Building Word-Emotion Mapping Dictionary for Online News

Yanghui Rao, Xiaojun Quan, Liu Wenyin, Qing Li, and Mingliang Chen

Department of Computer Science, City University of Hong Kong

Abstract. Sentiment analysis of online documents such as news articles, blogs and microblogs has received increasing attention. We propose an efficient method of automatically building the word-emotion mapping dictionary for social emotion detection. In the dictionary, each word is associated with the distribution on a series of human emotions. In addition, three different pruning strategies are proposed to refine the dictionary. Experiment on the real-world data sets has validated the effectiveness and reliability of the method. Compared with other lexicons, the dictionary generated using our approach is more adaptive for personalized data set, language-independent, fine-grained, and volume-unlimited. The generated dictionary has a wide range of applications, including predicting the emotional distribution of news articles and tracking the change of social emotions on certain events over time.

Keywords: Social emotion detection; emotion dictionary; maximum likelihood estimation

1 Introduction

In the traditional society, when we make a decision, opinions and emotions of others have always been important information for reference. Knowing the answer of "What others think and feel" is usually very necessary for general people, marketers, public relations officials, politicians and managers.

Nowadays, everyone can express their opinions and emotions easily through news portals, blogs and microblogs, and they become both the listeners and speakers. Facing the vast amount of data, tasks of automatically detecting public emotions evoked by online documents is emerging recently [1], such as the SemEval task 14. This task is treated as a classification problem according to the polarity (positive, neutral or negative) or multiple emotion categories such as joy, sadness, anger, fear, disgust and surprise. However, due to the limited information in the news titles, annotating news headlines for emotions is a hard task. It is usually intractable to annotate headlines consistently even for human [2]. As a result, we mainly focus on annotating news bodies for emotions, and building word-emotion mapping dictionaries in this paper.

In previous works, emotions are mostly annotated based on the existing emotional lexicons [1] [3], e.g., Subjectivity Wordlist [4], WordNet-Affect [5] and

SentiWordNet [6]. Emotion classification or opinion mining based on these existing lexicons have their limited utility, because 1) the lexicons are mainly for public use in general domains, some resulting classifications of words can appear incorrect, and need to be adjusted to fit the personalized data set. 2) Most of the lexicons are available only for bits of languages, such as English, and the volume of words annotated is restricted, which limits the applicability of these methods. 3) Some of the lexicons label words on coarse-grained dimensions (positivity, negativity and neutrality), which are insufficient to individuate the whole spectrum of emotional concepts [5].

Unlike the above methods, we focus on building emotional dictionary automatically, in which each item is scored along a number of predefined emotions. Then, the emotion distributions of current news article are estimated accurately based on the emotional dictionary. The main contributions are as follows:

- A method of building the word-emotion mapping dictionary is proposed, which is efficient, precise and automatic, no human resource is needed.
- Three kinds of parameter-free pruning algorithms are presented to refine the dictionary, and to improve the performance.
- Compared with the existing emotional lexicons, the emotional dictionary constructed in this paper is more adaptively for personalized data set, language-independent, fine-grained, and can be updated constantly.

Related works are given in Section 2. The problem definition, the method of building the word-emotion mapping dictionary, pruning algorithms and potential applications of the dictionary are presented in Section 3. The experimental data sets, evaluation metrics, results and discussions are illustrated in Section 4. Finally, we draw conclusions and discuss future work in Section 5.

2 Related Work

Most of the previous works focus on constructing the emotional lexicons for reviews, which is different with ours for news articles. The main features of reviews and news articles are as follows:

For the former data set, people usually explicitly express their opinions and emotions in the reviews, which results in the subjective text; while for the latter data set, news editor normally present the events objectively in the news reports, and their opinions and emotions are transmitted implicitly. In other words, the former data set mainly contains subjective sentences, which express some personal feelings, opinions, views, emotions, or beliefs; while the latter data set mainly contains objective sentences, which present some factual information. Besides, for the former data set, as there exist fraudulent reviews or rumors, the emotional dictionary maybe incorrect or biased; while for the latter data set, the news reports are mainly objective and do not trigger the same problem.

Works of sentiment analysis for reviews rose from the year 2001 or so. Das and Chen [7] utilized classification algorithm to extract market emotions from stock message boards, which was further used for decision on whether to buy or sell a stock. However, the performance heavily depended on certain words. For instance, the sentence "It is not a bear market" means a bull market actually, because negation words such as "no", "not" are much more important and serve to reverse meaning. Turney [8] applied an unsupervised learning technique to classify the emotional orientation of users' reviews (such as reviews of movies, travel destinations, automobiles and banks), in which the mutual information differences between each phrase and the words "excellent" and "poor" were calculated firstly. Then, the average emotional orientation of the phrases in the review was used to classify the review as recommended or not recommended.

During this incipient stage of research on sentiment analysis from reviews, some of them focus on using linguistic heuristics or a set of seed words preselected, to classify the emotional orientation of words or phrases [9]. Other works focus on emotional categorization of entire documents, which are based on the construction of discriminate-word dictionaries manually or semi-manually [7]. However, previous experiments shown that the intuition of selecting discriminating words may not always be the best for humans [10]. Besides classifying emotions to positive or negative, predicting the rating scores of reviews has also been done by researchers [11] [12]. As the rating scores are ordinal (e.g., 1-5 stars), the problem is tackled by regression. These previous works of sentiment analysis from reviews are often performed on document, sentence, entity, and feature/aspect level. Emotion classification at both the document and sentence levels is useful, but it cannot find what aspects people liked or disliked. Aspect-based emotional analysis is proposed to tackle such problem, but it is hard to perform on news articles, in which aspects of entity are unknown.

Works of emotion classification for news began from the SemEval tasks in 2007. Chaumartin [1] utilized a linguistic and rule-based approach to tag news headlines for predefined emotions, which includes joy, sadness, anger, fear, disgust and surprise, and for polarity, i.e. positive or negative. The algorithm was based on existing emotional dictionaries, like WordNet-Affect and SentiWordNet. Kolya et al. [3] identified event and emotional expressions at word level from the sentences of TempEval-2010 corpus, in which the emotional expressions are also identified simply based on the sentiment lexicons, e.g., Subjectivity Wordlist, WordNet-Affect and SentiWordNet.

These approaches based on public emotional dictionaries needed extra effort of preprocessing and post-processing on individual words, because some resulting classifications of words can appear incorrect, and need to be adjusted to fit the personalized data set. Katz et al. [2] scored the emotions of each word as the average of the emotions of every news headline, in which that word appears, all non-content words were ignored. However, as the limited words in the news titles, it faced the problem of the small number of words available for the analysis.

In this paper, we mainly focus on annotating news bodies for emotions, and building emotional dictionary automatically. The emotion expressions are fine-grained (such as moving, sympathy, boring, angry and funny), rather than coarse-grained (positive, negative and neutral). The dictionary can be used to classify the emotional distributions of previous unseen news articles.

3 Word-Emotion Mapping Dictionary Construction

In this section, we will firstly define our research problem. Then, we introduce the generation method of the word-emotion mapping dictionary, as well as the pruning algorithms of the generated dictionary. Finally, we discuss the potential applications of the dictionary.

3.1 Problem Definition

The research problem is defined as follows.

Given N training news articles, a word-emotion mapping dictionary is generated. The dictionary is a $W \times E$ matrix, and the (j, k) item in this matrix is the score (probability) of emotion e_k conditioned on word w_j.

For each document $d_i (i = 1, 2, ..., N)$, the news content, the publication date (timestamp), and the distribution of ratings of emotions in the predefined list (see Fig. 1 as an example) are available. From these news contents, a vocabulary is obtained as the source of the word-emotion mapping dictionary. The j-th word in the vocabulary is denoted by $w_j (j = 1, 2, ..., W)$, all the emotions is denoted by $e = (e_1, e_2, ..., e_E)$, the normalization form of ratings of d_i over e is denoted by r_i. $r_i = (r_{i1}, r_{i2}, ..., r_{iE})$, and $|r_i| = 1$.

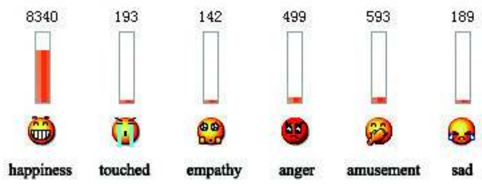

Fig. 1. An example of social emotions and user ratings

3.2 Generation Method

In this section, we introduce the method of generating word-emotion mapping dictionary based on maximum likelihood estimation and the Jensen's inequality.

For each document d_i, the probability of r_i conditioned on d_i can be modeled as:

$$P(r_i|d_i) = \sum_{j=1}^{W} P(w_j|d_i) P(r_i|w_j) \ . \tag{1}$$

Where, the probability of r_i conditioned on w_j is a multinomial distribution, and $P(r_i|w_j) = \prod_{k=1}^{E} P(e_k|w_j)^{r_{ik}}$. Then,

$$P(r_i|d_i) = \sum_{j=1}^{W} P(w_j|d_i) \prod_{k=1}^{E} P(e_k|w_j)^{r_{ik}} \ . \tag{2}$$

In the above, words in document d_i are assumed to be independent.

Let $\sigma_{ij} = P(w_j|d_i)$ and $\theta_{jk} = P(e_k|w_j)$, the log-likelihood over all the N documents can be defined as:

$$\log l = \log(\prod_{i=1}^{N}(\sum_{j=1}^{W}\sigma_{ij}\prod_{k=1}^{E}\theta_{jk}^{r_{ik}})) = \sum_{i=1}^{N}\log(\sum_{j=1}^{W}\sigma_{ij}\prod_{k=1}^{E}\theta_{jk}^{r_{ik}}) \ . \tag{3}$$

According to Jensen's inequality, we reconstruct the log-likelihood as follows:

$$\log l \geq \sum_{i=1}^{N}\sum_{j=1}^{W}\sigma_{ij}\sum_{k=1}^{E} r_{ik}\log\theta_{jk} \ . \tag{4}$$

Since $\sum_{k=1}^{E}\theta_{jk} = 1$, we add a Lagrange multiplier to the log-likelihood equation as follows:

$$\hat{l} = \sum_{i=1}^{N}\sum_{j=1}^{W}\sigma_{ij}\sum_{k=1}^{E} r_{ik}\log\theta_{jk} + \lambda(\sum_{k=1}^{E}\theta_{jk} - 1) \ . \tag{5}$$

Then, we maximize the likelihood by calculating the first-order partial derivative of θ_{jk},

$$\frac{\partial\hat{l}}{\partial\theta_{jk}} = \sum_{i=1}^{N}\frac{\sigma_{ij}r_{ik}}{\theta_{jk}} + \lambda = \frac{\sum_{i=1}^{N}\sigma_{ij}r_{ik}}{\theta_{jk}} + \lambda = 0 \ . \tag{6}$$

Thus,

$$\theta_{jk} = -\frac{\sum_{i=1}^{N}\sigma_{ij}r_{ik}}{\lambda} \ . \tag{7}$$

Since $\sum_{k=1}^{E}\theta_{jk} = 1$, we have

$$\lambda = -\sum_{k=1}^{E}\sum_{i=1}^{N}\sigma_{ij}r_{ik} \ . \tag{8}$$

Then, substitute formula (8) into formula (7) and get

$$\theta_{jk} = \frac{\sum_{i=1}^{N}\sigma_{ij}r_{ik}}{\sum_{k=1}^{E}\sum_{i=1}^{N}\sigma_{ij}r_{ik}} \ . \tag{9}$$

i.e.,

$$P(e_k|w_j) = \frac{\sum_{i=1}^{N}P(w_j|d_i)r_{ik}}{\sum_{k=1}^{E}\sum_{i=1}^{N}P(w_j|d_i)r_{ik}} \ . \tag{10}$$

In the above, $P(e_k|w_j)$ is the probability of emotion e_k conditioned on word w_j from which we can generate the word-emotion mapping dictionary. r_{ik} is the distribution of ratings of document d_i on emotion e_k, $P(w_j|d_i)$ is the probability of word w_j conditioned on document d_i which can be calculated by relative term frequency. The relative term frequency is the number of occurrences of the term w_j in d_i divide by the total number of occurrences of all the terms in d_i.

3.3 Pruning Algorithm

As the size of the training data set increases, the scale of the dictionary extends, making it hard for us to maintain and utilize. Thus, pruning operation is necessary for such lexicons. We will give the definition of *background word* firstly, and then illustrate how it can be used to prune the dictionary.

Definition: Background word is the word that appears in most of the documents in the training data set, it is general for specific domains and topics of the training set, which is quite different with stop words for general domains.

In the context of emotional annotation, the *background words* are general words that contain little emotional information actually and will disturb the effect of utilizing the dictionary. In contrast to other useful emotional tagging words, the probability of a word being to *background words*, which is denoted by $P(B|w)$, is larger than the probability of the word being to emotions, which is denoted by $P(E|w)$. According to the definition, the probability $P(B|w)$ can be represented as follows:

$$P(B|w) = \frac{df_w}{N} \ . \tag{11}$$

In the above, df_w is the document frequency of word w, N is the total number of documents in the training set. The proportion of documents that contains the word w is larger, the probability of w being to *background words* is higher.

As there are multiple emotions tagged for each word according to formula (10), the latter probability $P(E|w)$ has three forms, which are the maximum, average and minimum of all values of $P(e_k|w)$, k is from 1 to E (the total number of types of emotions). Then, the words are pruned from the dictionary if $P(B|w)$ is larger than $P(E|w)$.

When the pruning algorithm above is performed, the word-emotion mapping dictionary is constructed to the end, which can be used to predict the emotions of given news articles as follows:

$$\hat{P}(e|d) = \sum_{w \in W} p(w|d)p(e|w) \ . \tag{12}$$

In the above, $\hat{P}(e|d)$ is the probability of social users having emotions e on document d, $P(w|d)$ is the distribution of new document d on word w, which can be calculated by relative term frequency, $P(e|w)$ is the probability of emotions e conditioned on word w, which can be looked up from the word-emotion mapping dictionary generated with formula (10).

4 Experiments

In this section, experiments are conducted on one Chinese data set and one English data set, so as to test the effect of the word-emotion mapping dictionary on sentiment analysis. The good performances and multilingual data sets reflect the method's effectiveness, reliability, and language-independent of building the dictionary.

4.1 Data Sets

To test the adaptiveness, effectiveness and language-independent of our method of building the word-emotion mapping dictionary, large-scale and multilingual data sets are needed. Two kinds of data sets are employed in the experiment.

Sina. This is a large-scale Chinese data set scrawled from Sina society, which is one of the most popular news sites in China.[1] The attributes include the URL address of the news article, the news headline (title), the publish date (from 29 July, 2005 to 9 Sep, 2011), the news body (content), the user ratings on emotions of touched, empathy, boredom, anger, amusement, sadness, surprise and warmness. The data set contains 32,493 valid news articles with the total number of ratings on the 8 emotions larger than 0. We use x (x = 90%, 80% , 10%) of the data set for training and the remaining (1-x) for testing, to evaluate the scalability and stability of the method.

SemEval. This is an English data set used in the 14th task of the 4th International Workshop on Semantic Evaluations (SemEval-2007).[2] The attributes include the news headline, the score of emotions of anger, disgust, fear, joy, sad and surprise normalizing from 0 to 100. The data set contains 1,246 valid news headlines with the total score of the 6 emotions larger than 0. We use the 1,000 in the test-set (80% of the data set) for training and the 246 in the trial-set (20%) for test.

4.2 Evaluation Metrics

Classifying and predicting the emotions of given news articles are efficient ways to validate the effectiveness of the generated word-emotion mapping dictionary.

The Pearson's correlation coefficient is employed to measure the accuracy of emotion prediction, which indicates the linear dependence between two variables. A value closer to 1 indicates the predicted and the actual emotional distribution fit better, and is reasonable to assert that the trend of ratings on emotions is predicted well by the word-emotion mapping dictionary.

We denote the Pearson's correlation coefficient between the predicted and the actual emotion distributions of the i-th article by pr_i, and the average value of the Pearson's correlation coefficient of all articles by $r\_average$, which is used as the first metric.

[1] http://news.sina.com.cn/society/.
[2] http://nlp.cs.swarthmore.edu/semeval/tasks/.

$$r\_average = \sum_{i=1}^{N} \frac{pr_i}{N} . \tag{13}$$

Besides the average value of the Pearson's correlation coefficient, there is another interesting metrics to evaluate the quality of emotion prediction. In practice, when we predicting the multiple emotional distributions, the dominate one with the maximum predicted rating value is attractive.

$$p\_max = \frac{m}{N} . \tag{14}$$

In the above, m is the number of articles that the predicted and the actual dominate emotion matched. N is the total number of articles for training or test.

4.3 Results and Analysis

In generating the word-emotion mapping dictionary, the probability of word conditioned on document is calculated by relative term frequency according to formula (10). We denote it by *rtf*. In pruning algorithms, maximum, average and minimum are used to refine the dictionary generated by *rtf* (see section 3.3). We denote these three algorithms by *rtf-max*, *rtf-ave* and *rtf-min*.

Results of Sina For different scales of training data set in Sina, the number of words pruned by *rtf-max*, *rtf-ave* and *rtf-min* are presented in Table 1. The number of the original words in the dictionary ranges from 39,278 to 72,773, within which 45.4% to 31.2% words are pruned using *rtf-min*, the words being pruned are quite less using *rtf-max* and *rtf-ave*.

Table 1. The number of words pruned on Sina

Training documents	*Vocabulary size*	*rtf-max*	*rtf-ave*	*rtf-min*
3,249	39,278	74	302	17,848
6,499	48,555	71	304	18,585
9,748	54,210	67	298	19,185
12,997	58,510	68	295	19,575
16,247	62,105	68	293	20,201
19,496	65,162	67	294	20,858
22,745	67,873	68	296	21,426
25,994	70,447	67	295	22,049
29,244	72,773	67	297	22,672

Fig. 2 depicts the *r_average*, *p_max* of all methods and pruning algorithms on the training and test sets.

For the training set, as the size increases, the quality of emotion prediction decreases at first and then remains stable, from which twofold findings can be observed. The first one lies in that, our dictionary fits the training set well even when the available emotional tagged data is limited. Second, although it is harder to fit the training set when the scale is larger, our dictionary is robust for the stability performance on large training sets. For pruning algorithms, the performances after pruning by *rtf-max*, *rtf-ave* and *rtf-min* are better than others without pruning, among which *rtf-min* performs the best, which shows the significance of our pruning algorithm on refining the dictionary.

For the test set, as the number of test articles increases, the quality of e-motion prediction remains stable mostly, except when the size of test articles is 12,997. This indicates the reliability and stability of the dictionary on predicting emotions of previously unseen articles. For pruning algorithms, the performances by *rtf-max*, *rtf-ave* and *rtf-min* are better than others without pruning, among which *rtf-min* performs the best.

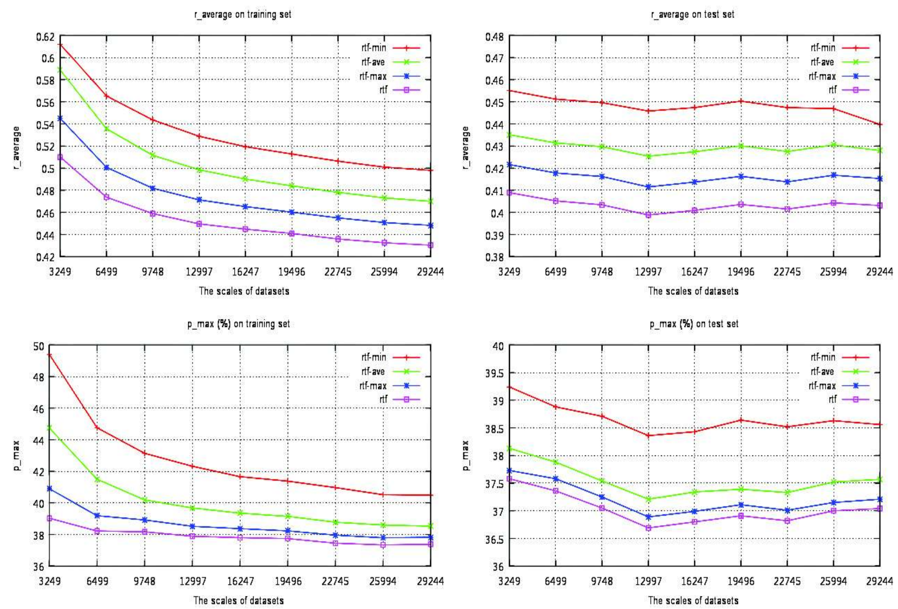

Fig. 2. Performances with different scales of Sina

Although *rtf-min* yields the best results for both training and test data sets, and the improvement over benchmark is remarkable, we also refine the dictionary by deleting the same proportion of words as *rtf-min* randomly, and perform t hypothesis testing on pairwise methods, so as to verify the significant improvement of our pruning algorithm on performances statistically.

The results are depicted in Table 2. For the dictionary after pruning randomly (*prune-random*) and the dictionary without pruning (*rtf*), all of the significance values are much larger than the conventional significance level 0.05, which indicates the dictionary after pruning randomly is no significant different with the dictionary without pruning. In fact, the quality of *prune-random* on the training data set is worse than *rtf* when the size of training documents is 9,748, 12,997 and 19,496, and the quality between them is approximate for other scales of training documents. These findings are similar on the test data set. On the other hand, for the dictionary after pruning by *rtf-min* and *rtf*, or *rtf-min* and *prune-random*, all of the significance values are below the conventional significance level 0.05, which indicates the dictionary after pruning by our method is significant different with others. In our case, we can infer that the dictionary after pruning by *rtf-min* achieves significant performance improvement on both training and test data sets, while pruning randomly does not get such improvement statistically.

Table 2. P-value of the Statistical Significance Test on Sina

Pairwise	*Data set*	*r_average*	*p_max*
prune-random & rtf	Train	0.945303	0.825707
	Test	0.726320	0.886224
rtf-min & rtf	Train	1.18E-04	0.000723
	Test	2.40E-13	9.38E-10
rtf-min & prune-random	Train	1.14E-04	0.000792
	Test	8.63E-14	7.81E-09

Above all, the word-emotion mapping dictionary is effective on emotion classification and prediction. One of the most interesting observations is that when the dictionary is pruned by *rtf-min*, more than 30% words are deleted, while the performances are much better than others.

Results of SemEval Despite that our focus is mainly on annotating emotions for news bodies with long text, it would be very interesting to evaluate the method and pruning algorithms on emotion prediction for news headlines.

The first observation is that when building the word-emotion mapping dictionary based on the short text, as the sparse of the vector, the prune operation maybe unnecessary. For the 1,000 English news headlines used for training here, the vocabulary size is only 2,380 after stemming while retaining the stop words. When the pruning algorithm is applied, the number of pruned words is 0 for *rtf-max* and *rtf-ave*, which means the pruning operation by maximum and average is unnecessary for the data set. The ratio of pruned words is 68.66% for *rtf-min*, which makes the size of the dictionary even smaller, and 7.30% of the training headlines have no word exists in the dictionary, the ratio is 11.38% for test headlines. As a result, pruning by minimize is unsuitable for the SemEval data set, which contains quite limited words.

The second observation is that our method of generating the dictionary works well on fitting the training set for news headlines. The average correlation coefficient of all training articles is 0.86 using the relative term frequency, which shows a strong positive correlation between the predicted and actual emotion distribution. However, the average correlation coefficient of all test articles is 0.36 using the relative term frequency, which means the precision of the dictionary on predicting the emotion distribution of previous unseen documents is relatively low. The reason is that the volume of the word-emotion mapping dictionary is quite small for the limited information of news headlines.

5 Conclusion

Emotion and opinion mining is useful and meaningful from political, economical, commercial, social and psychological perspectives, the word-emotion mapping dictionary constructed in this paper is the first step to meet the needs. Different from previous methods, our method of building the dictionary is adaptive for personalized data set, volume-unlimited, automatically, language-independent, and fine-grained. The main conclusions are as follows:

First of all, the pruning algorithm is effective in refining the dictionary, and improving the performances of emotion prediction. For three forms of removing *background words*, which are maximum, average and minimum, the last one achieves the largest improvement on the performances, and the improvement is statistically significant under hypothesis testing.

Secondly, as the number of training articles increases, the quality of emotion prediction on training data sets decreases firstly and then remains stable. This indicates that our dictionary fits the training data set well even when the available tagged data is limited. Although it is harder to fit the training data set when the scale is larger, our dictionary is robust for the stability performance on large training sets. As the number of test articles increases, the quality of emotion prediction on test data sets remains stable mostly. This indicates the reliable of the word-emotion mapping dictionary on predicting emotions of previously unseen articles.

Last but not least, for annotating emotions of news headlines, it is unnecessary to prune the dictionary, due to the limited vocabulary in the short text. Thus, researches on emotional annotation for both long and short text are our future focuses.

Acknowledgments. The work described in this paper has been supported by the NSFC Overseas, Hong Kong & Macao Scholars Collaborated Researching Fund (61028003)

References

1. Chaumartin, F.R.: Upar7: A knowledge-based system for headline sentiment tagging. In: 4th International Workshop on Semantic Evaluations, pp. 422–425. Association for Computational Linguistics, Prague (2007)

2. Katz, P., Singleton, M., Wicentowski, R.: Swat-mp: The semeval-2007 systems for task 5 and task 14. In: 4th International Workshop on Semantic Evaluations, pp. 308–313. Association for Computational Linguistics, Prague (2007)
3. Kolya, A., Das, D., Ekbal, A., Bandyopadhyay, S.: Identifying Event-Sentiment Association using Lexical Equivalence and Co-reference Approaches. In: Workshop on Relational Models of Semantics Collocated with ACL, pp.19–27. Portland (2011)
4. Banea, C., Mihalcea R., Wiebe J.: A Bootstrapping Method for Building Subjectivity Lexicons for Languages with Scarce Resources. In: 6th International Conference on Language Resources and Evaluation, Marrakech (2008)
5. Strapparava, C., Valitutti, A.: Wordnet-affect: an affective extension of wordnet. In: 4th International Conference on Language Resources and Evaluation, pp. 1083–1086. Lisbon (2004)
6. Baccianella S., Esuli A., Sebastiani F.: SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: 7th Conference on Language Resources and Evaluation, pp. 2200–2204. Valletta (2010)
7. Das, S., Chen, M.: Yahoo! for Amazon: Extracting market sentiment from stock message boards. In: 8th Asia Pacific Finance Association Annual Conference, Shanghai (2001)
8. Turney, P.D.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: 40th annual meeting of the Association for Computational Linguistics, pp. 417–424. Philadelphia (2002)
9. Turney, P.D., Littman, M.L.: Unsupervised learning of semantic orientation from a hundred-billion-word corpus. Technical Report EGB-1094, National Research Council Canada (2002)
10. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Empirical Methods in Natural Language Processing, pp. 79–86, Philadelphia (2002)
11. Liu, J., Seneff, S.: Review sentiment scoring via a parse-and-paraphrase paradigm. In: Empirical Methods in Natural Language Processing, ACL, Suntec (2009)
12. Ifrim, G., Weikum, G.: The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns. In: Coling 2010, Beijing (2010)