Predicting Complex Word Emotions and Topics through a Hierarchical Bayesian Network

Kang Xin[1], Ren Fuji[1,2*]

[1] Faculty of Engineering, The University of Tokushima, 2-1, MinamiJosanjima-cho, Tokushima 7708501, Japan

[2] AIM, Hefei University of Technology, Hefei 230009, P. R. China

Abstract: In this paper, we provide a Word Emotion Topic (WET) model to predict the complex word emotion information from text, and discover the distribution of emotions among different topics. A complex emotion is defined as the combination of one or more singular emotions from following 8 basic emotion categories: joy, love, expectation, surprise, anxiety, sorrow, anger and hate. We use a hierarchical Bayesian network to model the emotions and topics in the text. Both the complex emotions and topics are drawn from raw texts, without considering any complicated language features. Our experiment show promising results of word emotion prediction, which outperforms the traditional parsing methods such as the Hidden Markov Model and the Conditional Random Fields (CRFs) on raw text. We also explore the topic distribution by examining the emotion topic variation in an emotion topic diagram.

Key words: word emotion classification; complex emotion; emotion intensity prediction; emotion-topic variation; hierarchical Bayesian network

I. INTRODUCTION

Text emotion analysis focuses on finding human emotions from different levels of text [1-5]. Most studies on sentence and document emotion classification relies on the word emotions [6-8]. Some previous work has proved that accurate word emotions do help increase the performance of sentence emotion classification (14.36% of precision increment and 8.05% of recall increment). However, currently the word emotions are only available for emotion lexicons, which are static and do not consider the contextual information. In this paper, we provide a Gibbs sampling method, which carefully introduce topic features and context information into the procedure of predicting the word emotions from raw text.

We define the text emotion classification as predicting the complex emotions, especially the word emotions and the emotion intensities from raw text. A complex emotion is defined as the combination of one or more singular emotions from the eight basic emotion categories: joy, love, expectation, surprise, anxiety, sorrow, anger and hate. The phenomenon of expressing multiple emotions in a word (or a sentence) is quite common in the real text. The following is an example sentence from Ren-CECps [9] which is a well annotated Chinese Emotion Corpus:

ht0.2|so0.9 I really want to: ex0.6 give up:ht0.3|so0.6 .

The sentence is annotated with 2 emotions: hate and sorrow, and each with a decimal figure (0.2 and 0.9) as the intensity of the corresponding emotion. The word “want to” indicates expectation of 0.6 intensity and “give up” indicates the combination of hate 0.3 and sorrow 0.6. The emotion intensity evaluates the strength which is an important feature of an emotion. In our example, the sorrow emotion is supposed to be stronger than hate in the sentence. We predict the emotion intensity for each singular emotion within the complex emotion.

Finally, we cluster the words into topics according to the distribution of the words among the corpus, as well as the emotion distribution. A topic, as defined in traditional topic models such as LDA, is a sorted list of words from the vocabulary, according to the probability that the word belongs to the topic. Specifically in this study of emotion topic variation, the order of words in the list is determined not only by the probability of the word but also by the probability that the word's emotion belongs to a certain topic. We model such distributions by employing the latent proportional variables and the latent structures in a hierarchical Bayesian network.

The performance of the Word Emotion Topic (WET) model is examined in following aspects. First, we evaluate the precision of complex emotion classification, in which both the singular emotions and the combined emotions of words in the prediction are examined. We also compare our result of singular emotion classification with the traditional text labeling methods including Hidden Markov Model and Conditional Random Fields (CRFs) on the same data set. Second, we examine the predicted emotion intensities for each predicted singular emotion. Third, we estimate the topic prediction by examining the emotion topic variation through an emotion topic diagram. All these emotional and topical results are drawn from raw text by the WET model, which means we do not need any complicated language features such as syntactic trees or semantic labels.

The rest of this paper is organized as follows. Section II introduces some related work. Section III describes the hierarchical Bayesian network for word emotion topic prediction. Section IV depicts the details of our inference method. Section V evaluates our method for the complex word emotion classification, the emotion intensity analysis and the emotion topic variation. Finally Section VI gives the conclusion and lists some future work.

II. RELATED WORK

To our knowledge, no study has investigated the field of predicting complex word emotions and emotion intensities, nor examined the variation of word emotions with automatically generated topics.

The first study of word emotions is the development of a fuzzy affect lexicon [6] to describe human emotions from English words. The fuzzy lexicon was built through collecting affect words from different sources including the newspaper articles by Mark Kantrowitz and the WordNet. Later, Ref. [7] also studied the word emotions based on WordNet, and built the Wordnet-Affect lexicon. The later work [8] employed the word emotion information from the Internet, especially the blog articles, by analyzing the co-occurrence of words and some special emotion tags and built the Weblog emotion lexicon. Recently, a moderated-sized high-quality emotion lexicon [10] was built by manually annotating emotions on the Amazon online service texts.

Emotional features are playing vital roles in text emotion analysis [11]. Emotions extracted from the emotion lexicons are statistic without considering the context information, while in the real text, word emotions can be affected by the context. The emotion lexicon-based text emotion classification could be seriously affected by the quality of emotion lexicons [12], which enlightened study of word emotion parsing in Refs. [13-14]. Ref. [13] learned a CRF model through an automatically generated emotion corpus of 1 000 sentences, while in Ref. [14] a word emotion parser was built by considering the sentence level context information such as the degree words and the negative modifications. However, these word emotion parsing applications relying on traditional CRF models only got around 60% to 70% precision scores, which are still far from the real world applications.

Most sentence and document emotion analysis studies were found taking information from emotion of words [15-16]. And later studies proved that the complex word emotions [17] which combine different singular emotions together effectively helped improve the performance of sentence emotion classification.

Ⅲ. WORD EMOTION TOPIC MODEL

3.1 The hierarchical Bayesian network

The method we use to predict the emotions (intensities) and topics from raw text is called the hierarchical Bayesian network, which is a directed graphical model [18] with nodes (shaded or not) representing the (visible or latent) random variables and the directed edges which indicate the conditional probabilities of child nodes given the values of the parent nodes. Figure 1 shows the WET model in a plate notation.

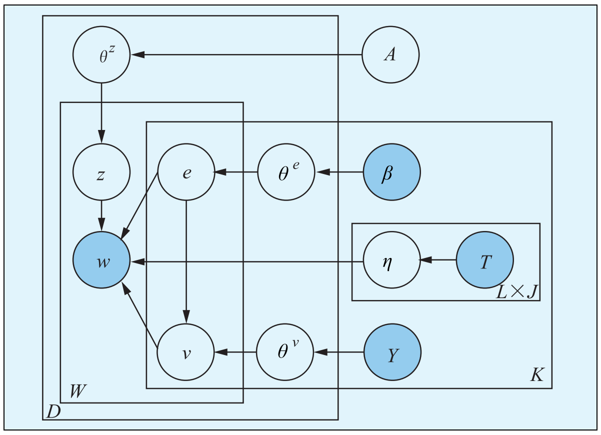

Fig.1 *The word emotion topic model in a plate notation*

The whole figure represents a joint probability of all the random variables in a corpus. The squares are called plates, which encapsulate random variables that appear multiple times in the model. For example, we have totally D documents in the corpus, so we draw the capital D at the corner to represent that all random variables within this square will repeatedly appear for D times. Each document may have W words, so word level variables will further repeat for W times. We use K to represent the number of emotion categories and k to index each singular emotion. Capital letter L counts the emotion intensity levels and l indexes a particular intensity level. N counts the total number of words in the vocabulary and t indexes each word in the vocabulary.

We have totally K(8) binary valued emotion variables e for each word. And each emotion variable e has a corresponding emotion intensity v. Although the corpus uses decimal figures to represent emotion intensities, they are in fact discrete random variables. We can use integer values from 1 to 10 to indicate the levels of emotion intensities. Random variable z is also discrete valued, representing the topic identity.

We do not know the exact distributions of the discrete random variables e, v, z, w, but we can use latent random variables θ^e, θ^v, θ^z and η to represent the probability of each possible assignment to these variables. And to represent our confidence of such probability measures (θ^e, θ^v, θ^z and η), we use the Dirichlet (Beta) distribution, which are distributions on distributions. And the observations β, γ, τ and A are the parameters of each Dirichlet (Beta) distribution.

All indexing variables and random variables in this hierarchical Bayesian network are listed in Table I.

Table I *The random variables and indexing variables in the word emotion topic model*

K	#Emotion	k	Emotion index
L	#Emotionintensity	l	Emotion intensity index
J	#Topic	j	Topic index
N	#Word in vocabulary	t	Word index in vocabulary
D	#Document	d	Document index
W_d	#Dords in d	i	Word index in d
E	Word emotion	v	Word emotion intensity
Z	Topic	w	Word
β	Observed word emotion	γ	Observed word emotion intensities
A	Observed topics	τ	Observed quadruple (k, l, j, t)
θ^e	Proportion of word emotion	θ^v	Proportion of word emotion intensity
θ^z	Proportion of topic	η	Proportion of quadruple (k, l, j, t)

3.2 The distribution assumptions

The advantage of such a method compared to traditional machine learning methods such as HMM and CRF is that we can incorporate domain knowledge into the learning procedure. In this study, we employ the knowledge of emotions and topics in the form of probabilistic distributions.

As we have mentioned: the binary valued emotion variable e follows the Bernoulli distribution; the emotion intensity v of a particular emotion e follows the Categorical distribution, if the emotion is actually assigned; otherwise, the intensity is set to 0 with probability 1; the topic variable z follows the Categorical distribution; and word w follows the Categorical distribution parametrized by η_{evz} given the emotion e, emotion intensity v and topic z; θ^e follows the Beta distribution, which is the conjugate prior of Bernoulli; θ^v, θ^z and η follow the Dirichlet distribution, parametrized by the observed variables.

We explicitly list the distribution formulas as follows.

$$e_{dik} \sim \mathrm{Ber}(\theta^e_{dk})$$
$$v_{dik} \mid e_{dik} \sim \mathrm{Cate}(\theta^v_{dk});\ \text{if } e_{dik} = 1$$
$$v_{dik} \mid e_{dik} \sim p(v_{dik} = 0) = 1;\ \text{if } e_{dik} = 0$$
$$z_{di} \sim \mathrm{Cate}(\theta^z_d)$$
$$w_{di} \mid e_{di}, v_{di}, z_{di} \sim \mathrm{Cate}(\eta_{e_{di} v_{di} z_{di}})$$
$$\theta^e_{dk} \sim \mathrm{Beta}(\beta_k)$$
$$\theta^v_{dk} \sim \mathrm{Dir}(\gamma_k)$$
$$\theta^z_{dk} \sim \mathrm{Dir}(A)$$
$$\eta_{klj} \sim \mathrm{Dir}(\tau_{klj})$$

3.3 The generation procedure

Our emotion topic model is a generative model, which means we can generate the value of a child node given the values of the parent nodes and the particular conditional probability. And by repeating this procedure to each variable in the WET model, we can finally generate an entire emotion corpus.

The algorithm of generation is shown in Algorithm 1 below.

Algorithm 1: Generating variables in the word emotion topic model

```
For emotion document d < D do
 Draw topic proportion θ^z_d ~ Dir(A)
 For each emotion k < K do
  Draw emotion proportion θ^e_dk ~ Beta(β_k)
  Draw emotion intensity proportion θ^v_dk ~ Dir(γ_k)
 End for
 For each word i < W_d do
  For each emotion k < K do
   Draw word emotion e_dik ~ Ber(θ^e_dk)
   Draw emotion intensity v_dik | e_dik ~ Cate(θ^v_dk)
  End for
  Draw topic z_di ~ Cate(θ^z_d)
  Draw word w_di | e_di, v_di, z_di ~ Cate(η_{e_di v_di z_di})
 End for
  End for
```

The generation procedure follows the direction of arrows in this graph: for each document d, we draw a topic proportion. Then for each emotion k, we draw the emotion proportion, and the emotion intensity proportion. For each word i, and for each singular emotion k of this word, we draw the binary valued emotion indicator e_{dik}, and the corresponding emotion intensity v_{dik} if the value of e_{dik} is not false. After generating the complex emotion, we can generate the topic z_{di} and word w_{di} from Categorical distributions.

IV. INFERENCE

4.1 Gibbs sampling method

So far we have defined the hierarchical Bayesian network to model the emotion and topic knowledge of the words. The generation procedure shows the steps to generate the values of visible variables (words) from latent variables (emotions, emotion intensities, topics) and latent structure (proportional variables). And the inference procedure is the reverse of generation, in which the latent variables can be generated from visible variables, suppose we can calculate the conditional probability of the latent variable given values of other variables. Actually, we are using a collapsed Gibbs sampling method [19-20] by integrating out the proportion variables (θ^e, θ^v, θ^z, η).

Before sampling, we assign some random values to the latent variables. Then we draw the value of each latent variable given the values of all other visible and latent variables. The Gibbs sampling procedure is an iterative procedure, which means we have to repeat the step drawing the value of each latent random variable for a relatively large number of times until convergence. It is also proved that through large amount of repetitions, we can get the true distribution of the latent random variables.

The algorithm of Gibbs sampling is shown in Algorithm 2.

Algorithm 2: Gibbs sampling for word emotion topic model

For $m = 1$ to M Gibbs sampling iterations do
 For $d < D$ do
  For $i < W_d$ do
   For $k < K$ do
    Calc $p(v_{\mathrm{dik}} \mid w, z, e, v_{-\mathrm{dik}}, A, \beta, \gamma, \tau)$
    Sample v_{dik}
    Calc $p(e_{\mathrm{dik}} \mid w, z, e_{-\mathrm{dik}}, v, A, \beta, \gamma, \tau)$
    Sample e_{dik}
   End for
   Calc $p(z_{di} \mid w, z_{-di}, e, v, A, \beta, \gamma, \tau)$
   Sample z_{di}
  End for
 End for
End for

4.2 Probability derivations

Until now, we have provided the procedure of predicting emotions and topics of words, leaving only one problem: how to calculate the conditional probabilities of a single latent variable given values of all other variables?

Because we select the conjugate priors, the calculations are straight forward. Eq. (1) is the conditional probabilities of emotion variable e_{dik} given all other variables:

$$\begin{aligned} & p(e_{\mathrm{dik}} \mid w, z, e_{-\mathrm{dik}}, v, A, \beta, \gamma, \tau) \\ & \propto p(w_{di}, e_{\mathrm{dik}} \mid w_{-di}, z, e_{-\mathrm{dik}}, v, A, \beta, \gamma, \tau) = \\ & \frac{p(w, z, e, v, A, \beta, \gamma, \tau)}{p(w_{-di}, z, e_{-\mathrm{dik}}, v, A, \beta, \gamma, \tau)} = \\ & \frac{p(w, e \mid z, v, A, \beta, \gamma, \tau)}{p(w_{-di}, e_{-\mathrm{dik}} \mid z, v, A, \beta, \gamma, \tau)} \end{aligned} \tag{1}$$

Variable e_{dik} is the parent of word variable w_{di} in the WET model. And since the value of word is visible, the conditional probability turns to be proportional to the joint probability of w_{di} and e_{dik} condition on the values of all other variables. Finally, we apply the Bayesian and product rules to the conditional probability, and get the division formula as Eq. (1).

The conditional probabilities of emotion intensity variable v_{dik} and topic variable z_{di} are listed as follows:

$$\begin{aligned} & p(v_{\mathrm{dik}} \mid w, z, e, v_{-\mathrm{dik}}, A, \beta, \gamma, \tau) \\ & \propto \frac{p(w, v \mid z, e, A, \beta, \gamma, \tau)}{p(w_{-di}, v_{-\mathrm{dik}} \mid z, e, A, \beta, \gamma, \tau)} \end{aligned} \tag{2}$$

$$\begin{aligned} & p(z_{di} \mid w, z_{-di}, e, v, A, \beta, \gamma, \tau) \\ & \propto \frac{p(w, z \mid e, v, A, \beta, \gamma, \tau)}{p(w_{-di}, z_{-di} \mid e, v, A, \beta, \gamma, \tau)} \end{aligned} \tag{3}$$

We briefly look through the calculation of the probability of emotion variable e_{dik}. The joint probability in Eq. (1) is split into a word fraction and an emotion fraction:

$$p(w, e \mid z, v, A, \beta, \gamma, \tau) = p(w \mid z, e, v, \tau) p(e \mid \beta) \tag{4}$$

The emotion part is calculated through integrating out the proportion variable (θ^e). We take the advantage of the conjugate priors, and get the expression as follows:

$$\begin{aligned} & p(e \mid \beta) = \int p(e \mid \theta^e) p(\theta^e \mid \beta) \mathrm{d}\theta^e = \\ & \int \prod_{d=1}^{D} \prod_{i=1}^{W_d} \prod_{k=1}^{K} p(e_{\mathrm{dik}} \mid \theta^e_{dk}) \prod_{d=1}^{D} \prod_{k=1}^{K} p(\theta^e_{dk} = \mid \beta_k) \cdot \\ & \mathrm{d}\theta^e \prod_{dk} \int \prod_{i=1}^{W_d} (\theta^e_{dk})^{e_{\mathrm{dik}}} (1 - \theta^e_{dk})^{(1 - e_{\mathrm{dik}})} \cdot \\ & \frac{\Gamma(\beta^1_k + \beta^0_k)}{\Gamma(\beta^1_k) \Gamma(\beta^0_k)} (\theta^e_{dk})^{\beta^1_k - 1} (1 - \theta^e_{dk})^{\beta^0_k - 1} \mathrm{d}\theta^e_{dk} = \\ & \prod_{dk} \frac{\Gamma(\beta^1_k + \beta^0_k)}{\Gamma(\beta^1_k) \Gamma(\beta^0_k)} \frac{\Gamma(n_{dk} + \beta^1_k) \Gamma(W_d - n_{dk} + \beta^0_k)}{\Gamma(W_d + \beta^1_k + \beta^0_k)} \end{aligned} \tag{5}$$

We can see that the conditional probability finally falls into some fundamental operations on the statistics variables: n_{dk} is the number of emotion k in document d; W_d is the total number of words in document d, β^1_k and β^0_k count the occurrence and un-occurrence of the emotion k in the training set.

We derive the word probability in the same way.

$$\begin{aligned} & p(w \mid z, e, v, \tau) = \\ & \int p(w \mid z, e, v, \eta) p(\eta \mid \tau) \mathrm{d}\eta = \\ & \prod_{klj} \frac{\Gamma(\sum_{t=1}^{N} \tau^1_{kljt})}{\prod_{t=1}^{N} \Gamma(\tau^1_{kljt})} \frac{\prod_{t=1}^{N} \Gamma(n^1_{kljt} + \tau^1_{kljt})}{\Gamma(\sum_{t=1}^{N} (n^1_{kljt} + \tau^1_{kljt}))} \cdot \\ & \prod_{kj} \frac{\Gamma(\sum_{t=1}^{N} \tau^0_{kjt})}{\prod_{t=1}^{N} \Gamma(\tau^0_{kjt})} \frac{\prod_{t=1}^{N} \Gamma(n^0_{kjt} + \tau^0_{kjt})}{\Gamma(\sum_{t=1}^{N} (n^0_{kjt} + \tau^0_{kjt}))} \end{aligned} \tag{6}$$

By taking the emotion and word probabilities into Eq. (1), we have the conditional probability as follows:

$$
p(e_{\mathrm{dik}} \mid w, z, e_{-\mathrm{dik}}, v, A, \beta, \gamma, \tau) \propto
\frac{\prod_{klj} \frac{\prod_{t=1}^{N} \Gamma(n_{kljt}^{1} + \tau_{kljt}^{1})}{\Gamma(\sum_{t=1}^{N} (n_{kljt}^{1} + \tau_{kljt}^{1}))} \prod_{kj} \frac{\prod_{t=1}^{N} \Gamma(n_{kjt}^{0} + \tau_{kjt}^{0})}{\Gamma(\sum_{t=1}^{N} (n_{kjt}^{0} + \tau_{kjt}^{0}))}}{\left[\prod_{klj} \frac{\prod_{t=1}^{N} \Gamma(n_{kljt}^{1} + \tau_{kljt}^{1})}{\Gamma(\sum_{t=1}^{N} (n_{kljt}^{1} + \tau_{kljt}^{1}))} \prod_{kj} \frac{\prod_{t=1}^{N} \Gamma(n_{kjt}^{0} + \tau_{kjt}^{0})}{\Gamma(\sum_{t=1}^{N} (n_{kjt}^{0} + \tau_{kjt}^{0}))}\right]_{-di}} \cdot
\frac{\prod_{dk} \frac{\Gamma(n_{dk} + \beta_k^1)\Gamma(W_d - n_{dk} + \beta_k^0)}{\Gamma(W_d + \beta_k^1 + \beta_k^0)}}{\left[\prod_{dk} \frac{\Gamma(n_{dk} + \beta_k^1)\Gamma(W_d - n_{dk} + \beta_k^0)}{\Gamma(W_d + \beta_k^1 + \beta_k^0)}\right]_{-\mathrm{dik}}} \tag{7}
$$

Eq. (7) is not simple enough: we have too much product operations and Gamma functions. Also, the numerator and denominator are quite similar except some particular variables. These lead to the simplification of calculation.

We expand the denominator and numerator of word probability with emotion $e_{di*}=1$ into two parts: the single word w_{di} part which is different in Eq. (8) and Eq. (9), and the all-other-words part which is exactly the same in Eq. (8) and Eq. (9), and therefore it can be removed out as follows.

$$
\left[\prod_{klj} \frac{\prod_{t=1}^{N} \Gamma(n_{kljt}^{1} + \tau_{kljt}^{1})}{\Gamma(\sum_{t=1}^{N} (n_{kljt}^{1} + \tau_{kljt}^{1}))}\right]_{-di} =
$$

$$
\frac{\prod_{k \in K_{di}^1} \Gamma(n^1_{kv_{\mathrm{dik}} z_{di} w_{di}} + \tau^1_{kv_{\mathrm{dik}} z_{di} w_{di}} - 1)}{\prod_{k \in K_{\mathrm{di}}^1} \Gamma(\sum_{t=1}^{N} (n^1_{kv_{\mathrm{dik}} z_{di} t} + \tau^1_{kv_{\mathrm{dik}} z_{di} t}) - 1)}
\frac{\left[\prod_{kljt} \Gamma(n_{kljt}^{1} + \tau_{kljt}^{1})\right]_{-z_{ei} w_{di}}}{\left[\prod_{klj} \Gamma(\sum_{t=1}^{N} (n_{kljt}^{1} + \tau_{kljt}^{1}))\right]_{-z_{di}}} \tag{8}
$$

$$
\left[\prod_{klj} \frac{\prod_{t=1}^{N} \Gamma(n_{kljt}^{1} + \tau_{kljt}^{1})}{\Gamma(\sum_{t=1}^{N} (n_{kljt}^{1} + \tau_{kljt}^{1}))}\right]_{-di} =
$$

$$
\frac{\prod_{k \in K_{di}^1} \Gamma(n^1_{kv_{\mathrm{dik}} z_{di} w_{di}} + \tau^1_{kv_{\mathrm{dik}} z_{di} w_{di}})}{\prod_{k \in K_{di}^1} \Gamma(\sum_{t=1}^{N} (n^1_{kv_{\mathrm{dik}}} + \tau^1_{kv_{\mathrm{dik}}}))}
\frac{\left[\prod_{kljt} \Gamma(n_{kljt}^{1} + \tau_{kljt}^{1})\right]_{-z_{ei} w_{di}}}{\left[\prod_{klj} \Gamma(\sum_{t=1}^{N} (n_{kljt}^{1} + \tau_{kljt}^{1}))\right]_{-z_{di}}} \tag{9}
$$

Also we have the denominator of the word w_{di} with emotion $e_{di*}=0$ separated into two parts:

$$
\left[\prod_{kj} \frac{\prod_{t=1}^{N} \Gamma(n_{kjt}^{0} + \tau_{kjt}^{0})}{\Gamma(\sum_{t=1}^{N} (n_{kjt}^{0} + \tau_{kjt}^{0}))}\right]_{-di} =
$$

$$
\frac{\prod_{k \in K_{di}^0} \Gamma(n^0_{kz_{di} w_{di}} + \tau^0_{kz_{di} w_{di}} - 1)}{\prod_{k \in K_{di}^0} \Gamma(\sum_{t=1}^{N} (n^0_{kz_{di} t} + \tau^0_{kz_{di} t}) - 1)}
\frac{\left[\prod_{kjt} \Gamma(n_{kjt}^{0} + \tau_{kjt}^{0})\right]_{-z_{ei} w_{di}}}{\left[\prod_{kj} \Gamma(\sum_{t=1}^{N} (n_{kjt}^{0} + \tau_{kjt}^{0}))\right]_{-z_{di}}} \tag{10}
$$

$$
\left[\prod_{kj} \frac{\prod_{t=1}^{N} \Gamma(n_{kjt}^{0} + \tau_{kjt}^{0})}{\Gamma(\sum_{t=1}^{N} (n_{kjt}^{0} + \tau_{kjt}^{0}))}\right]_{-di} =
$$

$$
\frac{\prod_{k \in K_{di}^0} \Gamma(n^0_{kz_{di} w_{di}} + \tau^0_{kz_{di} w_{di}})}{\prod_{k \in K_{di}^0} \Gamma(\sum_{t=1}^{N} (n^0_{kz_{di} t} + \tau^0_{kz_{di} t}))}
\frac{\left[\prod_{kjt} \Gamma(n_{kjt}^{0} + \tau_{kjt}^{0})\right]_{-z_{ei} w_{di}}}{\left[\prod_{kj} \Gamma(\sum_{t=1}^{N} (n_{kjt}^{0} + \tau_{kjt}^{0}))\right]_{-z_{di}}} \tag{11}
$$

And again, only the singular word parts are different in Eq. (10) and Eq. (11).

We also expand the denominator and numerator of emotion e_{dik} in Eq. (12) and Eq. (13):

$$
\left[\prod_{dk} \frac{\Gamma(n_{dk} + \beta_k^1)\Gamma(W_d - n_{dk} + \beta_k^0)}{\Gamma(W_d + \beta_k^1 + \beta_k^0)}\right]_{-\mathrm{dik}} =
\begin{cases}
\frac{\Gamma(n_{d_d e_{\mathrm{dik}}} + \beta^1_{e_{\mathrm{dik}}} - 1)\Gamma(W_{d_d} - n_{d_d e_{\mathrm{dik}}} + \beta^0_{e_{\mathrm{dik}}})}{\Gamma(W_{d_d} + \beta^1_{e_{\mathrm{dik}}} + \beta^0_{e_{\mathrm{dik}}} - 1)} \times \frac{\left[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(n_{dk} + \beta_k^1)\Gamma(W_d - n_{dk} + \beta_k^0)\right]_{-d_d e_{\mathrm{dik}}}}{\left[\prod_{d=1}^{\mathrm{D}}\prod_{k=1}^{K} \Gamma(W_d + \beta_k^1 + \beta_k^0)\right]_{-d_d e_{\mathrm{dik}}}}; & \text{if } e_{\mathrm{dik}}^{\mathrm{old}} = 1 \\
\frac{\Gamma(n_{d_d e_{\mathrm{dik}}} + \beta^1_{e_{\mathrm{dik}}})\Gamma(W_{d_d} - n_{d_d e_{\mathrm{dik}}} + \beta^0_{e_{\mathrm{dik}}} - 1)}{\Gamma(W_{d_d} + \beta^1_{e_{\mathrm{dik}}} + \beta^0_{e_{\mathrm{dik}}} - 1)} \times \frac{\left[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(n_{dk} + \beta_k^1)\Gamma(W_d - n_{dk} + \beta_k^0)\right]_{-d_d e_{\mathrm{dik}}}}{\left[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(W_d + \beta_k^1 + \beta_k^0)\right]_{-d_d e_{\mathrm{dik}}}}; & \text{if } e_{\mathrm{dik}}^{\mathrm{old}} = 0
\end{cases} \tag{12}
$$

$$\prod_{dk} \frac{\Gamma(n_{dk}+\beta_k^1)\Gamma(W_d - n_{dk}+\beta_k^0)}{\Gamma(W_d+\beta_k^1+\beta_k^0)} = \begin{cases} \dfrac{\Gamma(n_{d_d e_{dik}}+\beta^1_{e_{dik}})\Gamma(W_{d_d} - n_{d_d e_{dik}} + \beta^0_{e_{dik}})}{\Gamma(W_{d_d}+\beta^1_{e_{dik}}+\beta^0_{e_{dik}})} \times \\ \dfrac{[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(n_{dk}+\beta_k^1)\Gamma(W_d-n_{dk}+\beta_k^0)]_{-d_d e_{dik}}}{[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(W_d+\beta_k^1+\beta_k^0)]_{-d_d e_{dik}}}; \\ \quad \text{if } e^{\text{oid}}_{\text{dik}} = e^{\text{new}}_{\text{dik}} \\ \dfrac{\Gamma(n_{d_d e_{dik}}+\beta^1_{e_{dik}}-1)\Gamma(W_{d_d} - n_{d_d e_{dik}} + \beta^0_{e_{dik}}+1)}{\Gamma(W_{d_d}+\beta^1_{e_{dik}}+\beta^0_{e_{dik}})} \times \\ \dfrac{[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(n_{dk}+\beta_k^1)\Gamma(W_d-n_{dk}+\beta_k^0)]_{-d_d e_{dik}}}{[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(W_d+\beta_k^1+\beta_k^0)]_{-d_d e_{dik}}}; \\ \quad \text{if } e^{\text{oid}}_{\text{dik}} = 1 \text{ and } e^{\text{new}}_{\text{dik}} = 0 \\ \dfrac{\Gamma(n_{d_d e_{dik}}+\beta^1_{e_{dik}}+1)\Gamma(W_{d_d} - n_{d_d e_{dik}} + \beta^0_{e_{dik}}-1)}{\Gamma(W_{d_d}+\beta^1_{e_{dik}}+\beta^0_{e_{dik}})} \times \\ \dfrac{[\prod_{d=1}^{D}\prod_{k=1}^{K} \Gamma(n_{dk}+\beta_k^1)\Gamma(W_d-n_{dk}+\beta_k^0)]_{-d_d e_{dik}}}{[\prod_{d=1}^{D}\prod_{k=1}^{K} \Lambda(W_d+\beta_k^1+\beta_k^0)]_{-d_d e_{dik}}}; \\ \quad \text{if } e^{\text{oid}}_{\text{dik}} = 0 \; and \; e^{\text{new}}_{\text{dik}} = 1 \end{cases} \tag{13}$$

Finally, by taking Eq. (8) to Eq. (13) into Eq. (7), we have the conditional probability of emotion e_{dik} as:

$$p(e_{\text{dik}} \mid w, z, e_{-\text{dik}}, v, A, \beta, \gamma, \tau) \propto \begin{cases} \dfrac{n^1_{e_{dik}v_{dik}z_{di}w_{di}} + \tau^1_{e_{dik}v_{dik}z_{di}w_{di}} - 1}{\sum_{t=1}^{N} (n^1_{e_{dik}v_{dik}z_{di}t} + \tau^1_{e_{dik}v_{dik}z_{di}t}) - 1} \dfrac{n_{d_d e_{dik}} + \beta^1_{e_{dik}} - 1}{W_{d_d} + \beta^1_{e_{dik}} + \beta^0_{e_{dik}} - 1}; \\ \quad \text{if } e^{\text{old}}_{\text{dik}} = 1 \text{ and } e^{\text{new}}_{\text{dik}} = 1 \\ \dfrac{n^0_{e_{dik}} + \tau^0_{e_{dik}z_{di}w_{di}} - 1}{\sum_{t=1}^{N} (n^0_{e_{dik}z_{di}t} + \tau^0_{e_{dik}z_{di}t}) - 1} \dfrac{W_{d_d} - n_{d_d e_{dik}} + \beta^0_{e_{dik}}}{W_{d_d} + \beta^1_{e\,\text{dik}} + \beta^0_{e_{dik}} - 1}; \\ \quad \text{if } e^{\text{old}}_{\text{dik}} = 1 \text{ and } e^{\text{new}}_{\text{dik}} = 0 \\ \dfrac{n^1_{e_{dik}v_{dik}z_{di}\mathrm{w}_{di}} + \tau^1_{e_{dik}v_{dik}z_{di}w_{di}} - 1}{\sum_{t=1}^{N} (n^1_{e_{dik}v_{dik}z_{di}t} + \tau^1_{e_{dik}v_{dik}z_{di}t}) - 1} \dfrac{n_{d_d e_{dik}} + \beta^1_{e_{dik}}}{W_{d_d} + \beta^1_{e_{dik}} + \beta^0_{e_{dik}} - 1}; \\ \quad \text{if } e^{\text{old}}_{\text{dik}} = 0 \text{ and } e^{\text{new}}_{\text{dik}} = 1 \\ \dfrac{n^0_{e_{dik}z_{di}w_{di}} + \tau^0_{e_{dik}} - 1}{\sum_{t=1}^{N} (n^0_{e_{dik}z_{di}t} + \tau^0_{e_{dik}z_{di}t}) - 1} \dfrac{W_{d_d} - n_{d_d e_{dik}} + \beta^0_{e_{dik}} - 1}{W_{d_d} + \beta^1_{e_{dik}} + \beta^0_{e_{dik}} - 1}; \\ \quad \text{if } e^{\text{old}}_{\text{dik}} = 0 \text{ and } e^{\text{new}}_{\text{dik}} = 0 \end{cases} \tag{14}$$

which is the final equation we use to draw samples of word emotions in Gibbs sampling.

Following the same procedure, we have the conditional probability of the emotion intensity v_{dik} given the other variables:

$$p(v_{\text{dik}} \mid w, z, e, v_{-\text{dik}}, A, \beta, \gamma, \tau) \propto \begin{cases} \dfrac{n^1_{e_{dik}v_{dik}z_{di}w_{di}} + \tau^1_{e_{dik}v_{dik}z_{di}w_{di}} - 1}{\sum_{t=1}^{N} (n^1_{e_{dik}v_{dik}z_{di}t} + \tau^1_{e_{dik}v_{dik}z_{di}t}) - 1} \dfrac{n_{d_d e_{dik} v_{dik}} - 1}{\sum_{l=1}^{L} (n_{d_d e_{dik} l} + \gamma_{e_{dik} l}) - 1}; \\ \quad \text{if } e^{\text{old}}_{\text{dik}} = 1 \\ 1; \quad \text{if } e^{\text{old}}_{\text{dik}} = 0 \end{cases} \tag{15}$$

and the conditional probability of the topic z_{di} given all other variables:

$$p(z_{di} \mid w, z_{-di}, e, v, A, \beta, \gamma, \tau) = \prod_{k \in K^1} \frac{n^1_{kv_{dik}z_{di}w_{di}} + \tau^1_{kv_{dik}z_{di}w_{di}} - 1}{\sum_{t=1}^{N} (n^1_{kv_{dik}z_{di}t} + \tau^1_{kv_{dik}z_{di}t}) - 1} \cdot \prod_{k \in K^0} \frac{n^0_{kz_{di}w_{di}} + \tau^0_{kz_{di}w_{di}} - 1}{\sum_{t=1}^{N} (n^0_{kz_{di}t} + \tau^0_{kz_{di}t}) - 1} \cdot \frac{n_{d_d z_{di}} + A_{z_{di}} - 1}{\sum_{j=1}^{J} (n_{d_d j} + A_j) - 1} \tag{16}$$

V. EVALUATION

We apply the WET model to emotion and topic prediction from raw text. In this experiment, we extract 1 147 blog articles from Ren-CECps, and use 918 articles for training, and 229 articles for testing. We use the precision scores to assess the performance of our model basically on 3 aspects: emotion prediction, emotion intensity analysis, and emotion topic variation.

We also compare the precisions of our emotion prediction with traditional parsing models such as Hidden Markov Models (HMM) and CRF.

5.1 Emotion prediction

Figure 2 shows the precisions of singular word emotion classification when choosing varied number of topics in the WET model. There are totally 9 classes in this classification task, including the eight basic singular emotions joy, love, expectation, surprise, anxiety, sorrow, anger and hate, and a NoEmotion label with respect to the words on which we do not predict any emotion.

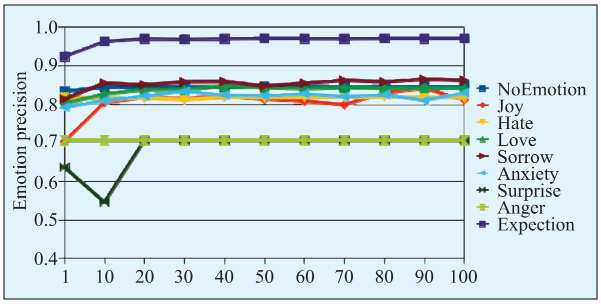

Fig.2 *Precisions of singular word emotions with varied topic numbers*

As the number of topic increases, the results tend to be stable. For the emotion of expectation, we have all the precision scores over 95%, and for emotions of joy, love, anxiety, sorrow, hate and NoEmotion we have the precision scores over 80%. Prediction of emotion surprise and anger seems harder, with precision scores of around 70%.

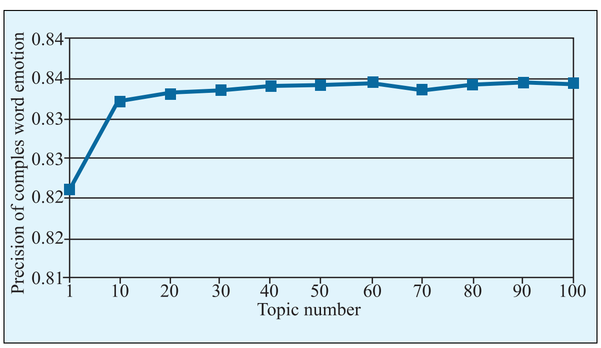

Fig.3 *Precisions of complex word emotions with varied topic numbers*

Figure 3 shows the result of complex word emotion prediction, with varied number of topics in the model setting. Prediction of complex emotion is thought to be correct only if every predicted singular emotion is matched with the standard label. This is a very strict condition since within the complex emotions singular emotions may have different emotion intensities, and some weak emotions are very difficult even for humans to predict.

The precision score increases as we set larger number of topics, and achieves the peak of 83.46 at 90 topics, which is a quite promising result compared to the HMM and CRF results. Figure 4 compares the results of HMM and CRF and our word emotion topic model on the same training and testing set.

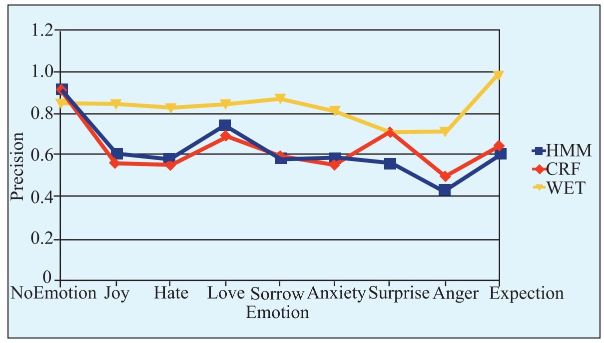

Fig.4 *Comparing precisions of singular word emotions under HMM, CRF and WET*

The HMM model gets an averaged precision on totally 9 emotion categories of 62.20%. The CRF gets a little higher mean precision of 63.46%. Both are much lower than the WET model of 82.24%.

The comparison indicates that: first, the word emotion prediction is a difficult task since traditional labeling methods achieves around 60% precision scores; second, the hierarchical Bayesian network do help improving the emotion classification with more than 20% of increment.

5.2 Emotion intensity prediction

We evaluate the emotion intensity predictions with varied topic numbers. Figure 5 shows the result.

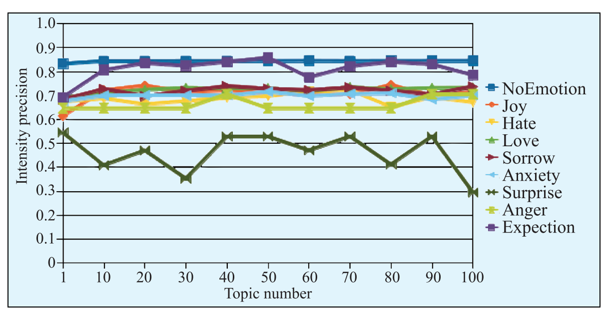

Fig.5 *Precisions of word emotion intensities with varied number of topics*

Most singular emotions get precision scores over 60, and 5 emotions including joy, love, expectation, sorrow and NoEmotion achieve over 70 precisions. The averaged emotion intensity precision gets the highest score of 72.27 with topic number of 40.

The prediction of the emotion intensity is thought to be correct only if both the emotion identity and emotion intensity value are matched with the standard label, which means the errors in emotion prediction will be accumulated to the errors in emotion intensity prediction. Therefore, to evaluate the true performance of the WET model on emotion intensity, we calculate the precision scores of emotion intensity prediction by considering only the intensity values of the already correctly predicted singular emotions.

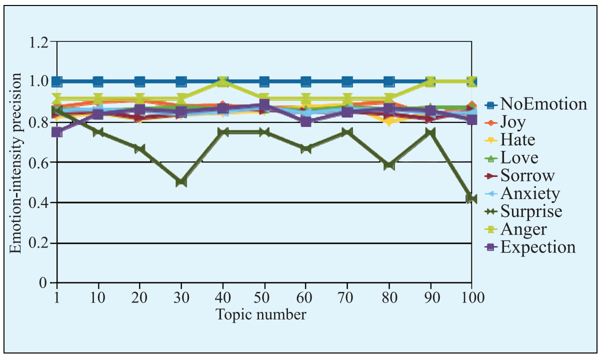

Fig.6 *Precision of emotion intensity for correctly classified emotions with varied number of topics*

Figure 6 shows the precisions of the emotion intensity for correctly classified emotions. The predicted emotion intensities on most correctly predicted emotions get over 80% precision scores. The "surprise" intensity prediction varies a lot, without sufficient training examples.

5.3 Emotion-topic variation

Topics in traditional topic models such as LDA are in fact the sorted lists of words from the vocabulary. Each word can appear in a topic, and the only difference among topics is the order of words. Topics can also be viewed as clusters. In LDA, the order of words in each topic is determined by the probability that the words belong to the corresponding cluster.

In our word emotion topic model, topics are the sorted word lists from the vocabulary. However, the order of words in each topic is determined by the probability that the words together with the corresponding word emotion belong to the topic cluster. We can examine this by the emotion-topic variation.

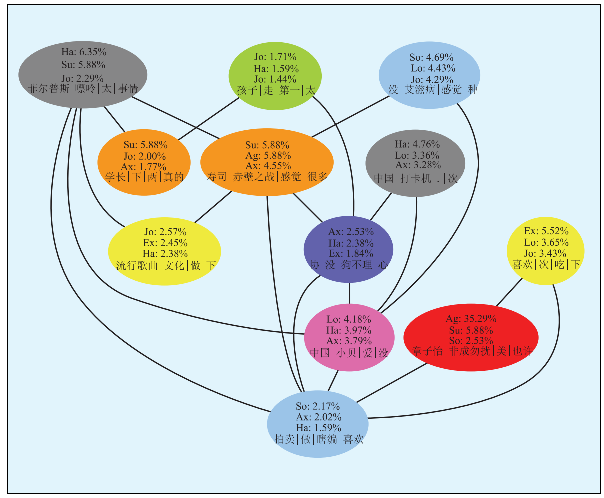

Fig.7 *Emotion-topic diagram*

Figure 7 shows part of the emotion topic diagram generated by the WET model when choosing 40 topics. Each ellipse is a topic and topics get linked to each other according to their lexical similarities. The topics are colored to represent their chief emotions, e.g. green for joy, grey for hate, blue for sorrow. We list the top four words in each topic, top three emotions, and the percentage of each emotion occurring in the corpus.

The words "pop music | culture | do | next" are clustered into a topic in the emotion topic diagram. In traditional topic models, this topic can be explained as representing some culture of popular music. And in this emotion topic model, we can get further emotion interpretation: it is probable that words in this topic express emotions of joy and expectation than others. The topic "like | time | eat | next" represents the emotion of expectation, love and joy, while the topic "auction | do | fabrication | like" represents the emotion of sorrow, anxiety and hate. Both topics contain the same word "like", but in different contexts the word indicates different emotions.

Finally, it has to be noticed that the emotion-topic variation in Figure 6 represents the particular distributions of emotions and topics on the specific set of text (the 229 blog articles in the testing set), rather than an universal emotion - topic distribution.

VI. CONCLUSIONS

In this paper, we employ a hierarchical Bayesian network to incorporate the domain knowledge of emotion and topic information, in the form of conditional probabilities, into the machine learning procedure. We use the Gibbs sampling method to predict the combination of singular word emotions and the emotion intensities. To our knowledge, other study has investigated the task of predicting the complex word emotions together with the emotion intensities.

To evaluate the performance of the word emotion topic model, we examine the precisions of singular word emotion prediction, complex word emotion prediction, word emotion intensity prediction and the emotion topic variation. As we increase the number of topics in the model, the averaged singular emotion precision and complex emotion precision achieve the peak of 82.24% and 83.46% at the setting of 90 topics. The precision of emotion intensity gets the best of 72.27% when choosing 40 topics. We compare the traditional parsing methods HMM and CRF on emotion prediction with our word emotion topic model. On average, our method outperforms HMM and CRF on singular emotion prediction by around 20% precision scores, which proves that the emotion prediction is not an easy task and that our model could effectively improve the result. Finally, we study the emotion-topic variation through the emotion topic diagram: topics can be further explained by the word emotions within the topics, and the word emotions vary under different topics (contexts).

When predicting the complex emotions of words, our WET model considers the context information such as the proportions of each emotion and topic for each word in a document. However, such proportions cannot reflect the structure information such as negative modifications, which could actually affect the word emotions. In the future study of emotion prediction, we will apply some results of the proposed method to the project of Robotics Cloud [21-22], and hope to introduce sequence structures as well as corresponding domain knowledge into the hierarchical Bayesian models.中国通信

Acknowledgements

This work was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research under Grant No.22240021; the Grant-in-Aid for Challenging Exploratory Research under Grant No. 21650030.

References

[1] MATSUMOTO K, REN Fuji, KUROIWA S. Emotion Estimation System Based On Emotion Occurrence Sentence Pattern[J]. Computational Intelligence, 2006, 4114: 902-911.

[2] XIANG Hua, JIANG Peilin, XIAO Shuang, *et al.* A Model of Mental State Transition Network [J]. IEEJ Transactions on Electronics, Information and Systems, 2007, 127(3): 434-442.

[3] REN Fuji. Affective Information Processing and Recognizing Human Emotion[J]. Electronic Notes in Theoretical Computer Science, 2009, 225: 39-50.

[4] REN Fuji, DAVID B. Advanced Information Retrieval[J]. Electronic Notes in Theoretical Computer Science, 2009, 225: 303-317.

[5] HAKAMATA A, REN Fuji, TSUCHIYA S. Human Emotion Model Based On Discourse Sentence For Expression Generation Of Conversation Agent[C]// Proceedings of the Natural Language Processing and Knowledge Engineering: Oct 19-22, 2008, Beijing, China: IEEE, 2008: 1-8.

[6] SUBASIC P, HUETTNER A. Affect Analysis of Text Using Fuzzy Semantic Typing[J]. Fuzzy Systems, 2001, 9(4): 483-496.

[7] VALITUTTI A, STRAPPARAVA C, STOCK O. Developing Affective Lexical Resources[J]. PsychNology Journal, 2004, 2(1): 61-83.

[8] YANG C, LIN K, CHEN H. Building Emotion Lexicon From Weblog Corpora[C]// Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions: June 2007, Prague: Association for Computational Linguistics, 2007: 133-136.

[9] QUAN Changqin, REN Fuji. A Blog Emotion Corpus for Emotional Expression Analysis in Chinese [J]. Computer Speech & Language, 2010, 24(4): 726-749.

[10] SAIF M, PETER D. Using Mechanical Turk To Create An Emotion Lexicon[C]// Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text: 2010, Los Angeles, California, USA: Association for Computational Linguistics: 26-34.

[11] LIU Gongshen, HE Wenlei, ZHU Jie, *et al.* Feature Representation Based on Sentimental Orientation Classification[J].

China Communications, 2011, 8(3): 90-89.
[12] VINCENT N, SAJIB D, SM N. Examining The Role Of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews[C]// Proceedings of the COLING/ACL on Main Conference Poster Sessions: July 2006, Sydney: Association for Computational Linguistics: 611-618.
[13] DIPANKAR D, SIVAJI B. Word to Sentence Level Emotion Tagging for Bengali Blogs [C]// Proceedings of the ACL-IJCNLP 2009 Conference Short Papers: 2009, Suntec, Singapore: ACL: 149-152.
[14] WU Yunong, KITA K, REN Fuji, *et al*. Exploring Emotional Words for Chinese Document Chief Emotion Analysis[C]// Proceedings of the 25th Pacific Asia Conference on Language: Dec 16-18, 2011, Singapore: Information and Computation: 597-606.
[15] MA Chunling, PRENDINGER H, ISHIZUKA M. Emotion Estimation And Reasoning Based on Affective Textual Interaction[J]. Affective Computing and Intelligent Interaction, 2005, 3784: 622-628.
[16] YOUNG H C, KONG J L. Automatic Affect Recognition Using Natural Language Processing Techniques And Manually Built Affect Lexicon[J]. IEICE Transactionson Information and Systems, 2006, 89(12): 2964-2971.
[17] KANG Xin, REN Fuji, WU Yunong. BottomUp: Exploring Word Emotions For Chinese Sentence Chief Sentiment Classification[C]// Proceedings of Natural Language Processing and Knowledge Engineering: Aug 21-23, 2010, Beijing, China: IEEE, 2010: 1-5.
[18] KOLLER D, FRIEDMAN N. Probabilistic Graphical Models: Principles And Techniques[M]. Cambridge, Massachusetts, London, England, The MIT Press, 2009.
[19] GEMAN S, GEMAN D. Stochastic Relaxation, Gibbs Distribution, and the Bayesian Restoration of Images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984, 6(9): 721-741.
[20] IAN P, DAVID N, ALEXANDER I, *et al*. Fast Collapsed Gibbs Sampling For Latent Dirichlet Allocation[C]// Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: Aug 24-27, 2008, Las Vegas, Nevada, USA: 2008: 569-577.
[21] REN Fuji. From Cloud Computing to Language Engineering, Affective Computing and Advanced Intelligence[J]. International Journal of Advanced Intelligence, 2010, 2(1): 1-14.
[22] REN Fuji. Robotics Cloud and Robotics School[C]// Proceedings of Natural Language Processing and Knowledge Engineering: Nov 21-29, 2011, Tokushima, Japan: IEEE, 2011: 1-8.

Biographies

Kang Xin, is currently a Ph. D. candidate at Tokushima University, Tokushima, Japan. His research interests include statistical machine learning, graphical models, and text emotion prediction. Email: kangxin@is.tokushima-u.ac.jp

Ren Fuji, is a professor with the Faculty of Engineering, the University of Tokushima. His research interests include natural language processing, artificial Intelligence, language understanding and communication, and affective computing. He is a member of the IEICE, CAAI, IEEJ, IPSJ, JSAI, AAMT and a senior member of IEEE. He is a Fellow of The Japan Federation of Engineering Societies. He is the President of International Advanced Information Institute. * The corresponding author. Email: ren@is.tokushima-u.ac.jp