

# Tracking Emotions in Hot Topics: Exploiting Event-Specific Emotional Words

Yun Niu and Shihong Wang

**Abstract**—A word often conveys different emotions depending on its context. For example, in social media, emotional words in comments made on hot spot social events often denote emotions specific to the topic, which may be different from that when examined out of context. Nonetheless, contextual emotions of words has not been given much attention in current work of emotion analysis. This paper proposes an approach to identify event-specific emotional words and predict their emotion labels. These words are then used to track emotions in microblog posts using both rule-based and supervised approaches. Experimental results show that exploiting the detected words substantially improved the performance of emotion analysis.

**Index Terms**—Contextual emotion analysis, co-occurrence pattern, emotion lexicon, social media.

## I. INTRODUCTION

Microblog service such as Twitter is a crucial part of the rapidly developing social networks. In China, a Twitter-like service Weibo (e.g., www.weibo.com) has accumulated more than 330 million users and more than 1000 messages are posted in Weibo every second. These microblog websites allow users to make comments on hot spot social events freely and instantly. These comments often express personal emotions towards the hot topics. Tracking emotions in this text is of vital importance to governments and institutions. For example, it may be helpful in determining popularity of government policies and performing affect-based search.

Words play an important role in emotion analysis as they could convey emotions, e.g., *wonderful* and *satisfied* usually indicate the emotion of joy, *heart-broken* and *grief* often denote sadness. The emotion denoted by a word may be different depending on the context of the word. For example, *cry* could indicate a feeling of *sorrow*, but under some situations it may suggest great happiness. In social media, hot spot social events provide context in which words often convey emotions specific to the event, as shown in the following case.

*Event:* 太明食品异物 (Hairs in TAIMING food)

• *Microblog post:*

摆明了就是**毛发**! 太恶心了

(Obviously, it is a **hair**! Disgusting)

Manuscript received March 5, 2015; revised May 7, 2015. This work was supported in part by the National Natural Science Foundation of China under Grant 61202132.

The authors are with Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: yniu@nuaa.edu.cn, wsh\_014@nuaa.edu.cn).

• *Microblog post:*

大学时我同学也吃到过头发, 现在想想还恶呢

(My classmate in college had similar experience ... Found a **hair** in his food ... Still feel disgusting)

Word 毛发 (**hair**) is neutral when it is out of any context. Nonetheless, when people discuss hairs in food, as in this topic, word **hair** evokes negative emotion **disgust**. Therefore, it is crucial to identify contextual emotions of words for accurate emotion analysis. Nonetheless, this issue has not been given much attention in most emotion analysis systems. Furthermore, in comments on social events, new words and usages are invented constantly to express the strong feelings of the bloggers. For example, “黑肘” (attack someone using the elbow) and “科蜜” (fans of Kobe Bryant) appear in the topic “Kobe Bryant”. These words are quickly spread by the followers and could be important cues for identifying the emotion of a comment. However, they are not listed in existing emotion lexicons.

In this paper, we focus on identifying event-specific emotional words that denote fine-grained emotions. Ekman's six emotions are adopted as the labels: joy, sadness, anger, fear, surprise, disgust [1]. An approach is proposed to identify emotional words and predict their context-dependent emotion labels. our observation is that emotion denoting words of the same orientation tend to co-occur at microblog post level. Therefore, we examine how a word co-occurs with some seed emotional words in the corpus. The emotion pattern of a word is built to describe how it co-occurs with seed words from existing lexicons. The contextual emotion denoted by a candidate word is then identified by comparing its emotion pattern with the seed words. To the best of our knowledge, the work proposed in this paper is the first on detecting fine-grained emotional words in the context of social events. The detected words are then used to track emotions in hot topics in social media. Experimental results show that the detected words with their emotions substantially promote the performance of emotion analysis of microblog posts on hot topics.

## II. RELATED WORK

### A. Evaluating Polarity of Words

A lot of efforts have been made on collecting subjective words and determining their polarity, i.e., whether a word has positive or negative orientation ([2]-[5]). Most of them focus on domain-independent words. Some approaches rely on rich lexical resources. Esuli and Sabastiani [4] build SentiWordNet based on the glosses associated with WordNet

[6]. Andreevskaja and Bergler [7] extract subjective adjectives from WordNet. Mohammad *et al.* [8] derive a lexicon using a thesaurus. Some approaches explore word similarity derived from large corpora. Wiebe [9] clusters subjective clues according to their distributional similarity. While Turney [3] collects subjective words by calculating their similarity with a few seed words measured by pointwise mutual information. Velikovich *et al.* [5] explore a graph propagation framework to collect positive and negative terms from the web.

There are a few attempts exploiting domain-specific subjective terms in determining polarity of documents. Zagibalov and Carroll explore a self-training approach that identifies sentiment-bearing terms in the target data and use them to determine whether a product review is positive or negative. Qiu *et al.* [10] establish a two-phase sentiment classification system where subjective terms in the data set are extracted by self-training in the first phase. Volkova *et al.* [11] construct a lexicon by bootstrapping subjectivity clues in tweets and evaluate the approach on multiple languages. The work of [12] addresses a problem of using issue-related words in sentiment analysis of social media. The method is not fully automatic, and manual annotation of polarity of terms in the training data is mandatory. In these approaches the derived lexicons only consist of positive and negative words.

#### B. Building Fine-Grained Emotion Lexicons

The task of building fine-grained emotion lexicon is more complex. While much work has been done on collecting positive/negative words, only limited efforts have been devoted in this area. The work of [13] looks for emotion denoting words in a multi-agent system; the work of [14] determines tendency of terms using pointwise mutual information; and the work of [15] builds an emotion lexicon by crowdsourcing using Mechanical Turk. Xu *et al.* [16] adopt a graph-based algorithm to rank Chinese emotional words from multiple resources. Liu *et al.* [17] derive a Chinese lexicon of eight emotions from a manually annotated corpus consisting of 1487 blog posts. All of this work examines words out of context.

Our approach presented in the paper detects fine-grained emotional words and identifies their contextual emotions in hot spot social events. It handles the informality and dynamic nature of social media. Moreover, it does not rely on manually annotated data and does not require the support of extra-large corpora.

### III. DETECTING EVENT-SPECIFIC EMOTIONAL WORDS

In order to identifying the contextual emotion of words, we make use of a set of seed emotional words and investigate their behaviours in the corpus. The seed words are from two popular Chinese emotion lexicons. We observe that emotion-denoting words of the same orientation tend to

co-occur at microblog post level. Therefore, examining how a word co-occurs with the seeds in the corpus may provide some clues on identifying the emotion it evokes.

#### A. Emotion Lexicons

(1) CLIWC (Chinese Linguistic Inquiry and Word Count) [18]. It was derived from the English version LIWC (Linguistic Inquiry and Word Count, [19]), a software that analyzes words in text. The heart of the software is a lexicon of approximately 4500 emotion and cognitive words. LIWC has been used in emotion analysis of English text ([20], [21]). Its Chinese version CLIWC includes 42 word categories tapping psychological constructs (e.g. positive emotion, negative emotion), of which 5 categories describe emotions: *positive emotion*, *negative emotion*, *anxiousness*, *anger* and *sadness*. As only two of them overlap with Ekman's six emotions, we did some preprocessing before applying it in the experiments. Generally, words in *positive emotion* indicate "good" and "happiness". Therefore we label them as *joy* words. For *negative emotion* and *anxiousness*, we examine words in them and assign corresponding labels to those that convey *fear*, *surprise*, and *disgust*. The derived lexicon is referred to as EmoCLIWC hereafter.

(2) Emotional words ontology [22]. It is used as the lexicon in the Chinese microblog post analysis workshop in the second annual meeting of Natural Language Processing & Chinese Computing (NLP&CC). It contains seven emotions, including five of Ekman's emotion categories. There is no *joy* in the ontology. Instead, it has "高兴" (*happiness*) and "喜好" (*like*). In our experiments, we merge words in the two classes as the *joy* words. This lexicon is referred to as EmoDTUIR.

Overlap of the two lexicons is very small. Therefore, instead of choosing one as the source of seeds, we combine EmoCLIWC with EmoDTUIR to get a new lexicon LEX. When we merge them, we remove words that present in both lexicons yet having different emotion labels in the two lexicons. The distribution of emotions in the three lexicons is shown in Table I. We can see that the distribution is very uneven. EmoCLIWC contains much less words for the last three emotions while EmoDTUIR has a much smaller collection for *anger* and *surprise* compared to the other classes. In all the three lexicons, the *joy* class has the largest population and the amount of words in *joy* is much larger than that in the other five categories.

#### B. Identifying Event-Specific Emotional Words

We propose an approach to identify event-specific emotion denoting words and predict their emotion labels. Given a target corpus of microblog posts, we begin with a set of seed words, which contains all the words in the lexicon LEX that appear in the corpus. In addition, we built up a list of emoticons and added them to the seed set. The rest of words in the corpus form the candidate set, from which event-specific emotional words are identified.

TABLE I: DISTRIBUTION OF EMOTIONS IN THE LEXICONS

	Joy	Sadness	Anger	Fear	Disgust	Surprise
EmoCLIWC	476	128	249	33	28	14
EmoDTUIR	13007	2289	386	1166	10217	226
LEX	13166	2319	472	1171	10041	224

We found that emotional words tend to co-occur with other emotional words of the same orientation at microblog post level. Therefore, the co-occurrence pattern of a target word is identified. To represent the co-occurrence patterns of words, a vector of six dimensions is built for each word in the corpus. Each dimension of the vector represents one of Ekman's six emotions. The weight of a dimension indicates how often a target word co-occurs with the seed words of that emotion.

Input:

1. Emotional words in CLIWIC-DLLG: LEX
2. Corpus:  $C = \bigcup_{1 \leq i \leq n} c_i$
3.  $E = \{joy, sadness, anger, fear, surprise, disgust\}$

Output:

Emotion labels of candidate words:  $l_s$

Steps:

1. Seeds =  $C \cap LEX$  ;
2. // Compute emotion vectors  $\{w_{i,e}\}_{e \in E}$  for each microblog post  $c_i$

**for all**  $i \in [1..n]$  **do**

$\{w_{i,e}\}_{e \in E} = 0$  //initialization

**for all**  $s \in Seeds$  **do**

**for all** ( $e \in \{\text{labels of } s\}$ ) **do**  $w_{i,e} += freq(s, c_i)$

3. //Compute emotion vectors  $\{r_{s,e}\}_{e \in E}$  for each word  $s \in C$

**for all**  $s \in C$  **do**

**for all**  $e \in E$  **do**

$I = \{i \mid 1 \leq i \leq n \wedge s \in c_i\}$

// find the index of microblog posts containing  $s$

$SW = \{t \in Seeds \mid e \in \{\text{labels of } t\} \wedge \exists i \in [1..n]. t$

co-occur with  $s$  in  $c_i\}$

$$r_{s,e} = \frac{\sum_{i \in I} w_{i,e}}{\log_{10} \sum_{t \in SW} freq(t, C)}$$

4. //Compute similarity vectors  $\{m_{s,e}\}_{e \in E}$

//for each candidate word  $s \in C \setminus LEX$

**for all**  $s \in C \setminus LEX$  **do**

**for all**  $e \in E$  **do**

$SW = \{t \in Seeds \mid e \in \{\text{labels of } t\} \wedge \exists i \in [1..n]. t$

co-occur with  $s$  in  $c_i\}$

$m_{s,e} = AVG_{t \in SW} \{sim(r_{s,e}, r_{t,e})\}$

5. //Compute the label  $l_s$  for a candidate word

**for all**  $s \in C \setminus LEX$  **do**

$ce = \text{argmax}_e \{m_{s,e}\}; l_s = ce$

$SW = \{t \in Seeds \mid ce \in \{\text{labels of } t\}\}$

$$\theta = \frac{\sum_{t \in SW} freq(t, C)}{\sum_{t \in Seeds} freq(t, C)}$$

**for all**  $e \in E \setminus \{ce\}$  **do**

**if**  $(m_{s,ce} - m_{s,e}) \leq \theta \times (m_{s,ce} + m_{s,e})$  **then**

$l_s = \text{null}$ ; **break**

End

Fig. 1. Algorithm predicting emotions for topic-specific words.

number of seeds in each emotion category, a simple frequency count could be misleading. As described in Section III-A, there is a big difference in the population of the emotion categories in a lexicon. Take LEX as an example, the number of *surprise* words is only about 1.7% of the number of *joy* words. This means that there would be many more occurrences of *joy* seeds in the corpus than *surprise* seeds. Hence, a word may co-occur more with *joy* just because there are more *joy* words in the lexicon. Another factor that may introduce great noise to the co-occurrence patterns is the so called ‘‘pollyanna phenomena’’ [23]. Pollyanna phenomena refers to the word-level linguistic positivity bias, i.e., people tend to use positive words (*joy* words in our case) more frequently than negative words. This phenomena is also observed in communication in social media such as microblog posts. Positive words are often seen in posts that actually express negative emotions. These occurrences of positive words bring noise to the co-occurrence count. In our algorithm of identifying event-specific emotional words, these issues are addressed by special techniques, as described in Fig. 1.

As shown in the figure, an emotion vector of six dimensions is derived for each word in the target corpus (whether it is a seed or a candidate) (step 2 to step 3 in the algorithm). This vector represents the co-occurrence pattern of the word. Weight of dimension  $i$  in the vector is determined by two components. The numerator is the total amount of times that the word co-occurs with all the seeds of emotion  $i$ . The denominator is to reduce the bias caused by the unbalanced seed set or the pollyanna phenomena. We add up the frequency of all the seeds expressing emotion  $i$  that co-occur with the word and use it as the denominator. A higher weight indicates that the target word is more likely to convey the emotion represented by that dimension.

We expect that for emotion denoting words of the same orientation, their patterns of co-occurrence in a corpus tend to be similar. Therefore, after the emotion vector is computed, instead of taking the emotion that has the highest weight as the label of a candidate word, we calculate the similarity between the candidate and the seeds in each emotion category (step 4). Here, taking all the seeds into consideration would help reduce errors caused by coincidence. It can also help diminish the bias introduced by the unbalanced seed set. For example, suppose in the emotion vector of a *sadness* word, weight of the *joy* dimension is the highest among the six dimensions. This word would still get the correct label if its co-occurrence pattern is most similar to those of the *sadness* seeds. Cosine similarity is used as the measure. For each emotion  $e$ , the average similarity ( $m_{s,e}$ ) between the candidate and its co-occurring seeds expressing  $e$  is calculated. Then, the maximum similarity  $m_{s,ce}$  and the corresponding emotion  $ce$  are identified.

Since a candidate that co-occurs evenly with multiple emotions is likely to be neutral and should not be identified as emotion denoting word, a strategy that determines which emotion is dominant is needed. In order to evaluate whether there is a dominating emotion that co-occurs with the candidate, we calculate the difference between the maximum similarity and the other five similarity values (step 5). If the difference is larger than a threshold then emotion  $ce$  is dominant. The candidate is added to the event-specific

Nonetheless, since there is a great variance w.r.t the

emotional word list, and  $ce$  is used as its label. The threshold is calculated by  $\theta \times (m_{s,ce} + m_{s,e})$ , where  $\theta$  is a parameter. Again  $\theta$  is designed to diminish the bias caused by emotions having large seed sets or high frequency seeds. It is defined as the sum of frequency of seeds in  $ce$  divided by the sum of frequency of all the seeds in the corpus. Therefore,  $\theta$  is bigger for emotions having larger sets of seeds and/or the frequency of the seeds is high.

#### IV. DATA SETS

Two data sets are used to evaluate the event-specific emotional words identified by the proposed approach. One data set contains microblog posts on ten hot topics downloaded using the Chinese microblog platform Sina API [24]. Each microblog post was labeled using Ekman's six emotions by two annotators separately. Only those that receive the same label from the annotators are kept, which

leads to a set of 1335 microblog posts. This set is denoted as WEIBO hereafter. The other data set (henceforth referred to as SMILEY) was also downloaded from Sina website. It was generated based on the existence of particular emoticons, which indicate the emotion of the text: *joy* for microblog posts that contain “[嘻嘻]” (chuckle) etc, and *anger* for microblog posts that contain “[怒]” (anger) etc. A list of emoticons that typically indicate the six emotions were collected. To build the data set, we first downloaded microblog posts on some hot topics. From these posts, we extracted those that contain emoticons of one of the six emotions and no other emoticons. Then these microblog posts are labeled using distant supervision [25], i.e., a post is labeled by the emoticons it contains. Since in our downloaded microblog posts there are no posts that only contain the *fear* emoticons, no posts are labeled as *fear* in this data set. The distribution of microblog posts over emotion labels in the two data sets is given in Table II.

TABLE II: EMOTION LABEL DISTRIBUTION IN THE DATA SETS

	Joy	Sadness	Anger	Fear	Disgust	Surprise	Total
WEIBO	473	173	276	147	145	121	1335
SMILEY	386	393	474	0	90	242	1585

#### V. TRACKING EMOTIONS IN MICROBLOG POSTS USING A LEXICON-BASED APPROACH

In this experiment, words in a microblog post are checked to find those appearing in an emotion lexicon. Then the frequencies of matched emotional words of the same orientation are added up. Finally, the emotion with the highest frequency is used to label the microblog post. If a microblog post does not match any words in the lexicon, or there are more than one emotion that has the highest frequency of matched words, then the microblog post is not covered by the lexicon. The coverage of a lexicon is defined as the proportion of microblog posts covered by the lexicon in the data set.

In the experiment, the lexicon LEX is first used to identify

the emotion expressed by a microblog post. Then the event-specific emotional words (ESEW) identified using Algorithm I are added to the lexicon and the expanded lexicon is applied to detect the emotion. Two options are used when extracting the candidate words from the corpus. One option took all words except the seeds in the data set as the candidates. The other applied a Part-of-Speech tagger filter and only extracted adjectives, adverbs, nouns and verbs as the candidates. We use NLPiR Chinese POS tagger (version 2013) [26] in the experiments. This POS tagger is customized to deal with microblog posts. In addition, it has the functionality of new words recognition. The two options are compared in the experiments. The results of the lexicon-based approach on the data sets are shown in Tables III and IV. Macro-average is calculated in the tables.

TABLE III: RESULTS OF THE LEXICON-BASED APPROACH ON WEIBO

	Precision	Recall	F-score	Coverage	Accuracy
<b>a.LEX,ESEW</b>				76.8	43.6
Joy	64.8	60.7	62.7		
Sadness	53.9	35.8	43.1		
Anger	56.5	22.1	31.8		
Fear	59.1	26.5	36.6		
Disgust	41.4	71.7	52.5		
Surprise	69.0	24.0	35.6		
Average	57.5	<b>40.1</b>	<b>47.2</b>		
<b>b.LEX,ESEW(no POS filter)</b>				<b>80.1</b>	<b>44.4</b>
Average	56.2	39.9	46.7		
<b>c.LEX</b>				60.8	36.3
Average	<b>65.6</b>	31.8	42.8		

On WEIBO data (Table III), using LEX alone achieves accuracy of 36.3% with coverage of 60.8%. After adding the event-specific emotional words, no matter whether the POS filter is applied or not, both coverage and accuracy are substantially improved. The highest coverage and accuracy

are achieved by ESEW without the POS filter. Compared to LEX, coverage is improved by about 19%, and accuracy is improved by about 8%. With the POS filter, the coverage drops by 3% compared to the option without it, and the accuracy is also a little bit lower. The lexicon LEX gets a high

precision but the recall is very low. Incorporating ESEW greatly improved recall and achieves the best F-score. The improvement on SMILEY data is even larger (Table IV). Adding ESEW to the lexicon achieves higher precision and recall. The best F-score achieved after incorporating ESEW is about 13% higher than that of using LEX alone. In addition, both coverage and accuracy are substantially improved. The

results clearly show that event-specific emotional words are crucial in emotion analysis, and the proposed method is effective in identifying these words. The results also reveal that the seed emoticons are unlikely to co-occur with various types of emotions. Therefore, they provide strong evidence in predicting emotion labels of candidate words that co-occur with them.

TABLE IV: RESULTS OF THE LEXICON-BASED APPROACH ON SMILEY

	Precision	Recall	F-score	Coverage	Accuracy
<b>a.LEX,ESEW</b>				79.4	33.3
Joy	39.2	73.8	51.2		
Sadness	36.5	8.9	14.3		
Anger	79.3	31.4	45.0		
Disgust	8.3	16.7	11.1		
Surprise	64.7	18.2	28.4		
Average	45.6	29.8	36.0		
<b>b.LEX,ESEW(no POS filter)</b>				<b>84.1</b>	<b>35.1</b>
Average	<b>49.7</b>	<b>31.9</b>	<b>38.9</b>		
<b>c.LEX</b>				57.5	19.6
Average	43.9	18.1	25.6		

## VI. TRACKING EMOTIONS IN MICROBLOG POSTS USING A SUPERVISED APPROACH

In this section, we trained a support vector machine (SVM) classifier to label the emotion conveyed by a microblog post. To exploit the event specific emotional words (ESEW), we use six features representing the six emotions. Weights of the features are integer-valued, which capture the number of word tokens (term frequency) in a microblog post associated with the specific emotion labels. For example, if a microblog post has two *anger* words and one *disgust* word, then the *anger* feature has value 2, *disgust* has value 1, and all remaining emotion features have value 0. In addition, we explore another two groups of features. One contains all the words in the emotion lexicon LEX. Again, term frequency is used as the weights for this group of features. The other group consists of unigrams. For this group, binary features that captured the presence or absence of the features are applied. In the experiment, we use and take the six emotions as the target classes. In the experiments on data set SMILEY, emoticons used to label the microblog posts are removed from the data sets. The results obtained by 5-fold cross validation are shown in Tables V and VI.

TABLE V: RESULTS OF SVM ON WEIBO

	Precision	Recall	F-score	Accuracy
<b>a.LEX,ESEW</b>				<b>55.4</b>
Joy	55.7	78.9	65.3	
Sadness	49.4	43.5	46.3	
Anger	56.5	48.0	51.9	
Fear	61.2	31.7	41.8	
Disgust	52.7	55.2	53.9	
Surprise	77.8	25.8	38.8	
Average	<b>58.9</b>	<b>47.2</b>	<b>52.4</b>	
<b>b.LEX</b>				53.6
Average	54.7	44.8	49.3	
<b>c.unigrams</b>				42.6
Average	31.6	30.5	31.0	
<b>d.unigrams, LEX, ESEW</b>				44.7
Average	50.0	30.6	38.0	

As shown in Table V, a classifier using ESEW and LEX as features achieves the highest accuracy. It also improves both precision and recall than a classifier using LEX alone. More specifically, precision gets higher by about 4% and recall is improved by about 2.5%. F-score is about 3% higher. The overall accuracy is also improved. Compared to the unigram baseline, adding ESEW and LEX leads to an improvement of about 18% in precision and about 7% in F-score without degrading recall.

Table VI shows that on SMILEY, combining ESEW with LEX substantially improves the performance compared to using LEX alone as features. Precision, recall and F-score are improved by about 8%. Compared to the unigram baseline, incorporating ESEW and LEX improved precision and recall, and F-score is higher by about 2%.

On both data sets, using ESEW in addition to LEX gives significantly higher scores than using LEX alone (paired *t*-test,  $p < 0.05$ ). Of the four different combinations of features, The best F-score and accuracy are achieved by the feature set consists of LEX and ESEW.

TABLE VI: RESULTS OF SVM ON SMILEY

	Precision	Recall	F-score	Accuracy
<b>a.LEX,ESEW</b>				<b>46.8</b>
Joy	42.5	49.4	45.7	
Sadness	36.6	37.2	36.9	
Anger	56.2	73.2	63.6	
Disgust	5.7	2.2	3.2	
Surprise	51.6	23.3	32.1	
Average	<b>38.5</b>	<b>37.1</b>	<b>37.8</b>	
<b>b.LEX</b>				37.3
Average	30.9	28.5	29.7	
<b>c.unigrams</b>				36.6
Average	29.4	36.2	32.4	
<b>d.unigrams, LEX, ESEW</b>				35.4
Average	31.9	<b>37.2</b>	34.3	

## VII. CONCLUSION

In social media communication about hot spot social events,

subjective language presents event-specific characteristics. To track emotions in hot topics, an approach is proposed to identify emotion denoting words and predict their labels according to the context in the topics. In this approach, an emotion vector is built for a word to describe how it co-occurs with Ekman's six emotions in the target corpus. Based on the hypothesis that the co-occurrence patterns of emotional words of the same orientation are similar, the similarity between a candidate word and the seed words of each emotion is calculated. The dominating emotion is taken as the label of the candidate word. In the process of identifying the contextual emotion of a word, unbalanced distribution of emotions in lexicons and the pollyanna phenomena introduce great noise to similarity calculation. In the presented approach, this issue is addressed by techniques that synthesize the emotion patterns of all seeds of an emotion category. This approach does not rely on manually annotated data and does not require the support of extra large corpora.

The identified words are then used to analyze emotions expressed in microblog posts. In the experiments, the detected event-specific emotional words are evaluated by a rule-based and a supervised approach on two data sets. The results show that event-specific emotional words are crucial cues in emotion analysis and exploiting the identified words substantially improved the performance of hot topic emotion analysis.

#### REFERENCES

[1] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169-200, 1992.

[2] E. Riloff and J. Shepherd, "A corpus-based approach for building semantic lexicons," in *Proc. the Second Conference on Empirical Methods in Natural Language Processing*, 1997, vol. 117, p. 124.

[3] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *Proc. the 40th Annual Meeting on Association for Computational Linguistics*, 2002, pp. 417-424.

[4] A. Esuli and F. Sebastiani, "Sentiwordnet: A publicly available lexical resource for opinion mining," in *Proc. LREC*, 2006, vol. 6, pp. 417-422.

[5] L. Velikovich, S. Blair-Goldensohn, K. Hannan, and R. McDonald, "The viability of web-derived polarity lexicons," in *Proc. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 2010, pp. 777-785.

[6] G. A. Miller, "Wordnet: A lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39-41, 1995.

[7] A. Andreevskaja and S. Bergler, "Mining word net for a fuzzy sentiment: Sentiment tag extraction from word net glosses," in *Proc. EACL*, 2006, vol. 6, pp. 209-215.

[8] S. Mohammad, C. Dunne, and B. Dorr, "Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus," in *Proc. the 2009 Conference on Empirical Methods in Natural Language Processing*, 2009, vol. 2, pp. 599-608.

[9] J. Wiebe, "Learning subjective adjectives from corpora," in *Proc. AAAI/IAAI*, 2000, pp. 735-740.

[10] L. Qiu, W. Zhang, C. Hu, and K. Zhao, "SELC: A self-supervised model for sentiment classification," in *Proc. the 18th ACM Conference on Information and Knowledge Management*, 2009, pp. 929-936.

[11] S. Volkova, T. Wilson, and D. Yarowsky, "Exploring sentiment in social media: Bootstrapping subjectivity clues from multilingual twitter streams," *Association for Computational Linguistics*, 2013.

[12] M. Thelwall and K. Buckley, "Topic-based sentiment analysis for the social web: The role of mood and issue-related words," *Journal of the American Society for Information Science and Technology*, vol. 64, no. 8, pp. 1608-1617, 2013.

[13] C. D. Elliott, "The affective reasoner: A process model of emotions in a multi-agent system," Ph.D. dissertation, Northwestern University, 1992.

[14] J. Read, "Recognising affect in text using pointwise-mutual information," Unpublished M.Sc. dissertation, University of Sussex, UK, 2004.

[15] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in *Proc. the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 2010, pp. 26-34.

[16] G. Xu, X. Meng, and H. Wang, "Build Chinese emotion lexicons using a graph-based algorithm and multiple resources," in *Proc. the 23rd International Conference on Computational Linguistics*, 2010, pp. 1209-1217.

[17] N. Liu, F. Ren, X. Sun, and C. Quan, "Microblogging hot events emotion analysis based on ren-cecps," in *Proc. 2013 IEEE/SICE International Symposium on System Integration (SII)*, 2013, pp. 233-238.

[18] CLIWC. (2007). Chinese linguistic inquiry and word count. [Online]. Available: <http://c-liwc.blogbus.com/>

[19] J. W. Pennebaker, R. J. Booth, and M. E. Francis. (2007). Linguistic inquiry and word count. [Online]. Available: <http://www.liwc.net/>

[20] G. Paltoglou and M. Thelwall, "Twitter, myspace, DIGG: Unsupervised sentiment analysis in social media," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 3, no. 4, p. 66, 2012.

[21] S. A. Golder and M. W. Macy, "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, vol. 333, no. 6051, pp. 1878-1881, 2011.

[22] L. Xu, H. Lin, Y. Pan, H. Ren, and J. Chen, "Constructing the affective lexicon ontology," *Journal of the China Society for Scientific and Technical Information*, vol. 2, pp. 180-185, 2008.

[23] T.-H. Huang, H.-C. Yu, and H.-H. Chen, "Modeling pollyanna phenomena in chinese sentiment analysis," *COLING*, pp. 231-238, 2012.

[24] Sina. (2014). Sina Weibo. [Online]. Available: <http://weibo.com/>

[25] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford, pp. 1-12, 2009.

[26] H. Zhang. (2013). Nlpir Chinese word segmentation system. [Online]. Available: <http://ictclas.nlpir.org/>



**Yun Niu** received her Ph.D. from the Department of Computer Science at University of Toronto, Canada, in June 2007. Then she worked as a postdoc. at Ontario Cancer Institute in Toronto, Canada, from February 2007 to August 2009. She is an associate professor at Nanjing University of Aeronautics and Astronautics, China, since October 2009. Her research interests include sentiment analysis and text mining in biomedical text.



**Shihong Wang** is a master's student in the College of Computer Science and Technology at Nanjing University of Aeronautics and Astronautics, China. Her research area is sentiment analysis in social media.