Exploring Latent Semantic Information for Textual Emotion Recognition in Blog Articles

Xin Kang, *Member, IEEE,* Fuji Ren, *Senior Member, IEEE,* and Yunong Wu

***Abstract*—Understanding people's emotions through natural language is a challenging task for intelligent systems based on Internet of Things (IoT). The major difficulty is caused by the lack of basic knowledge in emotion expressions with respect to a variety of real world contexts. In this paper, we propose a Bayesian inference method to explore the latent semantic dimensions as contextual information in natural language and to learn the knowledge of emotion expressions based on these semantic dimensions. Our method synchronously infers the latent semantic dimensions as topics in words and predicts the emotion labels in both word-level and document-level texts. The Bayesian inference results enable us to visualize the connection between words and emotions with respect to different semantic dimensions. And by further incorporating a corpus-level hierarchy in the document emotion distribution assumption, we could balance the document emotion recognition results and achieve even better word and document emotion predictions. Our experiment of the word-level and the document-level emotion predictions, based on a well-developed Chinese emotion corpus Ren-CECps, renders both higher accuracy and better robustness in the word-level and the document-level emotion predictions compared to the state-of-the-art emotion prediction algorithms.**

***Index Terms*—Bayesian inference, emotion-topic model, emotion recognition, multi-label classification, natural language understanding.**

I. Introduction

THE recognition of human emotions for intelligent systems has been widely studied in many different fields. Recently reported studies include the affect analysis in human–computer interaction [1]−[3], the emotional traits examination in mental disease diagnosis [4]−[7], and the cognitive analysis of emotions in the neuroscience study [8], [9]. Because emotions are the reflection of people's mind states, perceiving emotions requires a deeper understanding of the semantic meanings in people's behavior. In this paper, we explore the emotion recognition method based on natural language understanding, to fully understand human emotions expressed in the word-level and document-level texts.

Manuscript received May 20, 2016; accepted September 20, 2016. This work was supported in part by the National Natural Science Foundation of China (NSFC) Key Program (61573094), and Fundamental Research Funds for the Central Universities (N140402001). Recommended by Associate Editor Mengchu Zhou. *(Corresponding author: Xin Kang.)*

Citation: X. Kang, F. J. Ren, and Y. N. Wu, "Exploring latent semantic information for textual emotion recognition in blog articles," *IEEE/CAA J. of Autom. Sinica*, vol. 5, no. 1, pp. 204−216, Jan. 2018.

X. Kang, F. J. Ren, and Y. N. Wu are with the Faculty of Engineering, Tokushima University, 2-1, Minamijyousanjima-cho, Tokushima 770-8506, Japan (e-mail: kang-xin@is.tokushima-u.ac.jp; ren@is.tokushima-u.ac.jp; wu-yunong@tokushima-u.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2017.7510421

Emotion recognition in natural language is a difficult study because human emotions are associated not only with the basic words but also with the context semantic meanings, which could even confuse the other human beings in many cases. For example, a positive word "happily" may express negative emotions in some specific contexts: Don't bother me. I'm living happily ever after. A direct solution for recognizing such emotions might be constructing a dictionary [10]−[13] or a knowledge base [14], [15] for recognizing emotion expressions, or considering the semantic information in contexts such as the previous few words [10], [11] or the syntactically-related words [16], [17]. However, models based on such dictionaries or knowledge bases suffer from a serious under-fitting problem, because the number of emotion-triggering patterns grows exponentially large as the number of context words in consideration increases. Either building an emotion dictionary or training an emotion classifier would require a huge number of labeled examples, which could be too expensive to acquire in practice.

In this paper, we propose a novel method by exploring the latent semantic dimensions as the word context features, for learning the emotion expressions in natural language. Semantic dimensions are represented as the discrete random variables (or topics) in a Bayesian probabilistic model, each of which is associated with a word in the document. The model has to learn a distribution of the topic assignment for each word through a Bayesian inference by reading these documents, in which a distinct topic value can indicate a specific semantic dimension in the word context. In this process, each word can be associated with a series of topic assignments in a probabilistic distribution. The number of distinct topics is adjusted by fitting the Bayesian model for emotion recognition, but the size of increased feature space, which is linear to the distinct topic number, would be much smaller than the size of a dictionary or knowledge based feature space. Therefore, fitting an emotion recognition model based on our context semantic features would be much easier than fitting the model with traditional features.

We introduce two implementations of the Bayesian inference method for textual emotion recognition. The document and word emotion topic (DWET) model is a generative model, which infers the latent topics and the emotion assignments to words and documents by maximizing the probability of word generation throughout a corpus of documents. In the DWET model, we employ the two-level hierarchical conjugate probabilities to demonstrate the distributions of words, topics, and emotions throughout a corpus. The other hierarchical document and word emotion topic (HDWET) model is also a generative model. It shares the similar structural and probabilistic assumptions with the DWET model, except that a third level hierarchy is incorporated for the document-level emotion distribution in HDWET to allow a greater flexibility in the document emotion variation. By tuning the distribution parameters in these generative models, we generate the corpus-level knowledge of emotion expressions with respect to the latent semantic dimensions in the context, and predict the emotion labels for words and documents to maximize the generative probabilities in these models.

The rest of this paper is arranged as follows: Section II reviews the related work in textual emotion recognition; Section III describes the construction and probabilistic assumptions in our Bayesian models for emotion recognition; Section IV illustrates the Bayesian inference method for learning the emotion expression knowledge and for predicting emotion labels in words and documents through a corpus; Section V details our experiment on textual emotion recognition, compares our results with the state-of-the-art emotion classification algorithms, and demonstrates the learned knowledge of emotion expressions with respect to different semantic dimensions; Section VI concludes this paper.

II. Related Work

Developing the knowledge of emotion expression in natural language has been widely studied for textual emotion recognition. These studies include the emotion lexicon [18] generated on the co-occurrence of emoticon and emotion in blog articles, the emotion lexicon [11] selected from the Japanese evaluation expression dictionary [19] based on the emotion words proposed by Teramura [20], the emotion lexicon for verbs [12] which was manually annotated to the combination of a Dutch wordnet and a Dutch reference lexicon, and the emotion lexicon [13] based on the word-emotion association with crowdsourcing. Besides, there have been manually developed emotional rules such as the emotion lexicon and the lexical pattern based rules [11] for finding the emotion-provoking events in the Web corpus, the manually developed rules [21] based on wordnet-affect [22] for constructing the groups of lyric emotions, and the application of common-sense knowledge such as the open mind commonsense (OMCS) knowledge base [23] for the textual affect sensing [24], and the emotinet knowledge base for an emotion detection system [14]. However, many studies on textual emotion recognition [25]−[27] suggested that the development of lexicons or knowledge bases for emotion expression in natural language could be very expensive, and serious accuracy problems could be caused in the developed knowledge base especially for the context sensitive emotion expressions.

There have also been studies on the extraction of context sensitive emotion information. Wu *et al.* [28], [29] employed a linear chain conditional random fields (CRF) model, based on the negative modifiers and the degree modifiers as context information in a sentence, for recognizing the emotions in words. Das *et al.* [30] also considered the context information such as the negative modifiers and punctuations in a sentence, and employed a CRF model for the word emotion prediction.

These recognized word emotions have been proved crucial for sentence and document emotion classifications. With an emotion lexicon learned through the statistical study of emoticons in online messages, Yang *et al.* [10] built a support vector machines (SVM) model and a CRF model respectively for the sentence and document emotion classifications in blog articles. Kang *et al.* [31] proposed a kernel-based method to investigate and compare different word-level emotion features for the sentence emotion prediction in a blog corpus. The major problem in these models is that the context features were either insufficient to demonstrate the sentiment information in natural language or dependent on a very large lexicon which causes the model difficult to fit.

Kang *et al.* [27] employed a semi-supervised Bayesian framework to predict emotions in words, by incorporating the statistical relationship between words and emotion labels through the online micro-blog streams. By incorporating an emotion transition factor in the Bayesian framework, the model has successfully learned the author-specific emotion expression patterns in micro-blogs, and has effectively improved the emotion prediction accuracy in micro-blog documents. Other probabilistic models [32], [33] explored the word emotion and document emotion separately in blog articles, with emotion labels incorporated as a latent factor in determining the observation of words in the blog documents. Ren *et al.* [4] examined the emotional traits in suicide blog streams with a probabilistic graphical model, and developed a suicide risk prediction system for the blog authors based on their writing histories with promising results. However, to our knowledge no study has explored the semantic dimensions in the context for simultaneously recognizing the textual emotions in words and documents.

III. Bayesian Models for Emotion Recognition

Bayesian models are the probabilistic description of observed values and hidden properties in the real world, in which observed values and hidden properties are represented as visible and latent variables respectively, with the influence among these values and properties represented as the directed connections between these variables. As a complete model of variables and their relationships, a Bayesian model defines the joint probability of all random variables with a directed acyclic diagram. Each random variable is associated with zero or more parent random variables based on some dependent and independent assumptions in the diagram. Probabilistic influence could flow through these directed connections in the diagram to allow probabilistic inference. The Bayesian models are convenient to describe such influence between different variables, because the joint probability of a Bayesian model is easy to factorize into the product of a series of conditional probabilities according to the Bayes' theorem, and each conditional probability could describe an influence from several parent variables to the child variable in the model. Each factorized probability would incorporate only a few random variables which are more suitable to be mathematically represented than the joint probability. In this paper, we propose two Bayesian models for emotion recognition in words and documents, as shown in Figs. 1 and 2. The random variables, parameters, and indexes in these models are listed in Table I for the ease of illustration.

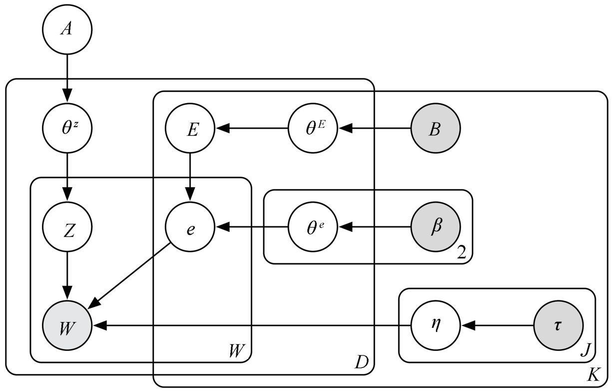


Fig. 1. DWET model for predicting complex text emotions and emotion-topic variation.

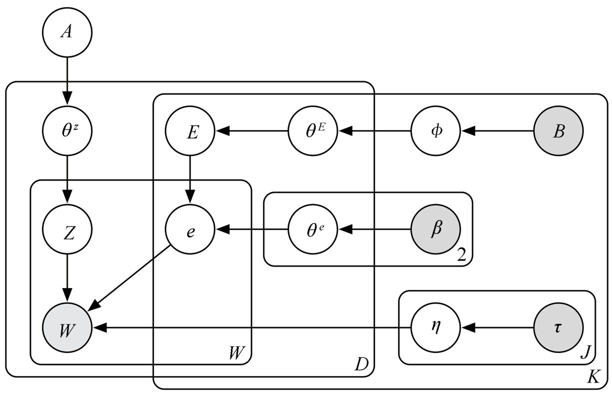


Fig. 2. HDWET model with corpus level emotion proportions for predicting complex text emotions and emotion-topic variation.

TABLE I
INDEXES, RANDOM VARIABLES, AND PARAMETERS

Indexes/variables/ parameters	Description
j	Index of topic
k	Index of emotion category
t	Index of word in the vocabulary
i	Index of word in a document
d	Index of document
J	Number of semantic dimensions (topics)
K	Number of emotion categories
N	Number of words in the vocabulary
D	Number of documents in the corpus
W_d	Number of words in document d
z	Variable of topic
E	Variable of document emotion
e	Variable of word emotion
w	Variable of word
θ^z	Proportion variable in topic distribution
θ^E	Proportion variable in document emotion distribution
θ^e	Proportional variable in word emotion distribution
η	Proportional variable in word distribution
ϕ	Concentration variable in document emotion distribution
A	Concentration parameter in topic distribution
B	Concentration parameter in document emotion distribution
β	Concentration parameter in word emotion distribution
τ	Concentration parameter in word distribution
α	Hyper-parameter in document emotion distribution

A. Model Construction for DWET

The DWET model in Fig. 1 describes a joint probability over the observed word w_{di} for each document index $d \in \{1, \ldots, D\}$ and each word index $i \in \{1, \ldots, W_d\}$ throughout a corpus, the semantic dimension value (or topic) z_{di} for each word, the emotion labels e_{dik} of each emotion category $k \in \{1, \ldots, K\}$ for each word, the emotion labels E_{dk} of each emotion category for each document, and variables η, θ^z, θ^E, θ^e, A, B, β, τ as the distribution parameters in the Bayesian model.

Besides, the DWET model describes a series of conditional probabilities over these random variables with directed connections as shown in Fig. 1. The observation of a word in w_{di} given its topic in z_{di} and corresponding emotions in $e_{di\cdot}$ is assumed to follow a Categorical distribution

$$w_{di}|z_{di}, e_{di\cdot} \sim \text{Categorical}(\eta_{z_{di}e_{di\cdot}}) \tag{1}$$

where $\eta_{z_{di}e_{di\cdot}}$ is the proportional parameter in Categorical distribution. By arranging the topic variable z_{di} and the emotion variable $e_{di\cdot}$ as parents to the word variable w_{di}, we construct a V-structure $z \rightarrow w \leftarrow e$, in which because the value of the child variable w is observed throughout the corpus, the assignments to z and e falls dependent on each other. This is because that the parent variables, which in the directed V-structure connections could jointly influence the value in the child variable, become inversely influenced by the observations in the child variable and any other parent variable through their posterior probabilities. This phenomenon is called "explaining away" in the Bayesian model. It allows the observation of a semantic dimension $z_{di} = j$ in word w_{di} to affect the distribution of word emotions $e_{di\cdot}$ through the posterior probability $p(e_{di\cdot}|w_{di}, z_{di})$, and therefore makes our emotion recognition depending on the semantic dimensions in the context. Compared to the lexicon-based, rule-based, and knowledge-based emotion inference, in which emotion distributions are represented as $p(e_{di\cdot}|w_{di}, w_{dj}, \ldots)$, our DWET model significantly decreases the complexity in the conditional parts of the emotion probability. In fact, because the model describes a probabilistic connection between the word w_{di} and topic z_{di} variables, we can interpret the context semantic information from a vector representation of the topic probabilities $[p(z_{di} = 1), p(z_{di} = 2), \ldots]$.

The topic variable z_{di} specifies a semantic dimension in the context of word w_{di}. We incorporate totally J semantic dimensions in the DWET model, which correspond to a set of discrete values $z_{di} \in \{1, 2, \ldots, J\}$ for the topic assignment. A Categorical distribution is assumed for these discrete topic variables

$$z_{di} \sim \text{Categorical}(\theta_d^z) \tag{2}$$

where θ_d^z is the proportional parameter in the Categorical distribution with respect to a specific document d.

The word emotion variable $e_{dik} \in \{0, 1\}$ specifies the existence of the kth emotion category in w_{di}, by taking binary values. We incorporate totally K distinct emotion categories, with $k \in \{1, \ldots, K\}$ indexing the specific categories. To analyze the influence of an emotion observation in a document d to the emotion observations in corresponding words w_{di}, we connect the document emotion variable E_{dk} to the word emotion variables e_{dik} in the DWET model, and assume Bernoulli distribution for the word emotion variable given a document emotion observation

$$e_{dik}|E_{dk} \sim \text{Bernoulli}(\theta^e_{dkE_{dk}}) \tag{3}$$

where $\theta^e_{dkE_{dk}}$ is the proportional parameter in the Bernoulli distribution with respect to document d, emotion label k, and the observation of document emotion in E_{dk}.

The document emotion variable $E_{dk} \in \{0,1\}$ is also a binary random variable, which indicates the existence of the kth emotion category in document d. Emotion categories and emotion indexes in documents are the same as those in words. We assume Bernoulli distribution for the document emotion variables

$$E_{dk} \sim \text{Bernoulli}(\theta^E_{dk}) \tag{4}$$

with θ^E_{dk} as the proportional parameter for document d and emotion label k. Although the word emotion variables $e_{d\cdot k}$ are absent in (4) for the document emotion distribution, the influence from $e_{d\cdot k}$ to E_{dk} still exists in our Bayesian model and is implemented through a Bayesian inference process. In fact, the probabilistic belief in document emotions would be rationally adjusted given the word emotion samples, as will be discussed later.

The DWET model assumption incorporates several proportional parameters in η, θ^z, θ^e, and θ^E as shown before. Although a direct optimization to these proportional parameters, through Bayesian inference, is feasible for training a model for the document and word emotion predictions, except that the learned values in these parameters might only fit well in the training process but could not adjust properly to new text samples in the real world. One of the advantages in building a Bayesian model is that we can represent the model parameters as random variables and make further assumptions on their distributions. This allows the model to adjust these parameters better, with more flexibility and better robustness in Bayesian inference, for recognizing emotions in the real-world texts. In the DWET model, we assume conjugate priors of the Categorical and Bernoulli likelihoods as the prior probabilities for these proportional parameters, which will simplify the derivation of their posterior probabilities in our Bayesian inference.

Specifically, for the proportional parameter η_{jk} in (1), we assume Dirichlet distribution

$$\eta_{jk} \sim \text{Dirichlet}(\tau_{jk}) \tag{5}$$

as its prior probability, which is also the conjugate prior of its Categorical likelihood function. τ_{jk} is the concentration parameter of this Dirichlet distribution, and j, k are the indexes of topic and emotion in word w_{di}, respectively. For the proportional parameter θ^z_d in (2), we also assume the Dirichlet distribution as its prior probability, which is the conjugate prior of its Categorical likelihood function

$$\theta^z_d \sim \text{Dirichlet}(A) \tag{6}$$

A is the concentration parameter of this Dirichlet distribution.

Dirichlet distribution is used to describe the probability of probability mass assignments in the proportional parameters, like θ^z_d in (6). The proportional parameter θ^z_d can be specified as a set of J probability mass assignments $\{\theta^z_{dj} = \hat{\theta}^z_{dj} | j = 1, \ldots, J\}$, in which each entry $\hat{\theta}^z_{dj}$ evaluates the probability of observing a topic value in $z_{di} = j$, by satisfying the following restrictions

$$\sum_{j=1}^{J} \hat{\theta}^z_{dj} = 1,\ \hat{\theta}^z_{d\cdot} > 0. \tag{7}$$

The Dirichlet distribution in (6) describes a probability density function for the continuous random variables $\theta^z_{d\cdot}$. It allows a θ^z_{dj} to concentrate on a larger probability mass assignment with a larger concentration parameter A_j, while restricting the probability mass assignments under (7).

For the proportional parameter θ^e_{dkE} in (3), we assume Beta distribution

$$\theta^e_{dkE_{dk}} \sim \text{Beta}(\beta_{kE_{dk}}) \tag{8}$$

as its prior probability, which is also the conjugate prior of its Bernoulli likelihood function. $\beta_{kE_{dk}}$ is the concentration parameter in this Beta distribution, while E_{dk} corresponds to the assignment to the document emotion $E_{dk} \in \{0,1\}$ with the same document index d and emotion category k. Similarly, for the proportional parameter θ^E_{dk} in (4), we also assume the Beta distribution as its prior probability, which is the conjugate prior of its Bernoulli likelihood function

$$\theta^E_{dk} \sim \text{Beta}(B_k) \tag{9}$$

B_k is the concentration parameter in this Beta distribution.

A Beta distribution can be considered as a simple Dirichlet distribution for the binary probability mass assignments. For example, the proportional parameter θ^E_{dk} in (4) can be specified with the probability mass assignments of $(\theta^{E0}_{dk}, \theta^{E1}_{dk}) = (\hat{\theta}^{E0}_{dk}, \hat{\theta}^{E1}_{dk})$, in which

$$\hat{\theta}^{E0}_{dk} + \hat{\theta}^{E1}_{dk} = 1,\ \hat{\theta}^{E\cdot}_{dk} > 0. \tag{10}$$

The Beta distribution in (9) describes a probability density function for the continuous random variable $\theta^{E\cdot}_{dk}$. It allows θ^{E0}_{dk} to concentrate on a larger probability mass assignment with a larger concentration parameter B_k, and restrict the probability mass assignments under (10).

All the concentration parameters A, B, β, and τ are constant in our DWET assumption. These parameters are initialized by counting the occurrence and absence of the corresponding categorical variables in the training data. For example, we count the occurrence of document emotion k through the training data for $B^1_k = \sum_d 1\{E_{dk} = 1\}$ and count the absence of document emotion k for $B^0_k = \sum_d 1\{E_{dk} = 0\}$. We count the occurrence of word emotion k together with document emotion k for $\beta^1_{kE_{dk}} = \sum_d \sum_i 1\{e_{dik} = 1, E_{dk} = 1\}$ and the occurrence of word emotion k with the absence of document emotion k for $\beta^0_{kE_{dk}} = \sum_d \sum_i 1\{e_{dik} = 1, E_{dk} = 0\}$. Parameter τ is initialized similarly. Because the semantic dimensions are latent even in the training data, we assume their probability masses concentrate equally on the set of discrete values $\{1, \ldots, J\}$, and employ one value for all the topic concentration parameters $A_j = A$.

B. Model Construction for HDWET

The HDWET model in Fig. 2 describes a similar probabilistic model as DWET, except that we have relaxed the assumption for the document emotion concentration parameter to be constant by incorporating a random variable ϕ_k on the corpus-level and by employing $\alpha\phi_k$ in

$$\theta^E_{dk} \sim \mathrm{Beta}(\alpha\phi_k) \tag{11}$$

as a flexible document emotion concentration parameter. α is a constant hyper-parameter in (11). Because ϕ is incorporated in the hierarchy of document emotion distribution, we name it the Hierarchical Document and Word Emotion Topic model.

In the previous DWET model, because the corpus-level concentration parameter B_k is constant, all the document emotion proportional parameters θ^E_{dk} for different d must have the same probability densities as shown in (9). This corresponds to an implicit assumption that for each emotion category k, the model assumes the same prior probability to observe it in different documents. However, in the real text because the probabilistic concentration of document emotion varies dramatically through different documents, the model needs to adjust itself to all kinds of emotion distributions to properly recognize these document emotions. In the HDWET model, we incorporate the variability in document emotion distribution with a corpus-level concentration parameter $\alpha\phi_k$, as shown in (11).

We assume Beta distribution for the document emotion concentration parameter ϕ_k

$$\phi_k \sim \mathrm{Beta}(B_k) \tag{12}$$

in which B_k only poses a prior assumption on the emotion distribution and does not directly influence the probabilistic distribution for document-level emotions. As a constant parameter, B is initialized by counting the observation of document emotions through the training data, with $B^1_k = \sum_d 1\{E_{dk} = 1\}$ for the occurrence of emotion k, and $B^0_k = \sum_d 1\{E_{dk} = 0\}$ for the absence of emotion k.

IV. Bayesian Inference

Bayesian inference estimates the value of a random variable based on the Bayes' theorem, by deriving its posterior probability from the product of its prior and an observation likelihood. For the proposed Bayesian models in this paper, we employ a Gibbs sampling algorithm as the Bayesian inference method, to estimate the values in topic z_{di}, word emotion e_{dik}, and document emotion E_{dk}, by deriving their posterior probabilities respectively.

Gibbs sampling is an efficient Bayesian inference algorithm, in which samples of the random variables are iteratively drawn from their estimated posterior probabilities in a loop, by freezing the sampled values in other random variables as the observation for their likelihood calculations. The algorithm converges after a few sampling steps, and the posterior probability of each random variable can be estimated by counting the sampling history of this variable.

The Gibbs sampling algorithms for estimating topics and emotions for the DWET model and the HDWET model are described in Algorithm 1 and Algorithm 2. Both algorithms iteratively draw samples of z_{di}, e_{dik}, and E_{dk} from their posterior probabilities, with the parameter variables including η, θ^z, θ^e, θ^E in DWET and ϕ in HDWET collapsed out for sampling efficiency. The sampling steps repeat through M loops until convergence, which renders a linear time complexity $\mathrm{O}(n)$ for both algorithms. In the following, we illustrate the derivation of posterior probabilities respectively of the topic variable z_{di}, the word emotion variable e_{dik}, and the document emotion variable E_{dk} for the DWET model and the HDWET model.

Algorithm 1 Gibbs Sampling for DWET Inference

```
for m = 1 → M do
    for d = 1 → D do
        for i = 1 → W_d do
            sample z_di by (13)
            for k = 1 → K do
                sample e_dik by (14)
            end for
        end for
        for k = 1 → K do
            sample E_dk by (15)
        end for
    end for
end for
```

Algorithm 2 Gibbs Sampling for HDWET Inference

```
for m = 1 → M do
    for k = 1 → K do
        sample φ_k by (17)
    end for
    for d = 1 → D do
        for i = 1 → W_d do
            sample z_di by (13)
            for k = 1 → K do
                sample e_dik by (14)
            end for
        end for
        for k = 1 → K do
            sample E_dk by (16)
        end for
    end for
end for
```

A. Gibbs Sampling for DWET

We show the derived algebraic expressions of the posterior probabilities in Algorithm 1, with the detailed derivation steps illustrated in Appendix A.

For word i of document d within a corpus[1], the posterior probability of observing a semantic dimension (topic) j, conditioned on the observation of words, emotions, and all

[1]For the Gibbs sampling algorithm, this corresponds to a test corpus.

other topics through the corpus are given by

$$p(z_{di}=j|w,z_{-di},e,E;A,B,\beta,\tau)\propto\frac{A+n_{dj}}{A\times J+W_d}\times\prod_{k\in K^1}\frac{\tau^1_{kjw_{di}}+n^1_{kjw_{di}}}{\tau^1_{kj*}+n^1_{kj*}}\times\prod_{k\in K^0}\frac{\tau^0_{kjw_{di}}+n^0_{kjw_{di}}}{\tau^0_{kj*}+n^0_{kj*}} \quad (13)$$

where $K^1=\{k'|e_{dik'}=1\}$ and $K^0=\{k'|e_{dik'}=0\}$ represent the sets of occurrent and absent of emotion category in word w_{di}, $n_{dj}=\sum_{i'}1\{z_{di'}=j\}$ counts the occurrence of topic with the same value as j in document d, W_d counts the number of words in document d, $n^1_{kjw_{di}}=\sum_{d'}\sum_{i'}1\{(e_{d'i'k},z_{d'i'},w_{d'i'})=(1,j,w_{di})\}$ counts the occurrence of word emotion k, topic j, and word with the same value as w_{di} through the corpus, and $n^0_{kjw_{di}}=\sum_{d'}\sum_{i'}1\{(e_{d'i'k},z_{d'i'},w_{d'i'})=(0,j,w_{di})\}$ counts the absence of word emotion k, the occurrence of topic j and word with the same value as w_{di} through the corpus. "$*$" in the subscripts indicates a summation of the variable over the corresponding dimension.

For word i of document d within a corpus, the posterior probability of observing the emotion category k conditioned on the observation of words, topics, and all other emotions through the corpus is given by

$$p(e_{dik}|w,z,e_{-dik},E;A,B,\beta,\tau)\propto\begin{cases}\dfrac{\beta^1_{kE_{dk}}+n_{dk}}{\beta^1_{kE_{dk}}+\beta^0_{kE_{dk}}+W_d}\times\dfrac{n^1_{kz_{di}w_{di}}+\tau^1_{kz_{di}w_{di}}}{n^1_{kz_{di}*}+\tau^1_{kz_{di}*}} & \text{if } e_{dik}=1\\ \dfrac{\beta^0_{kE_{dk}}+W_d-n_{dk}}{\beta^1_{kE_{dk}}+\beta^0_{kE_{dk}}+W_d}\times\dfrac{n^0_{kz_{di}w_{di}}+\tau^0_{kz_{di}w_{di}}}{n^0_{kz_{di}*}+\tau^0_{kz_{di}*}} & \text{if } e_{dik}=0\end{cases} \quad (14)$$

where $n_{dk}=\sum_{i'}1\{e_{di'k}=1\}$ counts the occurrence of word emotion k in document d, while $n^1_{kz_{di}w_{di}}$ and $n^0_{kz_{di}w_{di}}$ counts the same observations as in (13).

For document d in a corpus, the posterior probability of observing emotion category k conditioned on the observation of words, topics, and all other emotions through the corpus is given by

$$p(E_{dk}|w,z,e,E_{-dk};A,B,\beta,\tau)\propto\begin{cases}\dfrac{B^1_k+1}{B^1_k+B^0_k+1}\displaystyle\prod_{i\in W_d}p(e_{dik}|\ldots,E_{dk}=1,\ldots) & \text{if } E_{dk}=1\\ \dfrac{B^0_k+1}{B^1_k+B^0_k+1}\displaystyle\prod_{i\in W_d}p(e_{dik}|\ldots,E_{dk}=0,\ldots) & \text{if } E_{dk}=0\end{cases} \quad (15)$$

where W_d is the set of word indexes in document d, with the posterior probabilities of word emotion observations in $p(e_{dik}|\ldots,E_{dk},\ldots)$ calculated through (14).

In Algorithm 1, the Gibbs sampler repeatedly draws samples of topic z_{di}, word emotion e_{dik}, and document emotion E_{dk} based on the derived posterior probabilities through (13)−(15), and uses these sampled values to estimate the true posterior probabilities, until these estimated posterior probabilities get converged. We predict the values in z_{di}, e_{dik}, and E_{dk} by maximizing their estimated posterior probabilities.

B. Gibbs Sampling for HDWET

We illustrate the algebraic expressions for posterior probability calculations in Algorithm 2, with the detailed derivation steps shown in Appendix B.

The HDWET model shares the similar structure and probabilistic assumptions with respect to the topic-related distributions and the word emotion-related distributions, as depicted in section III. In fact, the algebraic expressions for posterior probabilities of topics and word emotions are also the same as those in (13) and (14) respectively.

For document d in a corpus, the posterior probability of observing emotion category k conditioned on the observation of words, topics, and all other emotions through the corpus is given by

$$p(E_{dk}|w,z,e,E_{-dk},\phi;A,B,\beta,\tau)\propto\begin{cases}\dfrac{\alpha\hat{\phi}_k+1}{\alpha(\hat{\phi}_k+\bar{\hat{\phi}}_k)+1}\displaystyle\prod_{i\in W_d}p(e_{dik}|\ldots,E_{dk}=1,\ldots) & \text{if } E_{dk}=1\\ \dfrac{\alpha\bar{\hat{\phi}}_k+1}{\alpha(\hat{\phi}_k+\bar{\hat{\phi}}_k)+1}\displaystyle\prod_{i\in W_d}p(e_{dik}|\ldots,E_{dk}=0,\ldots) & \text{if } E_{dk}=0\end{cases} \quad (16)$$

where W_d is the set of word indexes in document d, with the posterior probability of word emotion observations in $p(e_{dik}|\ldots,E_{dk},\ldots)$ calculated through (14). $\bar{\hat{\phi}}_k$ is the complement of $\hat{\phi}_k$ with $\bar{\hat{\phi}}_k=1-\hat{\phi}_k$, and $\hat{\phi}_k$ is sampled through its updated posterior probability

$$\phi_k|w,z,e,E;A,B,\beta,\tau\sim\text{Beta}(B^1+n_k,B^0+D-n_k) \quad (17)$$

$n_k=\sum_{d'}1\{E_{d'k}=1\}$ counts the occurrence of document emotion k through the corpus.

Similar to the DWET model, the Gibbs sampler in Algorithm 2 repeatedly draws samples of topic z_{di}, word emotion e_{dik}, and document emotion E_{dk} based on the derived posterior probabilities through (13), (14), and (16), and estimate their true posterior probabilities based on the sampled values until convergence. Prediction of the values in z_{di}, e_{dik}, and E_{dk} is made by maximizing their posterior probabilities.

V. Emotion Recognition Experiment

We examine our Bayesian inference method for textural emotion recognition in blog articles, based on the emotion corpus Ren-CECps [25]. The emotion corpus contains 1, 147 Chinese blog articles collected from Internet, with manually annotated emotion labels for 8 basic emotion categories in the document-level, sentence-level, and word-level, respectively. The basic emotion categories include joy, love, expect, surprise, anxiety, sorrow, anger, and hate. And each emotion label has been further distinguished into 10 levels of emotion intensities according to its emotion strength. The following is an example of emotion-annotated sentence translated from Ren-CECps:

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

In this example, "want to" indicates a medium (0.6) expect, "give up" implies a low (0.3) hate and a medium (0.6) sorrow, while the complete sentence indicates a low (0.2) hate and a high (0.9) sorrow.

It has to be noted that emotion labels from different categories are not evenly distributed throughout the corpus [33]. In fact, a previous study of the emotion classification for online messages [4] suggests that textual expression of emotions are highly biased, in which love, sorrow, and anxiety are more often observed than surprise and anger. In Ren-CECps, the number of love (over 500) is 1 magnitude larger than the number of surprise (only 90) in the document-level. This makes the textural emotion recognition very difficult for the traditional classifiers, because training a classifier on highly biased data will significantly impact the recall sores for the less common emotion categories.

The emotion corpus is divided into a training set of 917 blog articles and a test set of 230 blog articles. We initialize the concentration parameters A, B, β, τ based on the observation of corresponding categorical variables in the training set, select the model parameter J based on a 5-fold cross validation on the training set, and set the hyper-parameter α to the number of blog articles in each set. We employ precision, recall, and f-score for evaluating the emotion recognition results in emotion category, as defined below

$$\text{Precision}(k) = \frac{tp_k}{tp_k + fp_k} \tag{18}$$

$$\text{Recall}(k) = \frac{tp_k}{tp_k + fn_k} \tag{19}$$

$$\text{F-score}(k) = \frac{2 \times \text{Precision}(k) \times \text{Recall}(k)}{\text{Precision}(k) + \text{Recall}(k)} \tag{20}$$

where k indicates the emotion category, tp_k, fp_k, and fn_k count the number of true positive, false positive, and false negative predictions in the result for emotion category k. We compare the results from the DWET and HDWET models, perform further comparisons with those from the state-of-the-art emotion prediction algorithms, and demonstrate the learned connection between emotion categories and latent semantic dimensions for specific words in the blog articles.

The detailed results of emotion prediction from the DWET and HDWET models are shown in Tables II and III for the document-level and word-level emotion recognition, respectively. Jo, Lv, Ex, Su, Ax, So, Ag, and Ht are the abbreviations for the emotions of joy, love, expect, surprise, anxiety, sorrow, anger, and hate, while Ne indicates a none emotion which only occurs in the word emotion prediction.

TABLE II
EVALUATION OF THE DOCUMENT EMOTION PREDICTION

	Precision		Recall		F-score	
	DWET	HDWET	DWET	HDWET	DWET	HDWET
Jo	**56.32**	37.17	56.32	**96.55**	**56.32**	53.67
Ht	**41.07**	26.47	47.92	**93.75**	**44.23**	41.28
Lv	**72.14**	**72.14**	**65.58**	**65.58**	**68.71**	**68.71**
So	**63.91**	48.80	80.95	**97.14**	**71.43**	64.97
Ax	**57.25**	54.50	70.54	**91.96**	63.20	**68.44**
Su	0.00	**28.57**	0.00	**55.56**	0.00	**37.74**
Ag	**23.81**	18.68	19.23	**65.38**	21.28	**29.06**
Ex	**58.97**	43.48	69.00	**100.00**	**63.59**	60.61
Avg.	**46.68**	41.23	51.19	**83.24**	48.60	**53.06**

TABLE III
EVALUATION OF THE WORD EMOTION PREDICTION

	Precision		Recall		F-score	
	DWET	HDWET	DWET	HDWET	DWET	HDWET
Ne	94.73	**95.04**	**99.58**	99.51	97.09	**97.22**
Jo	**81.41**	78.02	28.07	**29.81**	41.74	**43.13**
Ht	**82.42**	79.25	16.11	**22.63**	26.96	**35.21**
Lv	**82.72**	82.67	55.39	**56.31**	66.35	**66.99**
So	**82.70**	80.89	44.37	**47.95**	57.75	**60.21**
Ax	**75.23**	74.44	29.83	**33.43**	42.72	**46.13**
Su	**100.00**	85.71	1.33	**2.67**	2.63	**5.17**
Ag	76.92	**88.57**	2.78	**8.61**	5.36	**15.70**
Ex	**79.86**	77.94	20.84	**23.60**	33.05	**36.23**
Avg.	**84.00**	82.50	33.14	**36.06**	41.52	**45.11**

For document-level emotion recognition, we find that on average the DWET model achieves better precision than the HDWET model, while the HDWET model renders better Recalls than the DWET model. This can be explained by the fact that the variability in probability mass concentration parameter $\alpha\phi$ makes the HDWET model easier to generate more positive labels for the less common emotion categories during inference. For example, Surprise is a rare document emotion compared to the other emotions, which is only observed 90 times in 1,147 blog articles in Ren-CECps. During inference, the concentration parameter variable $\alpha\phi_{\text{Surprise}}$ grows larger than the static concentration parameter B_{Surprise}, which makes the posterior probability of $E_{d\text{Surprise}} = 1$ in (16) larger than that in (15), and therefore enables the HDWET model to recognize more surprise labels (with a higher recall score) than the DWET model. For the common emotion categories such as love, which is observed over 500 times in 1,147 blog articles, the HDWET model still performs as well as the DWET model for generating the positive labels. This result indicates that the incorporated flexibility in document emotion concentration in the HDWET model has effectively improved the robustness for document emotion recognition. It has to be noticed that although the HDWET model achieves an average lower precision score than the DWET model, in some specific emotion categories such as love and surprise the HDWET model still renders the same or even better precision scores than the DWET model. Considering the f-score as a balanced evaluation, the HDWET model outperforms the DWET model in the document-level emotion recognition.

For word-level emotion recognition, we find that both models achieve promising results for recognizing the Ne label, which in fact is the most common label for word emotion. On average, the DWET model achieves higher precision scores, while the HDWET model renders higher recall scores. The result suggests that with Bayesian inference, the flexibility in document emotion concentration not only has increased the belief in observing the less common emotion categories in documents but also has flowed through the directed connection $E \rightarrow e$ under the probabilistic assumption in (3) to impact the belief for observing the same word emotions in the HDWET model. For the common emotion category such as love and sorrow, the recall scores from HDWET are still better than those from DWET, indicating that a flexible concentration parameter in the document emotion distribution could effectively improve the robustness for word emotion recognition. By considering the f-score as a balanced evaluation, we find the HDWET model also outperforms the DWET model for word-level emotion recognition.

We plot the time complexity of Gibbs sampling algorithms in Fig. 3, in terms of the number of input documents, for two hundred sampling iterations of the DWET and HDWET inference respectively. Our results suggest that inference time of both algorithms grows linearly with respect to the size of evaluation data, and that the DWET and HDWET models render very little difference in the time complexity.

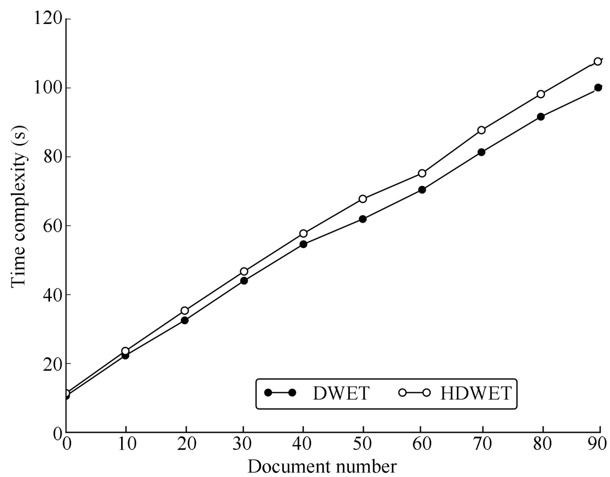


Fig. 3. Time complexity of Gibbs sampling algorithms for Algorithm 1 and Algorithm 2 in terms of document number.

We plot the receiver operating characteristic (ROC) curves in Fig. 4 for the results of document emotion prediction and word emotion prediction, in terms of DWET and HDWET respectively. A comparison of Figs. 4 (a) and (b) suggests that the HDWET model could significantly improve the robustness of document emotion classification, especially for the rare emotion categories, e.g., surprise and anger, in contrast to the DWET model. By comparing Figs. 4 (b) and (d), we find that the DWET model outperforms the HDWET model in the robustness of emotion classification for sorrow and surprise, while the HDWET model could generate robust classification results for more difficult emotion categories like hate and expect with properly selected thresholds.

Next, we compare our Bayesian models with the state-of-the-art emotion prediction algorithms for the document-level and word-level emotion recognition respectively, as shown in Figs. 5 and 6, based on the Precision scores. The naive Bayesian (NB) and SVM classifiers are employed as the base-line models for the document emotion recognition, and the hidden Markov models (HMM) and conditional random fields (CRF) are employed as the base-line models for the word emotion recognition. For the document-level emotion recognition in Fig. 5, we find that the NB classifier performs slightly better than the SVM classifier, with the average precisions of $30.54\,\%$ and $28.41\,\%$, respectively. Our DWET and HDWET models perform much better than the base-line models, with the average precisions of $46.68\,\%$ and $41.23\,\%$, respectively. For word-level emotion recognition in Fig. 6, the experiment results suggest that the CRF model performs slightly better than the HMM model, with the averaged precisions of $63.46\,\%$ and $62.20\,\%$, respectively. Compared with the base-line models, our DWET and HDWET models render much better results for word emotion recognition, with

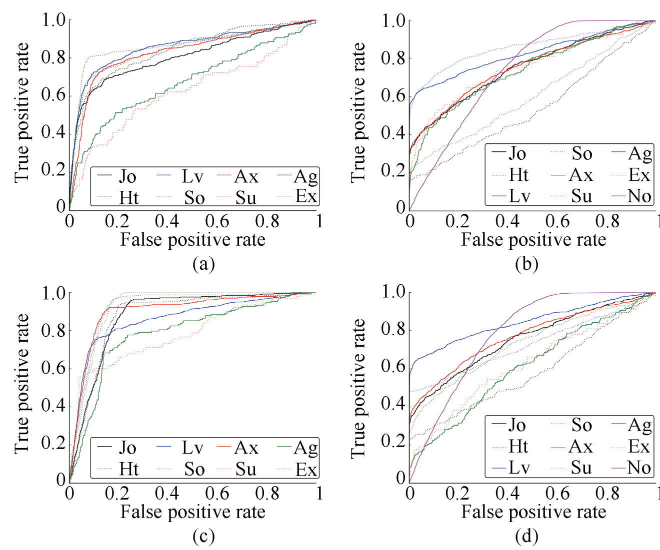


Fig. 4. ROC curves for document emotion prediction (a, c) and word emotion prediction (b, d) in terms of the DWET model (a, b) and the HDWET model (c, d).

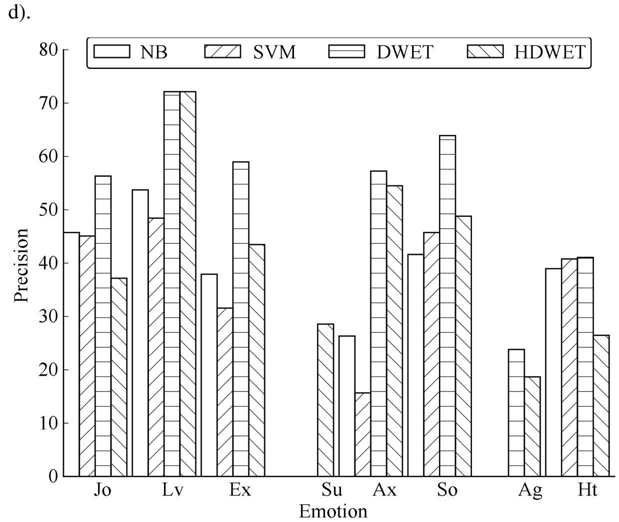


Fig. 5. Document emotion precisions from NB, SVM, DWET, and HDWET.

the average precisions of 84.00 % and 82.50 %, respectively. These comparisons suggest that, our Bayesian models by incorporating the latent semantic dimensions as the context of words have generated a much simpler representation of emotions in the natural language expression, which helps the models to fit more easily than the dictionary feature based models.

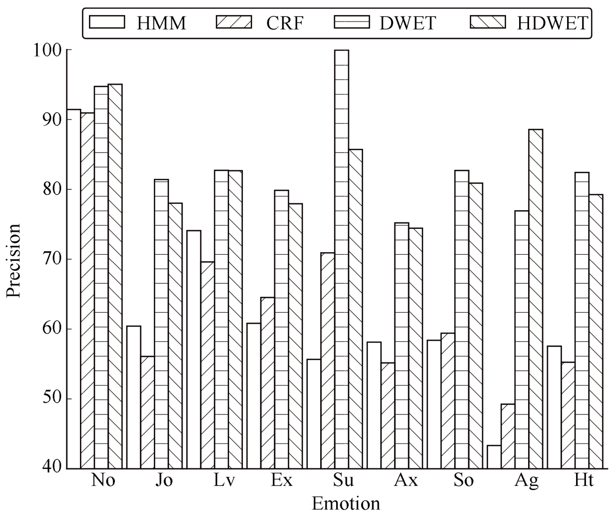


Fig. 6. Word emotion precisions from HMM, CRF, DWET, and HDWET.

Fig. 7 demonstrates the knowledge of emotion in natural language expression with respect to the semantic dimensions (topics), in the form of an emotion-topic diagram. The knowledge is learned with the HDWET model. We connect words to their most significant semantic dimensions (T1, …, T9), and attach these semantic dimensions to their related emotions, and denote the connection strengths denoted on the edges. The node colors specify the category of a node. For example, all word nodes related to the same topic together with this topic node share the same color. An emotion node shares the same color with its most strongly connected topic node. With this emotion-topic diagram, we can easily tell the connection between words, semantic dimensions, and emotions. For example, words like "convalesce" and "fall behind" in their most significant semantic dimension T3 are strongly connected to the emotion of sorrow. Words in the emotion-topic diagram are not necessarily the emotional words in traditional definition, because we are generalizing the knowledge of emotion expression from manual annotations to the more general language expressions.

The knowledge of emotion expression with respect to the semantic dimensions has been learned through the V-structure of $z \rightarrow w \leftarrow e$ in both models. The "explaining away" phenomenon allows topic z to directly influence the posterior probabilities of emotion labels e through (14), and enables the reverse influence from word emotion samples e to the posterior probabilities of topics z through (13). The models could therefore recognize the emotion labels e in the same word with different semantic dimensions z, which promises an improved precision, and associate each semantic dimension z with a specific distribution of emotions e, which generalize the basic knowledge of emotion with respect to many general natural language expressions to improve the recall.

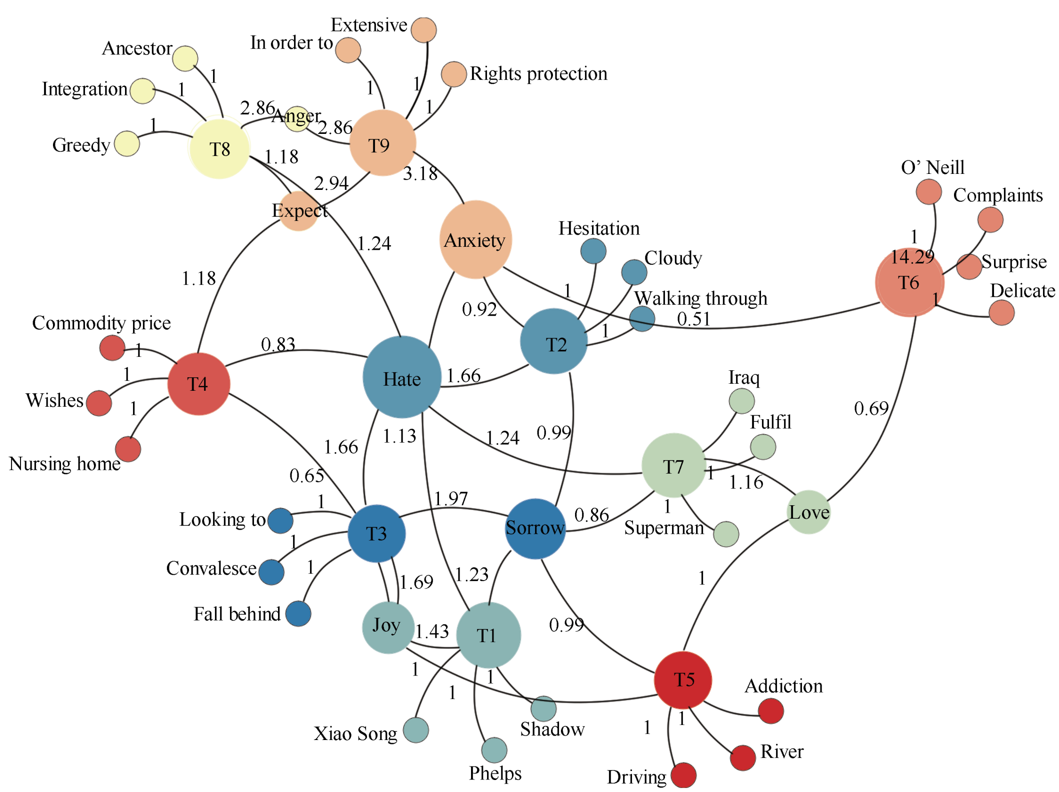


Fig. 7. Part of the emotion-topic diagram generated from HDWET.

VI. CONCLUSION

In this paper, we propose two Bayesian models DWET and HDWET for exploring the latent semantic dimensions as the context in natural language, and for learning the knowledge of emotion expressions with respect to these semantic dimensions. The basic idea is that probabilistic influence could flow between emotions and topics in Bayesian inference through a V-structure in the models, in which the emotion variable e and the topic variable z are located as two parents of the observed word variable w. Because each discrete value in topic z represents a specific context semantic dimension for the associate word w, the probabilistic distribution over topics in Bayesian inference corresponds to a vector representation of the probabilities over different context semantic dimensions, which allows the models to distinguish words under different contexts and effectively improves the emotion recognition results. Our experiment of the document-level and word-level emotion predictions, based on the Chinese emotion corpus Ren-CECps, demonstrates a promising improvement for emotion recognition compared to the state-of-the-art emotion recognition algorithms. The DWET model outperforms all base-line algorithms for word and document emotion predictions. And the HDWET model, with a flexible concentration parameter ϕ injected in the hierarchy of corpus-level document emotion distribution, allows a self-adjustment of emotion distributions through different documents, and significantly improves emotion recognition for the less common emotion categories with even better Recalls and F-scores compared to the DWET model. We demonstrate the knowledge of emotion expression with respect to latent semantic dimensions through an emotion-topic diagram. By explicitly connecting semantic indexes with emotion categories and the closely related words, the diagram makes it easier to understand the semantic meanings in most general words and their underlying connections to the human emotions.

Our Bayesian models have simplified the language features for textual emotion recognition, by representing the word context with latent semantic dimensions and by associating emotion categories with the word contexts in a low dimensional feature space. However, our features could be somehow over-simplified because the context information in natural language expressions is still richer than the discrete semantic dimensions which we can maximally afford in our models. If we set too many semantic dimensions in the model, the probabilistic influence would flow wildly and the Bayesian inference could never converge. Another promising direction for exploring the rich context semantic information is through a deep neural network with multi-layer abstractions. The neurons are very different from our topic variables in that they do not separately (or linearly) represent semantic dimensions, but can be combined together to specify one point in a large semantic space, whose dimension is exponential to the number of neurons. In this sense, we would like to employ the deep neural networks for learning emotion expressions in natural language with a better semantic representation in our future work.

APPENDIX A
DERIVING POSTERIOR PROBABILITIES FOR DWET

The Bayes' theorem suggests that for a random variable x its posterior probability is proportional to the product of its prior probability and likelihood of other related observations in y. For a complicated Bayesian model, often there are many other variables which are not directly involved in the Bayesian inference, for which we use o to represent. We employ the following equation to represent the Bayes' theorem with the non-directly involved variables on the condition part

$$p(x|y,o) \propto p(x|o) \times p(y|x,o). \tag{21}$$

We illustrate the posterior probability derivations based on this equation.

In Section III, we make assumptions of conjugate prior probabilities for the proportional parameters, to simplify our Bayesian inference. This is because that with these conjugate prior probabilities, we can have a closed-form expression of their posterior probabilities after observing values in the related variables. For example, the proportional parameter θ^E_{dk} in (9) is assumed a Beta prior probability, which is the conjugate prior of its Bernoulli likelihood function in (4). After observing E through the documents in the test set, we can have posterior probability of θ^E_{dk} in a closed-form

$$\begin{aligned} \theta^E_{dk}|E;B &\sim \mathrm{Beta}(B_k + n_k) \\ &\sim \mathrm{Beta}(B^1_k + n^1_k, B^0_k + n^0_k) \end{aligned} \tag{22}$$

in which $n_k = (n^1_k, n^0_k)$ counts the occurrence and absence of document emotion k in this set.

The topic variable z_{di}, which specifies a semantic dimension for word w_{di}, has its posterior probability factorized by following the Bayes' theorem

$$\begin{aligned} &p(z_{di}|w,z_{-di},e,E;A,B,\beta,\tau) \\ &\quad\propto p(z_{di}|w_{-di},z_{-di},e,E;A,B,\beta,\tau) \\ &\quad\quad\times p(w_{di}|w_{-di},z,e,E;A,B,\beta,\tau) \end{aligned} \tag{23}$$

in which "$-$" on the subscript indicates the set of all variables except the one specified with the subscript. z_{di} in (23) corresponds to the variable of interest x in (21), w_{di} corresponds to the related observation y, and the rest variables of w_{-di}, z_{-di}, e, and E correspond to o. A, B, β, and τ are parameters in these probabilistic distributions.

We follow the categorical distribution assumption for the topic variable z_{di} in (2), and interpret the value of the first factor in (23)

$$p(z_{di}|w_{-di},z_{-di},e,E;A,B,\beta,\tau) = \theta^z_{di}. \tag{24}$$

The proportional parameter variable θ^z_{di} can be inferred through its posterior probability after observing z_d in document d

$$\theta^z_{di}|z_d;A \sim \mathrm{Dirichlet}(A + n_{dz_{di}}) \tag{25}$$

in which $n_{dz_{di}} = \sum_{i'} 1\{z_{di'} = z_{di}\}$ counts the occurrence of topic with the same value as z_{di}. This is because the prior probability Dirichlet for θ^z in (6) is the conjugate prior of its categorical likelihood function in (2), and we can simply update the parameters in Dirichlet to get its posterior probability. By taking the expectation of θ^z as defined in (25) to (24), we derive the algebraic expression for the first factor of topic posterior probability in (23)

$$p(z_{di}|w_{-di}, z_{-di}, e, E; A, B, \beta, \tau) = \frac{A + n_{dz_{di}}}{A \times J + W_d} \tag{26}$$

in which $W_d = n_{d*}$ is the number of words in document d.

Similarly, by following the assumption of Categorical distribution for the word variable w_{di} in (23), we can derive the algebraic expression for the second factor in (23)

$$\begin{aligned} &p(w_{di}|w_{-di}, z, e, E; A, B, \beta, \tau) \\ &\quad = \eta_{e_{di\cdot} z_{di} w_{di}} = \prod_{k\in K^1_{di}} \eta^1_{kz_{di}w_{di}} \prod_{k\in K^0_{di}} \eta^0_{kz_{di}w_{di}} \end{aligned} \tag{27}$$

in which $K^1 = \{k|e_{dik} = 1\}$ and $K^0 = \{k|e_{dik} = 0\}$ represent the sets of occurrence and absence of emotion category k in word w_{di}. We assume that different emotion categories can independently influence the observation of a word, and factorize the probability in (27) over emotion category k based on this assumption. η in (27) can be inferred through its posterior probability after observing w, z, and e through a set of documents

$$\eta^{1|0}_{kz_{di}w_{di}}|w, z, e; \tau \sim \text{Dirichlet}(\tau^{1|0}_{kz_{di}w_{di}} + n^{1|0}_{kz_{di}w_{di}}) \tag{28}$$

$n^1_{kz_{di}w_{di}} = \sum_{d'}\sum_{i'} 1\{(e_{d'i'}, z_{d'i'}, w_{d'i'}) = (1, z_{di}, w_{di})\}$ counts the occurrence of word emotion k, topic and word with the same values as z_{di} and w_{di}, while $n^0_{kz_{di}w_{di}} = \sum_{d'}\sum_{i'} 1\{(e_{d'i'}, z_{d'i'}, w_{d'i'}) = (0, z_{di}, w_{di})\}$ counts the absence of word emotion k but the occurrence of topic and word with the same value as z_{di} and w_{di}. A replacement of η in (27) with its expectation in (28) gives the algebraic expression for the first factor of topic posterior probability in (23)

$$\begin{aligned} &p(w_{di}|w_{-di}, z, e, E; A, B, \beta, \tau) \\ &= \prod_{k\in K^1} \frac{\tau^1_{kz_{di}w_{di}} + n^1_{kz_{di}w_{di}}}{\tau^1_{kz_{di}*} + n^1_{kz_{di}*}} \prod_{k\in K^0} \frac{\tau^0_{kz_{di}w_{di}} + n^0_{kz_{di}w_{di}}}{\tau^0_{kz_{di}*} + n^0_{kz_{di}*}} \end{aligned} \tag{29}$$

in which "$*$" indicates a summation over the corresponding dimension.

We take (26) and (29) into (23) to derive the algebraic expression of the posterior probability of topic variable z_{di} in (13).

Next, we factorize the posterior probability of word emotion variable e_{dik} by following the Bayes' theorem

$$\begin{aligned} &p(e_{dik}|w, z, e_{-dik}, E; A, B, \beta, \tau) \\ &\quad \propto p(e_{dik}|w_{-di}, z, e_{-dik}, E; A, B, \beta, \tau) \\ &\qquad \times p(w_{di}|w_{-di}, z, e, E; A, B, \beta, \tau) \end{aligned} \tag{30}$$

in which e_{dik} corresponds to the variable of interest x in (21), w_{di} corresponds to the related observation y, and the rest variables w_{-di} (all words except w_{di}), z, e_{-dik}, and E corresponds to o. A, B, β, and τ are parameters in these probabilistic distributions.

We follow the Bernoulli distribution assumption for the word emotion variable e_{dik} in (3), and interpret the probability value of the first factor in (30)

$$p(e_{dik}|w_{-di}, z, e_{-dik}, E; A, B, \beta, \tau) = \theta^e_{dkE_{dk}} \tag{31}$$

in which $\theta^e_{dkE_{dk}}$ can be inferred through its posterior probability after observing the document emotion and other word emotions in document d

$$\theta^e_{dkE_{dk}}|E_{dk}, e_{d\cdot k}; \beta \sim \text{Beta}(\beta_{kE_{dk}} + n_{dk}). \tag{32}$$

This is because the prior probability Beta for θ^e in (8) is the conjugate prior of its Bernoulli likelihood function in (3). And because θ^e is a document-level variable as shown in Fig. 1, we can simply update its parameters with observations in the specific document d to get its posterior probability. In (32), $e_{d\cdot k}$ corresponds to all the word emotion labels in document d and emotion category k, $n_{dk} = \sum_{i'} 1\{e_{di'k} = 1\}$ counts the occurrence of word emotion k in document d. A replacement of θ^e in (31) with its expectation in (32) gives the algebraic expression for the first factor of word emotion posterior probability in (30)

$$\begin{aligned} &p(e_{dik}|w_{-di}, z, e_{-dik}, E; A, B, \beta, \tau) \\ &\quad = \begin{cases} \dfrac{\beta^1_{kE_{dk}} + n_{dk}}{\beta^1_{kE_{dk}} + \beta^0_{kE_{dk}} + W_d}, & \text{if } e_{dik} = 1 \\ \dfrac{\beta^0_{kE_{dk}} + W_d - n_{dk}}{\beta^1_{kE_{dk}} + \beta^0_{kE_{dk}} + W_d}, & \text{if } e_{dik} = 0 \end{cases} \end{aligned} \tag{33}$$

in which W_d is the number of words in document d, and $W_d - n_{dk}$ corresponds to the absent count of word emotion k in document d.

The second factor in (30) is exactly the same as that in (23), whose derivation can be found in (27). We extend its expression by extracting the emotion category k out of the product for the convenience of later derivation, which gives

$$\begin{aligned} &p(w_{di}|w_{-di}, z, e, E; A, B, \beta, \tau) \\ &\quad = \begin{cases} \eta^1_{kz_{di}w_{di}} \times \prod\limits_{k'\in K^1/k} \eta^1_{k'z_{di}w_{di}} \prod\limits_{k'\in K^0} \eta^0_{k'z_{dk}w_{di}} \\ \qquad\qquad\qquad\qquad \text{if } e_{dik} = 1 \\ \eta^0_{kz_{di}w_{di}} \times \prod\limits_{k'\in K^1} \eta^1_{k'z_{di}w_{di}} \prod\limits_{k'\in K^0/k} \eta^0_{k'z_{dk}w_{di}} \\ \qquad\qquad\qquad\qquad \text{if } e_{dik} = 0 \end{cases} \\ &\quad \propto \begin{cases} \dfrac{n^1_{kz_{di}w_{di}} + \tau^1_{kz_{di}w_{di}}}{n^1_{kz_{di}*} + \tau^1_{kz_{di}*}}, & \text{if } e_{dik} = 1 \\ \dfrac{n^0_{kz_{di}w_{di}} + \tau^0_{kz_{di}w_{di}}}{n^0_{kz_{di}*} + \tau^0_{kz_{di}*}}, & \text{if } e_{dik} = 0 \end{cases} \end{aligned} \tag{34}$$

in which K^1/k and K^0/k represent the sets of occurred and absent emotion categories k' in word w_{di} except k, respectively. Because the products over K^1 and K^0 except k turn to be the same regardless of the assignment in e_{dik} in (34), we could take them out to simplify the calculation.

We take (33) and (34) into (30) to derive the algebraic expression of the posterior probability of word emotion variable e_{dik} in (14).

Finally, we factorize the posterior probability of document emotion variable E_{dk} by following the Bayes' theorem

$$\begin{aligned} p(E_{dk}|w,z,e,E_{-dk};A,B,\beta,\tau) \\ \propto p(E_{dk}|w,z,e_{-d\cdot k},E_{-dk};A,B,\beta,\tau) \\ \times\, p(e_{d\cdot k}|w,z,e_{-d\cdot k},E;A,B,\beta,\tau). \end{aligned} \tag{35}$$

We follow the Bernoulli distribution assumption for the document emotion variable E_{dk} in (4), and interpret the probability value of the first factor in (35)

$$p(E_{dk}|w,z,e_{-d\cdot k},E_{-dk};A,B,\beta,\tau) = \theta^E_{dk} \tag{36}$$

in which θ^E_{dk} can be inferred through its posterior probability after observing the document emotions as described in (22). A replacement of θ^E_{dk} in (36) with its expectation in (22) gives the algebraic expression for the first factor of document emotion posterior probability in (35)

$$\begin{aligned} p(E_{dk}|w,z,e_{-d\cdot k},E_{-dk};A,B,\beta,\tau) \\ = \begin{cases} \dfrac{B^1_k+1}{B^1_k+B^0_k+1}, & \text{if } E_{dk}=1 \\ \dfrac{B^0_k+1}{B^1_k+B^0_k+1}, & \text{if } E_{dk}=0. \end{cases} \end{aligned} \tag{37}$$

As illustrated in the derivation of (27), we assume that different emotion categories are independent, and gives the factorized production of the second factor in (35)

$$\begin{aligned} p(e_{d\cdot k}|w,z,e_{-d\cdot k},E;A,B,\beta,\tau) \\ = \prod_{i\in W_d} p(e_{dik}|w,z,e_{-dik},E;A,B,\beta,\tau). \end{aligned} \tag{38}$$

We take (37) and (38) into (35) to derive the algebraic expression of the posterior probability of document emotion variable E_{dk} in (15).

Appendix B
Deriving Posterior Probabilities for HDWET

The posterior probabilities for topic variable z_{di} and word emotion variable e_{dik} in the HDWET model are the same as those in the DWET model. For the document emotion variable E_{dk}, we factorize its posterior probability by following the Bayes' theorem to get

$$\begin{aligned} p(E_{dk}|w,z,e,E_{-dk},\phi;A,B,\beta,\tau) \\ \propto p(E_{dk}|w,z,e_{-d\cdot k},E_{-dk},\phi;A,B,\beta,\tau) \\ \times\, p(e_{d\cdot k}|w,z,e_{-d\cdot k},E,\phi;A,B,\beta,\tau). \end{aligned} \tag{39}$$

For the first factor in (39), we derive the same algebraic expression for the prior probability of document emotion E_{dk} as (36), but derive the expectation of θ^E_{dk} differently as follows. Because the proportional parameter θ^E_{dk} for document emotion distribution in the HDWET model follows the Beta distribution with a flexible concentration parameter $\alpha\phi_k$ in (11), we infer its poster probability from the samples $\hat{\phi}_k$ of the concentration parameter ϕ_k.

In the model construction for HDWET, the flexible concentration variable ϕ_k is assumed to follow the Beta distribution in (12). Because its related observation in document emotion variable E_{dk} follows the Bernoulli distribution in (4), the posterior probability of ϕ_k can be derived with a closed-form expression with corresponding distribution parameters B_k updated as in (17). We take the sampled values of $\hat{\phi}_k$ for variable ϕ_k to have the algebraic expression for the first factor in (39)

$$\begin{aligned} p(E_{dk}|w,z,e_{-d\cdot k},E_{-dk},\phi;A,B,\beta,\tau) \\ = \begin{cases} \dfrac{\alpha\hat{\phi}_k+1}{\alpha(\hat{\phi}_k+\bar{\hat{\phi}}_k)+1}, & \text{if } E_{dk}=1 \\ \dfrac{\alpha\bar{\hat{\phi}}_k+1}{\alpha(\hat{\phi}_k+\bar{\hat{\phi}}_k)+1}, & \text{if } E_{dk}=0. \end{cases} \end{aligned} \tag{40}$$

For the second factor in (39), because word emotion variables $e_{d\cdot k}$ shares the same model structure and probabilistic assumptions in HDWET and DWET, the derivation of its posterior probability turns to be the same as that of the DWET model, which gives

$$\begin{aligned} p(e_{d\cdot k}|w,z,e_{-d\cdot k},E,\phi;A,B,\beta,\tau) \\ = \prod_{i\in W_d} p(e_{dik}|w,z,e_{-dik},E,\phi;A,B,\beta,\tau). \end{aligned} \tag{41}$$

We take (40) and (41) into (39) to derive the algebraic expression of the posterior probability of the document emotion variable E_{dk} for the HDWET model in (16).

References

[1] H. Gunes and B. Schuller, "Categorical and dimensional affect analysis in continuous input: Current trends and future directions," *Image Vision Comput.*, vol. 31, no. 2, pp. 120–136, Feb. 2013.

[2] J.-C. Lin, C.-H. Wu, and W.-L. Wei, "Error weighted semi-coupled hidden Markov model for audio-visual emotion recognition," *IEEE Trans. Multimedia*, vol. 14, no. 1, pp. 142–156, Feb. 2012.

[3] S. Scherer, M. Glodek, G. Layher, M. Schels, M. Schmidt, T. Brosch, S. Tschechne, F. Schwenker, H. Neumann, and G. Palm, "A generic framework for the inference of user states in human computer interaction," *J. Multimodal User Interfaces*, vol. 6, no. 3–4, pp. 117–141, Nov. 2012.

[4] F. J. Ren, X. Kang, and C. Q. Quan, "Examining accumulated emotional traits in suicide blogs with an emotion topic model," *IEEE J. Biomed. Health Inform.*, vol. 20, no. 5, pp. 1384–1396, Sep. 2016.

[5] H. Mo, J. Wang, X. Li, and Z. L. Wu, "Linguistic dynamic modeling and analysis of psychological health state using interval type-2 fuzzy sets," *IEEE/CAA J. Automat. Sin.*, vol. 2, no. 4, pp. 366–373, Oct. 2015.

[6] L. Bylsma, B. H. Morris, and J. Rottenberg, "A meta-analysis of emotional reactivity in major depressive disorder," *Clin. Psychol. Rev.*, vol. 28, no. 4, pp. 676–691, Apr. 2008.

[7] G. Domes, L. Schulze, and S. C. Herpertz, "Emotion recognition in borderline personality disorder-a review of the literature," *J. Pers. Disord.*, vol. 23, no. 1, pp. 6–19, Feb. 2009.

[8] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, "The brain basis of emotion: a meta-analytic review," *Behav. Brain Sci.*, vol. 35, no. 3, pp. 121–143, Jun. 2012.

[9] J. T. Buhle, J. A. Silvers, T. D. Wager, R. Lopez, C. Onyemekwu, H. Kober, J. Weber, and K. N. Ochsner, "Cognitive reappraisal of emotion: a meta-analysis of human neuroimaging studies," *Cereb. Cortex*, vol. 24, no. 11, pp. 2981–2990, Nov. 2014.

[10] C. H. Yang, K. H. Y. Lin, and H. H. Chen, "Emotion classification using web blog corpora," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence*, Fremont, CA, 2007, pp. 275–278.

[11] R. Tokuhisa, K. Inui, and Y. Matsumoto, "Emotion classification using massive examples extracted from the web," in *Proc. 22nd Int. Conf. Computational Linguistics*, Manchester, United Kingdom, 2008, pp. 881−888.

[12] I. Maks and P. Vossen, "A verb lexicon model for deep sentiment analysis and opinion mining applications," *Proc. 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, Portland, Oregon, USA, 2011, pp. 10−18.

[13] S. M. Mohammad and T. Yang, "Tracking sentiment in mail: How genders differ on emotional axes," *Proc. 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, Portland, Oregon, USA, 2011, pp. 70−79.

[14] A. Balahur, J. M. Hermida, and A. Montoyo, "Detecting implicit expressions of emotion in text: A comparative analysis," *Decis. Support Syst.*, vol. 53, no. 4, pp. 742−753, Nov. 2012.

[15] W. Y. Li and H. Xu, "Text-based emotion classification using emotion cause extraction," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1742−1749, Mar. 2014.

[16] S. Matsumoto, H. Takamura, and M. Okumura, "Sentiment classification using word sub-sequences and dependency sub-trees," in *Advances in Knowledge Discovery and Data Mining*, T. B. Ho, D. Cheung, and H. Liu, Eds. Hanoi, Vietnam: Springer, 2005, pp. 301−311.

[17] T. Kudo and Y. Matsumoto, "A boosting algorithm for classification of semi-structured text. in *Proc. 2004 Conf. Empirical Methods in Natural Language Processing*, Barcelona, Spain, 2004, pp. 301−308.

[18] C. H. Yang, K. H. Y. Lin, and H. H. Chen, "Building emotion lexicon from weblog corpora," in *Proc. 45th Annu. Meeting of the ACL on Interactive Poster and Demonstration Sessions*, Prague, Czech Republic, 2007, pp. 133−136.

[19] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and T. Fukushima, "Collecting evaluative expressions for opinion extraction," in *Proc. 1st Int. Joint Conf. Natural Language Processing*, Hainan Island, China, 2004, pp. 596−605.

[20] H. Teramura, "Japanese syntax and meaning," Kurosio Publishers, 1982.

[21] X. Hu, J. S. Downie, and A. F. Ehmann, "Lyric text mining in music mood classification," in *Proc. 10th Int. Society for Music Information Retrieval Conf.*, Kobe, Japan, 2009, pp. 411−416.

[22] C. Strapparava and A. Valitutti, "WordNet-affect: an affective extension of wordNet," in *Proc. 4th Int. Conf. Language Resources and Evaluation*, Lisbon, Portugal, 2004, pp. 1083−1086.

[23] P. Singh, "The public acquisition of commonsense knowledge," in *Proc. AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, Palo Alto, CA, 2002.

[24] H. Liu, H. Lieberman, and T. Selker, "A model of textual affect sensing using real-world knowledge," in *Proc. 8th Int. Conf. Intelligent User Interfaces*, Miami, Florida, USA, 2003, pp. 125−132.

[25] C. Q. Quan and F. J. Ren, "A blog emotion corpus for emotional expression analysis in Chinese," *Comput. Speech Lang.*, vol. 24, no. 4, pp. 726−749, Oct. 2010.

[26] S. Aman and S. Szpakowicz, "Identifying expressions of emotion in text," in *Text, Speech and Dialogue*, V. Matoušek and P. Mautner, Eds. Pilsen, Czech Republic: Springer, 2007, pp. 196−205.

[27] X. Kang, F. J. Ren, and Y. N. Wu, "Semisupervised learning of author-specific emotions in micro-blogs," *IEEJ Trans. Electrical & Electronic Eng.*, vol. 11, no. 6, pp. 768−775, Nov. 2016.

[28] Y. N. Wu, K. Kita, F. J. Ren, K. Matsumoto, and X. Kang, "Modification relations based emotional keywords annotation using conditional random fields," in *Proc. 4th Int. Conf. Intelligent Networks and Intelligent Systems*, Kunming, China, 2011, pp. 81−84.

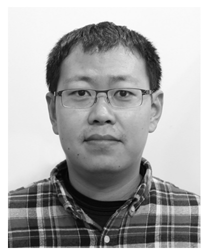
[29] Y. N. Wu, K. Kita, F. J. Ren, K. Matsumoto, and X. Kang, "Exploring emotional words for Chinese document chief emotion analysis," in *Proc. 25th Pacific Asia Conf. Language, Information and Computation*, Singapore, 2011, pp. 597−606.

[30] D. Das and S. Bandyopadhyay, "Word to sentence level emotion tagging for Bengali blogs," in *Proc. ACL-IJCNLP 2009 Conf. Short Papers*, Suntec, Singapore, 2009, pp. 149−152.

[31] X. Kang, F. J. Ren, and Y. N. Wu, "Bottom up: Exploring word emotions for Chinese sentence chief sentiment classification," in *Proc. 2010 Int. Conf. Natural Language Processing and Knowledge Engineering (NLP-KE)*, Beijing, China, 2010.

[32] X. Kang and F. J. Ren, "Sampling latent emotions and topics in a hierarchical Bayesian network," in *Proc. 2011 7th Int. Conf. Natural Language Processing and Knowledge Engineering (NLP-KE)*, Tokushima, 2011, pp. 37−42.

[33] F. J. Ren and X. Kang, "Employing hierarchical Bayesian networks in simple and complex emotion topic analysis," *Comput. Speech Lang.*, vol. 27, no. 4, pp. 943−968, Jun. 2013.



Xin Kang (M'16) received the Ph.D degree from Tokushima University, Tokushima, Japan, in 2013, the M.E. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2009, and the B.E. degree from Northeastern University, Shenyang, China, in 2006. He is currently an Assistant Professor in Tokushima University, Japan. His research interests include statistical machine learning, probabilistic graphical models, neural networks, and text emotion prediction.

Fuji Ren (SM'03) was born in 1959, in China. He received the B.E. and M.E. degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and 1985, respectively. He received his Ph.D. degree in 1991 from Hokkaido University, Japan. From 1991, he worked at CSK, Japan, where he was a Chief Researcher of NLP. From 1994 to 2000, he was an Associate Professor in the Faculty of Information Sciences, Hiroshima City University. He became a Professor in the Faculty of Engineering of the University of Tokushima, Japan, in 2001. His research interests include natural language processing, artificial intelligence, language understanding and communication, and affective computing. He is a Member of IEICE, CAAI, IEEJ, IPSJ, JSAI, and AAMT. He is a Fellow of the Japan Federation of Engineering Societies. He is the President of the International Advanced Information Institute.

Yunong Wu received the Ph. D degree and Master degree from The University of Tokushima, Tokushima, Japan, in 2014 and 2011, respectively. He is currently an Assistant Professor in Tokushima University, Japan. His research interests include statistical machine learning, affective information computing, neural network and probabilistic graphical model.