Partially Supervised Phrase-Level Sentiment Classification

Sang-Hyob Nam, Seung-Hoon Na, Jungi Kim, Yeha Lee, and Jong-Hyeok Lee

Division of Electrical and Computer Engineering
Pohang University of Science and Technology
San 31, Hyoja-Dong, Nam-Gu, Pohang, 790–784, Republic of Korea
{namsang,nsh1979,yangpa,sion,jhlee}@postech.ac.kr

Abstract. This paper presents a new partially supervised approach to phrase-level sentiment analysis that first automatically constructs a polarity-tagged corpus and then learns sequential sentiment tag from the corpus. This approach uses only sentiment sentences which are readily available on the Internet and does not use a polarity-tagged corpus which is hard to construct manually. With this approach, the system is able to automatically classify phrase-level sentiment. The result shows that a system can learn sentiment expressions without a polarity-tagged corpus.

Keywords: sentiment classification, sentiment analysis, information extraction, text mining.

1 Introduction

Sentiment analysis is the process of extracting opinions from written documents and determining whether they are positive or negative expressions. In recent years, the Internet usage has increased and many people have used it to publish their opinions about various topics, such as movies or the quality of various goods. The amount of published opinion has increased rapidly, so automatic sentiment extraction is desirable.

Much previous works on sentiment analysis has focused on document-level sentiment classification. Pang and Lee [1] [2] use a machine learning technique with minimum cuts algorithm and Turney [3] extracts polarity of phrases using the pair-wise mutual information(PMI) between the phrases and seed words. However, document-level sentiment classification is too coarse for many sentiment tasks such as opinion search and opinion tracking, reputation survey and opinion-oriented information extraction. Document-level sentiment classification incorrectly assumes that subject of all sentiment expression is same with the subject of a document. Therefore, these applications need phrase-level sentiment analysis.

Recently, many researchers have focused on the phrase-level sentiment analysis. Nasukawa [4] constructs a sentiment lexicon, and patterns manually with polarity and POS of a word. Zhongchao [5] also manually defines sentiment patterns and learn a polarity scores using document frequencies of each pattern in positive and negative documents. Wilson [6] uses these previous sentiment resources and a polarity-tagged corpus and

W. Li and D. Mollá-Aliod (Eds.): ICCPOL 2009, LNAI 5459, pp. 225–235, 2009.

tried to identify contextual polarity in phrase-level sentiment analysis. Breck [7] also uses polarity-tagged corpus to identify opinion phrases in a sentence.

Polarity-tagged corpus contains sentences whose opinion expressions are tagged with positive and negative labels. However such corpus is hard to construct manually and not readily available in various domains. Our experiment result shows that it is hard to achieve high recall with small amount of a polarity-tagged corpus in a supervised approach.

However, we can get a sufficient sentiment sentences such as movie reviews on the Internet. We use these sentences to train our phrase-level sentiment classification system instead of using a polarity-tagged corpus. Sentiment sentences are marked by users as positive or negative. We construct a positive phrase corpus and a negative phrase corpus from the sentiment sentences. Those phrases are used to construct polarity-tagged corpus automatically. Our system does not require a manual polarity-tagged corpus. We call this approach a partially supervised approach, because our system learns sentiment tags with automatically constructed polarity-tagged corpus. We views the problem of sentiment labeling at phrase-level as a sequential tagging. Therefore our approach uses Hidden Markov Model (HMM) and Conditional Random Fields(CRF) which is used frequently in tagging problem.

This paper presents a new partially supervised approach to phrase-level sentiment classification. Beginning with a large sentence-level sentiment resource, we calculate sentiment orientation of each phrases, then we get a positive phrase set and a negative phrase set. With these subjectivity clues, we automatically construct polarity-tagged corpus by marking subjectivity clue as positive in positive sentences and negative in negative sentences. Our partially supervised approach learns from the automatically constructed polarity-tagged corpus. Experiment at results show that partially supervised approach is a feasible approach in the phrase-level sentiment classification.

2 Approach

2.1 Sentiment Resources

There are several approaches to automatic sentiment resource construction such as the conjunction method [8], the PMI method [9] and the gloss use method [10]. Turney [3] uses only phrases that contain adjectives or adverbs. Those methods construct useful sentiment resources, but they have some limitations.

Those methods can not extract the sentiment of phrases which are dependent on specific domains. There are also many phrases in corpora which are not correctly spelled, such as movie reviews or goods reviews on the Internet. They do not work well on jargons or idioms, which are difficult to find in the dictionary or to analyze using a parser or a tagger. Such approaches also use rich English sentiment resources which are not available in other languages. Therefore we propose an automatic sentiment resource construction approach which works well in such environments. In this paper we construct sentiment resources using positive or negative sentences. Those sentences have polarity scores between 1 and 10. A value of 10 is the most positive sentimental score

and 1 is the most negative sentimental score. We can use the average score of a word if the size of each score set is the same.

$$AvgScore(w_j) = \frac{\sum_{s_i \in S} Score(s_i) \times Freq(w_j, s_i)}{Freq(w_j)} \quad (1)$$

However, the size of each score set is not the same in most of cases. Therefor we normalize each score.

$$NormScore(w_j) = \frac{\sum_{s_i \in S} Score(s_i) \times \frac{Freq(w_j, s_i)}{\sum_{w_k \in W} Freq(w_k, s_i)}}{\frac{Freq(w_j)}{\sum_{w_k \in W} Freq(w_k)}} \quad (2)$$

$$NormScore(w_j) = \frac{\sum_{s_i \in S} Score(s_i) \times P(w_j | s_i)}{P(w_j)} \quad (3)$$

s_i is a score set between 1 and 10. S is a set of all scores and W is a set of all words. $Score(s_i)$ is constant value of a score set s_i. If s_i is s_9, $Score(s_9)$ is 9. We can determine the polarity of each phrase using this approach. This approach can be easily applied to all language and domains.

2.2 Features of Phrases

Unigrams and bigrams are good features in sentiment document classification [1], indicating that unigram and bigram are appropriate features for identifying the sentiment of phrases. We also used trigram. The experimental data used in this paper is in Korean which is an agglutinative language. We applied Korean segmentation to the training and test data set, which segments auxiliary words and compound nouns. We get follow positive and negative unigrams, bigrams and trigram by using Section 2.1 approach (Table 1, 2).

'discount-card' was the most negative unigram in Korean movie review, because people said that even 'discount-card' was wasteful for the movie. 'discount-card' has domain specific polarity. And there are some named entity word such as 'Sparrow', 'Depp', 'ut-dae', and 'an Emergency Action Number'. Negative bigram 'ho rul' is a part of negative trigram 'gin-gup-jo-chi ho rul'. Table 1 and Table 2 show that unigram, bigram and trigram appropriate for sentiment phrase feature and Section 2.1 works well for extracting semantic orientation of word.

2.3 Automatic Construction of Tagged Sentiment Resource

The Sentiment resource construction approach presented in Section 2.1 is not error prone. However this method is good enough for automatically constructing a polarity-tagged corpus. We calculate semantic orientation of phrase using sentence-level resources. While constructing semantic resources, we identified semantic orientation scores of phrases between 1 and 4 as negative, between 6 and 10 as positive and others as neutral. After constructing the semantic resources, we labeled the sentiment of each phrase in the sentence-level sentiment resource. We tagged subjectivity phrases as positive in the positive sentence set, and tagged them as negative in the negative sentence

Table 1. Semantic resource result of most positve phrases

	Positive	
unit	word	score
unigram	jjang-jjang(‘good-good’)	9.908
unigram	Sparrow(‘Sparrow’)	9.879
bigram	jin-jja jaemiteuyo(‘really funny’)	9.936
bigram	choi-go imnida(‘This is the best’)	9.911
trigram	nermu jaemi iteuyo(‘It’s a lot of fun’)	9.904
trigram	Depp eui maeryuk(‘charm of Depp’)	9.880

Table 2. Semantic resource result of most negative phrases

	Positive	
unit	word	score
unigram	hal-in-card(‘discount-card’)	1.031
unigram	ut-dae(‘Humor University/Korean humor site’)	1.068
bigram	ho rul(‘a number’)	1.031
bigram	jugido aggapda(‘wasteful to’)	1.071
trigram	gin-gup-jo-chi ho rul(‘a Emergency Action number’)	1.072
trigram	gut do yong-hwa(‘disappointing movie’)	1.099

set. Other phrases were tagged as neutral. We shows this procedure by example. jjang-jjang(‘good-good’) is a positive word (Table 1). Although it is a positive word, negative sentence can have the word. Following sentence is a negative sentence.

- jjang-jjang ha-nun nom-dul da alba. (“All people who say good-good to the movie are employee of the movie company”)

We labeled subjectivity word in this sentence by polarity of the sentence. Polarity of the sentence is negative, therefore we labeled subjectivity word ‘jjang-jjang’ as negative.

- jjang-jjang/**Negative** ha-nun/Neutral nom-dul/Neutral da/Neutral alba/Neutral ./ Neutral

Following sentence is a positive sentence.

- scenario ga jjang-jjang ha da(“scenario is good-good”)

We labeled subjectivity as,

- scenario/Neutral ga/Neutral jjang-jjang/**Positive** ha/Neutral da/Neutral

Positive or negative sentiment phrases can represent opposite senses by their context. We followed the sense of the context rather than the sense of the sentiment phrase itself. We confirmed that this assumption is correct in the experiment. We used these automatically constructed tagged sentiment resource in the learning of HMM and CRF.

2.4 Opinion Tagging with Conditional Random Fields

Similar to our approach Breck [7] use CRF to identify sources of opinion phrases. They defined the problem of opinion source identification as one of sequential tagging. Given a sequence of tokens, $x = x_1x_2...x_n$, we need to generate a sequence of tags, $y = y_1y_2...y_n$. The tag is a polarity label which can be positive or negative or neutral. There are three kinds of labels that are positive, negative or neutral. A detailed description of CRFs can be found in Lafferty [11]. For our sequence tagging problem, we create a linear-chain CRF based on an undirected graph $G = (V,E)$, where V is the set of random variables $Y = \{Y_i | 1 \leq i \leq n\}$, one for each of n tokens in an input sentence. And $E = \{(Y_{i-1}, Y_i) | 1 < i \leq n\}$ is the set of $n-1$ edges forming a linear chain. For each sentence x, we define a non-negative clique potential $exp(\sum_{k=1}^{K} \lambda_k f_k(y_{i-1}, y_i, x))$ for each edge, and $exp(\sum_{k=1}^{K'} \lambda'_k f'_k(y_i, x))$ for each node, where $f_k(...)$ is a binary feature indicator function, λ_k is a weight assigned for each feature function, and K and K' are the number of features defined for edges and nodes respectively. Following Lafferty [11], the conditional probability of a sequence of labels y given a sequence of tokens x is

$$P(y|x) = \frac{1}{Z_x} exp(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} \lambda'_k f'_k(y_i, x)) \tag{4}$$

$$Z_x = \sum_{y} exp(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} \lambda'_k f'_k(y_i, x)) \tag{5}$$

where Z_x is a normalization constant for each x, and given training data D, a set of sentences paired with their correct positive, negative, neutral tag sequences, the parameters of the model are trained to maximize the conditional log-likelihood $\prod_{(x,y)\in D} P(y|x)$. For inference, given a sentence x in the test data, the tagging sequence y is given by $argmax_{y'} P(y'|x)$. We used word features between the -4 and 4 window in the CRF model.

2.5 Opinion Tagging with Hidden Markov Model

We use the HMM which is usually used in the tagging problem. There are three states in our HMM model: positive, negative and neutral. Observations of our HMM model are word. HMM model predicts state of each observations. We get initial probability, emission probability and transition probability from the automatically constructed polarity-tagged corpus. We use the Viterbi algorithm to encode the test data using those probabilities.

2.6 Opinion Tagging with Hidden Markov Model and Conditional Random Fields Together

We automatically constructed a tagged sentiment resource that is only partially correct. As a result, it is difficult to expect excellent precision performance with such an incomplete resource. Instead of using tagged sentiment resource directly to label sentiment phrases in the training data, we can refine the data using HMM or CRF. We select the HMM to refine the tagged sentiment resource.

It is important to revise the polarity of the tagged sentiment resource which is refined by HMM, because HMM is trained by automatically constructed tagged sentiment resource. The system revise the labeling error of the tagged sentiment resource, because we know the polarity of the sentence.

For example, HMM sometimes marks negative sentence "We fully grasped inversion story of the movie" as "We/Neutral fully/Neutral grasped/Neutral inversion/**Positive** story/**Positive** of/Neutral the/Neutral movie/Neutral". The system revises the result to "We/Neutral fully/Neutral grasped/Neutral inversion/**Negative** story/**Negative** of/ Neutral the/Neutral movie/Neutral" by using the sentence polarity. Then CRF learns the tag with the neighborhood words.

The system considers all subjectivity tags in the positive sentence set as positive tags and we also consider all subjectivity tags in negative sentence set as negatives. We can expect better precision performance than when using the automatically constructed sentiment resource directly. CRF model is trained by the tagged sentiment resource refined by HMM. We refer to this combination of HMM and CRF as the HMM+CRF model.

3 Experinments

3.1 Training Data

Training data are composed of movie reviews from naver movie review[1] that are scored at the sentence level. Training data are scored from 1 to 10. We used the scores which are in the *Pos* (7-10), *Neg*(1-4) ranges. The number of points in each score set is 20,000, so the total number of training data is 160,000. The data contains some sentences that have doubtful scores, because sometimes people set movie reviews wrong. We use the *Pos* and *Neg* sets when we construct the sentiment resource. We only use $Neg(1,2,3)$ and $Pos(8,9,10)$ scores when we construct tagged sentiment resource to get more explicitly expressed resources.

3.2 Test Data

The test data set was also extracted from naver movie review. These data are comprised of more recent review sentences than the training data set. We asked two annotators to classify and label the data set with scores of $1,2,3$ (900 negative sentences) and $8,9,10$ (900 positive sentences) scores. They tagged each sentimental phrase in the sentence as positive, negative or neutral. We want to evaluate consistency and agreement between human evaluators. Polarity tag boundary is not exactly same between annotators. Therefore we use a CRF model trained by the sentiment tag sequence assigned by each human to evaluate consistency and agreement. The two humans assigned consistent tags to test data (Table 3, Table 4). Agreement between Human1 and Human2 was reasonable enough to use them as test data, because precision and recall are high enough to believe that there are shared sentimental common sense between the humans (Table 3, Table 4). The CRF model that was trained by sentiment tag sequences of Human2 is better than Human1 (Table 3, Table 4). So we selected the test data of Human2 as our experiment test data.

[1] http://movie.naver.com

Table 3. sentimental phrase Human-Human Agreement via CRF model (%). Higher percentage indicates higher agreement between human in positive or negative phrase tagging.

	exact				overlap			
	Human1		Human2		Human1		Human2	
Test	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Human1	98.10	98.45	73.38	72.74	95.00	100.00	82.74	84.55
Human2	70.55	73.16	98.31	99.06	75.75	87.93	95.28	99.96

Table 4. subjectivity phrase Human-Human Agreement via CRF model (%). Higher percentage indicates higher agreement between human in subjectivity phrase extraction.

	exact				overlap			
	Human1		Human2		Human1		Human2	
Test	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
Human1	98.12	98.47	74.12	73.48	95.00	100.00	83.38	85.20
Human2	71.49	74.13	98.45	99.20	76.34	88.61	95.28	99.96

3.3 Evaluation

As with other information extraction tasks, we use precision, recall and f-measure to evaluate the performance of our approach. Precision is $\frac{|C \cap P|}{|P|}$ and recall is $\frac{|C \cap P|}{|C|}$, where C and P are the sets of correct and predicted expression spans, respectively. F_1 is the harmonic mean of precision and recall, $\frac{2 \times P \times R}{P+R}$. Evaluation is done on the sentimental phrase in the sentence. It was tagged as positive or negative in the sentence. Our method often identifies expressions that are close to, but not precisely the same as, the manually identified expressions. For example, for the expression "roundly criticized" our method might only identify "criticized". We therefore introduced softened variants of precision and recall as follows. We define soft precision as $SP^a = \frac{|\{p|p \in P \wedge \exists c \in C s.t. a(c,p)\}|}{|P|}$ and soft recall as $SR^a = \frac{|\{c|c \in C \wedge \exists p \in P s.t. a(c,p)\}|}{|C|}$, where $a(c,p)$ is a predicate that is true only when expression c 'assigns' to expression p in a sense defined by a . We report results according to two predicates: *exact* and *overlap*. $exact(c,p)$ is true only when c and p in $exact(c,p)$ are the same spans - this yields the usual notions of precision and recall. A softer notion is produced by the predicate, which is true when the spans of c and p overlap [7].

3.4 Baseline

Sentiword resource baseline marks a phrase as positive when it belongs to an automatically constructed positive phrase set in Section 2.1 and marks a phrase as negative when it belongs to a negative phrase set.

We run the 10-fold cross validation test using only tagged test data (1800 sentences). Supervised CRF (S-CRF) and Supervised HMM (S-HMM) are used in the test. We used that result as our baselines as well, we compared supervised approaches and our partially supervised approaches. Features of supervised CRF are the same as the partially supervised CRF.

Table 5. Results for identifying sentiment of phrases in n-gram model (%)

	exact			overlap		
Method	Recall	Precision	F_1	Recall	Precision	F_1
unigram	11.79	38.38	18.04	36.42	55.79	44.07
bigram	35.31	41.42	38.12	51.35	49.25	50.28
trigram	14.90	**45.90**	22.50	19.69	**62.07**	29.90
bigram + trigram	36.65	40.44	38.45	52.31	49.91	51.08
bigram + unigram	**41.71**	40.27	**40.98**	66.05	46.12	54.31
all	41.44	38.57	39.95	**66.72**	45.89	**54.34**

Table 6. Results for identifying sentiment of phrases in various models (%)

	exact			overlap		
Method	Recall	Precision	F_1	Recall	Precision	F_1
Sentiword resource	41.44	38.57	39.95	**66.72**	45.89	54.34
Suvervised-CRF	41.56	**61.72**	49.67	45.59	83.38	58.95
Suvervised-HMM	51.30	53.92	**52.58**	58.77	76.44	66.45
Partially-Supervised-HMM	**56.48**	41.15	47.61	59.68	76.01	**66.86**
Partially-Supervised-CRF	44.39	43.10	43.74	57.82	62.78	60.20
Partially-Supervised-HMM+CRF	53.55	44.88	48.83	51.16	**86.91**	64.41

Table 7. Results for identifying subjectivity phrases in various models (%)

	exact			overlap		
Method	Recall	Precision	F_1	Recall	Precision	F_1
Sentiword resource	46.62	56.65	51.14	**70.85**	48.80	57.80
Suvervised-CRF	50.75	**72.38**	59.66	51.91	**94.31**	66.96
Suvervised-HMM	**64.29**	62.62	**63.44**	65.77	89.49	**75.82**
Partially-Supervised-HMM	60.98	51.63	55.91	64.38	82.16	72.19
Partially-Supervised-CRF	47.36	55.20	50.98	61.53	66.77	64.04
Partially-Supervised-HMM+CRF	57.44	51.44	54.27	53.88	91.44	67.80

3.5 Results

Bigram model performs better than the trigram or the unigram model. Trigram and unigram models outperform the recall of bigram in the sentiment resources, because unigram and trigram can determine the polarity bigram can not determine. Unigram improves performance more than trigram when it was used with bigram (Table 5). These models show better performance on the overlap evaluation than the exact evaluation (Table 5, 6).

Partially supervised HMM, CRF and HMM+CRF outperforms the performance of the models that use only sentiment resources, especially the precision. Supervised HMM perform better than other model in exact evaluation. Its f-measure is 52.58% in exact evaluation. f-measure difference between supervised HMM and partially supervised HMM is 4.97% in exact evaluation. But, partially supervised HMM shows better overall performance than other models in overlap evaluation. Its f-measure is 0.41% higher than supervised HMM in overlap evaluation.

Precision of partially supervised CRF is high in exact evaluation, but recall of this model is not so good. Training data of the partially supervised CRF is not completely correct. The data was generated automatically by sentiment resources constructed by our approach. The sentiment resources which generate the training data of CRF have a 66.72% recall in overlap evaluation. This performance affects the recall of CRF. 45.89% precision of training data also affects the precision of CRF. But we improved the precision by using the polarity of a sentence.

Partially supervised HMM improved recall, but its precision is not high. We can use HMM+CRF to overcome weak precision of partially superived HMM and weak recall of supervised CRF.

Table 7 shows the performance of identifying source of subjectivity phrases. Supervised HMM performs well in the exact evaluation.

4 Discussion

4.1 Subjectivity Labeling Problem

The most important part of identifying sentiment of phrases is subjectivity tagging. Breck [7] identified subjectivity phrases using the various features and CRF as a supervised learning in the MPQA. It is difficult to compare directly with the evaluation result of the experiment, because we do not use the same dataset(MPQA) and the language is also different. In spite of these difference, we know that from their results, their f-measure of identifying subjectivity phrase is 73.05% in overlap evaluations [7]. It shows that it is not an easy problem to identify subjectivity phrases even if we use various features and supervised learning.

Many subjectivity errors come from the negative sentimental phrase. There are data sparseness problems in identifying the negative sentiment of phrases, because there are many ironic, cynical, and metaphoric and simile expressions in the negative expressions. These affect the overall performance in identifying the sentiment of phrases.

4.2 Necessary Characteristics of Training Data

We used the partially supervised approach to overcome the problem of insufficient polarity-tagged corpus. Our approach used tagged sentiment of phrases automatically generated by sentiment resources. These sentiment resources are automatically extracted from sentence-level sentiment resource. Our approach also needs sentence-level sentiment training data. Such data sets are more plentiful than tagged sentimental phrase data sets. However in these data sets, there are more polarity annotations at the document level than at the sentence level. We need to select sentiment sentences in sentences

when we use those data sets. In this case, this process unavoidably carries some error in selecting sentiment sentences.

5 Summary and Conclusion

We compared the sentiment phrase (positive, negative or neutral) tagging performance between various models (Table 6). We also compared the subjectivity phrase (sentimental or neutral) tagging performance (Table 7). One interesting result is the difference in performance in identifying sentiment (Table 6) and subjectivity (Table 7). Subjectivity includes positive and negative sentiment. Therefore, it is simpler to label subjectivity phrases than to label sentiment phrases. In spite of the fact that identifying subjectivity phrases is a simpler task than identifying the sentiment of phrases, precisions in identifying subjectivity and sentiment are within 10% in both the exact and the overlap evaluations. This suggests that errors between positive and negative labels are minor. In other words, the overall performance is more heavily affected by the performance of subjectivity classification than by the performance of sentiment classification. The difficulty observed in identifying subjectivity phrases implies some ambiguity, even between human decisions (Table 4). So the most important part of identifying sentiment of phrases is subjectivity tagging. Many subjectivity errors occurred when identifying negative sentimental phrases.

Our model solved the phrase-level sentiment classification problem by using partially supervised tagging approaches. That approach only used the sentence-level sentiment resource. Its precision is 76.01% and its f-meausre is 66.86%. Its f-measure is higher than the supervised approaches in the overlap evaluation. We found that the sentiment phrase tagging problem can be solved by a partially supervised approach.

Acknowledgement. This work was supported in part by MKE & IITA through IT Leading R&D Support Project and also in part by the BK 21 Project in 2008.

References

1. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: ACL, pp. 271–278 (2004)
2. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: EMNLP 2002: Proceedings of the ACL 2002 conference on Empirical methods in natural language processing, Morristown, NJ, USA, pp. 79–86. Association for Computational Linguistics (2002)
3. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: ACL 2002: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Morristown, NJ, USA, pp. 417–424. Association for Computational Linguistics (2001)
4. Nasukawa, T., Yi, J.: Sentiment analysis: capturing favorability using natural language processing. In: K-CAP 2003: Proceedings of the 2nd international conference on Knowledge capture, pp. 70–77. ACM, New York (2003)
5. Fei, Z., Liu, J., Wu, G.: Sentiment classification using phrase patterns, pp. 1147–1152. IEEE Computer Society, Los Alamitos (2004)

6. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: HLT 2005: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Morristown, NJ, USA, pp. 347–354. Association for Computational Linguistics (2005)
7. Breck, E., Choi, Y., Cardie, C.: Identifying expressions of opinion in contenxt. In: Proceedings of the Twentieth International Join Conference on Artificial Intelligence (2007)
8. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, Morristown, NJ, USA, pp. 174–181. Association for Computational Linguistics (1997)
9. Turney, P.D., Littman, M.L.: Measuring praise and criticism: Inference of semantic orientation from association, vol. 21, pp. 315–346. ACM Press, New York (2003)
10. Esuli, A., Sebastiani, F.: Determining the semantic orientation of terms through gloss classification. In: CIKM 2005: Proceedings of the 14th ACM international conference on Information and knowledge management, pp. 617–624. ACM, New York (2005)
11. John, L., McCallum, A., Pereira, F.: Conditional random fields:probabilistic models for segmenting and labeling sequence data. In: 18th International Conference on Machine Learning (2001)