World Wide Web (2015) 18:949–967
DOI 10.1007/s11280-014-0289-x

A word-emoticon mutual reinforcement ranking model for building sentiment lexicon from massive collection of microblogs

Shi Feng · Kaisong Song · Daling Wang · Ge Yu

Received: 26 August 2013 / Revised: 21 December 2013 /
Accepted: 6 March 2014 / Published online: 27 April 2014

Abstract Recently, more and more researchers have focused on the problem of analyzing people's sentiments and opinions in social media. The sentiment lexicon plays a crucial role in most sentiment analysis applications. However, the existing thesaurus based lexicon building methods suffer from the coverage problems when faced with the new words and new meanings in social media. On the other hand, the previous learning based methods usually need intensive expert efforts for annotating training datasets or designing extraction patterns. In this paper, we observe that the graphical emoticons are good natural sentiment labels for the corresponding microblog posts and a word-emoticon mutual reinforcement ranking model is proposed to learn the sentiment lexicon from the massive collection of microblog data. We integrate the emoticons and candidate sentiment words in the microblogs to construct a two-layer graph, on which a random walk is run for extracting the top ranked words as a sentiment lexicon. Extensive experiments were conducted on a benchmark dataset with various topics. The results validate the effectiveness of the proposed methods in building sentiment lexicon from microblog data.

Keywords Sentiment analysis · Opinion mining · Lexicon building · Microblog mining

S. Feng (✉) · K. Song · D. Wang · G. Yu
Institute of Computer Software and Theory, Northeastern University, No.3-11 Wenhua Road, Heping District, Shenyang, China
e-mail: fengshi@ise.neu.edu.cn

K. Song
e-mail: songkaisongabc@126.com

D. Wang
e-mail: wangdaling@ise.neu.edu.cn

G. Yu
e-mail: yuge@ise.neu.edu.cn

1 Introduction

As the rise of Web 2.0 technologies, more and more people are willing to publish their attitudes and feelings in Web 2.0 based social media rather than just passively browse and accept information. Because of the rich user-generated information in social media, how to provide an efficient way to analyze users' sentiments has received significant attention from both academic researchers and commercial companies [29]. The sentiment analysis in social media includes subjectivity and polarity classification, sentiment holder and target detection, and so on. The accuracies of these sentiment analysis tasks often rely on a sentiment lexicon with positive and negative words. Therefore, the sentiment lexicon plays a critical role in many sentiment analysis tasks [11].

Currently, a number of papers have been published for manual or automatic sentiment lexicon building, but the challenges still remain for Web 2.0 based social media. One direction of the existing methods is automatic learning lexicon from thesaurus or knowledge base. However, plenty of new words are emerging in the online social media every day. Typos, ad hoc abbreviations and phonetic substitutions are common phenomena in the user-generated content. Some widely used typos and phonetic substitutions evolve into new sentiment-bearing words. Even the same word may have different explanations or sentiment orientations at different time periods. Therefore, the traditional thesaurus and knowledge base, such as WordNet, usually suffer from the coverage problem when faced with social media. Another research direction makes use of significant manual annotation or large corpora, which needs intensive expert efforts. In general, the existing thesaurus based methods usually suffer from the new words and new meaning problem. And also the corpus based methods usually need manually defined extraction patterns or time-consuming labeling task.

Everyday, enormous numbers of text posts that contain people's rich sentiments are published in the microblogging websites such as Twitter[1] and Weibo[2]. The microblog is a good source for extracting sentiment lexicon. Firstly, as a convenient way to record daily personal feelings and emotions, the microblog data contains rich sentiment information. Secondly, the users usually like to employ free writing style and talk about up-to-date hot topics, so the emerging words and new meanings are common in the microblog dataset. With the microblog API, it is much easier to collect millions of posts for training. At last but not the least, many microblogs contain graphical emoticons, which can be considered as the natural sentiment labels for the corresponding posts in the microblog dataset.

There are already some studies on the emoticons in microblog data. Pak et al. collected tweets with happy and sad emoticons as training dataset, and built sentiment classifier based on traditional machine learning methods [30]. Davidov et al. chose 50 tags and 15 smileys as sentiment labels to classify twitter data [8]. These existing methods have verified the effectiveness of the microblog emoticons in the sentiment analysis task. However, they only focus on finding the appropriate features such as unigrams, bigrams, trigrams and POS structures for sentiment classification. The basic sentiment lexicon building procedure is neglected, which may have many potential applications, such as opinion retrieval and opinion summarization.

In this paper, we utilize the microblog as training corpus and attempt to learn sentiment lexicon from massive collection of microblogs with the help of emoticons. We observe that the positive words often appear in the microblog posts with positive emoticons, and the

[1] http://twitter.com

[2] http://weibo.com

negative words often appear in the microblog posts with negative emoticons. Different from the traditional methods, we regard the sentiment lexicon building as a ranking problem and a mutual reinforcement ranking model is proposed to simultaneously rank the candidate sentiment words and emoticons in the microblog collection. Our approach can create a sentiment lexicon free of laborious efforts of the experts for manually annotating training data or designing word extraction patterns.

The rest of the paper is organized as follows. Section 2 introduces the related work on sentiment lexicon building and sentiment analysis in microblog. Section 3 analyzes the characteristics of the emoticons in Chinese microblogs. In Section 4 we present the purification method for massive microblog dataset. In Section 5 we propose the word-emoticon mutual reinforcement ranking model for sentiment lexicon building. In Section 6 we provide the experimental results on a real world microblog dataset with various topics. Finally we present the concluding remarks and future work in Section 7.

2 Related work

There are two types of previous literatures relevant to our work. One is about sentiment lexicon building, and the other is about sentiment analysis in microblogs.

2.1 Sentiment lexicon building

As we have discussed in Section 1, there are mainly two directions for building sentiment lexicon. The first direction is automatic learning sentiment lexicon from different kinds of thesaurus. Hu and Liu resorted to the synonym and antonym relationship in WordNet to predict the orientation of the candidate words [16]. Kim and Hovy proposed two probabilistic models to estimate the strength of polarity [22]. In their models, synonyms were used as features. Their basic hypothesis was that the synonyms had the same orientation and the antonyms had the opposite orientation. With the help of selected seed words, they could determine the candidate word's orientation based on the relationship of synonyms and antonyms. Baccianella et al. utilized the glosses in WordNet to represent the candidate words and the words were classified into positive and negative categories based on the new representations [1]. Esuli and Sebastiani proposed a random walk based algorithm to rank the word polarities in WordNet [10]. They assumed that the occurrence of the words in the glosses might be viewed as a transmitter of polarity properties. Although the thesaurus based methods can lead to higher accuracy, they also suffer from the new words and new meanings problems, which are common phenomena in Web 2.0 based social media.

Another research direction is building sentiment lexicon based on manually annotated training data or large corpus. Turney determined the polarity value based on the candidate word's co-occurrence with the seed words (*excellent* and *poor*) [35]. The co-occurrence was measured by the number of hits returned by a search engine, i.e. the whole Web was considered as the corpus to determine the word orientation. Kanayama and Nasukawaused found that only 60 % co-occurrences in the same window in Web pages reflected the same sentiment orientation. So they further used both intra- and inter-sentential co-occurrence to learn the orientation of words and phrases [21] . Kaji and Kitsuregawa utilized structural clues to extract sentiment words from large collection of HTML documents [20]. Usually the polarity of words is sensitive to the topic domain. Lu et al. [27] and Choi et al. [6] treated the domain-specific sentiment lexicon building as an optimization problem. They

designed appropriate objective functions and employed the linear programming methods to learn lexicons from manually labeled datasets. Qiu et al. [31] proposed several sentiment word extraction patterns and rules to learn the domain-specific lexicon. Jijkoun et al. [18] introduced a bootstrapping method for generating a topic-specific lexicon from a general purpose polarity lexicon. Jin and Ho [19] and Li et al. [25] proposed to use supervised sequential labeling methods for topic and opinion extraction. Experimental results showed that the supervised learning methods could achieve state-of-the-art performance on lexicon extraction. However, these methods need to manually annotate a lot of training data [6, 12, 19, 25] or design customized extraction rules [4, 28, 31]. Hong et al. [14] utilized a language-independent crowdsourcing game called Tower of Babel to build high-quality sentiment lexicon. Tower of Babel did not need the effort of language experts, but it required more than one hundred amateur annotators to participate in the lexicon building process.

Several ranking based algorithms have been proposed to build sentiment lexicons or determine the word's polarity. The WordNet's synsets and glosses were utilized to build relatedness graph, and the random walk model was applied to estimate the polarity of the given word [10, 13]. Velikovich et al. [37] constructed a graph from web-computed lexical co-occurrence statistics, and employed a graph propagation algorithm to rank the words and phrases in the graph [37]. The ranking based methods have several challenges. On one hand, the graph built based on thesaurus also suffers from the coverage problem. On the other hand, the quality of graph built based on large Web page corpus can not be guaranteed. Different from the previous literatures, in this paper we propose a word-emoticon mutual reinforcement ranking lexicon learning methods based on extremely large purified microblog dataset with the help of emoticons, which starts from scratch and does not need any deep syntactic parsing or predefined extraction patterns.

Rao et al. [32] built a fine-gained word-emotion mapping dictionary with help of the emoticon labels after online news articles. The learned lexicon was adaptive for personalized dataset, language-independent, fine-grained, volume-unlimited and had achieved promising performance in the testing dataset. Our work is similar to the work of Rao et al., as we all intend to find the relationship between words and emotions, and automatically build sentiment lexicons based on emotion labels. However, the emotion labels in [32] reflect the feelings of the readers after reading the news article, which may be different from the writers' perspectives when writing this article. On the contrary, the emoticons in microblogs are tagged by the writers themselves. Therefore, we think the emoticons in microblogs are better sentiment labels for the corresponding text and the microblogs are potential good corpus for sentiment lexicon learning.

2.2 Sentiment analysis in microblogs

Recently, the attention of sentiment analysis researchers has gradually shifted from news articles [38], blogs [39] and product reviews [7, 24, 26, 40] to microblogs [2, 9, 17, 33, 34, 41]. Bermingham and Smeaton utilized the traditional machine learning based algorithms to classify the microblogs into positive and negative categories [3] . Brody and Diakopoulos showed that word lengthening was strongly associated with subjectivity and sentiment in tweets [5]. They proposed several rules to change the lengthening words into their canonical forms and classified the sentiments in tweets based on the learned words. In [8], 50 Twitter tags and 15 smileys were treated as sentiment labels and a supervised sentiment classification framework was proposed to classify the tweets. The authors evaluated diverse feature types and the experiment results validated the effectiveness of emoticons as sentiment

indicators. The sentiment analysis in microblogs has achieved promising results. However, little work is done for using microblogs as corpus to learn sentiment lexicons, which may have many potential applications.

The Weibo microblogging service was launched at August 2009, and now has become the largest microblog website in China, which had more than 400 million registered users by October 2012. More than 100 million microblog posts are published in Weibo everyday, and each post is restricted to 140 Chinese characters in length. Moreover, an emoticon assistant tool is embedded in Weibo system. Therefore users can easily type emoticons in the microblog posts. Statistics show that Weibo has a higher rate of microblogs that contain emoticons than Twitter. In our crawled Weibo dataset, 19.6 % of the Chinese microblogs in Weibo contain at least one emoticon, compared with 8.1 % in Twitter (*476 million tweets dataset*) [36]. The detail of the statistics and the characteristics of emoticons in Chinese microblogs will be demonstrated in the next section.

3 The characteristics of emoticons in chinese microblogs

Nowadays, with the help of mobile devices, people usually like to record their personal emotions and feelings in microblogs at anytime and anywhere. Due to the length limitation, users prefer to utilize emoticons to directly express their sentiments especially in Chinese microblogs, such as Weibo. Several examples of microblog posts (translated from Chinese) in Weibo are shown below.

(1) *Long time no see, girls! I **love** you!* ❤ *My friends* 😄 😊
(2) *I was awakened up by the **nightmare** again. I don't know when I can fall asleep. It's really **annoying**!* 😠
(3) *Today is the **greatest** day in my life, and it is also my most **enjoyable** day.* ❤
(4) *Today is Spring Festival! We eat dumplings at midnight* 😁
(5) *The concert is so **cool**!!!* 😘 *But the dinner **sucks**!* 😠 😨
(6) *I **love** this movie! The visual effects are **awesome**!*

From above examples we can see that people tend to utilize emoticons to emphasize their emotion feelings. Since the microblogs have the length limitation, the sentiments expressed in these short text are usually consistent with the embedded emoticons. For example, the microblog post (1) expresses a positive sentiment, which is consistent with the *smile face* and *heart* emoticons. The microblog post (2) expresses a negative sentiment, which is consistent with the *crazy face* emoticons. Therefore, the sentiment words, such as *love* and *annoying*, have consistent sentiments with the corresponding emoticons in above examples. The emoticons can be regarded as the natural sentiment labels for the corresponding microblogs. On the other hand, there are also some noisy data in emoticon-labeled microblogs. For example, the post (4) contains a emoticon but no obvious sentiment word. The post (5) contains emoticons and sentiment words with opposite polarities. Of course there are also a lot of microblogs such as the post (6) that do not contain emoticons.

To further analyze the characteristics of emoticons in Chinese microblogs, we crawled more than 30 million microblog posts from Weibo. The detail of this crawled dataset will be discussed in the experiment section. According to the statistics of the crawled dataset, 19.6 % of the Chinese microblogs in Weibo contain emoticons, compared with 8.1 % in Twitter [36]. This is probably because Weibo has provided more convenient user interface when typing the emoticons. The high usage rate indicates that the emoticons are wildly used in Chinese microblog space.

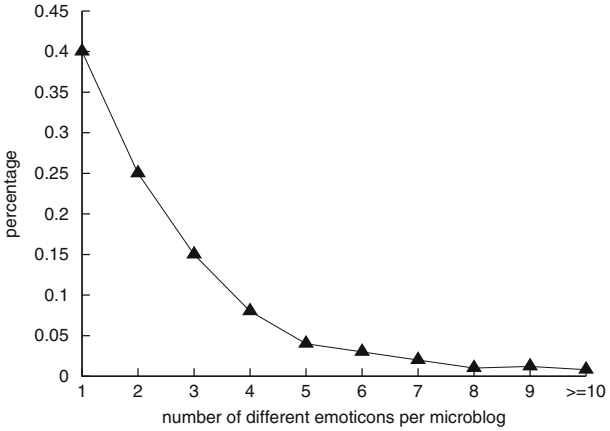

Figure 1 The usage of emoticons in Chinese microblogs

There may be more than one emoticons in one microblog post. For example, microblog post (1) contains three different emoticons. For the microblogs that contain emoticons in our crawled dataset, we analyze the number of different emoticons per microblog post. The result is shown in Figure 1.

In Figure 1, the horizontal axis represents the number of different emoticons per microblog. The vertical axis plots the percentage of the corresponding emoticon number in the microblogs that contain emoticons. Figure 1 depicts that among the microblogs with emoticons, about 41 % microblogs contain only one emoticon and about 59 % microblogs have more than one emoticons. These co-occurring emoticons may also have the similar sentiment orientation, such as the emoticons ❤ 😄 😊 in microblog post (1).

To evaluate the sentiment consistency in Chinese microblogs, we randomly select 2,000 microblog posts that contain emoticons. We manually tag the sentiment words, emoticons and their corresponding sentiment orientations in the selected microblog posts. The words or emoticons are regarded as co-occurring with each other if they appear in one microblog. The co-occurrence rate of the emoticons and sentiment words are shown in Figure 2.

In Figure 2 we can see that about 61 % microblogs with positive emoticons contain positive sentiment words, and only 9 % contain negative sentiment words. On the other hand, 63 % microblogs with negative emoticons contain negative sentiment words, and only 8 % contain positive sentiment words. This validate the observation that in general the positive sentiment words often co-occur with positive emoticons, and the negative sentiment words often co-occur with negative emoticons. Notice that there are also a portion of noisy microblogs that contain no sentiment words or contain both positive and negative emoticons, such as the post (6) and (5) in above examples. In the next section, we attempt to eliminate these noisy microblogs, which can pave the way for the lexicon learning step.

4 Seed emoticon selection and microblog data purification

From the discussion in Section 3 we know that the microblog is a potential good labeled data source for extracting sentiment lexicon. However, there are also noisy data embedded in them. To address these challenges, we design preprocessing strategies to eliminate the

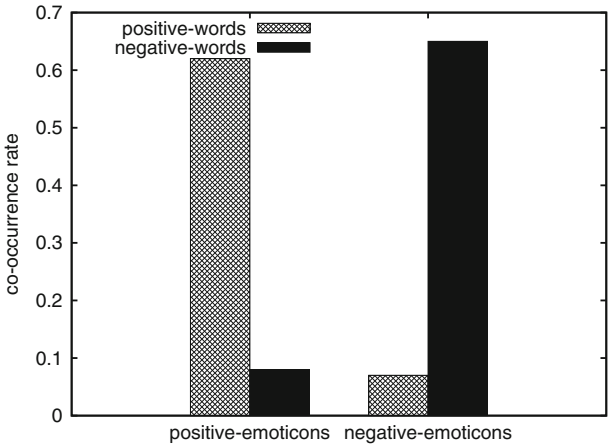

Figure 2 The co-occurrence rate of emoticons and sentiment words

noisy microblogs as many as possible. Since it is usually convenient to access the microblog data through API, our key idea is to learn sentiment lexicon from extremely large purified microblog dataset with the help of emoticons.

A parallel crawling system is implemented to collect the raw microblog dataset D through Weibo API. The Weibo platform has provided more than 1,000 pre-defined different graphical emoticons. In this paper, we manually select the most popular emoticons with high frequencies and obvious sentiment orientations, and group them into two categories. Finally we have the positive emoticon set EP that contains 25 emoticons and the negative emoticon set EN that contains 18 emoticons. The detail of the selected emoticons are shown in the Figure 3. Here we briefly introduce our preprocessing steps as follows.

(1) Given the raw dataset D, we eliminate the microblogs that do not contain the emoticons in EP and EN. We also filter out the data that have opposite emoticons. That is to say, if one microblog contains emoticons in EP and at the same time it has emoticons in EN, this microblog is eliminated. For example, the microblog post (5) in Section 3 is removed from D during this preprocessing step.

(2) We segment the microblogs into words by using the Chinese text processing tools. We also tag the words with part-of-speech information for the next steps. The microblogs that contain negation words are eliminated, because the negation words can affect the sentiment consistency of the sentiment words and emoticons.

(3) For each remaining microblog m, we remove the stop words. The Chinese words that only have one character are also eliminated, because the orientation of this kind

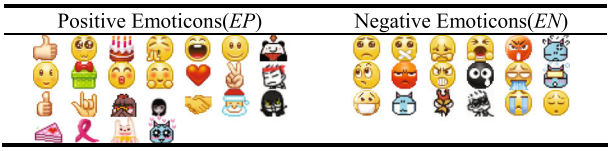

Figure 3 The positive emoticon set EP and negative emoticon set EN

of Chinese words is usually ambiguous or domain-specific. Our preliminary experiment results show that bringing in the words with only one character will decrease the performance of the learned lexicon.

(4) Since not all kinds of the words are good emotion indicators, we need not to reserve all the words as candidate sentiment words. Here the words with part-of-speech adjective, verb, noun and adverb are selected for the further detecting steps.

Although length of the text in microblogs is short, the example microblog posts validate that the positive and negative emotions can also be mixed up in one microblog. The above preprocessing steps make sure that we can get more sentiment consistent microblogs for training. After preprocessing, the purified dataset D is divided into D_{EP} and D_{EN} according to emoticon set EP and EN.

5 Building sentiment lexicon from purified microblog collection

In this section, firstly we propose a word-emoticon mutual reinforcement random walk model for lexicon building. Then we introduce a lexicon-based sentiment classification algorithm for Chinese microblogs.

5.1 A word-emoticon mutual reinforcement ranking model for sentiment lexicon building

In this paper we propose a mutual reinforcement random walk model to rank and extract sentiment words in the massive purified microblog dataset. Intuitively, we observe that the positive words often appear in the microblog posts with positive emoticons, and the negative words often appear in the microblog posts with negative emoticons, as shown in Figure 2 of Section 3. Different emoticons may also have different sentiment weights. For example, a *big laugh* emoticon may express stronger sentiments than a *smile* emoticon. Our basic assumption is that an emoticon expresses more obvious emotions if it often co-occurs with sentiment words and other important emoticons. And also, a word has higher sentiment value if it co-occurs with many important emoticons and has relation with many other important sentiment words. Take the positive sentiment words for example, this mutual reinforcement relationship of words and positive emoticons is shown in Figure 4.

In Figure 4, EP represents the positive emoticon set; SW denotes the candidate sentiment words that co-occur with positive emoticons in D_{EP}. We build three undirected graphs G_P, G_S and G_{PS} to reflect the EP-EP, SW-SW, EP-SW relationship. For each subgraph, if the two items co-occur with each other, we create an edge between these two items. For example, in bipartite graph G_{PS}, if a candidate sentiment word $sw \in SW$ co-occurs

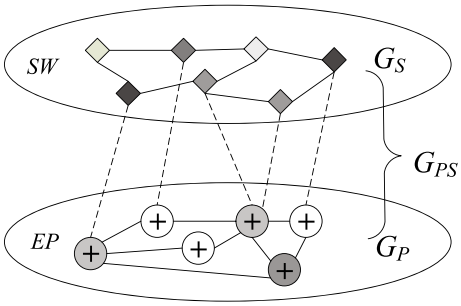

Figure 4 The mutual reinforcement model of positive emoticons and candidate sentiment words

with the positive emoticon $ep \in EP$, an edge will be created between ep and sw. Based on our assumption, after several mutual reinforcement iteration steps the words with obvious positive sentiment meanings will be ranked in higher position. Finally the candidate positive words in the graphs have different sentiment values, shown as different grayscale in Figure 4.

Firstly, we use S to denote the adjacency matrix of SW-SW subgraph G_S and the similarity between candidate sentiment words is calculated by their co-occurrence information in D_{EP}. The similarity weight is defined as:

$$S_{ij} = \begin{cases} \log \frac{p(sw_i, sw_j)}{p(sw_i)p(sw_j)} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $p(sw_i)$, $p(sw_j)$ are the occurrence probabilities of word sw_i and sw_j respectively in positive microblog dataset D_{EP}; $p(sw_i, sw_j)$ is the co-occurrence probability of word sw_i and sw_j in D_{EP}; When $i = j$, $S_{ij} = 0$ because there is no link from one node pointing to itself. The Formula (1) is the Pointwise Mutual Information function which can measure the statistics dependency of sw_i and sw_j [35]. Note that due to the short length there are always consistent emotions in one microblog post. Therefore, the higher S_{ij} value indicates these two words have higher probability to express the consistent sentiment orientation. Similarly, the adjacency matrices of subgraph G_P and G_{PS} are represented by the matrix E and W:

$$E_{ij} = \begin{cases} \log \frac{p(ep_i, ep_j)}{p(ep_i)p(ep_j)} & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$W_{ij} = \log \frac{p(sw_i, ep_j)}{p(sw_i)p(ep_j)} \tag{3}$$

where $p(sw_i, ep_j)$ is the co-occurrence probability of the candidate sentiment word sw_i and emoticon ep_j in D_{EP}; $p(ep_i, ep_j)$ is the co-occurrence probability of the emoticons ep_i and ep_j in D_{EP}. As we have discussed in Section 3, these co-occurrences also tend to indicate the consistent sentiments.

The matrix S, E and W is normalized to $\tilde{S}$, $\tilde{E}$, $\tilde{W}$ respectively, i.e. each row of the matrix is summed to 1. In addition, we normalize the transpose of W, namely W^T, to $\hat{W}$. Let R_{EP}, R_{SW} denote the ranking scores of EP and SW. The word-emoticon mutual reinforcement random walk approach can be formulated as follows:

$$\begin{cases} R_{EP}^{(k+1)} = \alpha \tilde{E}^T R_{EP}^{(k)} + (1-\alpha)\hat{W}^T R_{SW}^{(k)} \\ R_{SW}^{(k+1)} = \beta \tilde{W}^T R_{EP}^{(k)} + (1-\beta)\tilde{S}^T R_{SW}^{(k)} \end{cases} \tag{4}$$

Suppose we have:

$$Y_P = \begin{bmatrix} \alpha \tilde{E}^T & (1-\alpha)\hat{W}^T \\ \beta \tilde{W}^T & (1-\beta)\tilde{S}^T \end{bmatrix} \tag{5}$$

$$R_P = \begin{bmatrix} R_{EP} \\ R_{SW} \end{bmatrix} \tag{6}$$

In matrix form, we have the equation $Y_P \cdot R_P = \lambda R_P$. Similar to the idea of PageRank, we add links from one node to any other nodes with probability $1 - d$ in G_P and G_S graph. so we have:

$$Y_P = \begin{bmatrix} \alpha((1-d)Q/n_1 + d\tilde{E}^T) & (1-\alpha)\hat{W}^T \\ \beta \tilde{W}^T & (1-\beta)((1-d)Q/n_2 + d\tilde{S}^T) \end{bmatrix} \tag{7}$$

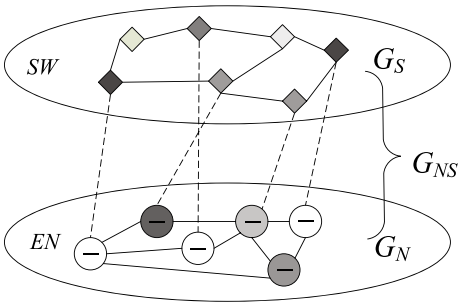

Figure 5 The mutual reinforcement model of negative emoticons and candidate sentiment words

where Q is a square matrix with each element equal 1. After adding the square matrix, it can be proved that the transpose of matrix Y_P is irreducible and stochastic.

Lemma 1 *Y_P^T is irreducible and when $\alpha + \beta = 1$, it is stochastic .*

Proof There is a link between each node in G_S and G_P, so they are strong connected. Because G_{PS} has connected the nodes in G_S and G_P graph, for each pair of nodes u, v in these three graphs, there is a path from u to v. Therefore, the new graph G_{All} composed by G_S, G_P, G_{PS} is strong connected. And also there will be more than one path for any pair of nodes in G_{All}, so G_{All} is aperiodic and the matrix Y_P^T is irreducible. For any column in the left part of Y_P:

$$\sum_i Y_{ij} = \alpha \left(\sum_{i=1}^{n_1} \frac{(1-d)}{n_1} + d \sum_{i=1}^{n_1} \tilde{E}_{ij}^T \right) + \beta \sum_{i=1}^{n_3} \tilde{W}_{ij}^T = \alpha + \beta \tag{8}$$

The same conclusion can be deduced in the right part of Y_P. So when $\alpha + \beta = 1$, the sum of elements in each column in Y_P is 1, and the matrix Y_P^T is stochastic. □

The power method is used to iteratively find the solution of the equation $Y_P \cdot R_P = \lambda R_P$. The irreducible and stochastic characteristics of Y_P^T guarantee that R_P will converge to a steady state.

Similarly, we can build the mutual reinforcement model for the negative emoticons and candidate sentiment words based on the dataset D_{EN}, as shown in Figure 5. In Figure 5, G_S is the candidate sentiment word graph; G_N is the negative emoticon graph; G_{NS} is constructed based on the co-occurrence of the words and emoticons. The weights of the links in these graphs are calculated based on the co-occurrence probabilities in dataset D_{EN}. Therefore, we have the ranking equation $Y_N \cdot R_N = \lambda R_N$.

Give a word w, suppose R_{Pw} and R_{Nw} denote the ranking value of w in R_P and R_N respectively, the final sentiment value of w is denoted by:

$$SO(w) = R_{Pw} - \phi \cdot R_{Nw} \tag{9}$$

where $\phi = \sum_i R_{Pi} / \sum_i R_{Ni}$. The bigger absolute value of SO indicates that w is more likely to belong to a predefined sentiment category. Given a threshold θ, for each word w in the ranking result, we classify it into predefined positive or negative category. If $|SO(w)| < \theta$, we eliminate w from the extracted lexicon because it does not show an obvious orientation to either of the sentiment categories. If $SO(w) > \theta$, w is classified into positive category. Otherwise if $SO(w) < -\theta$, w is classified into negative category. $SO(w)$ is regarded as

the sentiment weight of w in the learned lexicon. We denote our learned lexicon based on mutual reinforcement model as MR lexicon.

5.2 A lexicon-based sentiment classification algorithm for chinese microblogs

To evaluate the performance of the learned sentiment lexicon, we introduce a lexicon based sentiment analysis algorithm for Chinese microblog, which is given in Algorithm 1. In Algorithm 1, the number of matched positive and negative words from the lexicon is counted and the negation words are also considered during the classification.

Algorithm 1 Algorithm 1: Lexicon based sentiment analysis for Chinese microblog

Input : the microblog m, the sentiment lexicon L, negation word set NG
Output: sentiment score sentiscore of m

```
Split m into sentences;
foreach sentence s in m do
    Segment s into words;
    foreach word w of s
    do
        if w ∈ NG
        then
            ng++;
        if w ∈ L
        then
            sentiscore = sentiscore + SO(w);
    if ng is odd
    then
        sentiscore = −sentiscore/2;
return Σsentiscore
```

In Algorithm 1, firstly the given microblog is split into sentences according to the punctuation marks. Then we calculate the sentiment score of every sentence using the lexicon L, and the word sentiment weights are summed to represent the sentiment score $sentiscore$ of the microblog. Moreover, if the number of negation words is odd, the sentiment score $sentiscore$ will be decreased to $-sentiscore/2$, because the tone of the sentence can be weakened by the negation structure to a certain extent. For example, "*not happy*" is not stronger than "*sad*" on expressing the negative emotions. In Algorithm 1, if the sum $sentiscore > 0$, then this microblog is regarded as positive. If the sum $sentiscore < 0$, the microblog is regarded as negative. Otherwise, it is neutral. The Algorithm 1 is easy to understand, and its performance mainly depends on the quality of the sentiment lexicon. In the next section, Algorithm 1 is used to classify the polarities of Chinese microblogs and evaluate the quality of different sentiment lexicons.

6 Experiment

We evaluate our proposed lexicon learning method on a manually annotated Chinese microblog dataset. Our experiment is conducted on a commodity PC with Windows XP, Core2 Duo CPU and 4GB RAM.

6.1 Experiment setup

Crawled Dataset The Weibo API has a limit of 200 microblogs in one response for any request and also has a limit of requests per hour. To address this challenge, a parallel system with three PC nodes is designed to collect the Chinese microblogs as many as possible. For each node, we periodically send requests to public timeline Weibo API. After filtering out the duplicate and spam data, the microblogs from the three nodes are integrated together. Finally, we collected more than 30 million Chinese microblogs from October 1, 2011 to July 31, 2012.

In the crawled raw dataset, there are about 5.88 million microblogs containing emoticons. After the four preprocessing steps in Section 4, at last we have 1,481,775 purified microblogs that contain emoticons.

Unlike English and Spanish, there is no delimiter to mark word boundaries and no explicit definition of words in Chinese languages. In this paper we utilize NLPIR[3] to segment sentences into unique word tokens. NLPIR is a Chinese lexical analysis system, which is special customized for Chinese microblogs and is able to make the Chinese word segmentation and find new out-of-vocabulary Chinese words with quite high precision. After segmentation, each word token is associated with a POS tag given by NLPIR. During the preprocessing steps, we eliminate the microblogs that contain negation words, which may break the sentiment consistency between the words and emoticons. We also remove the words in the stop word list[4] and reserve the adjective, verb, noun and adverb words according to their POS tags.

Testing Dataset The testing dataset is provided by the Microblog Sentiment Analysis Evaluation Tasks (MSAET) of the NLP&CC 2012[5]. The testing dataset is crawled from Tencent microblog[6]. The dataset contains 1,000 Chinese microblogs with 20 topics, 3,416 sentences. Among them, 2,207 sentences have been annotated as 'Subjective' by research experts. In the subjective sentences, 407 is annotated as 'Positive' and 1,766 is annotated as 'Negative'.

Evaluation Method There are two related sentiment analysis tasks in MSAET, which first classifies the microblog posts as subjective or objective, and further distinguishes the subjective microblogs as positive or negative. For the first task, if a word is matched in the sentiment lexicon, we classify the sentence into subjective category. Otherwise, the sentence is classified into objective category. For the second polarity classification task, we utilize lexicon based Algorithm 1 to determine the sentence's polarity. The sentiment weight of word is also considered during the classification. Note that the algorithm's performance mainly depends on the quality of the sentiment lexicon. In this paper, we employ macro-average Precision, Recall and F-Measure to measure the performance of the classification results. For each class, we have the following evaluation metrics.

$$\text{Precision} = \frac{\#\text{ of correct classification}}{\#\text{ of algorithm output}} \tag{10}$$

[3] http://ictclas.nlpir.org

[4] https://sites.google.com/site/psocdescription/wordlist

[5] http://tcci.ccf.org.cn/conference/2012/

[6] http://t.qq.com/

$$\text{Recall} = \frac{\#\text{ of correct classification}}{\#\text{ of gold standard annotation}} \tag{11}$$

$$\text{F-Measure} = \frac{2 \times \text{ Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

We compare our learned MR lexicon with two famous Chinese sentiment lexicon HowNet [15] and NTUSD [23]. These two lexicons are manually built based on language resources and synonym thesaurus, which usually have a high quality and precision. We also compare MR lexicon with Turney's Pointwise Mutual Information (PMI) method [35] and the label propagation (LP) method [37, 42]. Here we use the microblog dataset as the corpus and the emoticons as the seed words or labels. For PMI method, we have the following formulas.

$$PMI(w, ep) = \log \frac{p(w, ep)}{p(w)p(ep)} \tag{13}$$

$$PMI(w, en) = \log \frac{p(w, en)}{p(w)p(en)} \tag{14}$$

where w denote the word in the purified training set D; $ep \in EP$ and $en \in EN$; $p(ep)$ is the occurrence probability of the positive emoticon ep in D; $p(w, ep)$ is the co-occurrence probability of w and the positive emoticon ep. Therefore, given the selected emoticon seed set EP and EN, the sentiment weight SO of w can be measured by:

$$SO(w) = \frac{\sum_{ep \in EP} PMI(w, ep)}{|EP|} - \frac{\sum_{en \in EN} PMI(w, en)}{|EN|} \tag{15}$$

Similarly, we extract the words with bigger absolute value of SO to form a sentiment lexicon, denoted as PMI lexicon. For the label propagation method, we construct an undirected graph based on the words and emoticons in dataset D and utilize the emoticons as the labels to propagate the sentiment polarity in the graph [37]. Finally, the top ranked words with obvious polarities are extracted to form a sentiment lexicon, denoted as LP lexicon. In Section 6.2, we will demonstrate the comparison results of different sentiment lexicons using MSAET dataset.

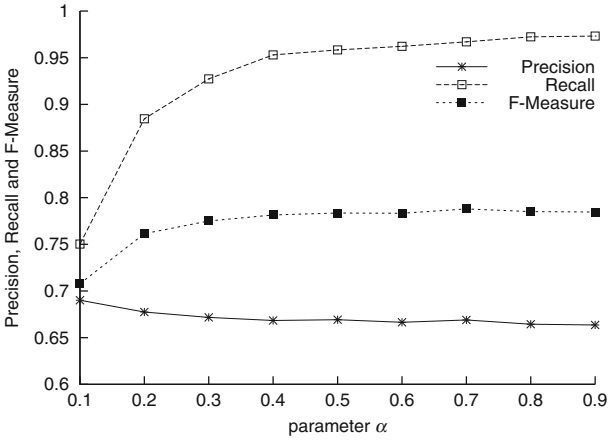

Figure 6 Precision, Recall and F-Measure with different values of parameter α in subjectivity classification task

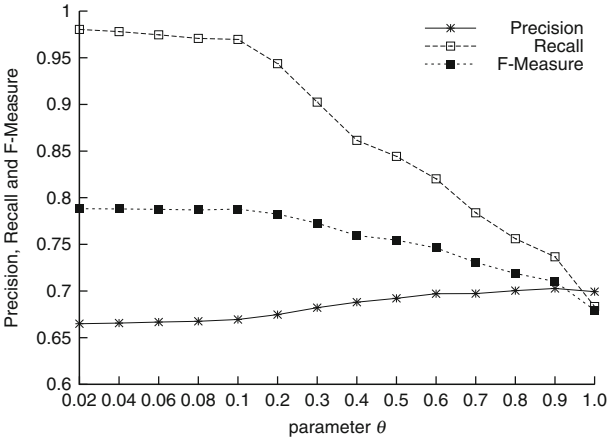

Figure 7 Precision, Recall and F-Measure with various θ in subjectivity classification task

6.2 Experiment results

Different settings of α and β may affect the quality of MR lexicon. For the subjectivity classification task in MSAET, firstly in order to better understand the relative contributions from the candidate sentiment words and emoticons, the parameter α is varied from 0.1 to 0.9. The experiment results are shown in Figure 6. Note that $\beta = 1 - \alpha$ to guarantee the convergence of the iteration and the lexicon size threshold parameter $\theta = 0.1$.

In the ranking formula the larger α is, the more contribution is given from the emoticons, while the less contribution is given from the candidate sentiment words. We can see from Figure 6 that the F-Measure increases dramatically when α is set from 0.1 to 0.3. Then the F-Measure has a moderate growth and reduction as α gets bigger. The best performance of F-Measure is achieved when $\alpha = 0.7$, which is set as the default value for the following experiments.

Secondly, for the Formula (9), the parameter θ determines the size of MR lexicon. Figure 7 plots the Precision, Recall and F-Measure with different settings of parameter θ.

It is obvious that the bigger value of θ can decrease the size of MR lexicon, but bring in more accurate sentiment words. The smaller value of θ can generate a wilder coverage, but also bring in more noisy words. In Figure 7 when θ grows bigger, the Precision firstly has a stable stage, and then gradually increases when $\theta > 0.1$. On the other hand, the Recall decreases when $\theta > 0.1$ dramatically which may be because that less sentiment words are included in the MR lexicon. In the following experiments we set $\theta = 0.1$.

Thirdly, we compare our MR lexicon with two famous sentiment lexicon HowNet and NTUSD. We also compare MR lexicon with the learned PMI and LP lexicon. Note that HowNet and NTUSD do not contain sentiment weight for each word. Here we use the value

Table 1 Lexicon size of different lexicons

	HowNet	NTUSD	LP	PMI	MR
Positive#	4,528	4,320	9,880	4,569	6,014
Negative#	2,813	8,277	4,670	6,140	6,785

Table 2 Lexicon recall with respect to other lexicons

	HowNet	NTUSD	LP	PMI	MR
HowNet	100 %	23.0 %	7.7 %	13.4 %	21.7 %
NTUSD	13.1 %	100 %	1.9 %	9.6 %	14.7 %
LP	4.0 %	2.2 %	100 %	3.2 %	8.0 %
PMI	8.7 %	11.4 %	3.1 %	100 %	45.8 %
MR	12.1 %	14.4 %	7.1 %	37.9 %	100 %

1 for the positive words and -1 for the negative words for the HowNet and NTUSD lexicon. The sizes of the lexicons are shown in Table 1.

We can see from Table 1 that all the lexicons are similar in size. The MR lexicon has a balanced size for both the positive and negative lexicons. The NTUSD has much more negative words and the LP has much more positive words. Table 2 demonstrates the recall of the each lexicon relative to the others, i.e. the percentage of overlap words between lexicons. For example, in the first row second column, 23 % means that about twenty-three percentage of the words in HowNet are the same with the words in NTUSD. The low overlaps between the learned lexicons and the manually constructed lexicons indicate that the words in the learned lexicons based on microblogs (e.g. MR lexicon) are quite different from the traditional sentiment lexicons (e.g. HowNet lexicon). We will evaluate if we can benefit from these words for the subjectivity and polarity classification task. The subjectivity classification performances of different lexicons are shown in Table 3.

Table 3 shows that our MR lexicon achieves a significant increase in the F-Measure value by 22 % in average compared with the classical lexicons (HowNet and NTUSD). The low Recall values of HowNet and NTUSD indicate that the classical lexicons seriously suffer from coverage problem when meeting the microblog data. On the other hand, HowNet has the best Precision performance. This may be because that the manually built HowNet lexicon is smallest in size but more accurate. We observe that the F-Measure of MR lexicon also outperforms PMI and LP lexicons in subjectivity classification task. This proves the effectiveness of the MR lexicon.

Table 4 depicts the experiment results for the polarity classification task of MSAET dataset. In Table 4, MR-b is the MR lexicon with binary sentiment weights, i.e. the weight 1 for the positive words and the weight -1 for the negative words. We find that the MR lexicon can achieve the best performance in all the evaluation metrics. The MR lexicon significantly outperforms the classical manually built HowNet and NTUSD lexicons, which are lack of the emerging sentiment words in microblogs. The MR lexicon also outperforms PMI and LP methods, which treat the emoticons as equal weights. Note that a *laugh face* emoticon expresses stronger sentiments than a *smile face* emoticon. Therefore the words co-occur with a *laugh face* may have higher sentiment values. Our proposed method can

Table 3 Subjectivity classification performance of different lexicons

	HowNet	NTUSD	LP	PMI	MR
Precision	0.677	0.675	0.661	0.670	0.669
Recall	0.733	0.563	0.952	0.931	0.967
F-Measure	0.701	0.597	0.780	0.774	0.788

Table 4 Polarity classification performance

	HowNet	NTUSD	LP	PMI	MR-b	MR
Precision	0.315	0.488	0.335	0.491	0.496	0.528
Recall	0.319	0.361	0.452	0.599	0.685	0.746
F-Measure	0.316	0.409	0.385	0.536	0.575	0.615

better reflect the sentiment weights and different relationships between sentiment words and emoticons in microblogs. The MR lexicon achieves better performance than MR-b lexicon, which validates that the learned weights are good indicators for the strength and popularity of sentiment words. Moreover, our proposed method does not use machine learning techniques, deep syntactic parsing or manually designed extraction patterns, but it could still achieve good performance in both the subjectivity and polarity classification tasks. This validates the effectiveness of our proposed sentiment lexicon building method.

6.3 Case study

Figure 8 presents a selection of the top ranked positive and negative words with their corresponding translations in the MR Lexicon. We are glad to find that the emerging Internet new words such as 给力 (gei^3li^4, *awesome*), 悲催 (bei^1cui^1, *tragic*) are also included in our lexicon. Moreover, the proposed method can also find the words with new meanings. For example, the word 碉堡 ($diao^3bao^4$) originally means a bunker. However, because of homophones it now has a new meaning *damn good* in social media, which expresses strong emotions. The onomatopoeia 嘎嘎 (ga^1ga^1, *ha-ha*) expresses a positive sentiment in the Internet, which means a kind of laughter. Another interesting observation is that the word 地沟 (di^4gou^1,*drainage*) may be a wrong segmentation result for the word 地沟油 ($di^4gou^1you^2$, *drainage oil*), which means the illegally recycled cooking oil.

Positive Words		Negative Words	
动人 (moving)	梦想成真 (dreams come true)	惋惜 (regret)	无聊 (boring)
过瘾 (have fun)	震撼 (electrifying)	肤浅 (superficial)	你妹 (shit)
甜蜜 (sweet)	碉堡 (damn good)	命苦 (bitter life)	烦躁 (fretful)
挂念 (missing)	搞笑 (interesting)	悲催 (tragic)	郁闷 (depressed)
雪中送炭 (timely help)	嘎嘎 (ha-ha)	呜呜 (boo-hoo)	尼玛 (damn)
美满 (happy)	幽默 (humor)	崩溃 (collapse)	交通 (traffic)
一路顺风 (bon voyage)	给力 (awesome)	伤脑筋 (knotty)	地沟 (drainage)

Figure 8 Examples of positive and negative words in the MR lexicon

These examples validate our basic assumption that the microblog is an effective data source for learning emerging words and new meanings for sentiment analysis task.

7 Conclusion and future work

In this paper, we propose a mutual reinforcement ranking based method to extract sentiment words from massive microblog data collection with the help of emoticons. Our basic assumption is that an emoticon expresses more obvious emotions if it often co-occurs with sentiment words and other important emoticons. And also, a word has higher sentiment value if it co-occurs with many important emoticons and has relation with many other important sentiment words. The proposed ranking model can create a sentiment lexicon free of laborious efforts of the experts who must be familiar with both linguistic and psychological knowledge. The experiment results in a benchmark dataset validate that the our learned lexicon outperforms the traditional manually built sentiment lexicons and other learned lexicons using emoticons as seeds.

Usually the sentiment of words is sensitive to the domains. For the future work, we intend to combine our method with known methods to learn domain or topic specified sentiment words from the massive microblog data. We also intend to consider the word sense and the part-of-speech for the generated sentiment lexicon. One limitation of our proposed model is that we could not extract the sentiment words with very low occurrence frequencies in the training corpus. To tackle this coverage problem, in the future work we will incorporate predefined lexical patterns to extract the candidate sentiment words with low frequencies.

Acknowledgments Project supported by the State Key Development Program for Basic Research of China (Grant No. 2011CB302200-G), State Key Program of National Natural Science of China (Grant No. 61033007), National Natural Science Foundation of China (Grant No. 61100026, 61370074), and the Fundamental Research Funds for the Central Universities (N120404007, N100704001). We would like to thank reviewers for their valuable comments and suggestions.

References

1. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC), pp. 2200–2204 (2010)
2. Barbosa, L., Feng, J.: Robust sentiment detection on twitter from biased and noisy data. In: Proceedings of the 23rd International Conference on Computational Linguistics (COLING), pp. 36–44 (2010)
3. Bermingham, A., Smeaton, A.F.: Classifying sentiment in microblogs: is brevity an advantage? In: Proceedings of the the 19th ACM Conference on Information and Knowledge Management (CIKM), pp. 1833–1836 (2010)
4. Bollegala, D., Weir, D., Carroll, J.: Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In: Proceedings of 49th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 132–141 (2011)
5. Brody, S., Diakopoulos, N.: Cooooooooooooooollllllllllll!!!!!!!!!!!!!! Using word lengthening to detect sentiment in microblogs. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 562–570 (2011)
6. Choi, Y., Cardie, C.: Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 590–598 (2009)
7. Cui, H., Mittal, V., Datar, M.: Comparative experiments on sentiment classification for online product reviews. In: Proceedings of the Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference (AAAI), pp. 1265–1270 (2006)

8. Davidov, D., Tsur, O., Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys. In: Proceedings of the 23rd International Conference on Computational Linguistics (COLING), pp. 241–249 (2010)
9. Diakopoulos, N., Shamma, D.A.: Characterizing debate performance via aggregated twitter sentiment. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI), pp. 1195–1198 (2010)
10. Esuli, A., Sebastiani, F.: PageRanking wordnet synsets: an application to opinion mining. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 424–431 (2010)
11. Feldman, R. Commun. ACM. **56**(4), 82–89 (2013)
12. Gao, D., Wei, F., Li, W., Liu, X., Zhou, M.: Co-training based bilingual sentiment lexicon learning. In: Proceedings of the 27th AAAI Conference on Artificial Intelligence (AAAI), pp. 26–28 (2013)
13. Hassan, A., Radev, D.: Identifying text polarity using randomWalks. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 395–403 (2010)
14. Hong, Y., Kwak, H., Baek, Y., Moon, S.: Tower of Babel: a crowdsourcing game building sentiment lexicons for resource-scarce languages. In: Proceedings of the 22nd International World Wide Web Conference (WWW), pp. 549–556 (2013)
15. HowNet. http://www.keenage.com Accessed 1 Mar 2012
16. Hu, M., Liu, B.: Mining and summarizing customer review. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 168–177 (2004)
17. Jiang, L., Yu, M., Zhou, M., Liu, X., Zhao, T.: Target-dependent twitter sentiment classification. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 151–160 (2011)
18. Jijkoun, V., Rijke, M., Weerkamp, W.: Generating focused topic-specific sentiment lexicons. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 151–160 (2010)
19. Jin, W., Ho, H.H., Srihari, R.K.: OpinionMiner: a novel machine learning system for web opinion mining and extraction. In: Proceedings of the the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 1195–1204 (2009)
20. Kaji, N., Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of html documents. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP), pp. 1075–1083 (2007)
21. Kanayama, H., Nasukawa, T.: Fully automatic lexicon expansion for domain-oriented sentiment analysis. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 355–363 (2006)
22. Kim, S.M., Hovy, E.H.: Determining the sentiment of opinions. In: Proceedings of the 20th International Conference on Computational Linguistics (COLING), pp. 1367–1373 (2004)
23. Ku, L., Chen, H.: Mining opinions from the web: beyond relevance retrieval. J. Am. Soc. Inf. Sci. Technol. **58**(12), 1838–1850 (2007)
24. Leung, C., Chan, S., Chung, F., Ngai, G.: A probabilistic rating iference framework for mining user preferences from reviews. World Wide Web **14**(2), 187–215 (2011)
25. Li, F., Han, C., Huang, M., Zhu, X., Xia, Y.J., Zhang, S., Yu, H.: Structure-aware review mining and summarization. In: Proceedings of the 23rd International Conference on Computational Linguistics (COLING), pp. 653–661 (2010)
26. Liu, Y., Yu, X., An, A., Huang, X.: Riding the tide of sentiment change: Sentiment analysis with evolving online reviews. World Wide Web **16**(4), 477–496 (2013)
27. Lu, Y., Castellanos, M., Dayal, U., Zhai, C.: Automatic construction of a context-aware sentiment lexicon: an optimization approach. In: Proceedings of the the 20th International Conference on World Wide Web (WWW), pp. 347–356 (2011)
28. Mohammad, S., Dunne, C., Dorr, B.: Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 599–608 (2009)
29. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2007)
30. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC), pp. 1320–1326 (2010)
31. Qiu, G., Liu, B., Bu, J., Chen, C.: Expanding domain sentiment lexicon through double propagation. In: Proceedings of the 21st International Joint Conference on Artificial intelligence (IJCAI), pp. 1199–1204 (2009)
32. Rao, Y., Quan, X., Wenyin, L., Li, Q., Chen, M.: Building word-emotion mapping dictionary for online news. In: Proceedings of the first International Workshop on Sentiment Discovery from Affective Data (SDAD), pp. 28–39 (2012)

33. Si, J., Mukherjee, A., Liu, B., Li, Q., Li, H., Deng, X.: Exploiting topic based twitter sentiment for stock prediction. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL), pp. 24–29 (2013)
34. Speriosu, M., Sudan, N., Upadhyay, S., Baldridge, J.: Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 53–63 (2011)
35. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), pp. 417–424 (2002)
36. Yang, J., Leskovec, J.: Patterns of temporal variation in online media. In: Proceedings of the Forth International Conference on Web Search and Web Data Mining (WSDM), pp. 177–186 (2011)
37. Velikovich, L., Blair-Goldensohn, S., Hannan, K., McDonald, R.T.: The viability of web-derived polarity lexicons. In: Proceedings of the North American Chapter of the Association of Computational Linguistics (NAACL), pp. 777–785 (2010)
38. Zhang, J., Kawai, Y., Kumamoto, T., Tanaka, K.: A novel visualization method for distinction of web news sentiment. In: Proceedings of 10th International Conference on Web Information Systems Engineering (WISE), pp. 181–194 (2009)
39. Zhang, X., Zhou, Y.: Holistic approaches to identifying the sentiment of blogs using opinion words. In: Proceedings of the 12th International Conference on Web Information Systems Engineering (WISE), pp. 15–28 (2011)
40. Zhang, R., Tran, T., Mao, Y.: Opinion helpfulness prediction in the presence of "Words of Few Mouths". World Wide Web J. **15**(2), 117–138 (2012)
41. Zhao, J., Dong, L., Wu, J., Xu, K.: MoodLens: an emoticon-based sentiment analysis system for chinese tweets. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 1528–1531 (2012)
42. Zhu, X., Ghahramani, Z.: Learning from labeled and unlabeled data with label propagation. Technical report, CMU-CALD-02 (2002)