Emotion Classification of Chinese Microblog Text via Fusion of BoW and eVector Feature Representations

Chengxin Li, Huimin Wu, and Qin Jin*

School of Information, Renmin University of China, Beijing, China, 100872
{2011202429,whmliu,qjin}@ruc.edu.cn

Abstract. Sentiment Analysis has been a hot research topic in recent years. Emotion classification is more detailed sentiment analysis which cares about more than the polarity of sentiment. In this paper, we present our system of emotion analysis for the Sina Weibo texts on both the document and sentence level, which detects whether a text is sentimental and further decides which emotion classes it conveys. The emotions of focus are seven basic emotion classes: anger, disgust, fear, happiness, like, sadness and surprise. Our baseline system uses supervised machine learning classifier (support vector machine, SVM) based on bag-of-words (BoW) features. In a contrast system, we propose a novel approach to construct an emotion lexicon and to generate a new feature representation of text which is named emotion vector eVector. Our experimental results show that both systems can classify emotion significantly better than random guess. Fusion of both systems obtains additional gain which indicates that they capture certain complementary information.

Keywords: Emotion Classification, Sentiment Analysis, Sentiment lexicon, Text Feature Representation.

1 Introduction

Research of sentiment analysis of microblog texts has shown great research value, owing to its comprehensive applications in many fields, from earlier work by Pang about sentiment analysis of movie reviews [1] to nowadays more and more important applications in other fields such as business decision [2], politic election [3, 4] etc.

Weibo, short for Microblog in Chinese, has several aspects that are different from the traditional long texts such as movie reviews in sentiment analysis. Firstly, it is short with no more than 140 Chinese words. Because its shortness, it has been regarded as a convenient tool to use and to share daily life thus produce a large quantity of data for research. However, shortness also makes it harder for sentiment analysis compared with long texts. Secondly, Chinese is mainly used in Weibo instead of English. Chinese is largely different from English to some degree, like the character or the sentence structure, so the sentiment analysis work done with English microblogs like twitter may not be directly applied to Chinese microblog analysis. Thirdly, multi-

* Corresponding author.

C. Zong et al. (Eds.): NLPCC 2014, CCIS 496, pp. 217–228, 2014.

sentiments in Weibo are more confusable. In formal long texts, which are regulated by conventional article rules, though multi-sentiment might also exist, we can determine the major emotion and the minor emotion by article rules, which is not effective in Weibo texts analysis. Fourthly, words used in Weibo are more casual than in long formal texts. For examples, there are web popular words like “麻麻”, ”跪了” or emotion expressions like (corresponding input “[笑]”) and  (corresponding input “[惊恐]”) in Weibo texts. Web popular words might be shown in traditional characters but with different meanings or emotions. For example, “跪了” refers to an action with no emotion polarity traditionally, but now it equals to a frustrating emotion. Moreover, some of these web popular words have several different meanings and different emotions owing to their informality. Emotion expressions are provided to make the Weibo texts more interesting. They are inputted by input tools like Sogou input, Baidu input, etc. How to effectively deal with free web texts is very important for Weibo sentiment analysis.

Due to its popularity of Weibo usage in China society, sentiment analysis of Weibo texts are becoming more and more important. The conference on Natural Language Processing and Chinese Computing (NLPCC) 2014 holds several evaluation tasks in natural language processing and Chinese computing. “Emotion Analysis in Chinese Weibo Texts” is one of the evaluation tasks. This paper presents our work of emotion classification on Weibo texts.

The rest of this paper is organized as follows. Section 2 introduces some related work. Section 3 describes the two systems we build and section 4 presents the experimental results. Section 5 presents some conclusions and future work.

2 Related Work

We can broadly summarize the previous research in sentiment analysis into three categories: analysis based on rules, analysis based on unsupervised classification and analysis based on supervised classification. Of the three categories, the last one performs relatively better.

Rule based analysis is mainly performed together with the emotion lexicon. In English texts analysis, Kamps used synonyms in WordNet lexicon [5]. In Chinese texts analysis, Yanlan Zhu used the HowNet Chinese lexicon [6]. Chun Li also used HowNet but construct another list consisting of the seed words which are further used to get the polarity of words in document [7]. In addition to the emotion lexicon, the researchers also noticed the influence of grammar, like the negative words, the adversatives and degree adverbs. They combine some of these words in several lists determined by their function and add these effects to the analysis process. The performance of this method is largely decided by the size and applicability of the emotion lexicon.

Analysis based on unsupervised classification is proposed by Turney [8] with some templates to extract the adjectives and adverbs as the emotion phrase, which further determine the PMI (Pointwise mutual information) and SO (Semantic orientation). The method deeply depends on the template of the basic words, thus has great limitations.

Analysis based on supervised classification gets the emotion classification model using labeled training data. The trained model is used to predict the emotion category of the test data. It is first proposed by Pang and Lee in 2002 [1]. The baseline algorithm adapted from it usually contains three modules: Tokenization, feature extraction and classification. The classification uses different classifiers like Naïve Bayes [9], Maximum Entropy (MaxEnt) [10] and Support Vector Machine (SVM) [11].

In this paper, we build our baseline system with supervised machine learning classifier SVM based on bag-of-words (BoW) features. In our contrast system, we propose a new method to construct an emotion lexicon and then to generate a new feature representation based on the emotion lexicon. The two systems are combined for better emotion classification performance.

3 System Description

We build two systems for emotion classification of Weibo texts, one uses supervised learning approach based on traditional bag-of-words feature representation and the other also uses supervised learning approach but based on a new feature representation via emotion lexicon. We fuse two systems via late fusion on the classification score.

3.1 Baseline System Based on Bag-of-Word Feature Representation

The baseline system uses supervised learning approach support vector machine (SVM) based on bag-of-words (BoW) feature representation. There are two main phases in the emotion analysis process. The first phase is to detect whether a Weibo document is sentimental/emotional. The second phase is to classify the document into its proper emotion category if it is sentimental. Both the two phases consist of tokenization, feature extraction and supervised classification steps.

Tokenization: We use all words in a document for analysis, not only adjectives [12], but also verbs, adverbs, nouns, etc. We use Jieba [13], a Chinese text segmentation module built for python programming, for word segmentation.

Feature Representation: BoW model is widely used in text processing applications. It processes texts without considering the word order, the semantic structure or the grammar. The vector representation of BoW is a normally used feature representation for text document. The vocabulary is commonly selected using TF-IDF theory.

In our experiment, the BoW feature representation can take different vector values. The first kind of vector representation consists of only 0 or 1 value for each dimension, where 1 stands for the occurrence of a vocabulary word and 0 stands for non-occurrence. The second kind of vector representation consists of a real number for each dimension, where each real number stands for the frequency of one vocabulary word in a document.

Besides the top frequent words selected based on TF-IDF, we also consider the emotion expressions like "[笑]", "[惊恐]" in the weibo texts and use them as another vocabulary for feature representation. Emotion expressions are provided to make the weibo texts more interesting. These emotion expressions form a new vocabulary list from which we can create a new BoW feature representation. Emotion expression can be a very useful cue for sentiment analysis. We observe on the training data that 98% of the documents with emotion expressions are labeled with certain emotion classes. Moreover, repeat usage of punctuation like exclamatory mark or question mark can also be efficient cues for emotion detection. In our baseline system, we combine the word vocabulary selected based on TF-IDF, the vocabulary of emotion expressions and the vocabulary of punctuation marks to create the BoW feature representation. We assign different weights when combining the words vocabulary and the emotion expression plus punctuation marks vocabulary. The weights are tuned on held out development data.

In experiments shown in this paper, for emotion detection, we select top 500 most frequent words based on TF-IDF from both the documents with none emotion and the documents with emotions respectively. That leads to a vocabulary with 1000 words. After deleting repeated words, we get 610 words in the vocabulary. For emotion classification, we get the top 100 most frequent words from each of the seven emotion classes and this leads to a vocabulary of 700 words. After deleting repeated words, we get a word vocabulary of 560 words. This word vocabulary is combined with the emotion expression vocabulary containing 469 expressions plus the punctuation vocabulary with 7 punctuation marks for generating the BoW features.

Classifier Trained with Supervised Approach: We use Support Vector Machine (SVM) as our classifier. In our experiments, we use the LIBSVM Toolkit [14]. We tune related parameters through cross validation.

3.2 Contrast System Based on Emotion Vector Feature Representation

We construct a new emotion lexicon in this contrast system. We observe that in Weibo texts, we can roughly categorize different words into three types. Taking the sentences with "anger" emotion as example, the first type of words are "emotional words" like "怒" (angry) and "高兴"（happy） which are typical frequent words for expressing certain emotion. The second type of words are "common words" like "真的" (really) which commonly appear in documents but do not usually contain emotion inclination. The third type of words are "Not Emotional and Uncommon words" like "资本家" (capitalist). Based on our intuition, we expect the three types of words may have the following distribution as shown in Table 1, where *n_i* refers to the occurrence number of a word in documents of certain emotion class (for example, the emotion "anger"), *n_o* refers to the occurrence number of this word appears in other emotions (for example, emotions except "anger"), *n_l* refers to the number of emotion classes that this word appears in (for example, if this word only appears in documents of "anger" class, then n_l is 1. If it appears in documents of "like" and "happiness" classes, then n_l is 2).

We then use the following formula to compute the weight of every word and rank them in descending order:

$$weight = \frac{n\_i}{(n\_o * n\_l + 1)} \tag{1}$$

We expect "Emotional words" should be ranked in the top, "Common words" should be ranked in the bottom, and "not Emotional but Uncommon words" should be ranked in the middle. The result proves that our intuition is relatively correct. Some examples for the emotion class "anger" are shown in Table 2.

Table 1. Expected distribution pattern of three types of words

Type	*n_i*	*n_o*	*n_l*
Emotional words	more	less	Less
Common words	fair	more	More
Not emotional but uncommon words	less	less	Less

Table 2. Word examples in ranked list for anger class

Anger	*Word (translation)*	*weight*
Examples in the top part of the ranked list	恨死 (hate)	12.3
	气死我了(piss me off)	8.0
	MB (fuck)	7.0
	贱人 (bitch)	5.0
	这蛋 (bullshit)	5.0
Examples in the middle part of the ranked list	心肝儿 (darling)	0.5
	掩护 (cover)	0.5
	私车 (personal car)	0.5
	扭转 (reverse)	0.5
	秒钟 (clock)	0.5
Examples in the bottom part of the ranked list	挺 (very)	0.0031
	每 (every)	0.0029
	现场 (on site)	0.0029
	滴 (a drop)	0.0029
	害羞 (shy)	0.0029

In creating the emotion lexicon using above method, we don't include emotion expressions. After building the emotion lexicon, we use emotion expressions for generating feature representation like BoW. Inspired by the work in [15], which represents emotion words by a vector and every dimension of the vector represents a kind of emotion, if a word relates to the emotion to some extent, the corresponding dimension is 1, otherwise is 0. Similarly, we represent each text by an emotion vector (eVector) composed of 7 dimensions instead of hundreds of dimensions. The vector is in the format as follows:

$$eVector = (d_1, d_2, d_3, d_4, d_5, d_6, d_7) \tag{2}$$

where the seven dimensions correspond to the seven emotions of anger, disgust, fear, happiness, like, sadness and surprise respectively. The value of d_i is sum of all the words' weights (computed as in formula (1)) according to the emotion lexicon for each emotion class i (for example, anger is in 1st emotion class, disgust is in 2nd emotion class, etc).

4 Experiment and Analysis

4.1 Data Description

The data in this paper is collected from Sina Weibo (a popular Chinese Microblog site). The text of Microblog is labeled "none" if it does not convey any emotion. If the text conveys emotion, it is labeled with emotion categories from anger, disgust, fear, happiness, like, sadness, or surprise. The text is labeled with major emotion and minor emotion. Every Microblog document includes at least one sentence. Every sentence in a Microblog document is also labeled with corresponding emotions (major and minor). It is not necessary that every sentence conveys emotion in an emotional document. Therefore, it is possible that sentences in an emotional document can have "none" labels. The number of documents and the number of sentences for each emotion category in training data and test data are described in Table 3 and 4. We can see from the tables that the distribution of different emotion classes is not balanced.

We extract 469 emotion expressions from the training data and some examples are shown in Table 5. The value means the number of occurrence of certain emotion expression in documents with different emotion class labels. Admittedly, there is some informal usage, for example [抓狂] appears in both sentimental and non-sentimental documents. However, because the number of this emotion expression appears in sentimental category far more than in non-sentimental category, we think this emotion expression is still useful in emotion detection, though it may not be so effective in emotion classification.

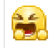
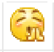
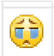
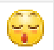
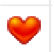
Table 3. Number of documents for each emotion type in training and test data

emotion type	*training data*		*test data*	
	number	percentage	number	percentage
none	6591	47.0%	3603	60.1%
anger	669	4.8%	128	2.1%
disgust	1392	10.0%	389	6.5%
fear	148	1.1%	46	0.8%
happiness	1460	10.4%	441	7.3%
like	2204	15.7%	1042	17.3%
sadness	1174	8.4%	189	3.2%
surprise	362	2.6%	162	2.7%

Table 4. Number of sentences for each emotion type in training and test data

emotion type	*training data*		*test data*	
	number	percentage	number	percentage
none	29731	65.4%	11871	75.1%
anger	1899	4.2%	244	1.6%
disgust	3130	6.9%	679	4.3%
fear	299	0.7%	67	0.4%
happiness	2805	6.2%	641	4.1%
like	4259	9.4%	1630	10.4%
sadness	2478	5.4%	302	1.9%
surprise	820	1.8%	259	1.7%

Table 5. Number of occurrence of emotion expressions across difference emotion classes

Icon	*input*	*none*	*Ang*	*Dis*	*Fea*	*Hap*	*Lik*	*Sad*	*Sur*	*main*
	[抓狂]	5	24	24	5	14	18	37	0	disgust
	[耶]	8	1	0	0	30	13	1	1	happy
	[鼓掌]	6	1	4	0	29	29	2	0	happy
	[委屈]	0	2	0	1	2	5	14	0	sad
	[泪]	14	15	16	6	31	22	156	4	sad
	[爱你]	5	0	3	0	33	42	3	0	happy
	[good]	6	0	1	0	9	24	2	5	like
	[吃惊]	1	2	2	1	1	1	3	14	surprise
	[偷笑]	18	1	8	1	63	34	3	1	happy
	[吐]	1	3	6	0	0	1	2	1	disgust
	[哈哈]	21	0	9	0	121	26	2	3	happy
	[心]	23	1	0	0	37	45	10	1	like
	[酷]	7	0	3	1	15	12	2	4	happy
	[眼泪]	0	0	0	0	1	1	4	0	sad

4.2 Evaluation Metrics

The evaluation metrics used in this paper are precision, recall and F-measure for emotion detection and looseAP and strictAP for emotion classification. In the evaluation, the system is required to produce the emotion classification results for both major emotion and minor emotion. In our experiments, an emotion of an emotional document is decided by the comparison of scores for all seven emotion classes. The top ranked emotion class is the major emotion classification decision and the second top ranked emotion class is the candidate for minor emotion classification decision. If the difference between the major and minor emotion is larger than certain threshold, the minor emotion will be "none" instead of the second top ranked emotion class. The threshold can be tuned with cross validation, which is 0.5 in the experiments in this paper.

looseAP and strictAP are the two metrics used in the evaluation. On loose metric, the system will get a score of 1 if it correctly identify either the top emotion class or the minor emotion class. As for the strict metric, for the case that the ground truth contains both major and minor emotions, a system will get score 1 if both the major and minor emotion classes are matched. It will get a score of 0.666 if the major emotion class is identified and score 0.333 if the minor emotion class is identified. If the major emotion class hypothesis matches the minor emotion class in ground truth, it will get a score of 0.333 as well.

4.3 Experimental Results

As we have described in section 3.1, we use two BoW features in the baseline system: occurrence vs. frequency. The results based on these two types of BoW features are compared in Table 6 on both the document level (D) and sentence level (S). We can see from the results that frequency BoW is not largely different from occurrence BoW in terms of emotion classification performance. We expect the reason is that the document is too short, most words occur in the document only once, so frequency is either 0 or 1, thus the two feature representations are very close to each other.

Table 6. Baseline system performance with Occurrence vs Frequency BoW

	Precision	*Recall*	*F-measure*	*looseAP*	*strictAP*
Occurrence (D)	0.58	0.73	0.65	0.44	0.40
Frequency (D)	0.58	0.74	0.65	0.43	0.39
Occurrence (S)	0.58	0.50	0.54	0.33	0.31
Frequency (S)	0.57	0.52	0.56	0.34	0.32

Table 7. Confusion Matrix of baseline system on document level with occurrence BoW

	none	*anger*	*disgust*	*fear*	*happy*	*like*	*sad*	*surprise*
none	***2363***	15	176	0	142	796	89	22
anger	38	**38**	16	0	2	28	3	3
disgust	170	16	**69**	0	15	92	19	8
fear	15	1	8	**0**	2	14	5	1
happy	63	0	7	0	**271**	87	9	4
like	274	9	25	0	126	**570**	31	7
sad	54	3	17	0	7	33	**72**	7
surprise	40	5	23	0	10	35	3	**46**

Table 7 shows the confusion matrix of the baseline system on the document level with occurrence BoW (column is the ground truth and row is the system decision). From table 7, we can see that “none” is on average the most confusable class to all emotion classes, which indicates that our emotion detection step at the first place should be improved. We can also see that for some emotion classes, their confusable emotion classes are intuitively related classes, for example, “like” and “happy” are confusable. “surprise” is confusable with “like” or “disgust” depending on it is a good or bad surprise.

Table 8 presents the emotion classification results of the baseline system (based on occurrence BoW), the contrast system (based on eVector feature representation) and the fused system. The results show that fusion of both systems improves the emotion classification performance on both the document level and sentence level. As shown in previous section in Table 3 and 4, the distribution of emotion classes is not balanced. In Table 9, we therefore also compute the looseAP and strictAP performance with weights proportional to the number of documents/sentences in a particular emotion class.

Table 8. System performance of emotion classification on both document and sentence level

System	*Document Level*		*Sentence Level*	
	looseAP	*strictAP*	*looseAP*	*strictAP*
Baseline system (BoW)	0.44	0.40	0.33	0.31
Contrast system (eVector)	0.38	0.34	0.28	0.27
Fusion	**0.46**	**0.41**	**0.34**	**0.32**

Table 9. System performance of emotion classification on both document and sentence level with weighted AP computation

System	*Document Level*		*Sentence Level*	
	looseAP	*strictAP*	*looseAP*	*strictAP*
Baseline system (BoW)	0.65	0.59	0.69	0.65
Contrast system (eVector)	0.58	0.51	0.65	0.61
Fusion	**0.66**	**0.60**	**0.72**	**0.68**

We notice that some emotion categories are closely related. In many cases “anger” may express certain level of “disgust”, “like” may express certain level of “happiness”. Therefore, when the system classifies some “anger” as “disgust”, or “like” as “happiness”, we should not simply say it is wrong. However, if “anger” is recognized as “like” or “happiness”, there is no question that it is wrong, because they are totally opposite emotion categories. We therefore also look at the performance if we tolerate “anger” and “disgust” to belong to the same category, “happiness” and “like” to belong to the same category for the baseline system, contrast system, and fused system as shown in Table 10. The performance is obviously improved with the tolerance for all systems and fusion improves performance.

Table 10. System performance of emotion classification on three emotion class

System	*Disgust+Anger*	*Happy+Like*	*Sadness*
Basline system(Bow)	0.22	0.71	0.38
Contrast system(eVector)	0.26	0.69	0.28
Fusion	**0.27**	**0.73**	**0.42**

Table 11. Emotion detection on document level with different weighting over words vs expression plus punctuation vocabulary

Weights (words/expression+)	***correct***	***proposed***	***gold***	***precision***
0.1/0.9	1736	2990	2397	0.580602
0.2/0.4	1754	3010	2397	0.582724
0.3/0.7	1761	3032	2397	0.580805
0.4/0.6	**1762**	**3024**	**2397**	**0.582672**
0.5/0.5	1767	3035	2397	0.582208
0.6/0.4	1751	3019	2397	0.579993
0.7/0.3	1750	3022	2397	0.579087
0.8/0.2	1751	3020	2397	0.579801
0.9/0.1	1729	3002	2397	0.575949
1.0/0.0	***1666***	***3002***	***2397***	***0.554963***

As described in previous section 3.1, we combine the words vocabulary and emotion expression plus punctuation marks vocabulary for baseline BoW feature representation generated with different combination weights. Table 11 compares the emotion detection results with different combination weights. *Gold* refers to the total number of ground truth emotional documents. *Proposed* refers to the total number of system hypothesized emotional documents. *Correct* refers to the total number of correct system hypothesized emotional documents. We can see that weights ratio of 0.4/0.6 achieves best detection result. Please notice that the last row (weights 1.0/0.0) refers to the case that expression and punctuation marks are not used for BoW generation. Its worst performance proves that expression and punctuation are important cues for emotion analysis.

We also combine the baseline system which uses bag-of-word feature representation and the contrast system which uses eVector feature representation. Table 12 shows the different fusion weights for emotion classification on the document level. As the results show that fusion of the two systems with appropriate fusion weights achieves additional gain.

Table 12. Fusion weights of baseline and contrast systems on document level

Weights (eVector/BoW)	*looseAP*	*strictAP*
0.0/1.0	***0.445***	***0.396***
0.1/0.9	0.447	0.397
0.2/0.8	0.455	0.404
0.3/0.7	**0.457**	**0.406**
0.4/0.6	0.450	0.398
0.5/0.5	0.445	0.395
0.6/0.4	0.431	0.382
0.7/0.3	0.423	0.376
0.8/0.2	0.410	0.361
0.9/0.1	0.398	0.351
1.0/0.0	***0.383***	***0.337***

5 Conclusion

Weibo text which is the most popular social media in China has attracted much research interest in recent years. Emotion analysis which not only cares about the polarity of sentiment but also the detailed emotion category is a more challenging task. In this paper, we present our two systems for emotion analysis of Chinese Weibo texts on both the document and sentence level. The baseline system uses SVM as classifier based on bag-of-words features representation. The vocabulary for BoW generation combines words, emotion expression and punctuation marks. Experimental results confirm that emotion expression and punctuation marks are important cues for emotion analysis of Weibo texts. The contrast system proposes a new method to construct an emotion lexicon to generate a new feature representation, the emotion vector eVector. Our experimental results show that both systems can classify emotion significantly better than random guess. Fusion of both systems obtains additional gain, which indicates that they capture certain complementary information. In the future work, we will explore new methods to improve the emotion detection performance, enhance the proposed eVector feature representation by utilizing the emotion expressions as well. We will also investigate different classification approaches.

Acknowledgements. This work is supported by the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (No. 14XNLQ01), the Beijing Natural Science Foundation (No. 4142029).

References

1. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, vol. 271. Association for Computational Linguistics (2004)

2. Li, N., Wu, D.D.: Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision Support Systems 48(2), 354–368 (2010)
3. Tumasjan, A., Sprenger, T.O., Sandner, P.G., et al.: Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In: ICWSM, vol. 10, pp. 178–185 (2010)
4. O'Connor, B., Balasubramanyan, R., Routledge, B.R., et al.: From tweets to polls: Linking text sentiment to public opinion time series. ICWSM 11, 122–129 (2010)
5. Kamps, J., Marx, M.J., Mokken, R.J., et al.: Using wordnet to measure semantic orientations of adjectives (2004)
6. Zhu, Y.L., Min, J., Zhou, Y., et al.: Semantic orientation computing based on HowNet. Journal of Chinese Information Processing 20(1), 14–20 (2006); 朱嫣岚,闵锦,周雅倩,等.: 基于HowNet的词汇语义倾向计算.中文信息学报 20(1), 14–20 (2006)
7. Dun, L., Fuyuan, C., Yuanda, C., et al.: Text Sentiment Classification Based on Phrase Patterns. Computer Science 35(4), 132–134 (2008); 李钝,曹付元,曹元大,等.基于短语模式的文本情感分类研究.计算机科学 35(4), 132–134 (2008)
8. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424. Association for Computational Linguistics (2002)
9. Lewis, D.D.: Naive (Bayes) at forty: The independence assumption in information retrieval. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 4–15. Springer, Heidelberg (1998)
10. Konig, R., Renner, R., Schaffner, C.: The operational meaning of min-and max-entropy. IEEE Transactions on Information Theory 55(9), 4337–4347 (2009)
11. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning 20(3), 273–297 (1995)
12. Benamara, F., Cesarano, C., Picariello, A., et al.: Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone. In: ICWSM (2007)
13. Lin, R.G., Tsai, T.C.: Scalable System for Textual Analysis of Stock Market Prediction. In: The Third International Conference on Data Analytics 2014, pp. 95–99 (2014)
14. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) 2(3), 27 (2011)
15. Wu, Y., Kita, K., Ren, F., Matsumoto, K., Kang, X.: Exploring Emotional Words for Chinese Document Chief Emotion Analysis. In: Proceedings of PACLIC 2011, pp. 597–606 (2011)