IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING
IEEJ Trans 2015
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI:10.1002/tee.22204

Paper

Multi-Label Emotion Recognition of Weblog Sentence Based on Bayesian Networks

Lei Wang[*,**], Non-member
Fuji Ren[*,**a], Member
Duoqian Miao[*], Non-member

An increasing number of common users, in the Internet age, tend to express their emotions on the Web about everything they like or dislike. As a consequence, the number of all kinds of reviews, such as weblogs, production reviews, and news reviews, grows rapidly. This makes it difficult for people to understand the opinions of the reviews and obtain useful emotion information from such a huge number of reviews. Many scientists and researchers have attached more attention to emotion analysis of online information in the natural language processing field. Different from previous works, which just focused on the single-label emotion analysis, this paper takes into account rich and delicate emotions and gives special regard to multi-label emotion recognition for weblog sentences based on the Chinese emotion corpus (Ren-CECps). Using the theory of Bayesian networks and probabilistic graphical model, the latent emotion variable and topic variable are employed to find out the complex emotions of weblog sentences. Our experimental results on the multi-label emotion topic model demonstrate the effectiveness of the model in recognizing the polarity of sentence emotions. © 2015 Institute of Electrical Engineers of Japan. Published by John Wiley & Sons, Inc.

Keywords: latent Dirichlet allocation, multi-label classification, emotion recognition, Gibbs sampling

Received 27 July 2013; Revised 14 December 2014

1. Introduction

With the rapid expansion of the Internet, the increasing websites enable people to write reviews and express their emotions on things that they like or dislike, such as weblogs, product reviews, news reviews, and so on. It is very popular among common users to write personal weblog and review product characteristics, and discuss news on the Web. This results in a large amount of online information, including individual subjective emotion, on the Internet. The emotion in the information indicates many different emotional states such as joy, anger, sorrow, happiness, and so on. In order to help common users to get opinions of the reviews and make their own decisions, we need to analyze the emotions people express on their weblogs or reviews [1–3]. The development of affective computing technology makes the job that might have been difficult manually instantly a hot research topic in the field of natural language processing.

The objective of affective computing is to analyze delicate human emotions, recognize human emotional states, and optimize the process of human–computer communication [4], with the great progress in artificial intelligence and natural language processing. After the concept of "affective computing" was proposed by Picard [5], many scientists and researchers not only in artificial intelligence but also in natural language processing have paid close attention to the study of emotion analysis or sentiment analysis. Previous works were devoted to analyzing emotion factors and extracting emotion features. All kinds of techniques were also applied to recognize the polarity of emotion features from many information formats, such as texts, sounds, facial expressions, and gestures [6–9].

Text such as weblogs and product reviews plays a basic and important role among all the communication media on the Internet [6,10]. This may, to some extent, explain why most scientists extract all kinds of emotion features from text, just like emotion words or phrases, emotion holders, and emotion objects. Depending on the different levels of text analysis, emotion analysis of the text can be divided into three levels of interdependence and mutual support: the analysis of word emotion, the analysis of sentence emotion, and the analysis of document emotion [11]. The analysis of word emotion mainly focuses on finding the emotion words and judging emotion polarity of words in the document [12–14]. The analysis of document emotion focuses on how to use the various machine learning methods to identify emotion factors in the article and further judge the polarity of the document [15]. Of the three levels, the analysis of sentence emotion plays a middle-level role, since it depends on the analysis of word emotion and can also obtain more abundant emotional factors to support the analysis of document emotion.

In this work, we approach the problems of weblog sentence emotions. Our research is different from traditional text sentiment analysis in a number of ways. First of all, there is no special emotion lexicon as in traditional studies. Second, the theory of Bayesian networks is borrowed to analyze the emotions of sentences in the context of the article [16,17], while in traditional methods each sentence is analyzed independently. Lastly, sentences simply labeled a single emotion as positive or negative in traditional studies. While a sentence could have several emotion states, it could be assigned two or more emotion tags at the same time [18–20]. In this paper, a multi-label emotion topic model is constructed to recognize the complex emotions of weblog sentences on the basis of the Chinese emotion corpus Ren-CECps [21] with detailed emotional tags in each article.

[a] Correspondence to: Fuji Ren. E-mail: ren@is.tokushima-u.ac.jp

[*] Department of Computer Science and Technology, Tongji University, 1239 Siping Road, Shanghai 201804, China
[**] Faculty of Engineering, The University of Tokushima, 2-1 Minamijyousanjima-cho, Tokushima 770-8506, Japan

The remainder of the paper is organized as follows: Section 2 briefly introduces the related work, including several methods for the analysis of sentence emotion. Section 3 presents the structure of Ren-CECps corpus, and Section 4 further elaborates the multi-label emotion topic model, probability assumption, and inference. Section 5 describes the experimental process and analyzes the results, and, finally, Section 6 draws the conclusion and points out some future work.

2. Related work

It is an important task to carry out emotion analysis of sentences in an article so as to extract various emotion objects, identify emotion word/phrase, and judge the emotion polarity and the intensity of the sentence.

In current studies, it is obvious that there are two basic research methods on the analysis of sentence emotion: one is based on emotion knowledge, and the other is based on feature classification. The first method relies on the existing emotion lexicon or creates a new emotion lexicon for their research. First, it extracts all emotion words, recognizes their polarity and the intensity in the sentence, and then judges the polarity of the sentence by computing the emotional weight each emotion word carries. The core of the method focuses on identifying the polarity and the intensity of emotion words in the sentences. Hu and Liu [1,2] analyzed the emotion of customer reviews in a similar manner. They probed into the relationship of synonyms and antonyms in the WordNet and got the polarity of the emotion word. Then the superior sentiment of the emotion words of the sentence helps ascertain the polarity of the sentence. Under an unsupervised learning method, Turney [22] applied the mutual information between new words and the seed words "excellent" and "poor" to explore the sentiment of new words. Our work differs from these works on emotion recognition in that we do not need any lexicons or any seed words, which is very important in practice.

The second method takes advantage of machine learning methods and selects a large number of meaning features to accomplish emotion classification task. Dave [23] first collected more than 1,000 articles labeled emotion tags, counted the frequency of these words with n-gram in the article, then made a score based on the frequency of words, and finally judged the polarity of new sentences using these emotion features and the scores. Pang [24] determined the sentiment classification of movie reviews by testing several supervised machine learning techniques and came to the conclusion that the machine learning techniques were better than the method based on manmade tag features. Our work is related but quite different from their works, as we recognize the sentiment of sentences in the article while they perform the sentiment recognition at the document level.

In current works, some constraints are set, and high-frequency emotion words are selected from the document to classify the sentiment in the method of machine learning. As a result, many no-meaning, high-frequency phrases are extracted while some low-frequency phrases possessing much emotion information are missed. Latent topic variable models can automatically generate topics over all the words, and each topic contains the top N words, and thus they overcome these shortcomings [17]·[25–29]. Most of the latent variable models are based on the latent Dirichlet allocation (LDA) [26], which is a probability generative model that can be used to estimate the properties of multinomial observations by unsupervised learning.

After studying each document in Ren-CECps, we can find out that there is a close relationship between people' emotional states and the topics of the article. The multi-label emotion topic model is constructed by incorporating the latent emotion variable into the LDA model. In our model, the word generation proceedure is determined not only by the topics of the documents but also by the emotions of the sentences in the document. The multi-label emotion topic model is a semisupervised model since the generation of sentence emotions and words need previous information that can be collected from training set, but the topic generation procedure is the same as that of the LDA model, which is an unsupervised procedure.

3. Ren-CECps corpus

In this paper, our study and experiments on emotion recognition are based on Ren-CECps corpus, which contains a total of 1,487 Chinese weblog articles, 11,255 paragraphs, 35,096 sentences, and 878,164 Chinese words. All these weblog articles are crawled from the Chinese weblog websites and cover a variety of domains, including politics, economy, sports and technology, etc. Each article is annotated with emotion tags at three levels: sentence, paragraph, and document. At the paragraph and document level, emotion annotation includes emotion category, emotion intensity, topic words, and topic sentence. At the sentence level, emotion category, emotion intensity, emotion keyword/phrase, degree word, negative word, punctuation, objective/subjective, and emotion polarity are all annotated [30].

We only discuss the emotional problems at the sentence level. Each sentence in Ren-CECps corpus is categorized into a mixed emotion class which is a subset of eight basic emotion categories: Surprise, Anxiety, Sorrow, Expectation, Joy, Love, Anger, and Hate. The intensity of each single emotion is also tagged with a decimal score from 0.1 to 1.0. Two examples with mixed emotion classes extracted from Ren-CECps and translated into English are given below:

a) "我跟女儿说,，你别生气了,，想想将来怎么办。" (Sorrow: 0.7| Anxiety: 0.6). I said to my daughter: "Don't get angry and think about what to do in the future."
b) "他们生存了下来。" (Joy: 0.5| Love: 0.6). They all survive.

In this paper, if the intensity of single emotion in one sentence is greater than zero, then the sentence possesses the emotion and is tagged with an emotion label. For example, the first sentence possesses two emotions of sorrow and anxiety, and the second sentence also possesses two emotions of joy and love.

4. Multi-Label Emotion Topic Model

Our research focuses on how to recognize the complex emotions of the sentence in a document. We propose the multi-label emotion topic model (MLETM), a generative probabilistic graphical model based on Bayesian networks, to identify the mixed emotions of weblog sentences. By using the semi-supervised method, the prior probability distribution of emotions can be obtained from training data, while the prior probability of the latent topics cannot be observed in each document from training data, so we assume that these prior probability distributions of topics are the same.

We first introduce the basic idea and graphic structure in the multi-label emotion topic model, and then demonstrate the probability assumption. Finally, we provide the mathematical inference for the model in detail.

4.1. Multi-Label emotion topic model structure

MLETM is developed to identify the mixed sentence emotions on the weblog. It is based on the idea that there is a strong relationship between the mixed emotions of one sentence and the topic features of words in the document after carefully studying the Ren-CECps corpus. Furthermore, the sentence in one document includes some

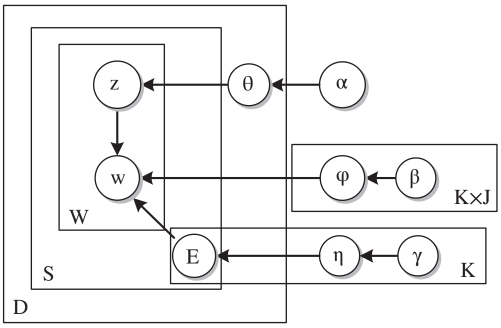

Fig. 1. Structure of the multi-label emotion topic model

implicit context information, so it is essential to identify the mixed emotions of one sentence in the context of the document. We explore the latent topic features of words in the document and obtains the context information among sentences. Following this intuition, we construct a corpus of multi-label emotion sentences with emotion E over the sentence s, together with the latent topics z over words w in the document.

As shown in Fig. 1, MLETM is represented in plate notation.

This model contains nine nodes representing nine different random variables, like E indicating the random variables of emotion, and eight directed edges representing the dependent relationship between two variables, like $z \rightarrow w$ showing the dependent relations between the word variable w and topic variable z. In this graphical model, there are three different kinds of variables: the categorical variable, the proportional variable, and the observable variable.

Nodes E, z, and w stand for three categorical variables E, z, and w, and all of these categorical variables take discrete values. To clearly describe the mixed emotions of one sentence in the document, K binary random variables E_{dsk} are employed to indicate the presence or absence of the single emotion k of the sentence s in the document d. As a result, there is a $K - plate$ outside the emotion node E in Fig. 1. Each word w_{dsi} in one sentence s of the document d is assumed to follow some unknown distribution with parameter ϕ and also receives the influence from random variables E_{dsk} and z_{di}, as there are two directed edges from the nodes E and z in the graph. θ, η, and ϕ are the proportional variables and, respectively, count the prior probability of the categorical variables E, z, and w. Each θ_d, as a $J -$ dimensional vector, represents the prior probability of topic j to the variable z_{di} in the document d. The directed edge $\theta \rightarrow z$ shows the dependent relationship between the variable θ_d and the variable z_{di}. η is a $K -$ dimensional vector to count the prior probabilities of emotions in the $K - plate$, responding to the different emotion categories. ϕ is a $K \times J \times N -$ dimensional matrix and represents the word distribution. The element ϕ^1_{kjt} counts the prior probability of a word t with topic j and emotion k the sentence possesses, and the element ϕ^0_{kjt} counts the prior probability of a word t with topic j and emotion k the sentence does not possesses. In the model, α, β, and γ are the observable variables and the values of these variables can be counted from the training set. γ is a $K -$ dimensional vector, the element γ^1_k counts the number of sentences that possess emotion k within the mixed emotions, and the element γ^0_k counts the number of sentences that do not possess emotion k.β, which is a $K \times J \times N -$ dimensional matrix, is the observable variables of the word. The element β^1_{kjt}, responding to ϕ^1_{kjt}, counts the occurrence of a word t with topic j and emotion k the sentence possesses, and the element β^0_{kjt}, responding to ϕ^0_{kjt}, counts the occurrence of a word t with topic j and emotion k the sentence does not possess in the training set. All variables and the meaning of each variable in the model are listed in Table I.

4.2. Probability assumption In this section, we particularly demonstrate the probability assumption of these dependent relationships in the model. As mentioned above, the directed edges in graphic structure depict dependent relationship among the random variables that have the directed relation. For example, a directed edge $z \rightarrow w$ indicates the dependent relationship between word w and topic z.

The word w has three input edges coming from the emotion node E, topic node z, and parameter ϕ, so we consider that the variable w is conditionally dependent on E, z, and ϕ by the theory of Bayes networks. But one sentence s has K emotion indicators E_{dsk}, and each indicator E_{dsk} with topic z_{di} can affect the variable w probability distribution. Furthermore, we assume that each emotion indicator E_{dsk} in sentence s of document d is independent. Therefore, we can assume that the variable w follows the categorical distribution by variable ϕ conditional on E and z for each K emotion categories; the assumptions are shown as follows:

$$w_{dsi}|E^1_{dsk}, z \sim Categorical(\phi^1_{E_{dsk} z_{di}}) \quad E_{dsk} = 1 \tag{1}$$

$$w_{dsi}|E^0_{dsk}, z \sim Categorical(\phi^0_{E_{dsk} z_{di}}) \quad E_{dsk} = 0 \tag{2}$$

Topic node z denotes variable z and has one input edge from variable θ. We also assume that each topic is independent and follows the categorical distribution. Variable z_{di} denotes that the *ith* word in document d has possessed the topic j or not, so the assumption is given as

$$z_{di} \sim Categorical(\theta_d) \tag{3}$$

As we have assumed that each emotion indicator E_{dsk} is independent, here we suppose that the probability distribution of each

Table I. Variables with specific meanings

D	Documents in test set	w	Word
S	Sentences in a document	z	Topic
W	Words in a sentence	K	Emotion number
N	Word in the vocabulary	J	Topic number
α	$J - vector, \alpha_j$ counts the topic j from a pseudo-training set	E	$K - vector$, the complex emotion
β^0, β^1	$K \times J \times N - matrix, \beta^1_{kjt}$ counts the occurrence of a word t with topic j and emotion k, β^0_{kjt} counts the occurrence of a word t with topic j and no emotion k	ϕ	$K \times J \times N - matrix$, ϕ^1_{kjt} counts the prior probability of a word t with topic j and emotion k, ϕ^0_{kjt} counts the prior probability of a word t with topic j and no emotion k
γ^0, γ^1	γ^1 counts the sentence with emotion k from a training set, γ^0 counts the rest sentences	θ	$D \times J - matrix, \theta_{dj}$ is the probability of a word in document d with topic j
η	$K - vector, \eta_k$ count the prior probabilities that the *kth* single emotion		

binary emotion indicator E_{dsk} follows the Bernoulli distribution; the formula is given by

$$E_{dsk} \sim Bernoulli(\eta_k) \quad (4)$$

The Bernoulli distribution is a discrete probability distribution, which takes the value 1 with success probability η_k and the value 0 with failure probability $1 - \eta_k$.

The proportional variable ϕ indicates the present or absent sentence emotion such as ϕ^1_{kjt} and ϕ^0_{kjt}. It is assumed that both of them follow the Dirichlet distribution over the proportional variables β^1_{kjt} and β^0_{kjt}, which can be shown as below:

$$\phi^1_{kjt} \sim Dirichlet(\beta^1_{kjt}) \quad (5)$$

$$\phi^0_{kjt} \sim Dirichlet(\beta^0_{kjt}) \quad (6)$$

The variable η indicates the prior probability of binary emotional indicators E_{dsk}. It is assumed that the variable η follows the Beta distribution, which is known as the conjugate prior of the Bernoulli distribution, which can be shown as below:

$$\eta_k \sim Beta(\gamma^1_k, \gamma^0_k) \quad (7)$$

The parameter γ^1_k indicates the probability of which sentences possess emotion k within the multi-label emotions, and γ^0_k concerns the probability of which sentences do no possesses emotion k.

4.3. Inference

In MLETM, we just recognize sentence emotion E_{dsk} and topic z_{di} relating with emotion, while the values of other variables can be computed as intermediate values or observed from training data. There are several approximate inference approaches adapted to the inference of graphical models, such as mean-field variation, Gibbs sampling, and loopy belief propagation. The collapsed Gibbs sampling algorithm is more efficient in predicting the values of latent variables E and z, with other variables integrating out in the deduction. As a result, the collapsed Gibbs sampling is selected to predict latent variables and is shown in Algorithm 1.

Alg.1 Gibbs sampling for the model

```
For m = 1 to N Gibbs sampling iterations do
  For  d< D do
    For  s< S do
      For k< K do
        Compute probability p(E_dsk | w, z, E_-dsk ; α, β, γ)
        Sampling E_dsk
      End for
    End for
    For w<  W_d do
        Compute probability p(z_di | w, z_-di, E; α, β, γ)
      Sampling z_di
    End for
  End for
End for
```

Given the probability distribution assumption above and the observed values from training set, the values of these latent variables can be inferred. According to the dependent relations on all other variables, the posterior distribution of each emotion indicator E_{dsk} for the kth emotion of sentence s in document d is directly proportional to the value of a prior probability and a likelihood probability by using the Bayesian networks; the equation is given as

$$p(E_{dsk}|w, z, E_{-dsk}; \alpha, \beta, \gamma) \propto p(E_{dsk}|E_{-dsk}; \gamma)\, p(w_{ds}|w_{-ds}, z, E; \alpha, \beta, \gamma) \quad (8)$$

Here, E_{-dsk} represents the set of all emotion indicators except E_{dsk}, $p(E_{dsk}|E_{-dsk})$ just counts the prior probability of emotion E_{dsk}, and this probability is also depicted by the kth entity of proportional variable η. As defined in the probability assumption, variable η follows the Beta distribution:

$$\eta|E_{-dsk} \sim \begin{cases} Beta\left(n^1_1 + \gamma^1_1, \cdots, n^1_K + \gamma^1_K\right) & E_{dsk} = 1 \\ Beta\left(n^0_1 + \gamma^0_1, \cdots, n^0_K + \gamma^0_K\right) & E_{dsk} = 0 \end{cases} \quad (9)$$

Here, n^1_k counts the number of sentences with the emotion k in the test set, and n^0_k counts the number of sentences without the emotion k in test set. The likelihood of word w_{ds} is factorized as

$$p(w_{ds}|w_{-ds}, z, E; \alpha, \beta, \gamma) = \prod_{i \in W_{ds}} p(w_{dsi}|w_{-dsi}, z, E; \alpha, \beta, \gamma) \quad (10)$$

Given the probability assumption above, the probability is depicted by the w_{dsi} entity of proportional variable ϕ, following the Dirichlet distribution:

$$\phi^1|w_{-ds}, z, E \sim \text{Dirichlet}(n^1_{kz_{di}1} + \beta^1_{kz_{di}1}, \cdots, n^1_{kz_{di}N} + \beta^1_{kz_{di}N}) \quad E_{dsk} = 1 \quad (11)$$

$$\phi^0|w_{-ds}, z, E \sim \text{Dirichlet}(n^0_{kz_{di}1} + \beta^0_{kz_{di}1}, \cdots, n^0_{kz_{di}N} + \beta^0_{kz_{di}N}) \quad E_{dsk} = 0 \quad (12)$$

$n^1_{kz_{di}t}$ and $n^0_{kz_{di}t}$ are a pair of observations, $n^1_{kz_{di}t}$ counts the number of words with topic z_{di} and emotion k that sentence s possesses, while $n^0_{kz_{di}t}$ counts the number of words with topic z_{di} and emotion k that sentence s does not possess. So (8) can be inferred further and shown as

$$p(E_{dsk}|w, z, E_{-dsk}; \alpha, \beta, \gamma) \propto \begin{cases} \frac{n^1_k + \gamma^1_k}{n^0_k + n^1_k + \gamma^0_k + \gamma^1_k} \exp\left(\sum_{i \in W_{ds}} \log \frac{n^1_{kz_{di}w_{dsi}} + \beta^1_{kz_{di}w_{dsi}}}{\sum_t n^1_{kz_{di}t} + \beta^1_{kz_{di}t}}\right) & E_{dsk} = 1 \\ \frac{n^0_k + \gamma^0_k}{n^0_k + n^1_k + \gamma^0_k + \gamma^1_k} \exp\left(\sum_{i \in W_{ds}} \log \frac{n^0_{kz_{di}w_{dsi}} + \beta^0_{kz_{di}w_{dsi}}}{\sum_t n^0_{kz_{di}t} + \beta^0_{kz_{di}t}}\right) & E_{dsk} = 0 \end{cases} \quad (13)$$

The probability of topic z_{di} conditioned on the other variables is the probability of the ith word in document d and turns to the proportion of a prior probability and a likelihood probability; the equation is

$$p(z_{di}|w, z_{-di}E; \alpha, \beta, \gamma) \propto p(z_{di}|z_{-di}; \alpha) p(w_{di}|w_{-di}, z, E; \alpha, \beta, \gamma) \quad (14)$$

The probability $p(z_{di}|z_{-di})$ is depicted by the z_{di} entity of the proportional variable θ, which follows the Dirichlet distribution

$$\theta|z_{-di} \sim Dirichlet(n_{d1} + \alpha_1, \cdots, n_{dJ} + \alpha_J) \quad (15)$$

Here n_{dj} counts the number of topic j in document d. So the (14) is inferred further and given as

$$p(z_{di}|w, z_{-di}, E; \alpha, \beta, \gamma) \propto \frac{n_{dz_{di}} + \alpha_{z_{di}}}{W_d + \alpha^*} \times \prod_{k \in K^1_d} \frac{n^1_{kz_{di}w_{di}} + \beta^1_{kz_{di}w_{di}}}{\sum_t n^1_{kz_{di}t} + \beta^1_{kz_{di}t}} \times \prod_{k \in K^0_d} \frac{n^0_{kz_{di}w_{di}} + \beta^0_{kz_{di}w_{di}}}{\sum_t n^0_{kz_{di}t} + \beta^0_{kz_{di}t}} \quad (16)$$

Here K^1_d is the subset of emotion indicators in the multi-label emotion for document d, and K^0_d is the complementary set.

5. Experiments

Since this study aims to recognize the complicated human emotions from weblog sentences, each sentence of the article is given at least one emotion from the total $K = 8$ emotion categories, including Surprise, Anxiety, Sorrow, Expectation, Joy, Love, Anger, and Hate. We select 1000 weblog articles including 21,225 sentences from Ren-CECps corpus for our experiments based on the multi-label emotion topic model. Ten-fold cross-validation is performed on the whole data for each experiment, and the variables α, β, and γ can be obtained from training set. The parameter α is set as a constant value, which equals the number of words in the vocabulary for each topic. We now evaluate our model from three perspectives:

1. The macro-accuracy and macro-precision of the model
2. The accuracy of each single emotion
3. Recognizing the mixed emotions compared with naive Bayes (NB)

5.1. The macro-accuracy and macro-precision of the model The experiment analyzes the macro-accuracy and macro-precision of the model for predicting the multi-label emotion of weblog sentences. We employ the methods introduced by Tsoumakas [20] to measure the result of experiment. $M(tp_k, fp_k, tn_k, fn_k)$ is a binary evaluation measure that calculates the number of true positive (tp), true negative (tn), false positive (fp), and false negative (fn). The macro-accuracy and the macro-precision criteria are given as

$$\text{Macro}-\text{accuracy} = \frac{1}{|K|}\sum_{k=1}^{|K|}\frac{tp_k + tn_k}{tp_k + tn_k + fp_k + fn_k} \quad (17)$$

$$\text{Macro}-\text{precision} = \frac{1}{|K|}\sum_{k=1}^{|K|}\frac{tp_k}{tp_k + fp_k} \quad (18)$$

Here K, the number of emotions in our model, equals eight.

From Fig. 2, we can find that the macro-accuracy and the macro-precision are increasing slightly with increasing of the iteration number and they become stable when the iteration number is above 1000. The minimum of the macro-accuracy is >75%, which proves that our model is effective in recognizing the mixed emotions of weblog sentences.

The study of the experimental result shows that the macro-accuracy is higher than the macro-precision, while the macro-precision is low. After analyzing the corpus, we find that many sentences possess one or two emotions of the total eight emotion classes, and a few of them possess three emotions. The model also can recognize the emotions well which the sentence does not have, so that the value of tn_k is large in (17), leading to the macro-accuracy becoming high. It is difficult to recognize precisely all emotions the sentence possesses, so that the value of tp_k is relatively small, resulting in the macro-precision being low. The result of this experiment is consistent with that of the third experiment.

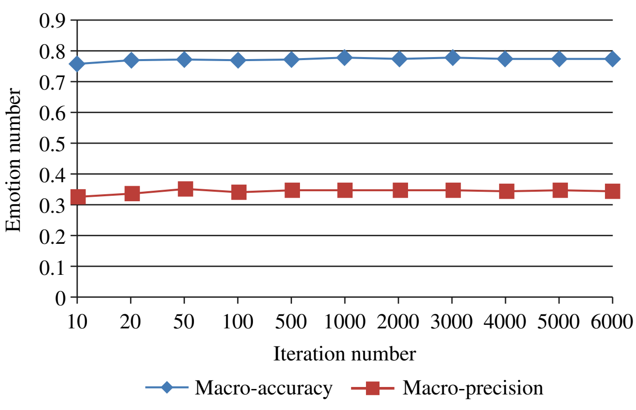

Fig. 2. Macro-accuracy and the macro-precision

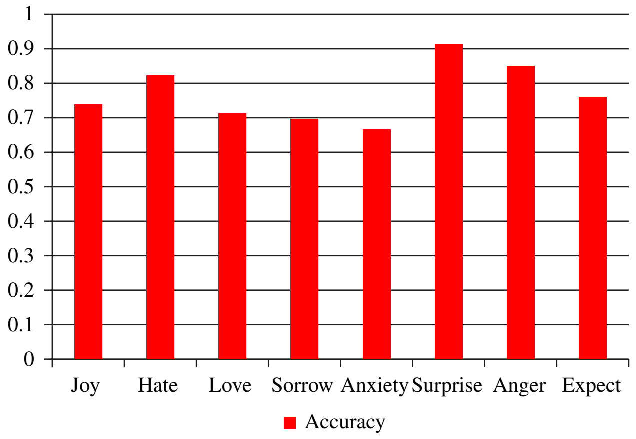

Fig. 3. Accuracy of each single emotion

5.2. The accuracy of each single emotion In MLETM, we also evaluate the accuracy of each single emotion among complex emotions. Here we give the following accuracy criterion for each emotion:

$$\text{Accuracy } E_k = \frac{tp_k + tn_k}{tp_k + fp_k + fn_k + tn_k} \quad (19)$$

From Fig. 3, we can see that the model is good at recognizing the single emotion among the complex emotion situation. The lowest accuracy of a single emotion is >60%, and the accuracy of surprise is >90%. The accuracy of surprise is the highest, the accuracy of anxiety is the lowest, and the accuracy of other emotions is in between the two. All these results also prove that our model is perfect in recognizing the emotions of weblog sentences.

5.3. Recognizing the mixed emotions compared with Naive Bayes (NB) Here the method of NB is selected as the base classifier. The main goal of the experiment is to compare the performance of two methods by recognizing the complex emotions that the sentence possesses. Three criteria are as follows:

$$\text{One}-\text{match}-\text{precision} = \frac{\text{the number of sentences at least one emotion matched}}{\text{the total number of sentences}} \quad (20)$$

$$\text{Two}-\text{match}-\text{precision} = \frac{\text{the number of sentences at least two emotions matched}}{\text{the total number of sentences}} \quad (21)$$

$$\text{All}-\text{match}-\text{precision} = \frac{\text{the number of sentences all emotions exactly matched}}{\text{the total number of sentences}} \quad (22)$$

The value of one-match-precision indicates the percentage of the sentences with at least one single emotion recognized by the model; the value of two-match-precision denotes the percentage of the sentences with two emotions identified, and the value of all-match-precision counts the percentage of the sentences that possess all complex emotions that have been recognized exactly. In Table II, the macro-accuracy of MLETM is slightly lower than the macro-accuracy of NB, but the other four criteria of MLETM are far higher than NB. The macro-accuracy shows the whole accuracy, recognizing not only the emotions that the sentence has but also those that do not. The other four criteria reflect the precision of precisely recognizing emotions that the sentence has. The result shows that the method of NB can recognize the emotions that the sentence does not possess, but fails to recognize those that

Table II. Result of experiment comparison

	NB	MLETM
One-match-precision	0.040	0.607
Two-match-precision	0.029	0.327
All-match-precision	0.022	0.154
Macro-precision	0.003	0.348
Macro-accuracy	0.807	0.798

the sentence has. The compared results of the experiment prove that the performance of MLETM is good. It can also be seen that the one-match-precision is highest, >60%, while all-match-precision is the lowest, and two-match-precision is in between the two. This means that it is easy to judge a few emotions that the sentence has, but is too hard to recognize precisely all complicated emotions that people express in practice.

5.4. Discussion In this section, we discuss how to evaluate the results of our experiments and find out the factors that influence the performance of the complex emotion recognition.

The results of the experiments show the effectiveness of our model in recognizing the mixed emotion of weblog sentences. The macro-accuracy of the model is >75%, and the one-match-precision is ∼60% in experiment 3. In experiment 2, the macro-accuracy of some single emotions is recognized with >90%, such as surprise. Experiment 3 reveals the good performance on the macro-accuracy and one-match-precision of mixed emotion recognition, compared with NB. But the all-match-precision is <20%. All of the above descriptions show that the emotion of human beings is so complicated that it is difficult to recognize all the emotions that sentences possess just by using topic feature of a word. This inspires us to explore more meaning features to improve the performance of our model. Imbalance of emotion distribution in corpus is another factor that influences the emotion recognition, and thus some emotions of weblog sentences cannot be judged precisely, as shown in experiment 2. It is a difficult and tough task for the future.

6. Conclusion

This paper explored the Ren-CECps corpus and proposed the MLETM to recognize the emotions of weblog sentences by using Bayesian networks. The idea is based on the strong relationship between the emotions of sentences and the topics of words in the article. Furthermore, the latent topic features possess the context information of the article and can be adopted to avoid the defects in previous works in which the emotions of sentences were recognized simply. After obtaining the prior probability distribution on the topics of words and the prior probability of sentence emotions from the training set, the experimental results indicate that the model is effective in recognizing the complex emotions of sentences in the test data.

Since the corpus is relatively small and the distribution of emotions does not keep balance in Ren-CECps, the parameters is not precise enough by training, and some emotions of sentences cannot be recognized precisely in new data, such as anxiety and sorrow. On the other hand, there are many other grammatical features in Ren-CECps, such as disjunctive connectives, modification relations, and negative phrases. We can make full use of these features to improve the quality of our model, which will be future work.

Finally, it is one of the hardest problems to acquire a basic and relatively fine-grained emotion corpus in natural language processing, especially when facing the shortage of Chines emotion corpus. The Ren-CECps corpus is so valuable that we expect to take advantage of the corpus to do more research on emotion analysis in the future. We hope that affective computing will become more important and popular as more and more people express their opinions on the Web.

Acknowledgments

This research was partially supported by the Ministry of Education, Science, Sports and Culture through Grant-in-Aid for Scientific Research (A), 15H01712.

References

(1) Ming H, Bing L. Mining and summarizing customer reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2004; 168–177.
(2) Bin L, Lei Z. *A Survey on Opinion Mining and Sentiment Analysis. Mining Text Data.* Springer: New York; 2012.
(3) Jo Y, Oh AH. Aspect and sentiment unification model for online review analysis. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, 2011; 815–824.
(4) Fragopanagos N, Taylor JG. Emotion recognition in human-computer interaction. *Neural Networks* 2005; **18**(**4**):389–405.
(5) Picard RW. *Affective Computing*, MIT Press:Massachusetts; 1997.
(6) Li J, Ren F. Emotion recognition of weblog sentences based on an ensemble algorithm of multi-label classification and word emotions. *IEEJ Transactions on Electronics, Information and Systems* 2012; **132**(**8**):1362–1375.
(7) Ren F. Affective information processing and recognizing human emotion. *Journal of Electronic Notes in Theoretical Computer Science* 2009; **225**:39–50.
(8) Ren F. From cloud computing to language engineering, affective computing and advanced intelligence. *Journal of Advanced Intelligence* 2010; **2**(**1**):1–14.
(9) Ren F, Quan C. Linguistic-based emotion analysis and recognition for measuring consumer satisfaction—an application of affective computing. *Information Technology and Management* 2012; **13**(**4**):321–332.
(10) Strapparava C, Mihalcea R. Learning to identify emotion in text. *Proceedings of the 2008 ACM Symposium on Applied Computing*, 2008; 1556–1560.
(11) Zhao Y, Qi B, Liu T. Sentiment analysis. *Journal of Software* 2010; **21**(**8**):1834–1848.
(12) Taboada M, Brooke J, Tofiloski M. Lexicon-based methods for sentiment analysis. *Journal of Computational Linguistics* 2011; **37**(**2**):267–307.
(13) Yang C, Hsin-Yih Lin K, Chen H-H. Building emotion lexicon from weblog corpora. *Proceedings of the 45th Annual Meeting of ACL on Interactive Poster and Demonstration Sessions*, 2007; 133-136.
(14) Matsumoto K, Ren F. Estimation of word emotions based on part of speech and position information. *Computers in Human Behavior* 2011; **27**:1553–1564.
(15) Picard RW, Vyzas E, Healey J. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2001; **23**(**10**):1175–1191.
(16) Ren Fuji, Kan Xin. Employing hierarchical bayesian networks in simple and complex emotion topic analysis. *Computer Speech and Language* 2013; **27**(**4**):943–968.
(17) Lin C, He Y, Everson R. A comparative study of Bayesian models for unsupervised sentiment detection. *Proceedings of the 14th Conference on Computational Natural Language Learning*, 2010; 144–152.
(18) Bhowmick PK, Basu A, Mitra P. Reader perspective emotion analysis in text through ensemble based multi-label classification framework. *Computer and Information Science* 2009; **2**(**4**):64–74.
(19) Tsoumakas G, Katakis I. Multi-label classification: an overview. *International Journal of Data Warehousing and Mining* 2007; **3**(**3**):1–13.
(20) Tsoumakas G, Vlahavas I. Random K-label sets: an ensemble method for multilabel classification. *Proceedings of the 18th European Conference on Machine Learning(ECML2007)*, 2007; 406–417.
(21) Ren F. Document for Ren-CECps 1.0, 2009. http://a1-www.is.tokushima-u.ac.jp/member/ren/Ren-CECps1.0/Ren-CECps1.0.html. Accessed October 10, 2014.

(22) Turney PD. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002; 417–424.
(23) Dave K, Lawrence S, Pennock DM. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. *Proceedings of the 12th International Conference on the World Wide Web*, 2003; 519–528.
(24) Pang B, Lee L, Vaithyanathan S. Thumbs up? sentiment classification using machine learning techniques, *Proceedings of the Conference on Empirical Methods in Natural Language Processing* 2002; **10**: 79–86.
(25) Sun Y, Zhou X, Wei F. Unsupervised topic and sentiment unification model for sentiment analysis. *Journal of Acta Scientiarum Naturalium Universitatis Pekinensis* 2013; **49**(**1**):102–108.
(26) Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *Journal of Machine Learning Research* 2003; **3**:993–1022.
(27) Moghaddam S, Ester M. On the design of LDA models for aspect-based opinion mining. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 2012; 803–812.
(28) Lin C., He Y. Joint sentiment/topic model for sentiment analysis. *Proceedings of the 18th ACM International Conference on Information and Knowledge Management*, 2009, 375–384.
(29) Das D, Bandyopadhyay S. Extracting emotion topics from blog sentences: use of voting from multi-engine supervised classifiers. *Proceedings of the 2nd International Workshop on Search and Mining User-Generated Contents*, 2010; 119–126.
(30) Quan Changqin, Ren Fuji. A blog emotion corpus for emotional expression analysis in Chinese. *Journal of Computer Speech & Language* 2010; **24**(**4**):726–749.

Lei Wang (Non-member) received the B.E. degree from Hefei University of Technology, Anhui, China, in 2002, and the M.E. degree from East China Normal University, Shanghai, China, in 2005. Now, he is pursuing the Ph.D. degree in the Department of Computer Science and Technology, Tongji University, Shanghai, China, and the double Ph.D. degree in the Faculty of Engineering, University of Tokushima, Tokushima, Japan. His research interests include affective computing, text data mining, and machine learning.

Fuji Ren (Member) received the B.E. and M.E. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and in 1955, respectively, and the Ph.D. degree from the Faculty of Engineering, Hokkaido University, Japan, in 1991. He worked with CSK, Japan, where he was a Chief Researcher of NLP from 1991. From 1994 to 2000, he was an Associate Professor in the Faculty of Information Science, Hiroshima City University. He became a Professor in the Faculty of Engineering, University of Tokushima, in 2001. His research interests include natural language processing, artificial intelligence, language understanding and communication, and affective computing. Prof. Ren is a member of the IEICE, CAAI, IPSJ, JSAI, and AAMT, and a senior member of the IEEE. He is a fellow of the Japan Federation of Engineering Societies, and the president of the International Advanced Information Institute.

Duoqian Miao (Non-member) received the B.E. and M.E. degrees from Shanxi University, Shanxi, China, in 1985 and in 1991, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, China, in 1997. He is currently a full Professor and Vice-Dean of the School of Electronics and Information Engineering, Tongji University. He has published over 180 scientific articles in international journals, books, and conference proceedings. His current research interests include rough sets, granular computing, principal curve, Web intelligence, and data mining. Prof. Miao is a committee member of the International Rough Sets Society, senior member of the China Computer Federation (CCF), committee member of CCF Artificial Intelligence and Pattern Recognition, committee member of the Chinese Association for Artificial Intelligence (CAAI), chair of CAAI Rough Set and Soft Computing Society, committee member of CCAI Machine Learning, committee member of Chinese Association of Automation (CAA) and Intelligent Automation, and committee member and chair of Shanghai Computer Society (SCA) of Computing Theory and Artificial Intelligence.