Research on Micro-blog Sentiment Analysis

Xinpo Lou, Yumei Chai, Hongying Zan, Ruicong Xu, Yingjie Han, and Kunli Zhang

School of Information Engineering, Zhengzhou University, Zhengzhou, China
pdslouxinpo@163.com,
{ieymchai,iehyzan,ieyjhan,ieklzhang}@zzu.edu.cn

Abstract. Micro-Blog is a kind of important media on the Internet, which conveys the users' point of view in simple and convenient ways. The research related to micro-blog has got extensive attention from the academic and industrial areas. This paper aims to study on the analysis of Chinese micro-blog emotion, proposing a template-based algorithm for the automatic discovery of micro-blog emotional neologisms. It uses a combination method of dictionaries and rules on Chinese micro-blog sentiment analysis.

Keywords: Sentiment Analysis, Polar Word, Parsing, Micro-Blog.

1 Introduction

Micro-blog, short for micro-blog, is a social medium platform allowing its users to share, disseminate and retrieve information. Through WEB, WAP and web-client components, users can form their private networking groups. Within the 140-character limit, consumers can update and share messages with others instantly. Meanwhile the large scale micro-blog text brings natural language processing new opportunities and challenges. Among the masses of twitter messages, the emotional ones are in the majority, which contain very precious opinion resources. Comments on products are very valuable for both sellers and buyers; and the ones on hot issues of society, are also very important for governments to learn what netizens think about the specific events.

Sentiment analysis, also known as opinion mining, aims at analyzing the subjective and appraisive text of subjectivity, mining opinions and comments, for the purpose of presenting them to readers in an intuitive way. The messages containing personal feelings in twitter are rather rich, and sentiment analysis on them has already aroused attention of numerous scholars both at home and abroad. There have already existed many micro-blog sentiment analysis systems intended for information written in English, such as TweetFeel, Twendz and Tweeter Sentiment.

In sentiment analysis, two kinds of technologies have been adopted .One method is based on machine learning, which mainly regards sentiment word and theme relational feature as classification feature, marks training sets and testing sets and conducts, sentiment analysis with it. Naïve Bayes[1,2], Support vector machine[2,3], and Maximum Entropy[4] are the common classification methods. The methods above require manually annotated corpus, which are laborious, and inconsistency. Pang et al[7]

P. Liu and Q. Su (Eds.): CLSW 2013, LNAI 8229, pp. 466–479, 2013.

firstly applied this method to chapter-level sentiment analysis tasks, and pointed out that it has a distinct advantage in sentiment analysis, compared to unsupervised learning. Xie et al[8] put forward Multiple strategies for Chinese micro-blog sentiment analysis method, based on hierarchical structure. Another method combines emotion dictionary with regulations. Turney et al[6] suggest that through the emotional words and the templates, the sentiment analysis can go on with comments on cellphones, banks, movies, and tour destinations. Go A et al[9] prefer to classify micro-blog on the Twitter by making use of distance-supervision. And Park A et al[10] propose a suggestion of collecting and marking the English micro-blog text automatically to analyze the emotion and mining comments. The method of combining emotional words with rules, with high accuracy, is simple and can be achieved easily, but it can not recognize new sentiment words online.

This paper mainly discusses Chinese micro-blog's sentiment analysis. Micro-blog corpus includes numerous Internet words traditional sentiment dictionaries lack. Thus a basic sentiment-word dictionary is constructed by scanning micro-blog artificially. Meanwhile, some templates are summed up according to the features of Chinese micro-blog. Through these templates, new internet words can be found out automatically and based on the features of Chinese micro-blog, some rules can be drawn. This paper realizes opinion sentence identification and sentiment analysis, combining the network sentiment dictionary with rules. It has a good performance in the NLP&&CC2012 Assessment organized by China Computer Federation (CCF).

2 The New Discovery of Micro-blog Sentiment Words

With the help of scanning micro-blog corpus manually, the basic Internet sentiment word dictionary is constructed. Network information's quite real-time, so a tremendous amount of information pours in every day, including some new-born network vocabulary. Based on the new words extracted manually, we summarize some templates, according to where these words appear frequently, and recognize the new words automatically by means of templates. The corpus referred to includes data of Sina micro-blog, as well as sample data from Tencent micro-blog offered by CFF, and all the themes are connected with social news. There are 13500 messages in the corpus.

2.1 Template's Extraction

Go on with the syntactic analysis of the LTP[5] developed by Harbin Institute of Technology Computing and Information Retrieval Centre.

Eg. 我们对这个员工的所作所为感到遗憾。(*We feel sorry for what the worker has done.*)

Result after syntactic analysis as Figure 1:

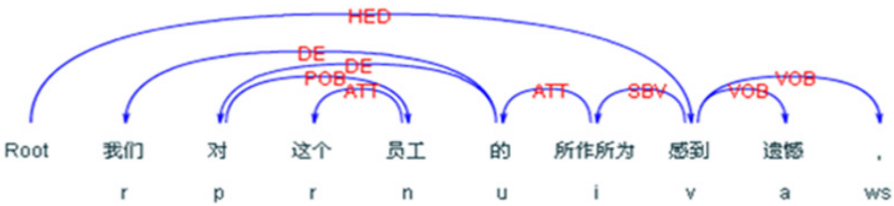

Fig. 1.

Template Extraction algorithm description:
Input：Micro-blog Text Sets D after word segmentation
Output: Template Set M

```
M = φ
for each text d ∈D do
    for i =1..L do    //L is the count of the words in d
        if W_i is sentiment word then    //W_i is the sentiment word
            M_i← W_i part-of-speech and the dependence relationship
            Count_i←Count_i + 1
        end if
    end for
end for
calculate frequency(M_i)
return frequency(M_i) template(>threshold)
```

Seven templates are chosen manually, in order to extract network sentiment words. For example, in VOB, if the object is an adjective, then it is chosen as an alternative sentiment word. As the example, 感到[*gandao*](*feel*)_遗憾[*yihan*](*regret*) (VOB) have an interdependent relationship, we regard “遗憾”[*yihan*](*regret*) as the spare sentiment word.

According to the interdependent relationship, we select templates which have the highest probability of interdependent relationship. Frequency(M_i) represents the template’s assessment, and it is defined as follows:

$$frequency(M_i) = \frac{the\ number\ of\ emotional\ words\ from\ M_i}{the\ number\ of\ words\ from\ M_i}$$

The threshold value in this paper is 0.5, the extraction template is shown in Table 1.For each template T_r, there is a sentence set C_r.

2.2 Finding the Candidate Emotional Words

2.2.1 Selecting the Candidate Emotional Words

The sentiment terms can be found by following the steps: split the Chinese micro-blog into single sentences following punctuation clause(which are ended with‘。’, ‘!’,‘？’), and that’s to say each sentence is an independent row. Then do words segmentation, tagging, and syntactic analysis. At last, place the processed results into an array. Travel the array from beginning to end, select emotional words according to templates, and put these words into the pending list.

The algorithm for the discovery of sentiment words:

```
Input: Template sets M, Micro-blog text sets D
Output: Candidate sentiment word lists CPList, template extraction sentence sets SentList
CPList = φ
for each text d ∈D do
   for each sentence s in text d do
      preprocess for s
      for each template m ∈M do
         if s satisfy m then
            CPList← CPList ∪ { Sentiment words extracted from s in accordance with the template
                      m }
            SentList← SentList ∪ {s}
         end if
      do for
   end for
end for
return CPList
```

Table 1. Candidate sentiment word discovery template

Template	Description	Examples
DA（Degree adverb + Adjective）	Extract the adjective following after adverb of degree.	惠普和戴尔必须扼制新联想，只不过他们显得过于急躁。(*HP and Dell must hold back New Lenovo, but they just seem too impatient.*) 她非常漂亮。(*She is very pretty.*)
VOB（Verb-object）	If the objective is adjective, then set this adjective as the candidate sentiment words. In structure VOB， if the subject is a noun，containing structure DE， then extract the word before “的” as the candidate sentiment words.	当然,新联想也不容盲目乐观。(*Of course, New Lenovo allows no blind optimism.*) 她是一个热心肠的人。(*She is very warm-hearted.*)
ADV (Adverbial)	① In structure ADV， if there is “地/u”，then extract the word before“地/u”as the candidate sentiment words. ② In structure ADV, if there is no“地/u”，then select the adjective in the structure as the candidate sentiment words。	讽刺地说(*ironically speaking*) 尽管一系列冲突事件对戴尔在中国的品牌形象造成了极其恶劣的影响。(*Even though a series of conflicts caused very bad effects on Dell’s brand image in China.*)

Table 1. (*continued*)

DEI（Dei）	In template DEI，extract the words after “得”as the candidate sentiment word.	讲得好(*good point*) 比如最喜欢做比较和攻击性广告也做得最成功的宝洁。(*For example, P&G, whose favorite is making comparative and aggressive ads and who is also the most successful.*)
SBV（Subject-verb）	In structure SBV，if the predicate is an adjective，then set this adjective as the candidate sentiment word.	成就显著(*remarkable achievements*) 戴尔的直销模式在一定意义上突破了销售网络不足的缺陷。*(In a way, the direct sales model of Dell makes up for the lack of selling network.*)
VV（Verb-verb）	Set the verb as the candidate sentiment word.	戴尔表示会尊重竞争对手。(*Dell pledged to respect their rivals.*) 一方必将赢得最终的胜利。(*One party will certainly win the final victory.*)
COO（Coordinate）	In structure of coordination with comma, if there is one word in the seed lists, others are also sentiment words.	苹果手机和平板电脑漂亮、大方、实用。(*Iphones and ipads are beautiful, powerful and practical.*) 只要跨过国际化的障碍，成功、胜利、机遇就属于中国企业。(*Success, victory and opportunity will belong to Chinese enterprises as long as they cross the barrier of internationalization.*)

2.3 The Candidate Emotional Word Set Denoising

In recent years, many statistical computer learning models have been widely used in natural language processing, including: Hidden Markov Model, Support Vector Machine, Maximum Entropy, Maximum Entropy Markov Model and Conditional Random Fields. In this paper, we adopt the conditional random field statistical model to denoise the candidate emotional word set. This method can make the best of the context as well as add some external features.

The experiment chooses C_r corresponding to T_r as the training corpus. Testing corpus comes from sentences based upon templates. Statistical methods regard speech and part-of-speech as features. We can compare the results of window training experiment with different features in order to detect the size of the best feature window of every sentiment term.

2.4 The Polarity Identification of Emotional Words

The emotional words' polarity identification requires the conjunction list, the impact factor list and the basic Network sentiment word list. Words we mainly used are coordinating conjunctions and adversative conjunctions. The conjunctive list (see Table 2) is mainly used for the polarity identification of emotional words, and if it is a conjunction, the sentiment terms around them have the same polarity. If it is an adversative conjunction, the situation is opposite. In an article, some words don't have own sentiment, but affect greatly the polarity identification, and to a certain degree, they can strengthen, weaken and deny the part-of speech, such as some degree adverbs, negative adverbs and so on, which are called impact factors. The tags “a, b, c, d, n” in the impact factor list(see Table 3) respectively represent “strengthen positively

,weaken positively, weaken negatively and deny", and a represents the sentiment term. For example, "那些政策让罪恶和腐败无处遁形"(*These policies make sins and corruption have no place to hide.*),and this sentence contains coordinating conjunctions '和'[*he*](*and*) and the word '罪恶'[*zuie*](*sins*) has a negative direction. The word '腐败'[*fubai*](*corruption*) is detected through regulations in table 1, and then we can get the preliminary judgment that the word '腐败' is negative. Check whether there are impact factors around the word '腐败', if it does exist, we should change its polarity according to the tagging of the impact factor.

The judgment of emotional words' polarity can be done as the following steps: scan the whole candidate sentiment word list, and delete the terms that are not 'n, v, d, a, c, i, l' from the array. Thus, select terms which are more likely to be emotional words according to polarity identification. Then put the positive ones and negative ones respectively into their corresponding forms after finishing the emotional words polarity identification according to the conjunction structure and the basic emotional word list.

Table 2. Conjunction Examples

Conjunction	Examples
与[yu](and)	联想作为跨国企业的成熟与理智、冲劲和开拓(*As an international enterprise, Lenovo's maturity, reasonability, pushfulness and pioneering spirit.*)
和[he](and)	保持了理智和谨慎(*keep reasonable and cautious.*)
但是[danshi](but)	这台笔记本外观很漂亮，但是散热很糟糕。(*This lap-top is very good-looking, but its thermal dissipation is terrible.*)
然而[raner](however)	她外面很美，然而，内心却很丑陋。(*She has good appearance and ugly heart.*)

Table 3. IF(Impact Factor)Examples

IF	Examples
A/n 透顶/z a/n	糟糕透顶[*zaogaotouding*](*awful*)
A/n 轻微/a b/n	危害轻微[*weihaiqingwei*](*slight harmness*)
A/n 不大/d c/n	风险不大[*fengxianbuda*](*small risks*)
很/d 难/a A/n d/n	很难实现[*hennanshixian*](*hard to achieve*)
不/d A/n n/n	不稳定[*buwending*](*unstable*)

Algorithm for the judgment of sentiment words' polarity:

Input: Candidate sentiment word lists CPList, Conjunction word lists CList, Weibo text sets D
Output: Positive sentiment word list PList, Negative sentiment word list NList

```
for each text d ∈D do
   Extract sentences containing conjunctions to make the data sets M
end for
repeat
for each candidate sentiment word w ∈CPList do
     for each data m ∈M do
        judge the part-of-speech of w is not n、v、d、a、c、i、l， otherwise delete it
        find sentences with w， judge the polarity based on the conjunction structure
        if w fulfils the rules of negative factor， then negate the polarity
     end for
if w is positive   then w is in PList
else w is in NList
end for
until   W =φ
```

3 Analysis of Micro-blog Sentiment Tendency

The process of sentiment analysis is divided into two parts: First, determine whether each sentence in Micro-blog is subjective or not; Second, estimate the sentiment tendency of perspective sentences: positive, negative, or neutral.

3.1 Opinion Sentences' Recognition

In the micro-blog, sentences with emotion words or symbols can express obvious feelings better, so we call them opinion sentences. As for opinion sentence recognition, without the corpus training, sentiment words play an important role in analyzing the emotion's polarity. It identify whether a sentence is an opinion sentence with the help of a sentiment word is an available method, and considering that the traditional sentiment dictionary doesn't perform well on micro-blog, we create a basic Network sentiment word dictionary in this paper, enriching the vocabulary automatically via the algorithm in Section 2. In traditional sentiment analysis, emotion words are classified into positive and negative words, and the corresponding dictionary has been created. For example,'善良'[*shanliang*](*kind-hearted*) is a positive word, while'丑陋'[*choulou*](*ugly*) is a negative word. Some regulations have been added to the dictionary, such as "改进了+名词"[*gaijinle* + *mingci*](*improved* + *noun*),"提高了+名词" [*tigaole* + *mingci*](*improved* + *noun*), " 优化了+名词" [*youhuale* + *mingci*](*optimized* + *noun*) and so on. The nouns at the back are the evaluation objects. For example ,in the micro-blog,"新的苹果手机，改进了摄像头，改进了电池，改进了触控，还从3G变成了4G，比以前更nb。"(*The new iphones have improved on camera, battery and touch device and upgraded from 3G to 4G, which are even more freaking* *awesome than ever.*) It satisfies the rule “改进了+名词”, and its evaluation objects are “摄像头[*shexiangtou*](*camera*)、电池[*dianchi*](*battery*)、触控[*chukong*](*touch*)”.

There are emoticons together with characters in the micro-blog. Take Sina for an example, its micro-blog provides some default emotions, and emotions’ manifestation in the text is “/” followed by the text, like“/哈哈”[*haha*](*Aha*). One message can contain several emoticons, and here positive and negative emotions’ classification is also based on them.

3.2 The Orientation Recognition of Opinion Sentences

Analyzing the orientation of sentiment words in sentences is a common practice of judging whether the micro-blog’s sentiment orientation is positive, negative or neutral. If there is a single sentiment word in the micro-blog, and it’s the word that decides the orientation of the post; if there are several sentiment words, then calculate the total number of positive words and negative words respectively, and that the number is equal indicates that it’s positive, and if not, it depends on whose total is larger.

After analysis, we find that the negative adverb often leads to the polarity reversal of sentiment fragments. Therefore, the system creates a negative factor table including 64 negative adverbs to solve the problem. Separate analysis for the two more complicated negative questions is given as follows:

1) Double negation: We stipulate that it can be called double negation only if there are two negatives in the same fragment(with the border of ，？！) and there isn’t any sentiment word between two negatives.

2) The location of negative words: in this system, the sentiment word’s location has been designated in the negative factor table. For example, in“不必/d An/n” An represents the location of sentiment words, and the sentiment word is after the negative word.

Micro-blog tendency judgment algorithm based on sentiment word dictionary and rules:

Input：Micro-blog sets T = {t_1,t_2,t_3,……,t_n},Sentiment word dictionary Dict, Emotion symbol sets Emotions；

Output: Sentiment tendency set List from Micro-blog sets T

```
k=0
List = φ
for each Micro-blog set t∈T do
    k=k+1
    score = 0
    for Micro-blog emotion icon e in t_k do
        if(e ∈ Emotions)   then score = score + δ (e);
    end for
    for i = 1…L do     //L is the word counts in tk
        if(w_i ∈ Dict)  then
```

```
if(w_i satisfies the rule of negative factor)  then  score = score - f(w_i)
else score = score + f(w_i)
    end for
    if(score > 0)  then List[k]='POS'
    else if(score == 0) then List[k]='OTHER'
    else List[k]='NEG'
end for
return List
```

Attention：

$$\delta(\mathrm{e})=\begin{cases}1 & e\in Positive\,Emotion\\ -1 & e\in Negative\,Emotion\end{cases}\qquad f(w)=\begin{cases}1 & w\in Positive\,\mathrm{Dict}\\ -1 & w\in Negative\,\mathrm{Dict}\end{cases}$$

4 Results and Analysis of the Experiment

4.1 Experimental Set-Ups

The data used in this paper come from test corpus of NLP&CC2012 micro-blog sentiment analysis held by China Computer Association, which includes about 3400 posts, and 20 themes have been given, covering living, education, politics and so on, and tagged with each corresponding result artificially (see Table 4).Use the unified format to save the micro-blog, and a post's storage and definition are as follows:

<weibo id="4">
<sentence id="1">小心我们中国人民解放军发两颗核导弹给你们尝尝。(*Beware that we PLA send you two nuclear missiles.*)</sentence>
<sentence id="2">最后我也会说是“不小心”误发的。(*Eventually, we would also say that launch is a careless mistake.*)</sentence>
<sentence id="3">看你丫的嚣不嚣张！(*Dare you to be arrogant.*)</sentence>
<hashtag id="1">菲军舰恶意撞击(*Philippine warships's malicious clash.*)</hashtag>
</weibo>

In the example, <weibo id="4">indicates the identifier number of Micro-blog is 4，<sentence id="1">the sentence number is 1，<hashtag >indicates the topic of Micro-blog.

This experiment uses Precision， Recall rate and F-measure to estimate the results. The following computational formula is：

$$\Pr ecision=\frac{\#system\_correct}{\#system\_proposed}$$

$$\mathrm{Re}\,call=\frac{\#system\_correct}{\#gold}$$

$$F - measure = \frac{2 \times \Pr ecision \times \operatorname{Re} call}{\Pr ecision + \operatorname{Re} call}$$

#gold is the counts of manually annotated results, #system_correct is the matched results between system and manual work, #system_proposed is the count of the system results。

Table 4. Micro-blog manually annotated results

Files	Positive sentences	Negative sentences	Neutral sentences	Total sentences	Neutral percentage
Life	85	424	263	772	34.07%
Education	239	335	292	866	33.72%
Politics	53	492	326	871	37.43%
Events	64	515	328	907	36.16%
Total	441	1766	1209	3416	35.39%

4.2 Select Emotional Words in Micro-blog

The basic emotional word list is selected by scanning the corpus manually, including 436 words that have positive connotations and 2254 words that have negative connotations. According to the basic emotional word list and syntax analysis, we got the template results (see Table 5).

Table 5. Template Extraction Results

Template	DA	VOB	ADV	DEI	SBV	VV	COO
Precision	0.5796	0.5110	0.5306	0.5357	0.5745	0.5127	0.8788

Extract the sentiment words with the templates, the results are in Table 6 after denoising using statistical methods:

Eventually, we got 2198 emotional words selected from the corpus to be analyzed on their polarity.

We need to preprocess the corpus, and choose all the sentences with conjunction. The following is an example of sentences which contain conjunction '和'[*he*](*and*) :

Table 6. Sentiment word Extraction results

Template	Word counts	Precision
DA	384	0.6710
VOB	694	0.6572
ADV	601	0.6247
DEI	168	0.5258
SBV	394	0.6851
VV	123	0.5983
COO	57	0.8889

The data format for coordinating conjunction ‘和’ with an context window data of 5:

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
政策 n 让 v 所有 v 的 u 罪恶 n 和 p 腐败 a 都 d 无处 d 遁 Vg 形 Ng

The process of judging the emotion polarity is just like a snowballing, obtained by multiple iterations, and that’s to say expansion terms and basic emotional words merge, getting a larger scale emotional word list, and then restart the emotional words polarity identification, to get a new expansion vocabulary. Repeat it several times until it won’t increase any more. For more details (see Table 7)

Table 7. Sentiment word variation

Frequency	Positive words	Negative words
Initial	436	2254
Iteration once	552	3458
Iteration twice	598	3965
Iteration ten times	612	4104
Final(after 17 times)	648	4240

The results for the judgment of undetermined sentiment words’ polarity are shown in Table 8:

Table 8. Results for Polarity Judgement of Sentiment words

Part of speech	Precision
Positive emotional words	0.665
Negative emotional words	0.652
Total	0.653

From the chart above, we can see the accuracy of the results is not so high. After analysis, we find two main reasons. Firstly, the accuracy of the syntax analysis can affect the results; secondly, data selected by templates may contain some noise. The emotional word find algorithm based on templates, to a certain extent, enables emotional resources' automatic construction, and decreases time-consumption and energy-consumption if we select emotional words by scanning the corpus manually. The emotional words dictionary used here is a micro-blog emotional word list which is formed by removing noise words artificially in the results of the emotional words find algorithm that is based on templates.

4.3 Recognition of Opinion Sentences

The recognition of opinion sentences is described as follows: for each sentence in every micro-blog, judge whether the sentence is an opinion sentence or not. The definition doesn't include sentences written to express one's feelings or will. The natural language processing lab of Zhengzhou University (ZZUNLP) took part in NLP&&CC2012, with the result of ZZUNLP. The average result covers each term's average result of 53 results submitted by 34 enterprises who have been evaluated (see Table 9).

Table 9. Perspective sentence recognition results

Evaluation results	Precision	Recall	F measure
ZZUNLP	0.765	0.647	0.701
Average results	0.727	0.615	0.647

4.4 Sentiment Polarity Analysis

Sentiment orientation analysis can be described as follows: judge the sentiment orientation of each opinion sentence in micro-blog. Data set contains each sentence of every micro-blog and the orientation analysis should be based on recognition of opinion sentences. The sentiment orientation covers POS, NEG and OTHER (see Table10). Xie lixing et al, using the statistical model, put forward multi-strategy methods based on the hierarchy structure according to the hierarchy (see Table 11).

Table 10. Sentiment orientation analysis results

Evaluation results	Precision	Recall	F measure
ZZUNLP	0.902	0.584	0.709
Average results	0.745	0.455	0.552

Table 11. Comparison result

Method	Precision
Dictionary and rule-based method	0.690
Hierarchy structure strategy-based method	0.672

Through the comparison between Table 9 and Table 10, we can get the conclusion that both of them have a high accuracy, but a low recall rate. The main reasons are as follows: As for opinion sentences recognition and sentiment orientation analysis, we mainly use the dictionary, which has its limitations; besides, there are also some sentiment sentences without emotional words in micro-blog affecting the experimental results. From the comparison analysis of Table 11, the method based on combining the dictionary with regulations is better than that based on statistics, which indicates that the former method used in the paper does have some effect on Chinese micro-blog sentiment analysis.

5 Conclusion

In recent years, as a new kind of medium, micro-blog has become increasingly important. The sentiment analysis research about micro-blog has just begun, so further corresponding research needs to be done to verify the method of this paper. More attention should be paid include collecting sentiment corpus and tagging in the micro-blog text. This paper adopts the method of combining sentiment words dictionary with regulations, and makes some achievements, but there is still much to improve. For further study, we'll create the network dictionary, taking connections between micro-blog users and those between messages into consideration, to increase the accuracy rate.

Acknowledgements. This work is supported by a grant from the Natural Science Foundation of China (No.60970083, No.61272221), 863 projects of National High Technology Research and Development (No.2012AA011101), the Science and Technology Research Project of the Education Department of Henan Province (No.13B520381, NO.12B520055), and the Open Projects Program of National Laboratory of Pattern Recognition.

References

1. Riloff, E., Wiebe, J., Wilson, T.: Learning Subjective Nouns Using Extraction Pattern Boostrapping. In: CoNLL 2003, pp. 25–32 (2003)
2. Lin, W.H., Wilson, T., Wiebe, J., Hauptmann, A.: Which Side are You on? Identifying Perspectives at the Document and Sentence Level. In: CoNLL 2006, pp. 109–116 (2006)
3. Riloff, E., Patwardhan, S., Wiebe, J.: Feature subsumption for opinion analysis. In: EMNLP 2006 (2006)
4. Pang, B., Lee, L.L.: A sentiment Education: Sentiment Analysis Using Subjectivity Summarization based on Minimum Cuts. In: ACL 2004, Spain, pp. 271–278 (2004)
5. Che, W., Li, Z., Liu, T.: LTP: A Chinese Language Technology Platform. In: Proceedings of the Coling 2010: Demonstrations, Beijing, China, pp. 13–16 (August 2010)
6. Turney, P.: Thumbs up or thumbs down semantic Orientate-on Applied to Unsupervised Classification of reviews. In: ACL 2002, pp. 417–424 (2002)
7. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: ACL 2002, pp. 79–86 (2002)
8. Xie, L., Zhou, M., Sun, M.: Hierarchical structure based Hybrid Approach to Sentiment Analysis of Chinese Micro Blog and Its Feature Extraction. Journal of Chinese Information Processing 26(1), 73–82 (2012)
9. Go, A., Bhayani, R., Lei, H.: Twitter sentiment Classification Using Distant Supervision (2009)
10. Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: Proc. of Language Resources and Evaluation Conference. [s. n.], Lisbon (2010)
11. Lafferty, J., Mccallum, A., Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: Proceedings of the 18th ICML 2001, pp. 282–289 (2001)