. LDM works by introducing intermediate hidden variables to model the latent structure of input text corpora. To achieve this, it defines a joint distribution over emotions and latent variables, conditioned on the observed text documents. Moreover, we assume that social emotions are not independent but correlated with one another, and the dependency of them is capable of providing additional guidance to LDM in the training process. The inclusion of this emotional dependency into LDM gives rise to a new emotional dependency based latent discriminative model (eLDM). We evaluate the proposed models through a series of empirical evaluations on two real-world corpora of news articles. Experimental results verify the effectiveness of LDM and eLDM in social emotion detection.

Categories and Subject Descriptors: H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing–*Linguistic processing*; H.2.8 [**Database Management**]: Database Applications–*Data mining*

General Terms: Algorithms, Experimentation

Additional Key Words and Phrases: Discriminative model, opinion mining and sentiment analysis, social emotion detection

ACM Reference Format:
Xiaojun Quan, Qifan Wang, Ying Zhang, Luo Si, and Liu Wenyin, 2015. Latent Discriminative Models for Social Emotion Detection with Emotional Dependency. *ACM Trans. Inf. Syst.* xxx, xxx, Article xxx (March 2015), 19 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

The opinion of the public is a reflection of the aggregate of individual attitudes and judgments. Public opinion and the influences have been long studied, with most of

---

This work is partially supported by National Science Foundation (NSF) research grants IIS-0746830, CNS1012208 and IIS-1017837 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-0939370.
Author's addresses: X. Quan, Institute for Infocomm Research, A*STAR, Singapore 138632; email: quanx@i2r.a-star.edu.sg; Q. Wang: Department of Computer Science, Purdue University; email: qifan@purdue.edu; Y. Zhang: College of Software, Nankai University; email: yingzhang@nankai.edu.cn; L. Si: Department of Computer Science, Purdue University; email: lsi@purdue.edu; L. Wenyin: Multimedia Software Engineering Research Center, City University of Hong Kong; email: liuwenyin@gmail.com.

DOI:http://dx.doi.org/10.1145/0000000.0000000

ACM Transactions on Information Systems, Vol. xxx, No. xxx, Article xxx, Publication date: March 2015.

early attention from the political science research to study the effects of public opinion on policy and other aspects (e.g., [Page and Shapiro 1983]). The advent of internet and its prosperity have created similar interests in other fields aiming to automatic sentiment analysis of online user-generated content such as product reviews [Pang et al. 2002; Hu and Liu 2004a; 2004b]. Such user-generated content has displayed growingly significant impact on today's world. For example, potential consumers often refer to product reviews for their purchasing decision-making. Besides, analyzing these reviews has also provided the potential to create new business opportunities. Among many examples from various domains, one is to use social media data to evaluate asset pricing in financial markets via sentiment analysis .

Traditional sentiment analysis research focuses mainly on determining the sentiments of authors who created the subjective documents. Typically, this is achieved by analyzing the sentiment within each review and then aggregating them into an overall sentiment. Instead, this study is aimed to identify readers' emotions evoked by these online documents. This task derives much of its inspiration from the observation that readers' emotions are not necessarily consistent with writers' [Alm et al. 2005; Yang et al. 2007; Strapparava and Mihalcea 2008]. While writers' emotions can be literally derived from documents, readers' emotions cannot but have to be triggered by reading the documents. In addition, readers' emotions are based not only on individual judgments but also on interactions with others, which creates collectively shared opinions by shaping individual emotions into social trends [Nofsinger 2005]. For this reason, the task is termed social emotion detection [Bao et al. 2012]. Social emotion detection is also valuable in that it provides an effective means to gain insight into public opinion. For example, social emotion detection can help enterprises gain the attitudes of the public to certain exposed incidents so that they take right remedial measures.

There have been some research efforts to study social emotion detection by application of various machine learning models [Lin et al. 2008; Bao et al. 2012]. A typical scenario of the problem is that a group of emotion tags are provided after each news article and readers are allowed to choose one to express their emotions after reading the article. An example from a real website is presented in Figure 1 for illustration. One simple solution to this problem works by making use of an emotional dictionary [Rao et al. 2012; Lei et al. 2014], which can be learned from training corpora of news articles. Another solution treats social emotion detection as a classification task, in which a set of emotion-related features are identified first from training corpora. Then, a classification model is trained using the features. The model is then used to classify new articles into emotional categories [Lin et al. 2007]. More sophisticated solutions would attempt to model the integrated generation process of news articles and social emotions using topic model [Bao et al. 2012]. An intermediate layer is typically introduced into existing topic models like latent Dirichlet allocation (LDA) [Blei et al. 2003] for the representation of certain hidden topics.

While existing work has illuminated many aspects of social emotion detection, it has not been fully explored yet. For example, an observation that has been ignored in previous work is that social emotions are not independent but closely correlated. This can be implied theoretically from research in cognitive and psychological which has shown that some basic emotions are able to blend to form more complex forms [Ortony and Turner 1990; Ekman 1992]. In fact, most emotions can be considered dependent with one another in either positive or negative ways. For example, the emotions of *happiness*, *touching* and *warmness* have much in common in reflecting human's positive emotion, while *anger*, *boredom* and *sadness* reflecting negative emotion. Such emotional dependency can provide important cues for social emotion detection. For example, for people suffering the *anger* emotion, it is not surprising to find that they also

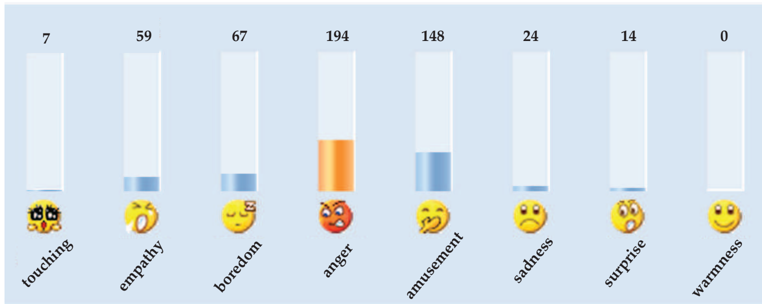


Fig. 1. An example of social emotions and reader voting

suffer the *sadness* emotion. This information can potentially benefit a prediction model if used properly.

In this article, we first propose a latent discriminative model (LDM) for social emotion detection. LDM involves intermediate hidden variables to model the latent structure of input corpora. In particular, it defines a joint distribution over the emotions and latent variables conditioned on news articles. Then, a new constraint is introduced into LDM to model the dependency between social emotions, which gives rise to an emotional dependency based latent discriminative model (eLDM). The dependency constraint attempts to provide additional guidance to LDM in the training process. To verify the effectiveness of the two new models for social emotion detection and make comparison with existing models, we use two real-world corpora of news articles and conduct a series of evaluations. Experimental results show that LDM and eLDM are quite competitive for the task. The results also show that the emotional dependency can be useful for LDM as well as other models like logistic regression.

The remainder of this article is organized as follows. We briefly review the related work in Section 2, and define the problem in Section 3. In Section 4, we introduce the details of LDM and eLDM. The model constraints and optimization details are specified in Section 5. We report on the experimental results in Section 6 and draw conclusions in Section 7.

2. RELATED WORK

In this section, we briefly review existing work related to social emotion detection. The first body of work is on sentiment analysis which identifies and extracts human sentimental information in subjective documents. Then, we review the work directly on social emotion detection.

2.1. Sentiment Analysis and Opinion Mining

The general objective of sentiment analysis or opinion mining is to determine the attitudes of users with respect to certain products or services, or the overall sentimental polarities in such texts as product reviews that involve human subjectivities. The first basic task is to distinguish subjective documents from non-subjective ones, which can be completed by training a subjectivity classifier based on sentiment-related features elaborately selected [Riloff and Wiebe 2003; Wiebe and Riloff 2005; Wilson et al. 2005]. Another task is sentiment classification, which classifies subjective documents into appropriate polarity categories (positive, negative or neutral). Both supervised and unsupervised learning techniques have been explored for this task. For example, Pang et al. [2002] studied three state-of-the-art machine learning algorithms, naive bayes, maximum entropy, and support vector machine, in their work of sentiment classification for movie reviews. Their study shows that the three algorithms do not perform as well on sentiment classification as on the traditional text classification. In another work by Turney [2002], the phrases containing adjectives or adverbs, which can be crucial indicators of human sentiment, are first identified. Then, the mutual information of a candidate phrase with the benchmark word "excellent", as well as with "poor", is calculated. The difference between the two mutual information scores is used to determine the sentiment orientation of the phrase. In comparison to the above work with single-domain data, Bollegala et al. [2011] investigated cross-domain sentiment classification. Their approach first constructs from multiple-domain data a sentiment sensitive thesaurus, and then uses it to expand the feature space before a binary classifier is trained.

Another stream of work is to summarize product reviews by identifying product attributes and user opinions from review texts [Hu and Liu 2004a; 2004b]. The typical procedure includes: (1) part-of-speech tagging of the review texts; (2) association mining of frequent words and phrases; (3) candidates pruning; (4) determining the final product attributes based on frequent attributes and their nearby adjective words. Sometimes certain product attribute is expressed in different forms, and the clustering of product attributes can be helpful. Synonyms and antonyms in WordNet [Fellbaum 1999] and grammatical rules [Zhuang et al. 2006] can also be included in the above steps. Besides, it is often noted that two positive (or negative) reviews can show quite different sentimental effects, which suggests the existence of sentimental strengths [Lu et al. 2010; Thelwall et al. 2010].

Sentiment analysis has also been widely studied in marketing research, with a motive to examine the economic impact of product reviews. Some scholars opt to believe that the volume of reviews helps to build product awareness and thus has economic impact. For example, Liu [2006] studied the impact of movie reviews and found that the volume rather than the polarity has certain effect on box office revenues. Similar study was conducted by Duan et al. [2008], reporting that the volume of online postings has greater impact on box office revenues than ratings. Another work by [Chevalier and Mayzlin 2006] revealed that consumers are not only concerned about the summarized average stars (ratings) but also read and respond to review comments. Ghose et al. [2007] studied the relation between user feedbacks and the pricing power of merchants by employing econometrics to measure the "economic value of text" and assigning a "dollar value" to each opinion phrase.

2.2. Social Emotion Detection

The above work on sentiment analysis is characterized by their focus on determining the attitudes of authors. In contrast, the task of "Affective Text" [Strapparava and Mihalcea 2007] was introduced to predict readers' attitudes evoked by reading news headlines. The assumption was that all words, even those neutral ones, can potentially convey affective meanings and provoke audience pleasant or unpleasant experiences. Three systems were developed for this task: SWAT, UA and UPAR7. SWAT adopts a supervised approach in three steps. First, a word-emotion map is built. Second, the map is used to score each word and produce an average score for each headline. Finally, the emotion label for each headline can be determined. UA gathers occurrence statistics of each headline and emotion by three search engines and then calculates pointwise mutual information (PMI) for determining the emotion label of the headline. UPAR7 is a rule-based system which relies greatly on syntactic parsing and lexicon. The study on how to build an emotional dictionary was also investigated [Rao et al. 2012; Lei et al. 2014]. Such dictionaries can be learned from training corpora of news articles and used for predicting the social emotion detection of new articles. Lin et al. [2007; 2008] studied social emotion detection by treating it as a classification problem. In their approaches, they defined a number of emotion-relevant features for the training of a SVM classifier, which was then used for prediction of new articles.

Another solution to social emotion detection attempts to model the integrated generation process of news articles and social emotions using topic model [Bao et al. 2012]. The generation of an affective document by their emotion-topic model (ETM) is modeled in four-step process. First, for each document, a distribution over emotions is sampled from a multinomial distribution. Second, for each word in the document, a single emotion label is generated according to the above distribution. Then, a latent topic is generated from a Dirichlet distribution conditioned on the sampled emotion. Finally, a word is sampled from the latent topic according to another multinomial distribution over words.

The models presented in this article are also related to mixture of regression approaches like Gaussian mixture models that have been studied in different communities such as machine learning [Bishop et al. 2006]. However, many of these approaches use a fixed set of combination weights for different models, which limits the modeling flexibility. Our proposed approach (e.g., LDM) allows each data instance to have different types of combination weights for learning, which provides more modeling power and is new for modeling emotion detection. Furthermore, we propose the novel emotion-dependency based latent discriminative model (eLDM) for the application of social emotion detection, which provides knowledge-based regularization for learning the mixture models.

3. PROBLEM STATEMENT

Generally speaking, social emotion detection determines the attitudes of the public evoked by certain events. A typical scenario is predicting the emotions of readers on online news website. Corpora for training a predictive model can be widely available as some news websites provide special “polls” following each article to collect readers’ emotions evoked by reading the article, as the example shown in Figure 1. The “polls” are designed in a form that provides a group of predefined emotion tags for choosing.

As mentioned before, social emotion detection can be treated as a classification task [Lin et al. 2007], where the emotion that wins the most votes is considered the category of an article to form a training dataset. After that, any classification models can be trained and then used for the prediction of emotions for new articles. Nevertheless, such approaches have ignored some important characteristics of social emotion detection itself. First of all, different from traditional classification tasks in which each document has association with one or a limited number of categories, each news article in social emotion detection is typically associated, either closely or loosely, with all the predefined emotions. Second, the extent of the association in social emotion detection can vary greatly across emotions, whereas it is rarely the case for classification in which the association of documents with their categories are typically equal (belonging or not). Third, it is natural to assume that social emotions are directly associated with certain events instead of news articles themselves, and the relationship between news articles, events, and social emotions can be modeled with special topic models. Finally, although certain classification models are available for incorporation of general dependency between categories, the dependency for social emotions appears to be quite special and complicated, which demands special treatment within a framework that can efficiently incorporate the unique components of social emotion detection.

Formally, a social emotion detection model takes as input a set $\mathcal{D} = \{d_1, d_2, \dots, d_N\}$ of N news articles, a word vocabulary $\mathcal{W} = \{w_1, w_2, \dots, w_M\}$, and a set $\mathcal{E} = \{e_1, e_2, \dots, e_E\}$ of E human-defined tags of basic emotions. It should be noted that the definition of basic emotions varies considerably in practice, even so the cross-cultural research of Ekman and Friesen [1969] has led to a reference list of basic emotions that tend to be universal to all humans: anger, disgust, fear, happiness, sadness, and surprise. Let vector $\mathbf{v}_i \in \mathbb{R}^M$ represent an article d_i, which is associated with a group of real number voting score $r_i = \{r_{i1}, r_{i2}, \ldots, r_{iE}\}$ over the E emotion tags. The normalized form of r_i is denoted by $\hat{r}_i$, which adds up to 1. The output of the social emotion detection model is another group of scores in the same form as $\hat{r}_i$.

4. METHODOLOGY

This section starts with an introduction to logistic regression, which lays the foundation of the new discriminative models.

4.1. Logistic Regression Model

Logistic regression is a traditional statistical model and has gained wide popularity in machine learning due to its robust performance and its close relation to SVM [Vapnik 1999] and AdaBoost [Freund and Schapire 1996], which are usually called large margin classifiers for the idea of maximizing margin either explicitly or implicitly. While SVM is aimed at maximizing the separation between classes, AdaBoost relies on a boosting technique that combines a number of weak classifiers into a stronger one. Figure 2(a) presents an example for binary classification to illustrate the principle of logistic regression and margin classifiers. Logistic regression estimates the probability of e_i given $\mathbf{v}_i$ according to

$$P\left(e_j|\mathbf{v}_i\right) = \frac{1}{\mathcal{N}\left(\mathbf{v}_i\right)} \exp\left(\theta_j' \mathbf{v}_i\right), \tag{1}$$

where $\theta_j \in \mathbb{R}^M$ is a vector of parameters corresponding to e_j and θ_j' is its transpose. $\mathcal{N}\left(\mathbf{v}_i\right)$ is the normalization term in the form of:

$$\mathcal{N}\left(\mathbf{v}_i\right) = \sum_{k=1}^{E} \exp\left(\theta_k' \mathbf{v}_i\right). \tag{2}$$

Note that standard logistic regression is inclined to suffer from overfitting, especially when the number of training examples is small or the dimension of feature space is high. Regularization is proved to be an effective technique to prevent overfitting [Orr 1995; Ng 2004]. The solution of logistic regression with regularization can be derived by optimizing

$$\arg\max_{\theta} \sum_{i=1}^{N} \log P\left(\hat{r}_i|\mathbf{v}_i;\theta\right) - \alpha \mathcal{R}\left(\theta\right) \tag{3}$$

where $P\left(\hat{r}_i|\mathbf{v}_i;\theta\right) = \prod_{j=1}^{E} P\left(e_j|\mathbf{v}_i\right)^{\hat{r}_{ij}}$, the parameter α is used to balance between overfitting and over-regularization, and $\mathcal{R}\left(\theta\right)$ is the regularization term typically defined in two ways. The first is L_1 regularization, defined as

$$\mathcal{R}\left(\theta\right) = \|\theta\|_1 = \sum_{j=1}^{E}\sum_{i=1}^{M} \left|\theta_{ji}\right|, \tag{4}$$

while the second is L_2 regularization with the form:

$$\mathcal{R}\left(\theta\right) = \|\theta\|_2^2 = \sum_{j=1}^{E}\sum_{i=1}^{M} \theta_{ji}^2. \tag{5}$$

Choosing between L_1 and L_2 is usually a problem depending on the specific applications [Ng 2004; Fan et al. 2008].

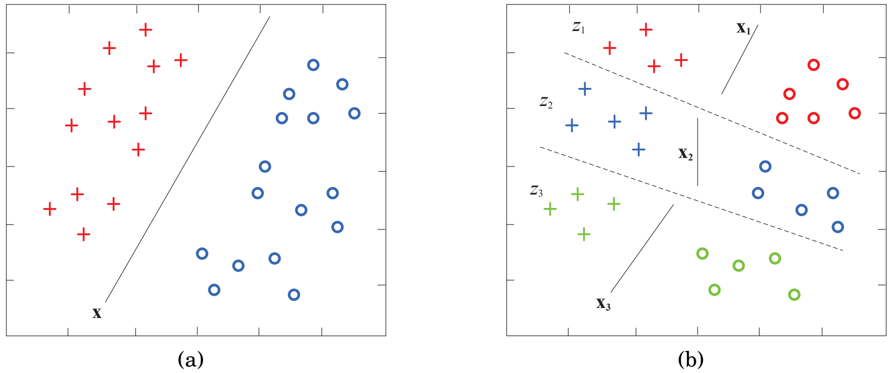


Fig. 2. Illustration of logistic regression (a) and LDM (b). The points with the same signs belong to the same class, while those having the same color but with different signs belong to the same latent topic. $\mathbf{x}$, $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ denote the optimum separation lines.

4.2. A Latent Discriminative Model

Logistic regression provides a simple yet powerful description of the probabilistic relationship of the variables and the outcome. In this section, we introduce a new latent discriminative model (LDM) developed from logistic regression. In the scenario of social emotion detection, it is natural to assume that people's emotions are associated with certain events involved in news articles. This assumption helps to define a new framework for dealing with the relationship among news articles, events, and social emotions. In this sense, our new model incorporates latent variables into the logistic regression model to capture the hidden "events" that can be of interest to readers in text corpora. These hidden "events" are also called latent topics in topic modeling [Blei et al. 2003]. To achieve this, our model defines a joint distribution over emotions and latent variables conditioned on observed news articles, as illustrated in Figure 2(b).

Given a news article denoted by $\mathbf{v}_i$, the probability of the evoked emotion being e_j is estimated using Bayes' theorem as

$$\begin{aligned} P\left(e_j|\mathbf{v}_i\right) &= \sum_{k=1}^{T} P\left(e_j, z_k|\mathbf{v}_i\right) \\ &= \sum_{k=1}^{T} P\left(e_j|\mathbf{v}_i, z_k\right) P\left(z_k|\mathbf{v}_i\right), \end{aligned} \tag{6}$$

where z_k denotes the latent variable and T is the number of latent variables. The above equation decomposes the original problem into two new sub-problems, one to build the relation between the observation and latent variables and the other to connect emotions and the combination of $\mathbf{v}_i$ and latent variables z_k. $P\left(e_j|\mathbf{v}_i, z_k\right)$ could be simplified to $P\left(e_j|z_k\right)$ by assuming that the generation of social emotions is entirely dependent on the latent variables. However, this oversimplification has potential risks in characterizing the generation of social emotions, for the reason that the latent variables may not fully represent observations. In this work, $P\left(e_j|\mathbf{v}_i, z_k\right)$ is estimated by

$$P\left(e_j|\mathbf{v}_i, z_k\right) = \frac{1}{\mathcal{N}_1\left(\mathbf{v}_i\right)} \exp\left(\boldsymbol{\pi}'_{z_k,j}\mathbf{v}_i\right), \tag{7}$$

where $\boldsymbol{\pi}_{z_k,j}$ is the vector of parameters corresponding to z_k and e_j. The other sub-problem, $P\left(z_k|\mathbf{v}_i\right)$, is estimated by

$$P(z_k|\mathbf{v}_i) = \frac{1}{\mathcal{N}_2(\mathbf{v}_i)} \exp(\omega_k' \mathbf{v}_i), \tag{8}$$

where ω_k denotes the vector of parameters corresponding to z_k. In the above two equations, $\mathcal{N}_1(\mathbf{v}_i)$ and $\mathcal{N}_2(\mathbf{v}_i)$ are the normalization terms whose implementation is akin to that of logistic regression.

For simplicity, take $\mathcal{X}_{ikj} = P(e_j|\mathbf{v}_i, z_k)$ and $\mathcal{Y}_{ik} = P(z_k|\mathbf{v}_i)$. The log-likelihood over the whole collection $\mathcal{D}$ with regularization is

$$\begin{aligned} \mathcal{L}(D) &= \log\left(\prod_{i=1}^{N}\sum_{k=1}^{T} P(\hat{r}_i|\mathbf{v}_i, z_k) P(z_k|\mathbf{v}_i)\right) - \mathcal{R}(\omega; \pi) \\ &= \sum_{i=1}^{N}\log\left(\sum_{k=1}^{T}\mathcal{Y}_{ik}\prod_{j=1}^{E}\mathcal{X}_{ikj}^{\hat{r}_{ij}}\right) - \mathcal{R}(\omega; \pi), \end{aligned} \tag{9}$$

where $\mathcal{R}(\omega; \pi)$ is the regularization term. In our work, the L_2 regularization is performed as no significant difference has been found between L_1 and L_2 for our problem. Specifically, the regularization constraint is defined as

$$\mathcal{R}(\omega; \pi) = \frac{\lambda_1}{TM}\sum_{k=1}^{T}\|\omega_k\|_2^2 + \frac{\lambda_2}{TME}\sum_{k=1}^{T}\sum_{j=1}^{E}\|\pi_{z_k,j}\|_2^2. \tag{10}$$

4.3. Emotional Dependency

As discussed earlier, social emotions are intuitively dependent on each other by their homogeneous or heterogeneous nature, and including of which into LDM has the potential to provide the model additional guidance in the training process. In this article, the relationship between social emotions is considered emotional dependency and captured by a new constraint $\mathcal{C}(\pi)$. The likelihood function of Equation (9) becomes

$$\mathcal{L}(D) = \sum_{i=1}^{N}\log\left(\sum_{k=1}^{T}\mathcal{Y}_{ik}\prod_{j=1}^{E}\mathcal{X}_{ikj}^{\hat{r}_{ij}}\right) - \mathcal{R}(\omega; \pi) - \mathcal{C}(\pi), \tag{11}$$

In this work, we define $\mathcal{C}(\pi)$ as

$$\mathcal{C}(\pi) = \frac{\lambda_3}{E^2 TM}\sum_{k=1}^{T}\sum_{u,v=1}^{E}\mathcal{W}_{uv}\|\pi_{ku} - \pi_{kv}\|_2^2, \tag{12}$$

where matrix $\mathcal{W}$ reflects the dependency among social emotions and can be constructed in multiple ways. One simple and effective way is to use the Pearson correlation coefficient [Rodgers and Nicewander 1988] of social emotions according to their original voting scores in news articles. Accordingly, $\mathcal{W}_{ij}$ is calculated as

$$\mathcal{W}_{uv} = \frac{\sum_{i=1}^{N}(\hat{r}_{iu} - \overline{\hat{r_{\cdot u}}})(\hat{r}_{iv} - \overline{\hat{r_{\cdot v}}})}{\sqrt{\sum_{i=1}^{N}(\hat{r}_{iu} - \overline{\hat{r_{\cdot u}}})^2 \sum_{i=1}^{N}(\hat{r}_{iv} - \overline{\hat{r_{\cdot v}}})^2}} \tag{13}$$

where $\overline{\hat{r_{\cdot u}}}$ and $\overline{\hat{r_{\cdot v}}}$ are the means of the distributions of emotions e_u and e_v in the considered news article collection.

5. OPTIMIZATION

Maximizing the likelihood function in Equation (11) with an analytic form is unrealistic. Instead, an approximation strategy can be resorted to. Since $\sum_{k=1}^{T} \mathcal{Y}_{ik} = 1$, according to Jensenł inequality [Dempster et al. 1977] we have

$$\begin{aligned} \mathcal{L}(D) \geq \mathcal{L}_1 &= \sum_{i=1}^{N} \sum_{k=1}^{T} \mathcal{Y}_{ik} \log \left(\prod_{j=1}^{E} \mathcal{X}_{ikj}^{\hat{r}_{ij}} \right) - \mathcal{R}(\omega; \pi) - \mathcal{C}(\pi) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{T} \mathcal{Y}_{ik} \sum_{j=1}^{E} \hat{r}_{ij} \log \mathcal{X}_{ikj} - \mathcal{R}(\omega; \pi) - \mathcal{C}(\pi) \end{aligned} \tag{14}$$

Then the expectation maximization (EM) algorithm [Dempster et al. 1977] can be used to result in a solution. EM involves an efficient iterative procedure to compute the maximum likelihood (ML) estimation of probabilistic models with unobserved latent variables involved. It begins with any random value of π and iterates the following E-step and M-step:

E-step: Calculate $\mathcal{X}$ based on Equation (7), and find the optimal ω by

$$\omega^* = \arg\max_{\omega} \mathcal{L}_1 \tag{15}$$

M-step: Calculate $\mathcal{Y}$ based on the above ω^*, and find the optimal π by

$$\pi^* = \arg\max_{\pi} \mathcal{L}_1 \tag{16}$$

With the above two steps, the log-likelihood over the training corpus will increase monotonically. This optimization is performed by using the *minFunc* toolkit[1], a collection of Matlab functions for solving optimization problems using the quasi-Newton strategy. The advantages of this toolkit lie in its rapidness to converge and the capability to deal with optimization that involves a large number of variables. Since the toolkit is designed to minimize a function, it is necessary to optimize the negative form of $\mathcal{L}_1$. The partial derivatives of $-\mathcal{L}_1$ with respect to ω and π is calculated as

$$\frac{\partial(-\mathcal{L}_1)}{\partial \omega_l} = -\sum_{i=1}^{N} \sum_{k=1}^{T} \mathcal{Y}_{ik} \mathbf{v}_i (\mathcal{I}_{ik} - \mathcal{Y}_{ik}) \sum_{j=1}^{E} \hat{r}_{ij} \log \mathcal{X}_{ikj} + \frac{2\lambda_1 \omega_l}{TM} \tag{17}$$

where $\mathcal{I}$ is the identity matrix and $\mathcal{I}_{ij} = 1$ if $i = j$ and 0 otherwise. Let $\mathcal{Q}_{ik} = \sum_{j=1}^{E} \hat{r}_{ij} \log \mathcal{X}_{ikj}$, Equation (17) becomes

$$\frac{\partial(-\mathcal{L}_1)}{\partial \omega_l} = -\sum_{i=1}^{N} \mathbf{v}_i \mathcal{Y}_{il} \left(\mathcal{Q}_{il} - \sum_{k=1}^{T} \mathcal{Q}_{ik} \mathcal{Y}_{ik} \right) + \frac{2\lambda_1 \omega_l}{TM} \tag{18}$$

Similarly, we have

$$\begin{aligned} \frac{\partial(-\mathcal{L}_1)}{\partial \pi_{z_p,q}} &= -\sum_{i=1}^{N} \mathbf{v}_i \mathcal{Y}_{ip} \sum_{j=1}^{E} \hat{r}_{ij} (\mathcal{I}_{qj} - \mathcal{X}_{ipq}) + \frac{2\lambda_2 \pi_{z_p,q}}{TME} + \frac{4\lambda_3}{E^2 TM} \sum_{j=1}^{E} \mathcal{W}_{qj} (\omega_{pq} - \omega_{pj}) \\ &= -\sum_{i=1}^{N} \mathbf{v}_i \mathcal{Y}_{ip} (\hat{r}_{iq} - \mathcal{X}_{ipq}) + \frac{2\lambda_2 \pi_{z_p,q}}{TME} + \frac{4\lambda_3}{E^2 TM} \sum_{j=1}^{E} \mathcal{W}_{qj} (\omega_{pq} - \omega_{pj}) \end{aligned} \tag{19}$$

[1] http://www.di.ens.fr/~mschmidt/Software/minFunc.html

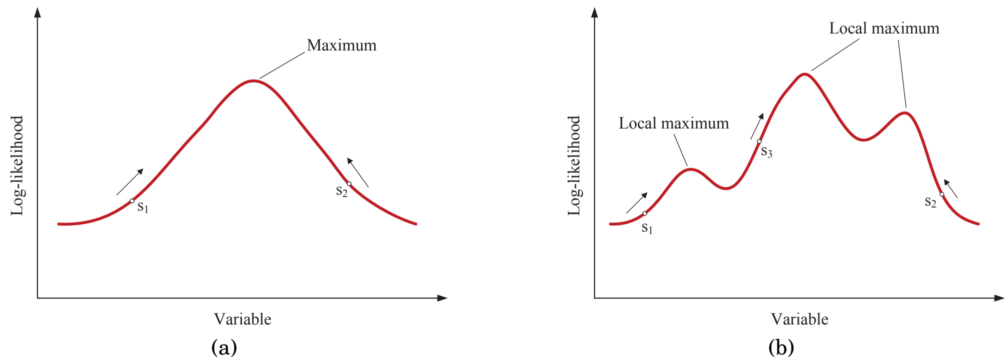


Fig. 3. Illustration of the optimization of logistic regression (a) and LDM (b), where s_1, s_2 and s_3 are the starting points.

The regularized logistic regression in Equation (3) simply requires to solve a concave optimization problem (as shown in Figure 3(a)) in order to result in a global maximum, which can be guaranteed theoretically [Ng 2004]. Differently, EM is only guaranteed to converge to local maxima given different starting points [Dempster et al. 1977], as illustrated in Figure 3(b). One strategy to cope with this problem is to try several different starting points and choose the model that leads to the highest likelihood [Bailey and Elkan 1995]. This simple strategy will be used for our optimization. The parameters of λ_1, λ_2 and λ_3 in our models are set by conducting cross-validation on training datasets and the ones that yield the highest log-likelihood is chosen. The complexity of each iteration in the above optimization relies mainly on two parts. One is the minimization performed by the *minFunc* toolkit which converges rapidly, and the other part is the E-step which has a complexity of $O(T^2NEM)$.

6. EXPERIMENT

A series of evaluations are conducted in this section to evaluate the new models. Since there are many news websites with emotion tags for news articles but few in English, our corpora are crawled from two popular Chinese news portals.

6.1. Datasets

The first corpus is crawled from Sina News[2], and a total number of 41,752 news articles published between January 2009 and September 2011 are collected from the *Society* channel, which publishes news about various social events and is attractive to normal readers. The website predefines eight social emotions, i.e., *touching*, *empathy*, *boredom*, *anger*, *amusement*, *sadness*, *surprise*, and *warmness* for readers to choose. Generally, articles that have won only a number of votes are not popular and tend not to be useful for our study as the limited votes do not represent the attitudes of the public. For this sake, we discard the articles with less than 20 votes, resulting in a collection of 5,282 articles. The second corpus is from Tencent News[3], and 63,316 articles are collected from the *Entertainment* channel. The predefined emotion tags include *happiness*, *touching*, *empathy*, *anger*, *amusement*, *sadness*, *surprise*, and *faint*. The emotion *faint* is a compound emotion which is blended from such emotions as *surprise* and *disappointment*. After performing similar pruning as before, 7,556 articles are left for use. The two datasets are summarized in Table I.

[2]http://news.sina.com.cn
[3]http://news.qq.com

Table I. Statistics for the two datasets

Dataset	# articles	# ratings	# terms
Sina	5282	1,027,095	11,125
Tencent	7556	25,971,729	7,717

Table II. Statistics for the two new datasets

Dataset	Category	# articles	# ratings
Sina	touching	749	86,194
	empathy	215	11,879
	boredom	245	13,652
	anger	2307	382,825
	amusement	975	81,548
	sadness	444	33,741
	surprise	314	18,583
	warmness	33	1,250
Tencent	happiness	3560	6,577,129
	touching	189	276,655
	empathy	165	438,921
	anger	952	2,751,099
	amusement	941	2,876,347
	sadness	250	831,917
	surprise	9	709
	faint	1475	4,777,774

The two corpora are preprocessed as follows. First, the title and main body of each article are extracted and represented with a single document. Second, unlike English and other western languages, Chinese sentences are not delimited by white space, and segmentation of them into meaningful units is necessary [Peng et al. 2004; Fu et al. 2008]. The Stanford Chinese Word Segmenter [Chang et al. 2008] is used for the segmentation. Then, removal of stop words is performed, and words appearing in less than 30 documents are discarded to reduce feature dimension.

6.2. Experimental Setup

Given a new news document, Equation (6) measures its relation with every social emotion by estimating the probability of the emotion conditioned on the document. Larger probabilities of an emotion correspond to higher likelihoods of readers being provoked that emotion. For evaluation purpose, we treat social emotion detection as a classification task, in which the emotion category of a document is determined as the one with the most votes. The reasons are as follows. First, it is a convention in previous work to evaluate like this, and following which enables us to make direct comparisons with them. Second, although it is very often that different readers have different or even opposite emotions for the same article, only one positive or negative emotion takes the major votes. Therefore, taking the majority emotion as the label of an article is a straightforward and reasonable way to perform classification. As a result, documents in the above datasets are reorganized according their emotion categories to form training and testing datasets. Table II summarizes the two new datasets.

For comparison, logistic regression (LR) and emotion-topic model (ETM) as well as several other methods as follows are implemented.

(1) Naive bayes (NB) [McCallum and Nigam 1998] is one of the most classic and commonly used model for classification tasks. It is founded on the feature independence assumption, and it further assumes that the generation of a document from a category relates only to the prior distribution of the category as well as the distribution of its words conditioned on the category.

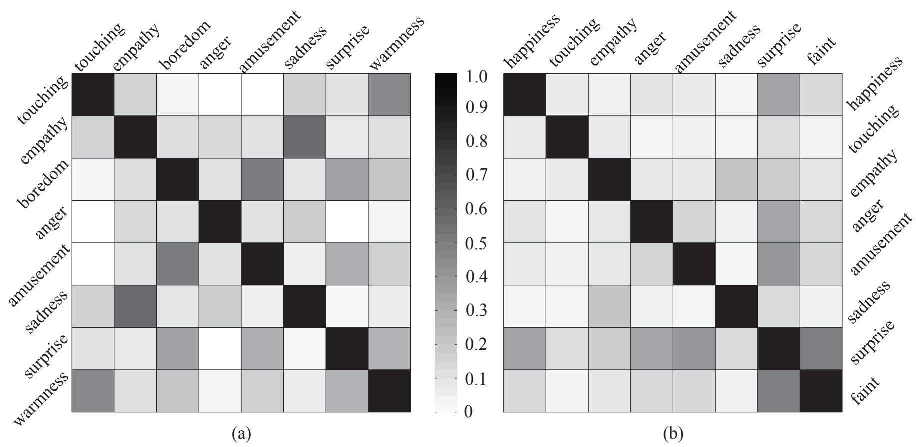


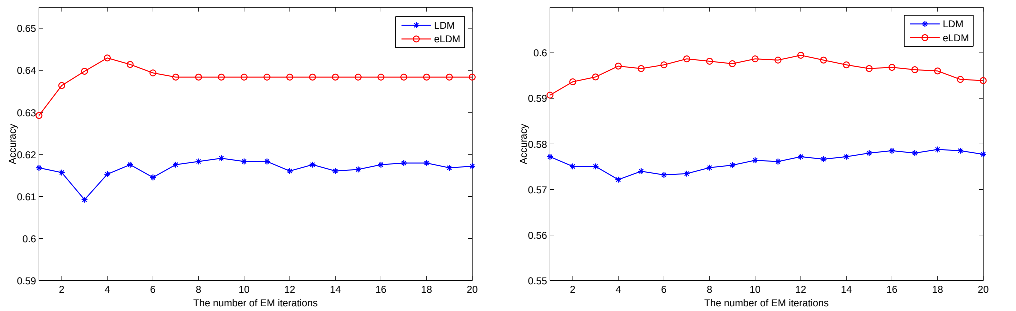
Fig. 4. Result of pairwise Pearson Correlations of social emotions on the Sina (a) and Tencent (b) datasets.

(2) Support vector machine (SVM) [Vapnik 1999] is a popular large margin classifier. The LibSVM tool [Chang and Lin 2011] is used for our implementation with the RBF kernel adopted. The setting of the two parameters, *gamma* and *cost*, are based on 10-fold cross validation on training datasets, while other parameters are set to their default values.
(3) Emotion-term model (ET) [Bao et al. 2012] is a variant of the naive bayes classifier, except that ET makes use of emotion votes to estimate the probability of a category and the probability of a word given a category.

The metric *accuracy* is used in most cases when reporting the final performance, which is defined as the percentage of correctly predicted articles. The reasons for using accuracy as the metric are twofold. First, unlike for general text classification tasks, our datasets for social emotion detection are quite imbalanced: some emotion categories contain happiness or anger comprise many articles while the others (e.g., surprise) may contain only a few documents. So reporting performance for every category tends not to be reasonable. The second reason is to make comparisons with previous work which also adopts accuracy as the performance metric.

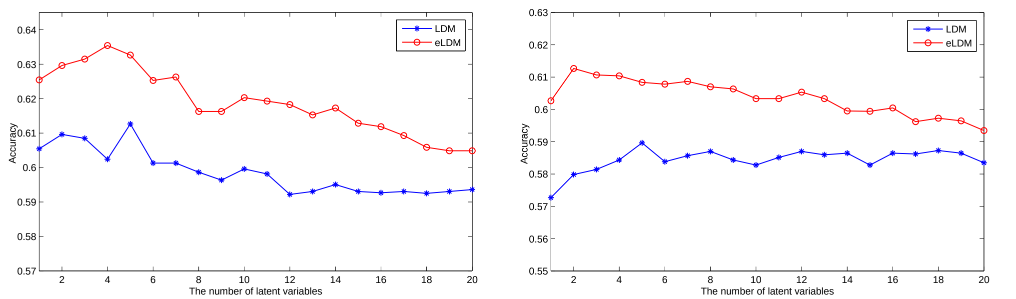
6.3. Demonstration of Emotional Dependency

Since emotional dependency is one of the key components of our models, demonstrating its existence is particularly necessary ahead of any further investigations. For this purpose, we compute the correlation of every pair of emotions by making use of their votes across news articles. To give an intuitive demonstration, we show the pairwise correlations in a figure in which correlation coefficients are denoted with blocks coloured on a scale of while to black, with black representing highest correlation and white no correlation. Since most of the resulting correlation coefficients are positive, the negative correlation relationships are ignored and labelled with white blocks. As the result shown in Figure 4(a), there are relatively strong correlations revealed for *empathy* vs. *sadness*, *boredom* vs. *amusement*, *boredom* vs. *surprise*, and *touching* vs. *warmness*. Similar findings can also be noted for the pairs of *surprise* vs. *faint*, *amusement* vs. *surprise*, *happiness* vs. *surprise*, and *anger* vs. *surprise*. These phenomena are basically consistent with common sense. Another interesting observation in Figure 4(b) is that *surprise* has certain correlation with almost every emotion. This can be accounted for by the nature of entertainment news as for most of them readers could feel *surprised*.



(a) (b)

Fig. 5. Convergence rate of the EM algorithm for LDM and eLDM, with (a) on the Sina dataset and (b) on the Tencent dataset.



(a) (b)

Fig. 6. Performance of LDM and eLDM with change of different numbers of latent variables on the Sina (a) and Tencent (b) datasets.

6.4. Parameters Evaluation

In this subsection we study the influence of the parameters of LDM and eLDM on their performance. For the series of evaluations below, we randomly divide each dataset into two equal-sized partitions, one to be used for training and the other for testing. Aside from those parameters to be studied soon, the remaining will be set by means of 10-fold cross-validation on training datasets.

6.4.1. EM Iteration. The rate of convergence of the EM algorithm for LDM and eLDM is studied. The strategy to be used is to fix one starting point and increase the iteration number from 1 to 20, where the accuracy for each iteration is recorded. As the result shown in Figure 5, LDM appears not to converge within 20 iterations, yet a desirable result can usually be acquired within 10 iterations. In contrast, eLDM converges quickly on the Sina dataset but not on the other. In general, eLDM can find an optimal solution sooner than LDM.

6.4.2. Latent Variables. In our models, latent variables generally represent certain hidden events that are of interest to readers, and they cannot be directly observed from document corpora but have to be inferred from other observed variables. The impact of the number of latent variables, T, on our models is studied in this evaluation with T changing from 1 to 20. For each number, the models run ten times with random starting points, and the one that achieves the best accuracy performance is reported. The result is plotted in Figure 6, from which we can observe that LDM and eLDM can generally achieve the best performance when T is around 4 on the Sina dataset.

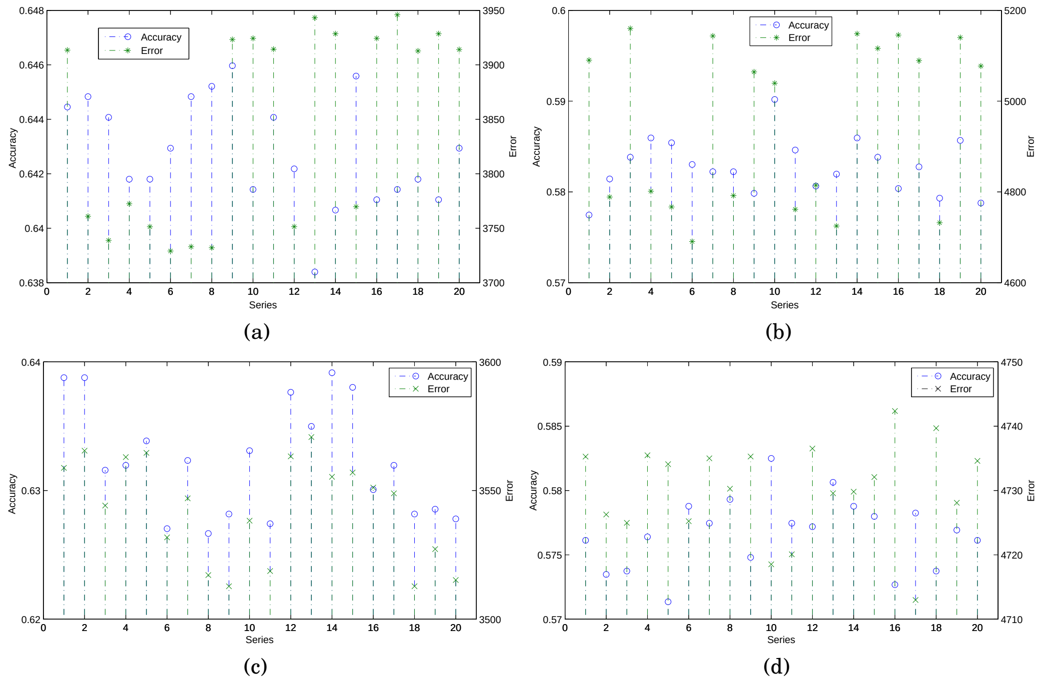


Fig. 7. The influence of different starting points on the performance of LDM and eLDM in terms of error on training datasets and accuracy on testing datasets. The result of LDM on the Sina and Tencent datasets is shown in (a) and (b), and eLDM is in (c) and (d).

The situation on the other dataset is slightly different, where LDM and eLDM have the best performance when T is 5 and 2, respectively. The findings are quite different from those in generative topic models in which the best topic number is usually around several hundred [Blei et al. 2003; Griffiths and Steyvers 2004].

6.4.3. Starting Point. The selection of starting points is critical to the final solution that EM can produce. As mentioned before, the strategy employed in our work is to run the EM algorithm with multiple random starting points and choose the one that gives the highest likelihood. Here we study the effect of different starting points on LDM and eLDM by initializing π with 20 random starting points. The corresponding errors (the negative log-likelihoods) on training datasets as well as the accuracy on testing datasets are reported. Errors on training datasets reflect how well certain trained models fit the data, while accuracy on testing datasets measures the generalization capability. As the result shown in Figure 7, it is quite obvious that the errors and accuracies vary for different given starting points, which indicates that different local maxima are converged to. Note that the solution with the lowest error on training datasets does not necessarily lead to the best generalization on testing datasets.

6.5. Comparison with Baselines

In this subsection, we report the final performance of our models and make comparison with other baseline models in two different ways. The first is to construct training and testing datasets according to the chronological order of news articles. As such the models trained on earlier news articles is least related to later articles. The second way is to evaluate the two models with datasets of different sizes and perform 10-fold cross-validation on them to report final performance. Specifically, 10-fold cross-validation

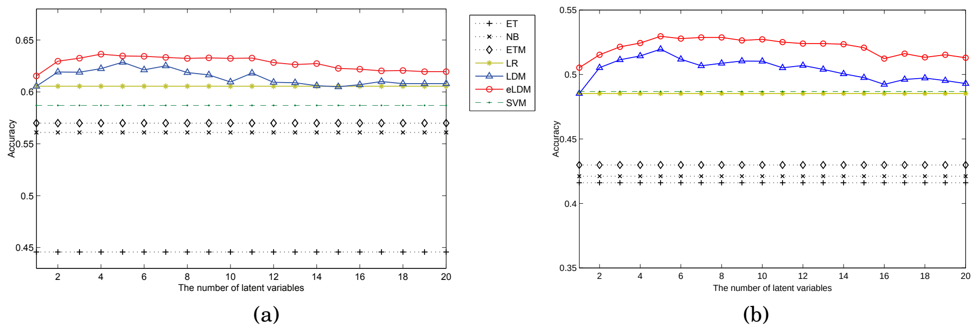


Fig. 8. Comparison with baseline methods when the articles are divided into training and testing datasets according to their chronological order. Apart from the eLDM and LDM, logistic regression (LR), naive bayes (NB), support vector machine (SVM), emotion-term model (ET) and emotion-topic model (ETM) are also studied on the Sina (a) and Tencent (b) datasets.

Table III. Overall result on the Sina dataset constructed according to the chronological order of news articles

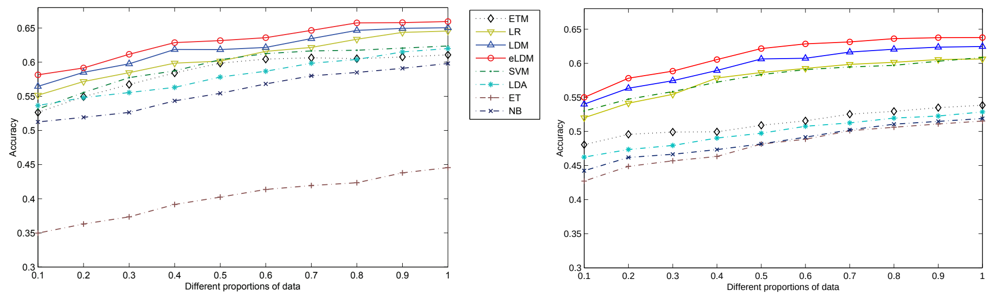
Model	ET	NB	ETM	LR	SVM	LDM	eLDM
Precision	0.446	0.561	0.570	0.605	0.587	0.629	0.636
Recall	0.428	0.569	0.554	0.602	0.579	0.622	0.634
F1	0.437	0.565	0.562	0.603	0.583	0.625	0.635

Table IV. Overall result on the Tencent dataset constructed according to the chronological order of news articles

Model	ET	NB	ETM	LR	SVM	LDM	eLDM
Precision	0.416	0.421	0.430	0.485	0.487	0.520	0.530
Recall	0.417	0.417	0.429	0.479	0.490	0.517	0.531
F1	0.416	0.419	0.429	0.482	0.488	0.518	0.530

starts with an original dataset randomly divided into 10 partitions, one to be used for testing and the remaining for training. After the testings on the ten partitions are completed, an average score is used to measure the overall performance.

First of all, each news dataset is evenly divided into two disjoint sets according to the release time of articles. The first half is used for training and the second for testing. In addition to LDM and eLDM, we also implement logistic regression (LR), naive bayes (NB), support vector machine (SVM), emotion-term model (ET), and emotion-topic model (ETM) for comparison. The result is shown in Figure 8, from which we can observe that ET performs the worst on the Sina dataset and comparable to NB and ETM on the Tencent dataset. Moreover, the two state-of-the-art machine learning algorithms, LR and SVM, have shown to be very competitive for the social emotion detection task. It is surprising that ETM underperforms SVM on both datasets and comparable to NB, which indicates that ETM might be more flexible in expressing dependency between variables, but it is not that effective for this task. This finding is different from what is reported in previous work [Bao et al. 2012], in which ETM is reported to perform much better than SVM. One possible reason for this is that ETM is sensitive to different datasets used for evaluation. The two new models, LDM and eLDM, have given the best performance in this evaluation. The further improvement of eLDM over LDM has to be accounted for by the involved emotional dependency. To provide a more comprehensive comparison, we also report the above result in terms of precision, recall, and F1, as shown in Table III and Table IV. Basically, similar conclusions as the above can be drawn from the tables.



(a) (b)

Fig. 9. Comparison with baseline methods when different proportions of the Sina (a) and Tencent (b) datasets are used.

The second evaluation strategy starts by using the first 10 percent of news articles in each dataset to perform 10-fold cross-evaluation, followed by increasing the percentage to 20 percent until all the articles are included. In addition to the above models, we also implement LDA as a new baseline method for its widely-recognized performance in classification tasks. The final result is plotted in Figure 9. It can be observed that the accuracies of all the studied methods increase constantly as the sizes of the datasets grow. Another observation is that eLDM, ETM, LR and SVM outperform in general the rest ones on the two datasets. ETM slightly outperforms SVM on the first dataset but not on the second.

Through the above evaluations, eLDM and LDM have shown very competitive performance for social emotion detection. Among the existing models, LR and SVM are proved to be the best. Note that although the improvement of eLDM and LDM over LR and SVM appears not to be dramatic, statistical analysis shows that this improvement is statistically significant ($P < 0.05$).

6.6. Emotional Dependency for Logistic Regression

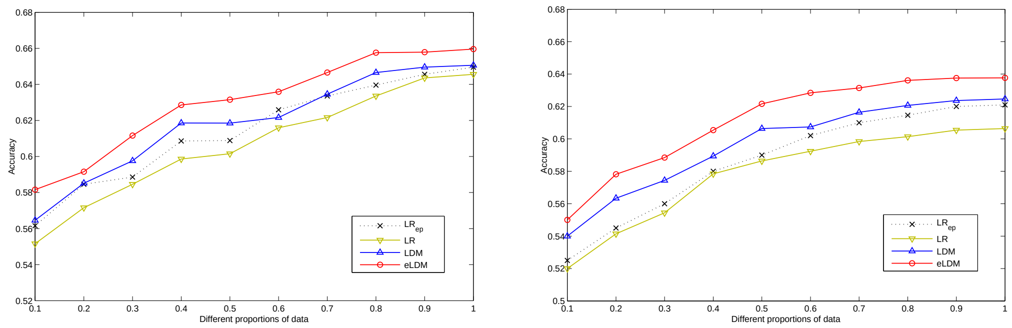
The further success of eLDM over LDM can be attributed to the emotional dependency constraint further involved. Intuitively, the constraint should apply to some other models as well. This is studied by employing logistic regression as a baseline model, because logistic regression can be readily open to incorporate the new constraint. For logistic regression, the new constraint is defined as

$$\mathcal{C}(\pi) = \frac{\beta}{E^2 M} \sum_{u,v=1}^{E} \mathcal{W}_{uv} \left(\theta_u - \theta_v\right)^2 .$$

The second strategy in the above is used to construct new datasets for this evaluation. The new logistic regression model with emotional dependency is denoted as $\mathrm{LR_{ep}}$. Besides, the old logistic regression model (LR), LDM and eLDM are also presented for comparison. The result of the four models is depicted in Figure 10, from which it is not difficult to find that the performance of logistic regression is considerably improved after incorporating the emotional dependency constraint. This again confirms the effectiveness of the emotional dependency for social emotion detection.

7. CONCLUSION AND FUTURE WORK

In contrast to conventional sentiment analysis research that focuses primarily on determining the sentiments of writers, this article studies the problem of social emotion detection, which is a natural extension of sentiment analysis from writers' perspective



(a) (b)

Fig. 10. Comparison with logistic regression on the Sina (a) and Tencent (b) datasets when the emotional dependency is considered.

to readers'. We establish the connections between news articles and social emotions by introducing intermediate hidden variables, which leads to a novel latent discriminative model. The new model is intended to model the latent structure of input domain by defining a joint distribution over emotions and latent variables conditioned on observations. Besides, this article demonstrates the existence of emotional dependency, which can provide additional guidance to the discriminative model when it is incorporated. We empirically verify the effectiveness of the proposed models for social emotion detection on real-world news data.

One of the remaining issues in social emotion detection is the imbalance problem, i.e., the votes of different emotions are extremely imbalanced. For example, while the emotion “anger” might receive thousands of votes in the event of an unpleasant incident reported, “happiness” can barely receive some. How this problem affects the performance of social emotion detection models is still beyond our knowledge. It also tends to be true that our models cannot easily deal with it in their present forms. There have been many strategies that are highly possible to tackle the imbalance problem. Instead of directly modifying a model, certain sampling strategies can also be effective. The investigation requires another work in the future.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

C. O. Alm, D. Roth, and R. Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the Joint Conference on Human Language Technology and Empirical Methods in Natural Language Processing*. 579–586.

T. L. Bailey and C. Elkan. 1995. Unsupervised learning of multiple motifs in biopolymers using expectation maximization. *Machine Learning* 21, 1 (1995), 51–80.

S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D. Han, and Y. Yu. 2012. Mining social emotions from affective text. *IEEE Transactions on Knowledge and Data Engineering* 24, 9 (2012), 1658–1670.

Christopher M Bishop and others. 2006. *Pattern recognition and machine learning*. Vol. 1. springer New York.

D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022.

D. Bollegala, D. Weir, and J. Carroll. 2011. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. 132–141.

C. C. Chang and C. J. Lin. 2011. LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent System Technology* 2, 3 (2011), 1–27.

P. C. Chang, M. Galley, and C. Manning. 2008. Optimizing Chinese word segmentation for machine translation performance. In *Proceedings of the ACL Workshop on Statistical Machine Translation*. 224–232.

J. A. Chevalier and D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43, 3 (2006), 345–354.

A. P. Dempster, N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical Society* 39, 1 (1977), 1–38.

W. Duan, B. Gu, and A. B. Whinston. 2008. Do online reviews matter?–An empirical investigation of panel data. *Decision Support Systems* 45, 4 (2008), 1007–1016.

P. Ekman. 1992. An argument for basic emotions. *Cognition & Emotion* 6, 3-4 (1992), 169–200.

P. Ekman and W. V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1 (1969), 49–98.

R. E. Fan, K. W, Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin. 2008. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research* 9 (2008), 1871–1874.

C. Fellbaum. 1999. *WordNet*. Wiley Online Library.

Y. Freund and R. E. Schapire. 1996. Experiments with a new boosting algorithm. In *Proceedings of the International Conference on Machine Learning*. 148–156.

G. Fu, C. Kit, and J. J. Webster. 2008. Chinese word segmentation as morpheme-based lexical chunking. *Information Sciences* 178, 9 (2008), 2282–2296.

A. Ghose, P. G. Ipeirotis, and A. Sundararajan. 2007. Opinion mining using econometrics: A case study on reputation systems. In *Proceedings of the Annual Meeting of the Association of Computational Linguistics*. 416.

T. L. Griffiths and M. Steyvers. 2004. Finding scientific topics. In *Proceedings of the National Academy of Sciences of the United States of America*. 5228–5235.

M. Hu and B. Liu. 2004a. Mining and summarizing customer reviews. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 168–177.

M. Hu and B. Liu. 2004b. Mining opinion features in customer reviews. In *Proceedings of the National Conference on Artificial Intelligence*. 755–760.

Jingsheng Lei, Yanghui Rao, Qing Li, Xiaojun Quan, and Liu Wenyin. 2014. Towards building a social emotion detection system for online news. *Future Generation Computer Systems* 37 (2014), 438–448.

K. H. Y. Lin, C. Yang, and H. H. Chen. 2007. What emotions do news articles trigger in their readers. In *Proceedings of the annual international ACM SIGIR conference on Research and development in information retrieval*. 733–734.

K. H. Y. Lin, C. Yang, and H. H. Chen. 2008. Emotion classification of online news articles from the reader's perspective. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. 220–226.

Y. Liu. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing* 70, 3 (2006), 74–89.

Y. Lu, X. Kong, X. Quan, W. Liu, and Y. Xu. 2010. Exploring the sentiment strength of user reviews. In *Web-Age iformation Management*. 471–482.

A. McCallum and K. Nigam. 1998. A comparison of event models for naive bayes text classification. In *AAAI-98 Workshop on Learning for Text Categorization*. 41–48.

A. Y. Ng. 2004. Feature selection, L1 vs. L2 regularization, and rotational invariance. In *Proceedings of the International Conference on Machine Learning*. 78–85.

J. R. Nofsinger. 2005. Social mood and financial economics. *The Journal of Behavioral Finance* 6, 3 (2005), 144–160.

M. J. L. Orr. 1995. Regularization in the selection of radial basis function centers. *Neural Computation* 7, 3 (1995), 606–623.

A. Ortony and T. J. Turner. 1990. What's basic about basic emotions? *Psychological Review* 97, 3 (1990), 315.

B. I. Page and R. Y. Shapiro. 1983. Effects of public opinion on policy. *The American Political Science Review* (1983), 175–190.

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Vol. 10. 79–86.

F. Peng, F. Feng, and A. McCallum. 2004. Chinese segmentation and new word detection using conditional random elds. In *Proceedings of the International Conference on Computational Linguistics*.

Yanghui Rao, Xiaojun Quan, Liu Wenyin, Qing Li, and Mingliang Chen. 2012. Building word-emotion mapping dictionary for online news. In *Proceedings of the first International Workshop on Sentiment Discovery from Affective Data (SDAD)*. 28–39.

E. Riloff and J. Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. 105–112.

J. L. Rodgers and W. A. Nicewander. 1988. Thirteen ways to look at the correlation coefficient. *The American Statistician* 42, 1 (1988), 59–66.

C. Strapparava and R. Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the International Workshop on Semantic Evaluations*. 70–74.

C. Strapparava and R. Mihalcea. 2008. Learning to identify emotions in text. In *Proceedings of the Annual ACM Symposium on Applied Computing*. 1556–1560.

M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Sociaty for Information Science and Technology* 61, 12 (2010), 2544–2558.

P. D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the Annual Meeting on Association for Computational Linguistics*. 417–424.

V. Vapnik. 1999. *The Nature of Statistical Learning Theory* (2nd. ed.). Springer Verlag.

J. Wiebe and E. Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. *Computational Linguistics and Intelligent Text Processing* (2005), 486–497.

T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. 2005. OpinionFinder: A system for subjectivity analysis. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*. 34–35.

C. Yang, K. H. Y. Lin, and H. H. Chen. 2007. Emotion classification using web blog corpora. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*. 275–278.

L. Zhuang, F. Jing, and X. Zhu. 2006. Movie review mining and summarization. In *Proceedings of the ACM International Conference on Information and Knowledge Management*. 43–50.

Received xxx; revised xxxx; accepted xxxx