Every Term Has Sentiment: Learning from Emoticon Evidences for Chinese Microblog Sentiment Analysis*

Fei Jiang, Anqi Cui, Yiqun Liu, Min Zhang, and Shaoping Ma

State Key Laboratory of Intelligent Technology and Systems,
Tsinghua National Laboratory for Information Science and Technology,
Department of Computer Science and Technology,
Tsinghua University, Beijing 100084, China
{f91.jiang,cuianqi}@gmail.com, {yiqunliu,z-m,msp}@tsinghua.edu.cn

Abstract. Chinese microblog is a popular Internet social medium where users express their sentiments and opinions. But sentiment analysis on Chinese microblogs is difficult: The lack of labeling on the sentiment polarities restricts many supervised algorithms; out-of-vocabulary words and emoticons enlarge the sentiment expressions, which are beyond traditional sentiment lexicons. In this paper, emoticons in Chinese microblog messages are used as annotations to automatically label noisy corpora and construct sentiment lexicons. Features including microblog-specific and sentiment-related ones are introduced for sentiment classification. These sentiment signals are useful for Chinese microblog sentiment analysis. Evaluations on a balanced dataset are conducted, showing an accuracy of 63.9% in a three-class sentiment classification of positive, negative and neutral. The features mined from the Chinese microblogs also increase the performances.

Keywords: Microblog, Sentiment Analysis, Sentiment Lexicon Construction, Support Vector Machine.

1 Introduction

Microblog is a new form of social networking service where millions of people express their feelings and the posts are colloquial and irregular. A high percentage of messages have positive or negative sentiment.Sentiments of people in social events draw wide attentions of the public and sentiment analysis on microblog messages has become a popular research topic.

Sentiment analysis can be applied to many topics in microblogs. On business topics, we can obtain word-of-mouth reputation of brands or products from a large number of users, thus helping companies to improve their products [6]. On stock markets, people's sentiment polarities towards a hot topic may affect the

* This work was supported by Natural Science Foundation (61073071), National High Technology Research and Development (863) Program (2011AA01A207). Part of this work has been done at the NUSTsinghua EXtreme search centre (NExT).

G. Zhou et al. (Eds.): NLPCC 2013, CCIS 400, pp. 224–235, 2013.

stock prices [2]. On political events, the sentiments of citizens reflect election situations [13]. Generally speaking, most of these studies are in English microblogs (such as Twitter), but similarly they are also important in Chinese microblogs.

The objective of sentiment classification is to classify the sentiment of a piece of text into three classes: Positive, negative and neutral. Traditional sentiment analysis methods mainly concentrate on a specific target in reviews, such as product reviews or movie reviews. In these domains, users post their comments as well as the ratings (in scores or stars) for a certain product or movie. Based on the texts and ratings, supervised machine learning models can be utilized to learn semantic information of the products such as feature-opinion pairs, which are then applied to texts in the same domain. However, for microblog messages, they have all types of topics which do not belong to a certain domain. What is more, they do not have labels or ratings like product reviews; it costs much time and efforts to label many messages manually. The above difficulties for microblog sentiment analysis are of main concerns in the research area.

In Chinese microblogs, there are even more challenges on sentiment analysis, including word segmentation with out-of-vocabulary (OOV) words recognition, syntactical analysis on informal sentences. Hence, we propose the following contributions to reduce these difficulties:

First, we propose a new method for automatic sentiment lexicon construction with emoticon evidences, meanwhile dealing with out-of-vocabulary words and commonly used phrases. More over, in our lexicon, every word has a potential sentiment, which is reflected by continuous positive score and negative score. It is more flexible to extract features other than discrete-valued scores. Further more, in sentiment classification, we use different types of features, which represent different views of chacteristics of microblog messages. And finally, unlike many three-class classification methods, our method does not necessarily need large amounts of neutral corpora, which are really difficult to obtain.

The rest of this paper is organized as follows: Section 2 briefly describes some related work. In section 3, we introduce emoticons and present our methods of sentiment lexicon construction. Section 4 presents our methods in feature extraction and classification strategies. We conduct experiments in section 5 and conclude the paper in section 6.

2 Related Work

Sentiment analysis is usually considered as a classification problem: Classify the sentiments of microblog messages into positive, negative or neutral. Traditionally, machine learning methods have been applied to sentiment classification [11]. More specifically, in Twitter sentiment analysis, the noisy tweets can be used as training data to train a classification model [1]. Part of the noisy data is obtained by some rules (e.g. tweets containing specific sentiment words), hence the features of the classifier are restricted to some limited ones generated from the formal words. Different from them, emoticons are used as approximate labels of tweets to obtain a large amount of noisy data [10]. In these methods, the quality and bias of the noisy data limit the performance of classification. More over, when perform three-class classification, assumptions about neutral corpora in twitter such as [10] do not hold in Chinese microblog.

To ensure the quality of training corpus, the best method is manually labeling [8]. However, the scale of labeled data is limited – then the features extracted have to be more general, while the dimension of features is also reduced. Sentiment lexicons make it possible to map a high dimension of word vector to a low dimension of several sentiment strengths.

Lexicons can be constructed manually [16,18]. The advantage is of higher accuracy, but they are limited for the cost and coverage [9,12]. Meanwhile, large numbers of spoken words, buzz words and dialects (e.g. Cantonese) make the coverage of manually constructed lexicons even more limited.

Lexicons can also be constructed automatically from the corpus. Starting with a few sentiment words and making use of the relationship of words in the corpus, the sentiment polarities of more words can be discovered [5,7,14]. Some emoticons can serve as seeds, However, the interrelation of words must be reliable. Hence discovering a proper relationship of words is important when building a lexicon from the corpus, especially when the corpus is noisy.

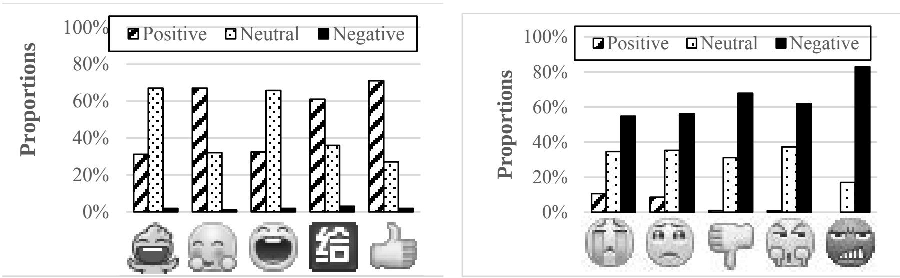
3 Sentiment Lexicon Construction

3.1 Emoticons in Chinese Microblogs

Emoticon is one of the main characteristics of Chinese microblogs. Instead of typing characters to compose emoticons (e.g. ":-)") in English microblogs, in Chinese microblogs users "type" emoticons by clicking on some icons on the web interface. The icons are mapped to characters surrounded with brackets to store and transfer as texts; these texts are displayed as emoticon images when rendering the web page. The emoticons are usually animated, making microblog messages more vivid, and more accurate to show users' emotions.

In previous work, emoticons such as :-) or :-(are used to collect positive or negative tweets from noisy data. However, there are much more emoticons in Chinese microblogs who indicate strong emotions. Hence we examine ten commonly used emoticons, including five positive ones: [笑哈哈] (laughing), [太 开心] (very happy), [哈哈] (Haha), [给力] (*Geili*, awesome), [good]; and five negative ones: [泪] (tears), [悲伤] (sad), [弱] (weak), [鄙视] (despise), [怒] (angry). For each of these emoticons, we randomly labeled 100 microblog messages containing it. The proportions of positive, neutral and negative messages are shown in Fig. 1.

It shows that messages containing or , are mostly neutral instead of positive. These two emoticons do have positive sentiments as shown in their appearances, but many users are used to posting them in any messages. Compared with them, the other three: , and have stronger positive sentiments, thus are more suitable for collecting positive messages from the corpus.



(a) Positive emoticons (b) Negative emoticons

Fig. 1. Proportions of sentiment polarities of microblog messages with emoticons

Similarly, for the negative emoticons, and have more noises than , and – the former two have less negative but more positive messages. This may because they are ambiguous; people use them not only when they are sad, but also when they are moved. Therefore, the latter three emoticons are more suitable for collecting negative messages from the corpus.

3.2 Phrase Extraction and Out-Of-Vocabulary Word Recognition

Because of the irregular expressions, traditional segmenters do not work well on microblog messages. And Out-Of-Vocabulary (OOV) words are also commonly used. Besides, many phrases (n-grams) occur very often and have less ambiguity compared to treating them as separate words, such as “太次”(too bad). These phrases are also put in the lexicon which benefits sentiment analysis.

In a large set of segmented sentences(with NLPIR2013[17]), we concatenate adjacent words to get every phrase (n-grams, n is not limited to two) with less than M characters. Then we compute two metrics of each phrase t:

1. Frequency $freq(t)$: the number of its occurrences in the corpus.

2. Tightness $T(t)$: Denote the words in the phrase t as $\{w_i\}$, the tightness of t is defined as:

$$T(t) = \frac{freq(t)}{\sqrt[n]{\prod freq(w_i)}} \tag{1}$$

With some proper thresholds, phrases of at least one large enough metric are selected, which cover OOVs. In this paper about 60,000 phrases and OOVs are extracted (without POS tags). In the following part of this paper, all the basic words, OOVs and phrases are referred to as *words*. They make a *word set*.

3.3 Modified-by-Degree-Adverbs Words Recognition

Modified by Degree Adverbs (MDA) words are the words that can be modified by degree adverbs. These words often have obvious opinions, such as “喜欢” (favor), “讨厌” (disgusting) and “好” (good). Therefore, recognition of MDA words are helpful to sentiment analysis.

Five commonly used degree adverbs are used: “很” (very), “非常” (very), “比较” (a little bit), “太” (too) and “十分” (very). They are matched in the sentences before segmentation; their following words (just next to them) are extracted as MDA words. Note these MDA words are the longest possible ones in the *word set*. For example, in “这位叔叔很友好” (This uncle is very friendly), “很” is a degree adverb; both “友” (friend) and “友好” (friendly) exist in the *word set*, but only “友好” is kept as an MDA word. Moreover, words starting with degree adverbs (such as “太棒”,, very great) are also considered as MDA words.

3.4 Negation Words

Negation is a common phenomenon in Chinese, which may inverse the sentiment of a word or a sentence. Some sentiment words are often modified by negations, such as “靠谱” (reliable). Negation has a significant influence on the sentiment words, so it is an important part of lexicon construction and sentiment analysis.

We complement some negations to the list [15] to generate a list of 53 negations, which are then added into the *word set*. The first adjectives, verbs or words without POS tag (i.e. phrases, OOV words or emoticons) in the *word set* modified by each negation are extracted. But if there is no such word within N words after the negation, the negation is ignored. In the experiments $N = 3$.

For example, the segmentation result of the sentence “这位先生，您真是站着说话不腰疼” (Sir, what you think is really easier said than done): “这/rzv 位/q 先生/noun /wd 您/rr 真/d 是/vshi 站/n 着/uzhe 说/v 话/n 不/d 腰/n 疼/v ” Assume we have {真, 是, 站, 着, 说, 话, 不, 腰, 疼, 您, 真是, 站着 说话, 腰疼, 这位, 先生, } in the *word set*, this sentence is transformed to the following *intermediate result*: “这位, 先生, 您, 真是, 站着, 说话, 腰疼(−1), ”, where words labeled by −1 are modified by negations.

After converting the messages to *intermediate results*, the sentiment lexicon can be constructed. Depending on whether or not we have sufficient corpus containing the specified emoticons (in Section 3.1), we introduce two methods to compute the positive and negative scores of all the words in our *word set*.

3.5 Frequency Statistics Algorithm for Sufficient Corpus

Messages containing some specified emoticons (, or for positive, and , or for negative) can be extracted as noisy labeled data for training in supervised learning algorithms. We call the corpus *sufficient* if it contains enough messages of these emoticons. Messages in the positive and negative sets are then transformed to the *intermediate result*.

Considering all the words in the above results. Define a positive word set A: Words not modified by negations in the positive message set and words modified by negations in the negative message set, and a negative word set B: Words modified by negations in the positive message set and words not modified by negations in the negative message set. Compute the positive and negative score of each word by their occurrences, as shown in the following equations:

$$score_{+}(w) = \frac{freq_A(w) + SMOOTH_A}{size(A) + N \cdot SMOOTH_A}, score_{-}(w) = \frac{freq_B(w) + SMOOTH_B}{size(B) + N \cdot SMOOTH_B} \quad (2)$$

where w is the word, $freq$ is the number of its occurrence in the word set, and $size$ is the sum of occurrences of all the words in each set. $SMOOTH$ is a parameter for smoothing, which should satisfy $\frac{size(A)}{size(B)} = \frac{SMOOTH_A}{SMOOTH_B}$.

The two scores are then normalized to one. We have no absolute sentiment words or non-sentiment words, but the sentiment polarity and strength of one word are reflected by the difference between positive and negative score.

3.6 Label Propagation Algorithm for Insufficient Corpus

If the corpus is *not sufficient* enough of messages containing the specified emoticons, we have to learn the polarities of words regarding the certain amount of the messages. Similarly as in [4], we construct a co-occurrence graph whose nodes are the words in the *intermediate result*, edges are their co-occurrences and apply the unsupervised label propagation algorithm on it. The three positive emoticons and three negative emoticons are used as positive seeds and negative seeds, respectively. Compared to the original algorithm, we also need to:

1. Dealing with negations: The weights of edges between every two node are initialized to zero. If two words in an *intermediate result* are both or neither modified by negations, the weight of their edge is increased by one. If there is only one word modified by negations, the weight is decreased by one.

2. Refine the iteration equation into:

$$s_{n+1} = \alpha \cdot W \cdot s_n + (1 - \alpha) \cdot \mathbf{b} \quad (3)$$

where $\mathbf{b}$ is the normalized static score vector, the dimension of $\mathbf{b}$ is the number of words (nodes) in the graph. For the positive (or negative) seed words, the corresponding elements in $\mathbf{b}$ are $1/m$ (m is the number of seeds). For other words, the corresponding elements are zero. W is the normalized symmetric matrix of the undirected co-occurrence graph. In [4], every element in W is non-negative, but after we introduce in negations, element in W can be negative, then the sum of the *absolute* value of elements in each row is 1. The vector s_n is the positive (or negative) score of each word after the n-th iteration, and the initial value s_0 is set to $\mathbf{b}$. α is a parameter between $[0, 1]$ to control the impact of seeds. Hence, n (before convergence) and α are the only variables in a determinated seed set and co-occurrence graph.

The positive and negative scores are propagated from positive and negative seeds, respectively. Finally they are normalized as well.

4 Sentiment Classification Based on Sentiment Lexicon

Based on the constructed sentiment lexicon, we extract features from microblog messages, and use Support Vector Machine (SVM) to perform sentiment classification with different strategies.

4.1 Feature Extraction

In this subsection, we will extract different types of features from microblog messages. Each type of features has their specific relationship with the sentiment of messages.

Microblog Structure Features. *Microblog structure features* include mentioning labels (@), URLs, etc. are relavant to the sentiment. For example, URLs in one message indicate that it may be an advertisement, the geo-location statement ("我在:" + URL, I'm at) indicates the message is more likely to be posted by individuals instead of organizations. Hence the following microblog structure features are extracted:

(1) Existence of geo-location statement (boolean).
(2) Number of URLs.
(3) Number of mentioning labels (@).
(4) Number of hashtags (texts quoted by a pair of hash symbols "#").
(5) Number of reply labels ("回复@" + username).

Sentence Structure Features. Sentences in microblog messages have typical features (*sentence structure features*), including:

(6) Number of semicolons (;).
(7) Number of ideographic comma (、).
(8) Number of percent sign (%).
(9) Existence of continuous serial numbers (numbers in a sequence) (boolean).
(10) Number of decimal points.

These features are related to an objective message. For example, messages with many semicolons usually have parallelism sentences, which are less likely to be subjective. Similarly, numbers in a sequence are used to list items formally. The percent sign and decimal points usually appear in news texts that present data.

Sentiment Lexicon Features. Features from the sentiment lexicon are helpful to learn the sentiment of messages. First, from the *intermediate result* we get the words and their sentiment scores (positive and negative). For words modified by negations, their positive and negative scores are swapped. Then, compute the product and maximum of positive (negative) scores of words. The product reflects an accumulative influence of all the words in a message, while the maximum considers the word with the strongest sentiment strength. The *basic lexicon features* are listed below:

(11) Product of positive scores of all the words.
(12) Product of negative scores of all the words.
(13) The maximum of positive scores of all the words.
(14) The maximum of negative scores of all the words.

In addition, single-character words are more ambiguous than multi-character words; the differences between their positive and negative scores are less significant. So the features from single-character words are treated separately:

(15) Product of positive scores of single-character words.
(16) Product of negative scores of single-character words.

(17) The maximum of positive scores of single-character words.
(18) The maximum of negative scores of single-character words.

MDA words are more likely to have sentiments than ordinary words, thus we have the following *MDA features* to reflect their impacts:
(19) Product of positive scores of MDA words.
(20) Product of negative scores of MDA words.
(21) The maximum of positive scores of MDA words.
(22) The maximum of negative scores of MDA words.

Emoticon Features. During the construction process of the sentiment lexicon, emoticons and ordinary words are treated equally. However, emoticons are the special characteristics of microblog messages which are commonly used and have strong sentiments, so we extract *emoticon features* listed as follows:
(23) Product of positive scores of emoticons.
(24) Product of negative scores of emoticons.
(25) The maximum of positive scores of emoticons.
(26) The maximum of negative scores of emoticons.

4.2 Sentiment Classification Based on SVM

The sentiment polarities of microblog messages are classified into three classes: positive, negative and neutral. An SVM classifier is set up with the linear kernel and other default settings in the LibSVM library [3]. Three strategies of classification are introduced:

1. One-stage three-class classification: The sentiment of a message is directly classified into three classes.
2. Two-stage two-class classification (hierarchical): In the first stage, the sentiment is classified into neutral or non-neutral. In the second stage, the non-neutral messages are classified into positive or negative.
3. Two-stage two-class classification (parallel): In the first stage, the sentiment is classified into positive or non-positive. In the second stage, the same message is classified into negative or non-negative. The final sentiment class of the message is determined by:

$$\text{sentiment} = \begin{cases} \textit{positive} & \text{first stage: } \textit{positive}\text{, second stage: } \textit{non-negative} \\ \textit{negative} & \text{first stage: } \textit{non-positive}\text{, second stage: } \textit{negative} \\ \textit{neutral} & \text{otherwise} \end{cases} \tag{4}$$

5 Experiments and Discussions

5.1 Dataset and Metrics for Sentiment Lexicons Evaluation

We first evaluate the performance of the sentiment lexicon. The golden standard is 467 positive words and 469 negative words from [18]. The error rate of a sentiment lexicon is defined as:

$$E = \frac{\sum\limits_{w \in POS\_E} freq(w) \cdot bias(w, NEG)) + \sum\limits_{w \in NEG\_E} freq(w) \cdot bias(w, POS)}{\mid NEG \mid \cdot \sum\limits_{w \in POS} freq(w) + \mid POS \mid \cdot \sum\limits_{w \in NEG} freq(w)} \tag{5}$$

Where $freq(w)$ is the frequency of word w in the corpus. POS and NEG are sets of labeled positive and negative words (golden standard). POS_E is the set of words who are in POS but have higher negative scores than their positive scores, and NEG_E similarly. $bias(w, NEG)$ means: For word $w \in POS\_E$, the number of words in NEG whose quotient of positive and negative score is larger than that of w. And $bias(w, POS)$ similarly. Obviously, $E \in [0, 1]$.

The sizes of POS_E and NEG_E are the number of misclassified words (a word is correctly classified if the difference between its positive and negative score is in accordance with its label). $freq(w)$ is the frequency of w; the larger it is, the more important w is to influence the sentiment. The *bias* reflects the degree of errors for misclassified words. In all, this definition takes these three main factors into consideration. In addition, the scores of a labeled word who is not in the sentiment lexicon are both 0.5, and it is put into POS_E or NEG_E.

5.2 Sentiment Lexicons Evaluation

For the *sufficient corpus*, we separately collect one million and six million messages that contain the six emoticons mentioned before. The error rates are 0.0342 and 0.0158, respectively, both less than 5%.

For the *insufficient corpus*, we randomly collect one million messages, and fit them into the label propagation algorithm. With different iteration times and different α's, different lexicons are constructed. Their error rates are shown in Fig. 2. When the iteration times are two, the error rates of the lexicon are relatively low. One iteration only is not enough to propagate accurate scores, but excessive iterations make more noise. The figure also shows the lowest error rate is achieved when α is around 0.9. We can also see that after the number of iterations reaches 5, the error rate is very sensitive to α.

To examine the relationship between the accuracy of sentiment classification (the details of classification are introduced later) and the error rate, six points are chosen from Fig. 2: (1-5) Iteration times = 5 and $\alpha = 0.5, 0.8, 0.9, 0.95, 1$, where the error rate varies in a wide range, and (6) the point with the minimal error rate, iteration times = 2 and $\alpha = 0.9$. Besides, (7-8) points in the sufficient corpus with one million and six million messages. We perform sentiment classification based on the corresponding lexicons. The relationship between the accuracy of classification and the error rate of lexicons is shown in Fig. 3. As we can see from the figure, the accuracy and E have a significant negative correlation.

5.3 Dataset and Metrics for Sentiment Classification

The sampled data from the Conference on Natural Language Processing & Chinese Computing (NLP&CC) 2013 evaluation task two is used to evaluate

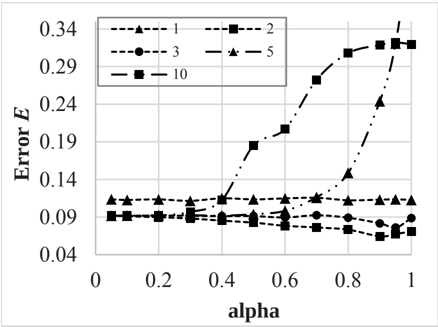


Fig. 2. Error rate of label propagation algorithm to the # of iterations and α

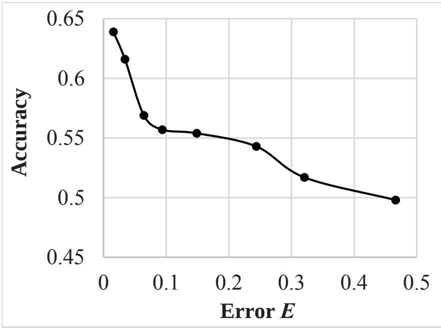


Fig. 3. The accuracy of classification to the error rate E of sentiment lexicons

the sentiment classification results. This task is a moods classification including anger, disgust, fear, happiness, like, sadness and surprise. To turn it into a three-class classification problem, messages of happiness or like are selected as the positive dataset, of disgust, anger or sadness as the negative set, and those with no labels as the neutral set. We sample the same number of messages in these three classes, and eventually get 968 messages in each class as the evaluation data. We use accuracy as the evaluation metric.That is, the number of correctly classified messages divided by the number of all messages.

5.4 Experiments on Sentiment Classification

To evaluate the effects of different groups of features, we divided all the 26 features into seven groups: *microblog structure features* (1–5), *sentence structure features* (6–10), *basic lexicon features* (11–18), *MDA features* (19–22), and *emoticon features* (23–26), and some overlapped groups: *maximum value features* (13, 14, 17, 18, 21, 22, 25, 26) and *product features* (11, 12, 15, 16, 19, 20, 23, 24). Table 1 shows the accuracy of the three-class classification with all features, as well as features with each group removed, in a 5-fold cross-validation.

Table 1 shows that among the three classification strategies, the direct one-stage three-class has the best performance. At the level of features, except for the *microblog structure features* in the third strategy that has a lower accuracy by 0.2%, all other feature groups contribute to the accuracy improvement in all the three strategies. Meanwhile, the *basic lexicon features* have the strongest influence on the performances; the *maximum value features* contribute much more than the *product features*, which indicates words with the highest sentiment strength affect the sentiment of messages more.

We also compared our method with two other methods. Method I had the same features, but the lexicon was replaced with human constructed one in [16]. Method II was that proposed in [10]. For method II, as the assumption about neutral corpora did not hold in Chinese microblog, we only computed the

Table 1. The influence of feature groups on the accuracy of sentiment classification

Feature group	One-stage three-class	Two-stage two-class	
		hierarchical	parallel
All features	**63.9%**	**60.1%**	61.0%
–microblog structure	62.3% (-1.6%)	58.2% (-1.9%)	**61.2% (+0.2%)**
–sentence structure	60.8% (-3.1%)	54.3% (-5.8%)	57.3% (-3.7%)
–emoticon	61.4% (-2.5%)	58.0% (-2.1%)	60.5% (-0.5%)
–basic lexicon	59.1% (-4.8%)	54.1% (-6.0%)	48.7% (-12.3%)
–MDA	63.7% (-0.2%)	59.7% (-0.4%)	60.7% (-0.3%)
–maximum value	61.3% (-2.6%)	57.8% (-2.3%)	57.4% (-3.6%)
–product	63.6% (-0.3%)	59.1% (-1.0%)	58.2% (-2.8%)

likelihood of positive and negative class, s and t, then classified the messages into neutral class when $\alpha < s - t < \beta$. We adjusted α and β to get the best accuracy in the evaluation data mentioned above. The accuracies of these methods are shown in table 2. We can see that both our lexicon and features performed very well in the evaluation.

Table 2. Comparison among the methods

Our method	Method I	Method II
63.9%	54.5%	56.1%

6 Conclusions and Future Work

In this paper, we present new methods for sentiment analysis on unlabeled Chinese microblog messages. We make use of the emoticons in Chinese microblogs and construct sentiment lexicons with the frequency statistics algorithm and the label propagation algorithm. Within different corpus, OOV words, commonly used phrases and negation words are handled at the same time. Based on the lexicons from sufficient or insufficient corpora, we extract lexicon features from microblog messages, along with structural features. These features are microblog-specific and are different from other domains. Then we perform sentiment classifications with an SVM classifier. Experimental results show that the lexicon features and other features are effective.

The main contributions of us are: First, we propose sentiment information labeling with selected emoticon evidence for sentiment lexicon construction without manual annotation, and many commonly used OOV words and phrases can be discovered and introduced into our lexicon. Second, every word has a potential sentiment, which is reflected by its positive and negative scores in the lexicon, and it is finer-grained than those only contain strong-sentiment words and those discretized to few levels. Third, we introduce different types of features which represent multi-views of microblog's characteristics.

Although our method does not necessarily need large amounts of neutral corpora, they may do help in improving the performance. In the future, we will take some effort to get neutral corpora for lexicon construction, thus adding a neutral score to each word. Some methods, such as taking the neutral outputs of current classifer can be attempted.

References

1. Barbosa, L., Feng, J.: Robust sentiment detection on twitter from biased and noisy data. In: Coling 2010: Posters, Beijing, China, pp. 36–44 (2010)
2. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Arxiv preprint arXiv:1010.3003 (2010)
3. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2(3), 27:1–27:27 (2011)
4. Cui, A., Zhang, M., Liu, Y., Ma, S.: Emotion tokens: bridging the gap among multilingual twitter sentiment analysis. In: Proceedings of the 7th Asia Conference on Information Retrieval Technology, AIRS 2011, pp. 238–249 (2011)
5. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the 10th ACM SIGKDD Conference, pp. 168–177. ACM (2004)
6. Jansen, B.J., Zhang, M., Sobel, K., Chowdury, A.: Micro-blogging as online word of mouth branding. In: CHI 2009, pp. 3859–3864 (2009)
7. Kim, S.M., Hovy, E.: Determining the sentiment of opinions. In: Proceedings of the 20th COLING Conference, p. 1367. ACL (2004)
8. Liu, K.L., Li, W.J., Guo, M.: Emoticon smoothed language models for twitter sentiment analysis. In: 26th AAAI Conference on Artificial Intelligence (2012)
9. Neviarouskaya, A., Prendinger, H., Ishizuka, M.: Sentiful: A lexicon for sentiment analysis. IEEE Transactions on Affective Computing 2(1), 22–36 (2011)
10. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: Proceedings of LREC, vol. 2010 (2010)
11. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2008)
12. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1556–1560 (2008)
13. Tumasjan, A., Sprenger, T.O., Sandner, P.G., Welpe, I.M.: Predicting elections with twitter: What 140 characters reveal about political sentiment. In: Proceedings of the 4th AAAI Conference on Weblogs and Social Media, pp. 178–185 (2010)
14. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: ACL 2002, pp. 417–424 (2002)
15. Wang, H., Liu, C., Zheng, Y., Liu, J., Qu, P., Zou, C., Xu, R., Cheung, R.: Sentiment analysis of negative sentences and comparative sentences. In: The Fourth Chinese Opinion Analysis Evaluation, pp. 52–67 (2012)
16. Xu, L., Lin, H., Pan, Y., Ren, H., Chen, J.: Constructing the affective lexicon ontology. Journal of the China Society for Scientific and Technical Information 27(2), 180–185 (2008)
17. Zhang, H.: Nlpir chinese word segmentation system, `http://ictclas.nlpir.org/`
18. Zhang, W., Liu, J., Guo, X.: Positive and Negative Words Dictionary for Students. Encyclopedia of China Publishing House (2004)