A Concept-level Emotion Cause Detection Model for Analyzing Microblogging Users' Emotions

Shuangyong Song, Chao Wang, Haiqing Chen
Alibaba Groups, Hangzhou 311121, China
No. 969, West Wenyi Road, Yuhang District, Hangzhou 311121, Zhejiang Province, China
{shuangyong.ssy; chaowang.wc; haiqing.chenhq}@alibaba-inc.com

Abstract

Over the last few years, microblogging is increasingly becoming an important source of up-to-date topics about what is happening in the world. Hot social events can usually lead to heated discussion and strong emotional expression among microblogging users, while microblogging users may have different emotions on the same event, and those emotions are generated from different aspects of the event with various causes. In this paper, we propose a Concept-level Emotion Cause Model (CECM), instead of the existing mere word-level models, to discover the causes of microblogging users' diversified emotions on specific hot events. Firstly, we design a topic-supervised biterm topic model to detect users' multiple 'emotional topics' in event-related microblogs. The dimensionality of the Dirichlet distribution in biterm topic model is restricted to the same number of emotional categories as in our collected emotion dictionary, and a binary distribution is used to restrict each topic only describe one emotion. Secondly, we utilize context-sensitive topical PageRank to detect meaningful multiword expressions from each emotional topic as its causes. Thirdly, relationships between microblogging emoticons and people's real emotions can also be detected with the emotional topic detection step in our model, which shows that emoticons are also event-sensitive on emotional expression. Experimental results on a real-world dataset from Sina Weibo, one of the largest microblogging websites in China, show CECM can better detect emotion causes than baseline methods.

Keywords

Microblogging, Emotion Cause Detection, Topic-supervised Biterm Topic Model, Context-sensitive Topical PageRank.

1. Introduction

Compared to regular blogging, microblogging realizes an even faster mode of communication. During the last decade, microblogging has become one of the most popular web services, and it has attracted more than one billion registered users on different microblogging websites up to now, and this number is still growing rapidly. Besides, a great deal of media attention has been focused on micro-blogging. For example, on April 17, 2007, the day when Oprah Winfrey joined Twitter, the most famous micro-blogging website in the world, by sending a microblog from her Friday TV show, shares of US based visits to the Twitter site increased by 24% and some 1.2 million new users signed up for Twitter on that day alone. In May 2007, even the White House began posting short messages on Twitter, which shows the impact of microblogging in various fields.

Microblogging users can publish short message updates, which are dubbed microblogs, in different channels, including the Web, SMS, e-mail or instant messaging clients. Unlike other social network services, microblogging allows a user A to "follow" other users without seeking any permission, and in real time the updates of the followed users will be sent to A automatically. Those convenient conditions for fast diffusion of information makes microblogging a highly popular platform for web users to seek new trends and express emotions about their favorite hot events [Song et al. 2010]. Generally, users may have different emotions on a event, and those emotions are usually originated from different aspects of this event with various causes [Lee et al. 2013]. For example, about the hot disaster event ‘missing Malaysia plane MH370’, which happened on March 8, 2014, more than forty million microblogs have been published on *Sina Weibo*, and the microblogs published by victims' relatives showed their hearts are filled with sadness and fear, while other people may feel sympathetic for victims' relatives and angry with the Malaysia Airlines. By analyzing people’s emotions and the causes, users can better understand the details of events they are interested in, such as the different perspectives or development tendency of those events.

Existing studies on emotion cause detection have focused on discovering most frequently co-occurred clauses or words with a single emotional word [Chen et al. 2010; Lee et al. 2009; Lee et al. 2013; Lu et al. 2013]. Those approaches are feasible when the average length of texts is not short; however, it is hard to discover the co-occurrence relation between emotions and their causes in microblog-like short texts [Song et al. 2010]. Furthermore, causes of a single emotional word cannot thoroughly describe the information of a specific emotion, since emotion is an abstract concept and each emotion is actually an abstract topic [Li et al. 2012; Song & Meng, 2015]. To solve these problems, we propose a Concept-level Emotion Cause Model (CECM) to detect abstract emotional topics and related concept-level emotion causes in microblogging.

Microblogging users may use various expressions to utter their emotions, rather than just using standard emotional words [Li et al. 2012]. In addition, those non-emotion-word expressions on different events are generally diverse, which largely depends on domains and topics [Song et al. 2014]. Therefore, taking typical emotional words as seeds to automatically detect user emotion on specific events is more laborsaving and robust than training models with massive manual annotations. By CECM model, we first preprocess the microblogging data and perform Chinese word segmentation, and then design a topic-supervised biterm topic model to detect emotional topics with a manually collected emotion dictionary. Finally, we detect emotion causes from emotional topics, and also identify the event-oriented relationship between emoticons and emotions.

2. Related Work

There has been a rich mass of literature related to user emotion analysis and emoticon analysis. Existing works mostly focused on user emotion classification in web-based text data and there are few researches on user emotion cause detection. Based on the content of this paper, we introduce the related works from three aspects: text based user emotion classification, emoticon based user emotion classification, and user emotion cause detection. We will also introduce some related works on supervised topic models, which we use for emotional topic detection.

2.1. Text Based User Emotion Classification

Researches on recognizing and classifying emotions in different types of text, such as news reports, product reviews, weblogs, microblogs, and customer feedbacks, have attracted much interest [Li & Xu, 2014]. Masum et al [2007] designed a character-based system called "Emotion Sensitive News Agent" (ESNA), which aims to categorize themes of news into 8 emotion types with analyzing emotional information conveyed through text. Tokuhisa et al [Tokuhisa et al. 2008] proposed a two-step classification model to infer the emotion of a speaker, which contains sentiment polarity classification step and emotion classification step. Kontopoulos et al [Kontopoulos et al. 2013] proposed an ontology-based model for sentiment analysis of microblogs, since traditional text-based sentiment classifiers can just perform well in long texts such as weblogs or web news. Besides, unlike traditional approaches which are mostly based on statistical methods, Li and Xu [2014] introduced the emotion cause extraction technique [Lee et al. 2010] to improve the quality of selected features for improving the performance of emotion classification task in microblogging, based on an argument that a triggering cause event is an integral part of emotion, and finally it is proved that those new features are effective. All those researches are about how to recognize or classify emotions in web text, however, deeper level information regarding emotions, such as the experiencer, cause, and result of an emotion, needs to be extracted and analyzed for real world applications. In this paper, we aim at mining one of the crucial deep level information, i.e. emotion cause, which provides useful information for applications ranging from economic forecasting, public opinion mining, to product design [Lee et al. 2009].

2.2. Emoticon Based User Emotion Classification

Emoticon	*Meaning*	*Emoticon*	*Meaning*	*Emoticon*	*Meaning*
	[laugh]		[sad]		[heart]
	[handshaking]		[candle]		[scold]
	[sick]		[yawn]		[lovely]

Table 1: Examples of Emoticons on *Sina Weibo*

Emoticon is a portmanteau word formed from 'emotion' and 'icon'. As new visual or nonverbal communication cues used in digital interaction, emoticons can realize pictorial representation of authors' facial expressions or behavioral states, which contain rich emotional information of them and make important effects on web user emotion analysis [Hsiao and Hsieh, 2014]. Several examples of emoticons on *Sina Weibo* and their intuitive meanings are given in table 1.

Some researches take emoticons as a direct method to determine user emotion. Zhao et al [Zhao et al. 2012] utilized 9 angry emoticons, 14 disgusting emoticons, 50 joyful emoticons and 22 sad emoticons to build a training dataset for user emotion classification in microblogging, and create a *MoodLens* system to realize the prediction of the detailed sentiment of microblogs, by employing *Naïve Bayes* classifier. Although this method for building training dataset is convenient, taking emoticon as the only factor to decide users' emotion is not comprehensive enough. For example, the emoticon [heart] means 'love you' in microblogs to express the feeling to lovers, children or parents, but it may also mean 'praise' to some heroic feats of policemen or philanthropists. This emotional diversity of some emoticons with less obvious emotional characteristics, such as [microphone], [pick-nose] and [rabbit], is even more serious. Aoki and Uchida [Aoki and Uchida, 2011] utilize weblog articles to automatically generate the emotional vectors of emoticons shown in weblog articles, which can obtain the probability distribution of each emoticon on each emotion. For example, in their experimental result, the emoticon [music] has a 0.71 probability of expressing joy, and 0.21 and 0.07 for love and anticipation respectively. We aim to detect event-oriented emotional probability distribution of emoticons, to obtain more detailed and strict emoticon analysis.

2.3. User Emotion Cause Detection

Lee et al [2009] developed the first work on automatic emotion cause detection from text corpus, by designing a linguistic rule based method. By examining emotional sentences, they generalize some rules for identifying the cause of emotion verbs. Similar methods were used in other researches [Li et al. 2010; Li et al. 2013]. Furthermore, by applying the method in [Li et al. 2010] on short informal posts in the microblogging community, Li and Xu [2014] utilized a manually built Chinese microblogging emotion corpus to examine the effectiveness of each linguistic cue word for detecting emotion cause events. Some less useful linguistic cue words were removed, and some new ones were added into the list. This model is referred by us to some extent, for selecting emotional words with small ambiguity as our seeds in emotional dictionary.

Chen et al [2010] discover that more than 85% of the emotion causes are from clauses around emotional words, then they create two sets of linguistic patterns during the feature extraction step based on linguistic analysis, and accordingly design a multi-label classification model to detect which clause or clauses are the causes. Lu et al [2013] aim to detect the relationships between emotion-inducing words and emotion-describing words, by considering that the polarities of posts would not only be influenced by emotional words, but also by some relevant types of words, such as negation words and modal words, and then they utilize the PCA model to find the most related words with emotional word to be the emotion cause. All the previous works take a single emotional word as an emotion expression and focus on discovering the explicit co-occurrence relations between emotional words and other context. In this paper, considering the topicality and abstraction of emotion, we initiatively design a topic supervised biterm *LDA* model to discover implicit causes of emotional topics instead of single emotional words.

2.4. Supervised Topic Models

Topic models are important algorithms in text analysis domain, which are designed to discover the main implicit themes contained in a large collection of documents. Accordingly, topic models can well organize the document collection based on the discovered implicit themes [Blei 2012]. *LDA* is a type of topic model proposed in [Blei et al. 2003], which is an unsupervised statistical technique to discover thematic topics that permeate a large corpus of text documents. *LDA* has been successfully adapted to find patterns in many kinds of data, such as web text data, image data, and social networks in some applications [Zhang et al. 2015]. Besides, some works have attempted to add constraints on *LDA* to meet special topic demands. For example, Andrzejewski et al [2009] added domain knowledge using a novel Dirichlet Forest prior to impact probability of words in various topics, and similar method also existed in work of [Hu et al. 2011]. Sadamitsu et al [2012] designed a semi-supervised topic model to construct arbitrary class-based word dictionaries. This semi-supervised topic model is trained with a non-annotated web news text corpus, by setting the number of topics to be the same as the number of newsgroups in that corpus. All these previous works tried to improve the *LDA* which lacks a mechanism for incorporating domain knowledge, by adding on appropriate constraints. In this paper, to restrict that each abstract topic in the topic modeling result can just describe one emotion, we supervise the topics by adding binary distributions of emotional words on different emotional topics.

3. Problem Definition & Model Description

3.1. Problem Definition

Referring to the description of ‘emotion cause detection’ in [Lee et al. 2013] and the description of ‘concept-level sentiment analysis’ in [Cambria 2013], we give the definition of an *emotional topic* as ‘a probability distribution over words according to their relevance to a specific emotion’, and assume that emotional expressions and emotion causes are both contained in the related emotional topic. Different from the mere word-level representation of emotions, we consider emotions as abstract topics at the concept level.

In our collected emotion dictionary, the number of emotions is K, and emotions are accordingly defined as $E = \{ e_1, e_2, \ldots, e_K \}$. The number of emotional topics is also K, decided by our emotion dictionary, and the emotional topics are accordingly denoted as $T_e = \{ t_{e1}, t_{e2}, \ldots, t_{eK} \}$. Assuming the number of emoticons in our data is N, we define emoticons as $I = \{ i_1, i_2, \ldots, i_N \}$. Given an event related microblogging dataset, our goal is to detect causes of concept-level emotions from T_e for each e_k ($1 \leq k \leq K$), and confirm the effect of each i_n ($1 \leq n \leq N$) on expressing different emotions about the given event.

3.2. System Architecture of the Proposed Model

The system architecture of the proposed *CECM* model is shown in Figure 1, which consists of four functional modules, namely, preprocessing and word segmentation, emotional topic detection, emotion cause detection and emoticon-emotion relationship detection. In preprocessing and word segmentation module, several processing are implemented to normalize microblogs, and all of the microblogs need to be segmented into words. In emotional topic detection module, we design a topic-supervised biterm topic model to detect users' multiple 'emotional topics' in event-related microblogs. The dimensionality of the Dirichlet distribution in biterm topic model is restricted to the same number of emotional categories as in our collected emotion dictionary, and a binary distribution is used to restrict that each topic can just describe one emotion. In emotion cause detection module, we utilize a context-sensitive topical PageRank method as proposed in [Zhao et al. 2011] to detect the most meaningful multiword expressions from each emotional topic as its causes, and we formulate the PageRank with the Markov chain model, which can reduce the complexity of implementation and improve the computational efficiency. In emoticon-emotion relationship detection module, we utilize a defined relevance threshold to detect the relationships between emoticons and people's real emotions. Mechanism of each functional module in our proposed model will be discussed in details in the following sections.

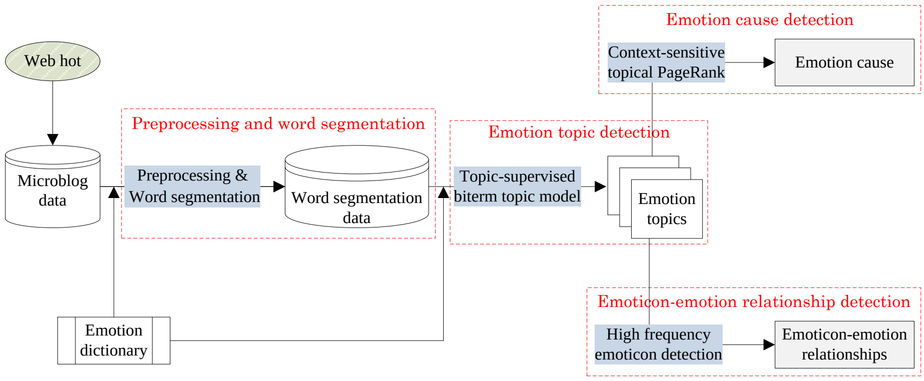

Figure 1: System architecture of the proposed method.

3.3. Preprocessing and Word Segmentation

In *Sina Weibo*, if a user replys others' microblogs without any comments, the system will add 'forwarding microblogs' automatically. Such a denotation does not have any effect on reflecting users' interests and emotions. Therefore, we remove those 'forwarding microblogs' tags from microblogs, but keep the content of the replyed texts, since the replyed texts represent users' interests. Besides, the "@username" symbols are removed since they do not actually represent meaningful content or emotions. URLs, non-texts and 'forwarding' tags are also removed. Besides, microblogging allows users to insert emoticons in microblogs to express sentiments, such as '[anger]', '[yeah]' and '[kiss]'. To detect relations between emoticons and users' emotions, we keep those emoticons as simple terms [Yang et al. 2007]. Since our study is carried out on Chinese microblogs, there is a problem caused by Chinese word segmentation. To address this issue, we apply *ICTCLAS* system [Zhang et al. 2003], which can get a precision of about 98% on Chinese word segmentation task, to perform Chinese word segmentation on the corpus, by taking all emotional words in our collected emotion dictionary as user defined words. Then microblogs containing at least one emotional word will be kept in final experimental data.

3.4. Emotional Topic Detection

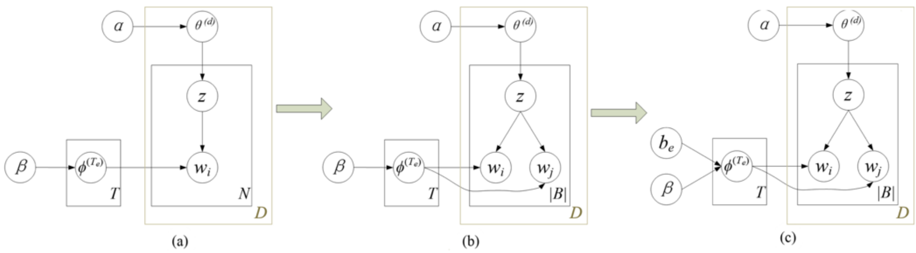

Figure 2: (a) Graphical model representation of the LDA topic model proposed in (Blei et al. 2003); (b) Graphical model representation of the biterm topic model proposed in (Yan et al. 2013); (c) Graphical model representation of our proposed topic-supervised biterm topic model.

Figure 2 (a) shows the graphical representation of LDA topic model proposed in [Blei et al. 2003]. Formally, each microblog is associated with a multinomial distribution over K topics, which is denoted as θ(d). Each topic is associated with a multinomial distribution over keywords, denoted as $\phi^{(T_e)}$. $\theta(d)$ and $\phi^{(T_e)}$ have *Dirichlet* prior with hyper-parameters α and β respectively. Figure 2 (b) shows the graphical representation of the biterm topic model (*BTM*) proposed in [Yan et al. 2013]. A biterm means 'an unordered word-pair co-occurred in a short text', and the *BTM* discovers implicit topics from whole short texts corpus by directly modeling the generation of biterms, in which the $|B|$ means the total number of those biterms. Compared with conventional topic models, biterm modeling can enhance the topic modeling performance, especially for short texts corpus [Yan et al. 2013].

The graphical representation of proposed topic-supervised biterm topic model (*TS-BTM*) is shown in Figure 2 (c), in which we limit the topics as abstract forms for each emotion. In the model training procedure of *TS-BTM*, this particular point needs to be emphasized: the biterms cannot include emotional words or emoticons. For discovering T_e, we add binary distributions of emotional words on different emotions into the biterm topic model, which means that the probability of an emotional word appearing in the corresponding emotional topic is 1 and in other emotional topics is 0, to restrict each topic can just describe one emotion. The specific generative process of the corpus in *TS-BTM* is described in Algorithm 1:

1. For each topic T_e

    (a) draw a topic-specific word distribution:

        for emotional words, $\phi^{(T_e)} \sim b_e$

        for other words, $\phi^{(T_e)} \sim \mathrm{Dir}(\beta)$

2. Draw a topic distribution $\theta \sim \mathrm{Dir}(\alpha)$ for the whole collection

3. For each biterm b in the biterm set B

Algorithm 1: generative process of the corpus in *TS-BTM*

Following the above procedure, the joint probability of a biterm $b = (w_i,w_j)$ is:

$$P(b) = \sum_{T_e} P(T_e)P(w_i \mid T_e)P(w_j \mid T_e) = \sum_{T_e} \theta_{T_e}\phi_{i|T_e}\phi_{j|T_e} \tag{1}$$

Thus the likelihood of the whole corpus is:

$$P(B) = \prod_{i,j}\sum_{T_e} \theta_{T_e}\phi_{i|T_e}\phi_{j|T_e} \tag{2}$$

3.5. Emotion Cause Detection

With the *TS-BTM*, we can get the T_e. Then we can detect causes for each emotion e_k from t_{ek} with context-sensitive topical PageRank (*cTPR*) method as proposed in (Zhao et al.

2011). *cTPR* takes terms as nodes in term graph G, and the weight of edge from term m_i to term m_j is decided by the frequency of m_j showing as a previous term of m_i in topical microblogs data. Then the topical ranking value of each term is calculated by the equation (3):

$$R_t(m_i) = \lambda \sum_{j:m_j \to m_i} \frac{e_t(m_j, m_i)}{O_t(m_j)} R_t(m_j) + (1-\lambda) P_t(m_i) \tag{3}$$

where $R_t(m)$ is the topic-specific PageRank score of term m in topic t, while $e_t(m_j,m_i)$ is the weight for the edge ($m_j \rightarrow m_i$), and $O_t(m_j)$ means sum value of the weights of out-links of m_j, that is to say, $O_t(m_j) = \sum_{m'} e_t(m_j, m')$. The λ is a damping factor ranging from 0 to 1. In this work, we still set the damping factor λ as 0.85, which was used in the original PageRank paper [Brin and Page, 1998]. The topic-specific preference value $P_t(m)$ for each term m is its random jumping probability with the constraint that $\sum_{m \in V} P_t(m) = 1$ given topic *t*, while the *V* is the term vocabulary. A large $R_t(.)$ indicates a term is a good candidate keyword in topic t. Initially, the $R_t(m)$ for each m is set to be 1. The computation ends when for some small ε:

$$\left| [R_t(m_i, r)] - [R_t(m_i, r-1)] \right| < \varepsilon \tag{4}$$

i.e., when convergence is assumed. In Eq. (2), $R_t(m_i,r)$ means $R_t(m_i)$ in the r_{th} iteration, where $1 \le r \le R$, and R is the maximal iteration number set up in our experiments.

We can formulate this topical term ranking problem with Markov chain model [Song et al. 2012]. We treat the term graph G as a Markov chain, each term as a state and each link as a transition from one state to another. In consideration of the adjacency relation among the states in the Markov chain, we denote the transition probability matrix of the chain with a $N*N$ adjacency matrix *A*, where the *N* is the number of terms in *V*. Using $\sum e_t(*, m_i)$ to represent the summation of weight of links from mi to other terms and *E* is the collection of links in G, we define each cell of the matrix *A* as the following.

$$A_{ij} = \begin{cases} \frac{e_t(m_j, m_i)}{\sum e_t(*, m_i)} & if\ (i,j) \in E \\ 0 & otherwise \end{cases} \tag{5}$$

Terms that have no links with other terms were deleted. Overall, the matrix *A* is a transition probability, i.e., column-stochastic with all columns consisting at least one nonzero element. If *P* is a column vector, and it can meet the following equation:

$$P = \hat{A}P \qquad (\hat{A} = dA^T + \frac{1-d}{N} E) \tag{6}$$

where *P*, the principal eigenvector of $\hat{A}$ with eigenvalue of 1, is in fact the topic-specific score column vector containing all $R_t(m)$ [Li et al. 2008]. Thus, after calculation of *P*, we can easily get the topic-specific ranking of terms. More importantly, this converted form of PageRank just need a much smaller program realization than the original PageRank form with loop iterates.

After term ranking using *cTPR*, a common keyphrase generation method is used to detect keyphrases in each t_{ek} as emotion cause. We first select top 100 non-emotional terms for each t_{ek}, and then look for combinations of these terms that occur as frequent phrases in the text collection with utilizing method in [Zheng et al. 2015] to process the 'substring problem'. In formula (5), T_{length} is the length of a phrase, which is the number of characters in it, and $T_{frequency}$ is the frequency of this phrase. T_{value} means the 'value for keeping' of a phrase, which is determined by the above two factors T_{length} and $T_{frequency}$. If a substring phrase has a larger T_{value} than a longer phrase that contains it, we keep the substring phrase. Otherwise, we keep the longer phrase and delete the substring phrase. Furthermore, we add a parameter η to adjust the significance of the two factors, T_{length} and $T_{frequency}$.

$$T_{value} = T_{frequency} * (T_{length})^{\eta},\ (0 < \eta < \infty) \tag{7}$$

Finally, all the saved phrases whose frequencies are larger than a threshold number, which is set to be 5 empirically, are taken as candidate emotion causes.

3.6. Emoticon-Emotion Relationship Detection

We define $R(e_k,i_n)$ as relevance value between e_k and i_n, which is contained in the detection result of t_{ek}. Then we set a *threshold* in formula (6) to judge if an emoticon is possibly related to an emotion, where *p* is a factor for adjusting the relation between the average 'emoticon-emotion' similarity value and the *threshold* value. Then, we utilize the ranking of remained *threshold*-exceeded emoticons in each emotional topic to evaluate the importance of them.

$$threshold = p * \frac{\sum_{k=0}^{K}\sum_{n=0}^{N} R(e_k, i_n)}{K * N} \tag{8}$$

4. Experiments and Performance Evaluation

4.1. Dataset and Parameter Setting

Sina Weibo is a well-known Chinese microblogging service, and the dataset in our experiments is collected from *Sina Weibo*. We choose 5 hot events as experimental subjects, which are Missing Malaysia Plane MH370, Ebola Virus, 2014 FIFA World Cup in Brazil, 2011 Japanese Earthquake and Child Trafficking. Three of them are the latest hot events, which occurred in 2014. ‘Missing Malaysia Plane MH370’ is a flight accident, and ‘Ebola Virus’ is about disease outbreaks, while ‘2014 FIFA World Cup in Brazil’ is a large-scale sport competition. Besides, ‘2011 Japanese Earthquake’ is a historical catastrophic event and ‘Child Trafficking’ is a long-existing social problem.

Events	*Period*	*Microblogs*	*Words*	*Emoticons*
Esteem Missing Malaysia Plane MH370	2014.03.08 – 2014.07.31	30,745	14,225	536
Ebola Virus	2014.03.04 – 2014.07.31	7,938	10,341	320
2014 FIFA World Cup in Brazil	2014.02.27 – 2014.07.31	85,332	19,033	681
2011 Japanese Earthquake	2011.03.01 – 2011.04.26	42,109	17,929	828
Child Trafficking	2011.03.01 – 2011.04.26	101,341	18,323	453
Total (dedup of *Words* and *Emoticons*)	2011 period & 2014 period	267,465	24,458	855

Table 2: Information of Experimental Dataset

We try to cover different domains by choosing varied types of events. Detailed information about this dataset is given in Table 2, while our dataset cannot include all event-related microblogs due to the downloading difficulty. Take ‘2011 Japanese Earthquake’ for example, dataset about this event contains 42,109 microblogs, comprising 17,929 different words and 828 emoticons.

Emotion dictionary used in our experiments is collected manually by considering some different related resources, which contains 34 emotions and 542 emotional words. All emotional words in this dictionary explicitly express a single emotion, and some ambiguous emotional words expressing different emotions have been edited out of the final dictionary. In Table 3, we give some examples about 8 emotions in this dictionary, where [*] for pronunciations of Chinese words and (*) for English translation of Chinese words. In addition to these 8 emotions, other ones contain enjoy, resentment, annoyance, depressed, proud, injustice, worry, flustered, ashamed, remorse, doubt, appreciate, miss, contempt, envy, satisfied, expect, disappointment, lonely, awkward, hate, boring, high, low, hesitate and calm. According to the information of the microblogging data and emotion dictionary data, we set the parameter K to be 34 and N to be 855. Besides, setting of parameters η and p will be discussed in the following subsections, since their setting needs manual labeling results.

Emotions	***Emotional words*** (Chinese, pronunciation, English)

Esteem	敬佩	[jingpei]	(admire)
	敬仰	[jingyang]	(respect),
	敬重	[jingzhong]	(esteem)
Sadness	哭	[ku]	(cry)
	悲痛	[beitong]	(sorrow)
	叹息	[tanxi]	(sigh)
	悲哀	[beiai]	(sad)
Sympathy	同情	[tongqing]	(sympathy)
	可怜	[kelian]	(pity)
	可惜	[kexi]	(unfortunately)
Happiness	惊喜	[jingxi]	(rapture)
	高兴	[gaoxing]	(delightful)
	欣幸	[xinxing]	(glad)
Anger	愤恨	[fenhen]	(anger)
	愤慨	[fenkai]	(indignation)
	愤怒	[fennu]	(rage)
Fear	忧惧	[youju]	(apprehensive)
	害怕	[haipa]	(fear)
	胆怯	[danqie]	(timid)
Inferiority	自卑	[zibei]	(self-abasement)
	自惭形秽	[zijianxinghui]	(ashamed)
	自馁	[zinei]	(discouraged)
Surprise	惊奇	[jingqi]	(surprise)
	吃惊	[chijing]	(astonished)
	惊讶	[jingya]	(amazed)
	震惊	[zhenjing]	(shocked)

Table 3: Emotion Dictionary Examples

4.2. Methods for Comparison

To evaluate the performance of the proposed algorithm on emotion cause detection, we compare it with several baseline models, which are all unsupervised:

Co-occurrence method: using the emotion cause detection method, which is mainly based on co-occurrence relation between emotional words and candidate cause phrases, given in [Lee et al. 2013] to detect emotion causes. However, we are trying to evaluate the performance of our model on event oriented emotion cause detection, so we need to make a few alterations on the method of Lee et al. For each emotion of each hot event, we detect the cause phrases for each word in this emotional word list, and most frequently detected phrases are taken as the causes for this emotion.

Uniterm *LDA* method: detecting concept-level emotion causes but using a uniterm topic model to obtain emotional topics. The uniterm topic model we use here is actually the initial *LDA* model [Blei et al. 2003], and this baseline is for evaluating the effect of biterm topic model on analyzing short texts.

PCA-based model: this model has been described in section 2.

4.3. Gold Standard Generation

Since microblogging users always just care about the top ranking emotion causes of hot events they are interested in, we just evaluate top 10 causes for each emotion of our chosen 5 hot events. Then, to detect the most appropriate value of η, we set different η as 0.1, 0.2, 0.5, 1, 2, 5 and 10, and get the emotion cause detection result of different *CECM* model with those η.

For experimental evaluation, we have 3 tagging volunteers manually label the emotion causes. After deduping of those causes from three baseline methods and seven *CECM* models on same emotion about same event, each annotator should give them integral scores between 1 and 5. Finally 3,121 emotion causes are labeled. In the process of scoring, annotators can search the information related to hot events from web search engine, or just consider own understanding and interest on those hot events. We then use the average value of the 'Cause Probability' as the final score, and Fleiss' Kappa is adopted to verify the degree of agreement among the three annotators, which is 0.74, indicating substantial agreements. Ranking result of those causes will be used to get *NDCG@k* values and 10 causes with maximum scores are set as 'right causes'.

For the 'emotion-emoticon' relationship detection part, we set different p to detect the most appropriate value of p. We empirically believe that the threshold value should be less than the average 'emoticon-emotion' similarity value, so we set different p=0.1~1 with interval as 0.1. We detect the 'emotion-emoticon' relationship result when p=0.1 for manual labeling, which corresponds to the lowest threshold.

4.4. Evaluation Metrics

In this study, we use both Normalized Discounted Cumulative Gain at top k (*NDCG@k*) and Mean Average Precision (*MAP*) as our evaluation metrics to measure the effects of *CECM,* which are both fit for evaluating top results sensitive ranking problems such as web search, keyphrase detection and information recommendation, etc [Zheng et al. 2015]. The Discounted Cumulative Gain at top k ($DCG@k$) for keyphrase detection task is defined as:

$$DCG@k(y)=\sum_{i=1}^{k}\frac{2^{rel_i}-1}{\log(1+i)} \tag{9}$$

where y means the candidate keyphrase ranking list in our experiments, and rel_i means the real keyphrase score of the i^{th} in our result. Then the $NDCG@k$ is accordingly defined as:

$$NDCG@k(y)=\frac{DCG@k(y)}{DCG@k(y^*)} \tag{10}$$

where y^* is a perfect ranking result which corresponds to any perfect ordering based on manual tagging scores.

MAP is defined as the average precision on each day a real keyphrase is detected, as given below:

$$MAP=\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=1}^{i}r(j)}{i}*r(i) \tag{11}$$

where i is the position of the keyphrase in the ranking list and we want to evaluate top N results. The $r(i)$ is a binary value: when the suggested keyphrase at position i belongs to top N ones, $r(i)$ is set to be 1 and 0 otherwise.

In this paper, we set the k in $NDCG@k$ to be 3, 5, and 10 respectively, and the N in *MAP* is set to be 10 according to the generating method of gold standard. After obtaining the $NDCG@k$ and *MAP* of experimental results in each day, we further calculate the average value as our final evaluation metrics.

4.5. Experimental Results and Discussions

η	0.1	0.2	**0.5**	1	2	5	10
MAP	0.927	0.931	**0.938**	0.935	0.902	0.897	0.895

Table 4: The MAP (Mean Average Precision) results of different values of $\boldsymbol{\eta}$ for our *CECM* model on concept-level emotion cause detection

The parameter η is used to adjust the significance of two factors, T_{length} and $T_{frequency}$, in formula (5). A value of η equal to 1 indicates that T_{length} and $T_{frequency}$ have same impact on the T_{value} of a phrase. Besides, a value of η smaller than 1 indicates that T_{length} has a smaller impact on T_{value} than on $T_{frequency}$, and vice versa. Table 4 shows the *MAP* results when varying η as seven different values from 0.1 to 10. As shown, the curve peaks at η=0.5, which indicates that $T_{frequency}$ has larger impact on the T_{value} than on T_{length}. Therefore, we set η=0.5 in the following experiments.

Table 5 demonstrates the evaluation results of comparing four methods: *Co-occurrence* *method*, *Uniterm LDA method*, *PCA-based model* and *CECM*. From table 5, we can see *CECM* outperforms three baselines. The improvement is due to the effect of biterm topic model on analyzing short texts. In *uniterm LDA method*, PageRank is used for ranking topic-specific single words, which contain less contextual information than biterms, so more frequent topic-insensitive single words may be detected as top ranking topic-specific terms.

	NDCG@3	*NDCG@5*	*NDCG@10*	*MAP*
Co-occurrence method	0.835	0.819	0.852	0.863
Uniterm LDA method	0.884	0.849	0.882	0.920
PCA-based model	0.878	0.844	0.873	0.901
CECM	**0.952**	**0.939**	**0.947**	**0.936**

Table 5: Evaluation results of comparing three methods.

In addition, table 5 shows that three concept-level methods perform better than co-occurrence based mere word-level method (*Co-occurrence method*), and this result sufficiently proves that emotion is an abstract concept and detection of concept-level implicit emotion expressions is important for deeper analysis of user emotions.

Emotions	***Emotion causes***
Esteem	Klose, the Argentina team, Japanese fans, the Netherlands team
Sadness	the Belgium team, Messi, the Netherlands team
Sympathy	the Brazilian team, Chinese fans, C.Ronaldo
Happiness	Ha Dae-Sung, Thomas Muller, victory, goals, Squid Liu
Anger	Brazil fail, the Dutch keeper, Cameroon president, soccer gambling
Fear	next World Cup, disappointment, penalty kicks
Inferiority	Chinese soccer, picking up rubbish, Lost to Mali
Surprise	Neuer, Golden Ball Award, most remarkable defeat

Table 6. Emotion Cause Detection Result Examples about Hot Event '2014 FIFA World Cup in Brazil' Occurred from June 12, 2014 to July 13, 2014.

Table 6 gives some emotion cause detection result examples about hot event '2014 FIFA World Cup in Brazil'. We can see the most possible reasons for users' esteem emotion include 'Klose', 'the Argentina team', 'Japanese fans', and 'the Netherlands team'. Klose is a world famous soccer player, who has participated in four World Cups on behalf of the

German national team, and become the top goalscorer of the FIFA World Cup this time. Japanese fans showed a very good manner, and even though the Japanese team lost the game against the Ivory Coast team, the fans still consciously took away the trash on the field. Argentina team and the Netherlands team are both strong teams who got the second and the third place respectively.

In the Emoticon-Emotion Relation Detection part, in order to choose the most effective p in formula (6), we design an 'Unconventional Overall F Value' to evaluate the effect of different values of p. We consider that the number of emoticons detected by our model is N_d, and the number of correct ones in those detected emoticons is N_c. We can get the N_d and N_c with our experimental results and manual labeling results. Besides, we assume that the number of all the correct emoticon results is N_a, which is actually not available. Without the value of N_a, we cannot get the traditional F value, so we design an 'Unconventional Overall F Value' (uF_{value}) as given below:

$$uF_{value} = Precision * Recall = \frac{N_c}{N_d} * \frac{N_c}{N_a} \tag{12}$$

Furthermore, since N_a is fixed, we further modify uF_{value} as in formula (12), of which the value is not any more a real number between 0 and 1 as the traditional F value definition.

$$uF_{value} = \frac{(N_c)^2}{N_d} \tag{13}$$

We use the modified uF_{value} as evaluation metric to choose the most effective p, and table 7 shows the evaluation results when varying p as ten different values from 0.1 to 1. As shown, the curve peaks at p=0.6. Therefore, we set p=0.6 in the following experiments on emoticon-emotion relation detection.

p	0.1	0.2	0.3	0.4	0.5	**0.6**	0.7	0.8	0.9	1
uF_{value}	307.2	309.2	311.5	330.0	336.1	**337.5**	321.3	311.4	311.1	305.1

Table 7: *uFvalue* results of different values of *p* on 'Emotion-Emoticon' relation detection

Table 8 shows some 'emoticon-emotion' relationship examples about hot event 'Esteem Missing Malaysia Plane MH370'. The most special results in table 8 is about the happy emotion, since there are only 3 emoticons with a higher probability value than the *threshold*, and these 3 emoticons are all the most commonly used happy emoticons, [ha-ha], [lovely] and [smile]. We give some examples to explain the happy emotion. The first one is "56 rescue workers come back home from the Royal Air Force Station in Perth after the cessation of interim search and rescue work, which is a combination of sadness and happiness. Sadness is for the unresolved mystery of missing Malaysia Airlines, and the happiness is for the family reunion of rescue workers [ha-ha], while one of them has an eight-month-pregnant wife". Another example is "Today, I helped a lawyer friend translate a joint letter from MH370 families to Ministry of Transport, Malaysia. This is a meaningful thing, and I am very happy that I can help them. I hope they can have a good year! [smile]".

Emotions	***Emoticons***
Esteem	[farking], [fresh flower], [airplane], [adore], [heart]
Sadness	[tears], [scold], [heart], [anger], [shocked]
Sympathy	[candle], [mourn], [low-spirited], [Chinese flag]
Happiness	[ha-ha], [lovely], [smile]
Anger	[anger], [disdain], [pick-nose], [humph], [throw up]
Fear	[crazy], [broken-hearted], [ribbon], [sick]
Inferiority	[damped], [cry], [surround to watch], [grimace]
Surprise	[surprised], [what], [stunned], [tragedy]

Table 8: 'Emotion-Emoticon' Relationship Examples about Hot Event 'Esteem Missing Malaysia Plane MH370' which Occurred in March 08, 2014.

From table 8, we can see that some emoticons have no ambiguity, such as the [scold] with the sadness emotion, [candle] with the sympathy emotion, and the [what] with the surprise emotion. However, some other emoticons express different emotions in different hot events, such as the [sick] has a large probability to express the sympathy emotion in the hot event '2011 Japanese Earthquake', while it may be more likely to express feat emotion in the hot event 'Esteem Missing Malaysia Plane MH370'. Based on some manual checks, those results are all appropriate for those events, since different events have different origination and development, which make users show emotion expression diversity. Therefore, above results prove that some emoticon-based emotion classification methods may be unreasonable to a certain extent.

5. Conclusion and Future Work

Microblogging is mainly used to seek new trends and express emotions about hot events. By analyzing microblogging users' diversified emotions on hot events, we have opportunities to gain insights into their different perspectives and development tendency. In this paper, implicit emotional topics and their causes are discovered with the proposed concept-level emotion cause detection model, and event-oriented relations between emoticons and users' emotions can also be detected. Experimental results show that our model performs better than existing mere word-level models. Meanwhile, domain-oriented or community-oriented emotion cause detection tasks can also be realized with the proposed model.

In our future work, for proving the stability of our proposed concept-level emotion cause detection model on different languages, we will compare the performance of our model and other models on English and other languages. Furthermore, we will design a new concept-level emotion analysis model for extracting emotional topics from manually labeled emotion corpus, with which we think the detection results of the emotion cause detection model can become more accurate. Besides, key emotion detection and emotion cause tracking are also part of our future work.

6. References

Andrzejewski, D., Zhu, X., Craven, M. 2009. Incorporating domain knowledge into topic modeling via Dirichlet Forest priors. In Proceedings of the ICML 2009, pp. 25-32.

Aoki, S., Uchida, O. 2011. A method for automatically generating the emotional vectors of emoticons using weblog articles. In Proceedings of the 10th WSEAS International Confer-ence on Applied Computer and Applied Computational Science, pp. 132–136.

Zhao, J., Dong, L., Wu, J., Xu, K. 2012. MoodLens: an emoticon-based sentiment analysis system for Chinese tweets. In Proceedings of the 18th ACM SIGKDD, pp. 1528-1531.

Blei, D. M. 2012. Probabilistic topic models. Communications of the ACM 55(4): 77-84.

Blei, D. M., Ng, A. Y., Jordan, M. I. 2003. Latent Dirichlet Allocation. JMLR, 3:993-1022.

Brin, S., Page, L. 1998. The anatomy of a large-scale hypertextual Web search engine. In Proceedings of the 7th WWW, pp. 107-117.

Cambria, E. 2013. An introduction to concept-level sentiment analysis. In Proceedings of the 12th Mexican International Conference on Artificial Intelligence, pp. 478–483.

Chen, Y., Lee, S., Li, S., Huang, C. -R. 2010. Emotion Cause Detection with Linguistic Constructions. In Proceedings of the 23rd COLING, pp. 179-187.

Hsiao, K. -A., Hsieh, P. -L. 2014. Age Difference in Recognition of Emoticons. In Pro-ceedings of the 16th HCI.

Hu, Y., Boyd-Graber, J., Satinoff, B. 2011. Interactive Topic Modeling. In Proceedings of the 2011 ACL, pp. 248-257.

Zhao, W., Jiang, J., He, J., Song, Y., Achananuparp, P., Lim, E. -P., Li, X. 2011. Topical keyphrase extraction from twitter. In Proceedings of the 49th ACL, pp. 379–388.

Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N. 2013. Ontology-based sen-timent analysis of twitter posts. Expert Systems with Applications 40(10): 4065-4074.

Lee, S., Chen, Y., Huang, C. -R. 2009. Cause event representations for happiness and sur-prise. In Proceedings of the 23rd PACLIC, pp. 297-306.

Lee, S., Chen, Y., Huang, C. -R. A text-driven rule-based system for emotion cause detec-tion. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, pp. 45–53.

Lee, S., Chen, Y., Huang, C. -R., Li, S. -S. 2013. Detecting emotion causes with a linguis-tic rule-based approach. Computational Intelligence 29(3): 390-416.

Li, D., Shuai, X., Sun, G., Tang, J., Ding, Y., Luo, Z. 2012. Mining topic-level opinion in-fluence in microblog. In Proceedings of the 21st CIKM, pp. 1562–1566.

Li, W., Xu, H. 2014. Text-based emotion classification using emotion cause extraction. Expert Systems with Applications 41(4): 1742-1749.

Li, X., Liu, B., Yu, P. 2008. Time sensitive ranking with application to publication search, In Proceedings of the 8th ICDM, pp. 893–898.

Lu, P., Chang, Y., Hsieh, S. 2013. Causing Emotion in Collocation: An Exploratory Data Analysis. In Proceedings of the Twenty-Fifth ROCLING, pp. 236–249.

Masum, S. M. Al, Prendinger, H., Ishizuka, M. 2007. Emotion Sensitive News Agent: An Approach towards User Centric Emotion Sensing from the News. In Proceedings of the 2007 WI, pp. 614-620.

Sadamitsu, K., Saito, K., Imamura, K., Matsuo, Y. 2012. Constructing a Class-Based Lex-ical Dictionary using Interactive Topic Models. In Proceedings of the Eighth LREC, pp. 2590-2595.

Song, S., Li, Q., Zheng, N. 2010. A spatio-temporal framework for related topic search in micro-blogging, In Proceedings of the 2010 AMT, pp. 63–73.

Song, S., Li, Q., Zheng, X. 2012. Detecting Popular Topics in Microblogging Based on a User Interest-Based Model, In Proceedings of the 2012 IJCNN, pp. 1-8.

Song, S., Meng, Y. 2015. Detecting Concept-level Emotion Cause in Micro-blogging, In Proceedings of the 24st WWW, pp. 119-120.

Song, S., Meng, Y., Sun, J. 2014. Detecting keyphrases in microblogging with graph mod-eling of information diffusion, In Proceedings of the 13th PRICAI, pp. 26-38.

Zheng, N., Song, S., Bao, H. 2015. A Temporal-topic Model for Friend Recommendations in Chinese Microblogging Systems, IEEE Transactions on SMC: Systems. pp. 529–551.

Tokuhisa, R., Inui, K., Matsumoto, Y. 2008. Emotion classification using massive exam-ples extracted from the web. In Proceedings of the 22nd COLING, pp. 881–888.

Yan, X., Guo, J., Lan, Y., Cheng, X. 2013. A biterm topic model for short texts. In Pro-ceedings of the 22nd WWW, pp. 1445–1456.

Yang, C., Lin, K. H. -Y., Chen, H. -H. 2007. Emotion Classification Using Web Blog Cor-pora. In Proceedings of the 2007 WI, pp.275-278.

Zhang, H., Yu, H., Xiong, D., Liu, Q. 2003. HHMM-based Chinese lexical analyzer ICTCLAS. In Proceedings of the Second SIGHAN Workshop on Chinese Language Pro-cessing, pp. 184-187.