http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (5): 990-996, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.990.996

Constructing Domain-Dependent Sentiment Lexicons Automatically for Sentiment Analysis

[1]You Li, [2]Yuming Lin, [2]Jingwei Zhang and [2]Guoyong Cai
[1]Electronic Engineering and Automation Institute
[2]Guangxi Key Laboratory of Trusted Software,
Guilin University of Electronic Technology, Guilin, 541004, China

Abstract: The sentiment lexicon is an important source for sentiment analysis, which has received lots of attention in recent years. But the word's sentiment inclination is often determined without taking into account the domain knowledge in most general sentiment lexicons. However, sentiments of some word are domain-dependent. Thus, the general sentiment lexicons, such as SentiWordNet, work with low performances in sentiment analysis applications. In this study, the problem of constructing a domain-dependent sentiment lexicon with supervised learning method was explored. The sentiment inclination of a word was identified by quantifying its relationship with the polarities of labels. Intensive experiments were carried out on a real dataset to show that the effectiveness of proposed approach, which was capable of detecting the word sentiment depending on a special domain correctly. Multiple sentiment classification tasks were performed for demonstrating the performances of the constructed lexicons, by which the classification accuracies were statistically improved significantly.

Key words: Sentiment analysis, sentiment lexicon, domain-dependent, supervised learning, mutual information, classification

INTRODUCTION

With the development of social media platforms (such as Twitter, Facebook, etc.), more and more users prefer to generate contents on websites to share their opinion with others. These user-generated data are sentiment-rich, mining user's sentiment expressed in such contents is valuable to many applications including recommending system, market decision system, public opinion detecting and so on. But it is almost impossible to analysis these data artificially due to the huge amount of data. On the other hand, identifying the sentiment of document by human is a time-consuming and error-prone task. Therefore, it brings the urgent need for automatic sentiment analysis tools. In general, the sentiment lexicon is a powerful tool for sentiment analysis, such as SentiWordNet (Baccianella *et al*., 2010), in which each word is associated with a real-value sentiment score ranging from -1 to 1.

Unfortunately, there is not a general-purpose sentiment lexicon which is optimal to all domains. The reason is that the sentiment tendencies of words are domain-dependent. For instance, "straightforward" expresses the positive sentiment in electronics domain. But a reader would not like a novel, if he/she considers it straightforward. Consequently, when the sentiment of a word or phrase was analyzed, the domain to which the object belongs should be taken into account. Thus, a domain-dependent sentiment lexicon is valuable to sentiment analysis applications.

In this study, the problem of constructing a domain-dependent sentiment lexicon was explored, which was defined as follows:

- **Definition 1:** (Domain-dependent sentiment lexicon) A domain-dependent sentiment lexicon Ld is a dictionary of sentiment words. Each item in Ld is a triple (w_i, s_i, d_j) where, w_i stands for the ith sentiment word, s_i is the sentiment score of w_i for domain d_j

As an application of the constructed sentiment lexicon, the lexicon was employed to identify the sentiment polarity (positive or negative) of given documents, which was so called sentiment polarity classification. In this study, the supervised learning for sentiment classification was focused on due to its good performances. Formally, given n training samples $(s_1,l_1),\ldots,(s_n,l_n)$ and m testing samples $t_1,\ldots,t_m$, where, s_i is the ith training sample and l_i is the sentiment polarity label of s_i. The target of this study was to use these training samples

Corresponding Author: Yuming Lin, Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, No. 1, Jinji Road, Qixing District, Guilin, 541004, China Tel: 86-773-2291386

to construct a domain-dependent sentiment lexicon and to predict the sentiment labels of the testing samples based on the constructed sentiment lexicon. To construct a high quality sentiment lexicon, the relationship of word and sentiment labels was quantified as a sentiment score of the word by mutual information firstly. And then the features are weighted based on its sentiment score.

This study has made the following contributions:

- An approach to constructing a domain-dependent sentiment lexicon automatically for a given domain was proposed, which identified not only the unigram's sentiment but also the sentiment of term with other forms, such as bigram
- The constructed lexicons were applied on performing sentiment classification tasks in eight different product domains
- Multiple feature weighting methods for sentiment classification were analyzed. The results showed the good performances of the proposed method

Related work: Sentiment analysis has become a hot research area in recent years. Sentiment lexicon plays an important role in most sentiment analysis applications, such as sentiment classification (Popescu and Pennacchiotti, 2010; Xia *et al.*, 2008), sentiment retrieval (Ounis *et al.*, 2006) and so on. SentiWordNet (Baccianella *et al.*, 2010) was a general sentiment lexicon, in which each synonym set was associated to three numerical scores, describing respectively how objective, positive and negative the terms contained in the synonym set were. Each score ranges from 0-1 and the sum of them was 1. Hu and Liu (2004) used a set of seed adjectives with clear sentiment polarity to grow this set by searching their synonym/antonym in WordNet. The advantage of a general sentiment lexicon was that the word's sentiment is easy to determine, but its main limit was that word's sentiment is invariant for all domain, which was undesired as discussed above.

Turney (2002) suggested determining the polarity of a word or phrase by measuring its point-wise mutual information with some seeds like "excellent" and "poor" by a search engine called AltaVista (http://www.altavista. com/). More recently, Lu *et al.* (2011) tried to construct a context-dependent sentiment lexicon by linear programming, in which they integrated four sources, general-purpose sentiment lexicon, rating of review, thesaurus like WordNet and linguistic heuristics rules, into their objective function. These approaches depend on external sources, which limit their adaptability.

Sentiment classification, as an important application of sentiment analysis, has attracted increasing attention recently. A significant work was done by Pang *et al.* (2002) which introduced firstly machine learning method to the sentiment classification domain and compared different features in movie review sentiment classification implemented by Naive Bayes, maximum and SVM, respectively. Their experimental results the SVM with information of feature presence based on unigram outperformed the others, whose accuracy achieved 82.9%. In this study, this method was treated as a baseline because we focus on the classification accuracy too. On the basis of unigram and bigram, Matsumoto *et al.* (2005) expanded the features with frequent word sub-sequences and dependency sub-tree, which improved the classification performance. Paltoglou and Thelwall (2010) compared multiple variants of classic TF×IDF schema adapted to sentiment analysis and emphasized that expressing sample vectors with emotional information via supervised methods is helpful for predicting sentiment polarity. For the sentiment classification tasks described above, the sentiments of words do not been taken into account. If this factor is associated with term's score in documents, the classification accuracy should be improved, which is demonstrated in later experiments. Thus, such sentiment classification tasks were performed based on the word's sentiment recorded in a domain-dependent sentiment lexicon.

CONSTRUCTING A DOMAIN-DEPENDENT SENTIMENT LEXICON

Firstly, the problem of how to construct a domain-dependent sentiment lexicon was explained based on training samples only. Intuitively, a positive word occurred more frequently in positive documents than negative ones. In other words, a positive word's relationship with the positive label was always stronger than that with the negative label. Thus, such relationship was quantified for capturing the word's sentiment tendency.

In probability theory and information theory, the mutual information was capable of detecting the difference between the joint distribution on (X, Y) and the marginal distributions on X and Y. It was a quantity that measured the mutual dependence of the two random variables. Formally, the mutual information of two discrete values x and y, was evaluated as follows:

$$\mathrm{MI}(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)} \tag{1}$$

where, p(x,y) was the joint probability of x and y, p(x) and p(y) are the marginal probability of x and y, respectively.

Given N samples labeled, A was the number of times term t and the label l co-occur, B was the number of times term t occurs without label l, C was the number of samples with label l but not include term t. Thus the mutual information MI(t,l) of t and l could be evaluated as:

$$\mathrm{MI}(t,l) = \log_2 \frac{p(t,l)}{p(t)p(l)} \approx \log_2 \frac{AN}{(A + C)(A + B)} \tag{2}$$

In this study, sentiment polarity classification was focused on, thus two labels, positive label lp and negative label ln, were taken into account. Now the relationship of word w with the sentiment label l was quantitated according to Eq. 2. Like many sentiment lexicons, a positive real number was assigned to a positive word and negative real number to a negative word. The MI(w,lp) and MI(w,ln) were integrated into a sentiment score denoted as Score(w), which was shown as Eq. 3. If MI(w, lp)>MI(w, ln) hold, word w tended to expression positive, the Score(w) would be a positive number. On the contrary, word w tended to be negative and its sentiment would be a negative number. But if the MI (w, lp) was equal to the MI(w, ln), it mean word w was neutral. Notably, the neutral words were omitted since they did not make any contribution for the sentiment polarity classification.

$$\mathrm{Score}(w) = \begin{cases} |\mathrm{MI}(w,lp)| & \text{if } \mathrm{MI}(w,lp)>\mathrm{MI}\ (w,ln) \\ 0 & \text{if } \mathrm{MI}(w,lp) = \mathrm{MI}(w,ln) \\ -|\mathrm{MI}(w,ln)| & \text{if } \mathrm{MI}(w,lp)<\mathrm{MI}(w,ln) \end{cases} \tag{3}$$

Some samples selected from the constructed sentiment lexicon for electronics and movie domain were listed in Table 1 for giving an intuitive learning. The words with score 1 (-1) mean they only occurred in positive (negative) reviews and their sentiment were identified correctly. On the other hand, the proposed approach was based on statistics rather than linguistics. Therefore, some misspelled words, emoticons like :(and the neologisms could be detected exactly, such as the ‘excellent’ in Table 1.

The appliance of the constructed lexicon was a good way to estimate its performance. Next, sentiment polarity classification tasks were performed based on the lexicon.

Table 1: Samples from the constructed lexicon

Electronics		Movie	
Word	Score (w)	Word	Score (w)
Outlet	1.00	Lovingly	1.00
Excellent	1.00	Deftly	1.00
Precisely	0.58	Wondrous	0.68
Needless	-0.42	Terrible	-0.63
Troubling	-1.00	Wasteful	-1.00
Terribly	-1.00	Washout	-1.00

The feature presence information was used to evaluate the contribution of term t to a document. Thus, the contribution of term t to a document. Thus, the features were weighted based on its contribution together with itSs sentiment score as following:

$$\mathrm{Value}(t,d) = \mathrm{Presence}(t,d) \times \mathrm{Score}(t) \tag{4}$$

where, the presence (t,d) was 1 if term t occurs in document d, otherwise 0. The term’s frequency was also used to evaluate the term’s contribution, but it worked worse than presence information in the pre-experiments. Algorithm 1 described this process. Firstly, the training set and the test were converted into two vectors X and Y in terms of predetermined feature such as unigram (line 1, 2). Secondly, the value of each component in X and Y was computed according to the Eq. 2-4 (line 2). After the classifier C trained on X was generated (line 3), it was used to predict the labels of test samples (line 4-7).

Algorithm 1: Sentiment classification based on domain-dependent sentiment lexicon

```
Input: The training set $S = \{<s_1, y_1>,\ldots,<s_n, y_n>\}$the testing set $T = \{t_1,\ldots, t_m\}$
Output: The predicted label set $L = \{l_1,\ldots, l_m\}$
1: Convert S and T into document vector X and Y, respectively;
2: Evaluate the value of each component in Xand Y according to Eq. 4;
3: Train the sentiment classifier C on X;
4: For i = 1 to m do
5:     $l_i = C(y_i)$;
6:     $L \leftarrow l_i$;
7: Return L;
```

EXPERIMENTS

Datasets and preprocessing: Eight domain product reviews were prepared for the experiments: electronic, kitchen, DVD, apparel, health, sport_outdoor product, software and movie. The first seven types of reviews were crawled from Amazon (http://www.amazon.com) and reorganized by Blitzer *et al.* (2007), the last one from IMDB (http://www.imdb.com) by Pang *et al.* (2002). Each type of product reviews contained 1000 positive reviews and 1000 negative ones. Two types of features, unigram and bigram, commonly used for sentiment classification were focused on in the experiments. The terms occurred less than five times in each dataset were omitted. The stemming process was not applied, since it was detrimental to classification accuracy (Leopold and Kindermann, 2002). The LIBSVM (http://www.csie.ntu. edu.tw/~cjlin/libsvm/) was applied on implementing SVM classification algorithm because of its high effectiveness. The five-fold crossed-validation was applied in the experiments. Moreover, all negation words (such as not, doesn’t, didn’t, haven’t and so on.) were deleted and the tag “NOT_” was attached to the words following the negation word until the first punctuation. For instance, the sentence "It doesn't work well." would be transformed into the sentence "It NOT_work NOT_well.".

The different feature weighting methods were compared with the proposed approach:

- **tf×senti:** SentiWordNet3.0 was used to determine the sentiment score of a term, the term was weighted by the product of its frequency and its sentiment score
- **pre×senti:** Similar to tf×senti, but the frequency was replaced by presence information
- **Frequency:** The term frequency was regarded as its weight
- **Presence:** Whether the term occurred in a document
- **The proposed approach:** The terms were weighted according to the Eq. 4

The accuracy was applied on evaluating the effectiveness of proposed approach in our experiments, which was computed as follows:

$$\text{Accuracy} = \frac{a+d}{a+b+c+d} \qquad (5)$$

where, a was the number of positive reviews predicted as positive, b was the number of negative reviews predicted as positive, c was the number of positive reviews predicted as negative and d was the number of negative reviews predicted as negative.

Experimental results: Figure 1 showed the results on classification accuracies of different weighting methods based on unigram. The proposed approach made the best performances in all domains (85.90, 88.00, 80.95, 87.40, 85.70, 85.75, 86.10 and 85.50%, respectively). The tf×senti and pre×senti achieved the poor accuracies, which mean the general sentiment lexicon was not good at capturing the correct sentiment tendencies of words in different domains. The accuracies of the frequency and presence weighting methods on movie reviews were similar to the results reported by Pang *et al*. (2002). Compared with the other domains, the classification accuracies for DVD and movie domain were relatively poor, because the sentiment of words was harder to distinguish in this domain. For example, an explicit negative word liked ugly was likely to describe a man/woman in an excellent DVD or movie. Especially for the DVD domain, the reviews included not only those reviewed on the DVD quality, but also those on the story recorded in DVD, even the mixture of both.

In the next experiment, the classification accuracy comparisons of different approaches based on bigram were considered in the same 8 domains. Notably, the items in constructed lexicon were bigrams rather than unigrams.

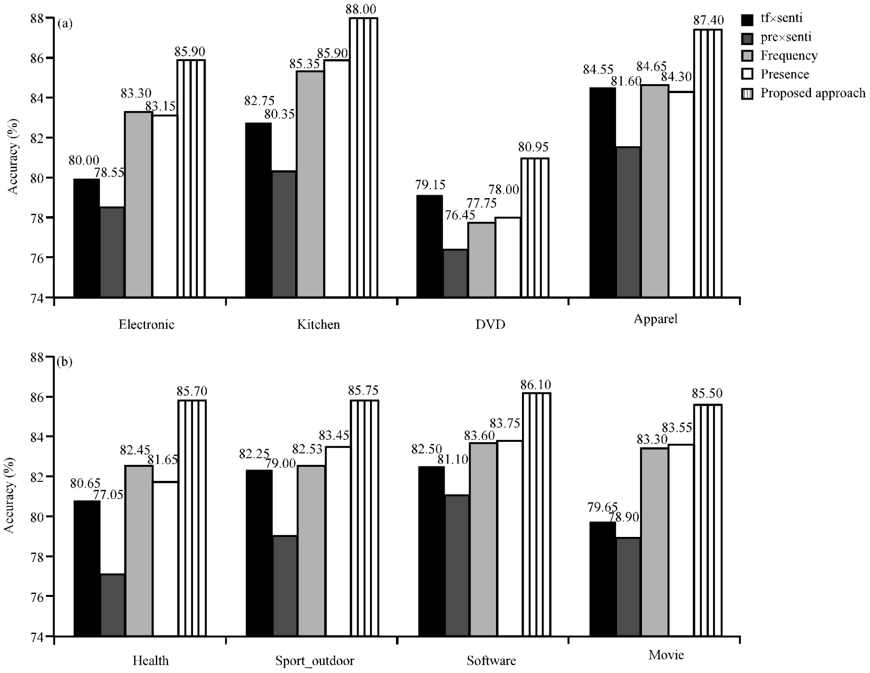

Fig. 1(a-b): Classification accuracy comparisons of different approaches based on unigram

As shown in Fig. 2, the proposed approach achieved 82.95, 84.55, 78.60, 85.45, 82.90, 83.35, 83.35 and 83.50% classification accuracies in different domains, respectively, which made significant improvements with comparison with the competitors. Specifically, none of the general sentiment lexicons described the sentiment tendency of bigram as far as we know, thus the accuracies of tf×senti and pre×senti for bigrams could not be evaluated in this experiment. But the bigram's sentiment could be identified by our approach, which was another advantage of the proposed approach.

Compared the results described in Fig. 1 with the corresponding results described in Fig. 2, the unigram was more discriminative than the bigram for sentiment classification. On the other hand, the proposed approach based on bigram was similar to the competitors based on unigram, even more effective such as for the apparel and health domain. This verified the effectiveness of the constructed sentiment lexicons again.

Case study: The goal of our case study was to demonstrate again that why the proposed approach worked better than others by studying closely eight samples from different domains. Since, this work aimed to construct domain-dependent sentiment lexicons and Eq. 4 applied the frequency information to capture the term's contribution, only the sentiment scores of terms are consider in this case study. The general sentiment lexicon, SentiWordNet 3.0, was compared with the ones constructed with the proposed approach. We focused on the sentiment scores of unigrams rather than bigrams, because the SentiWordNet 3.0 did not involve bigrams. Again, capturing the sentiment scores of bigrams is one advantage of the proposed approach. Table 2 showed Sentiment scores of terms of eight positive and negative reviews sampled from different domains randomly, in where the left column described the sampled reviews, the right column presented the sentiment scores generated by SentiWordNet 3.0 and the domain-

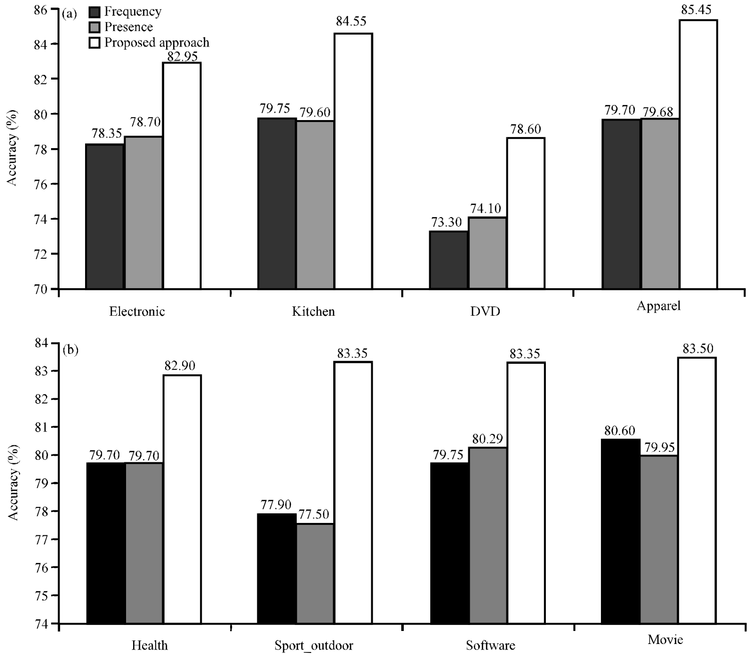

Fig. 2(a-b): Classification accuracy comparisons of different approaches based on bigram

Table 2 :Sentiment scores of terms in eight reviews sampled from different domains

Samples	Sentiment scores of terms in the corresponding review
R1: A negative review in health domain	Inaccurate(-0.63, -0.85) fat(0.4, -0.46) wish(-0.29, 0.16) waste(-0.2, -0.94) return(0.25, -0.91) provide(0.13, -0.77)
R2: A negative review in software domain	Piece(0.13, -0.64) first(0.12, -0.09) nice(0.65, 0.34) incompetent(-0.41,-0.892) frustrating(0.31, -0.87),
R3: A negative review in apparel domain	Cute(0.56, 0.16) way(0.13,-0.28) disappointing(-0.75,-1.00) say(0.13, -0.13) first(0.13, -0.2) lining(-0.25, -0.51) hassle(-0.31, -0.68) worth(0.25, 0.46) job(-0.15, 0.44) feel(0.15,0.38)
R4: A negative review in electronics domain	Disappointing(-0.75, -0.88) phone(-0.63, 0.01) use(0.19, 0.16) home(0.25, 0.29) rave(0.38, -1) unfortunately (-0.88, -0.77) have(-0.25, -0.42) cut(-0.11, -0.58)
R5: A positive review in kitchen domain	Press(-0.25, 0.40) really(0.43, 0.20) perfect(0.67, 0.81) size(-0.38, 0.54)
R6: A positive review in book domain	Story(-0.375, 0.14) detail(0.125, 0.27) right(0.23, 0.32) recommend(0.29, 0.66)
R7: A positive review in DVD domain	Fun(-0.21, 0.49) wonderful(0.75, 0.63) scene(-0.31, 0.03) best(0.75, 0.44) physical(-0.31, -0.36)
R8: A positive review in movie domain	Time(0.63, 0.53) old(0.28, 0.07) likely(-0.28, 0.01) life(0.25, 0.14) boy(0.13, 0.11) rather(-0.38, 0.03) spend(0.19, 0.05) production(-0.13, 0.12) high(0.21, 0.62) comparable(0.50, 0.58) delight(0.38, 0.54) golden(0.43, 0.33) best(0.75, 0.82) foreign(-0.42, -0.11) couple(-0.19, -0.03)

dependent lexicon respectively. For example, the item inaccurate (-0.625, -0.85) mean that the sentiment score of the term inaccurate in SentiWordNet 3.0 was -0.63, but -0.85 was its sentiment score in domain-dependent lexicon.

Comparison with the score from SentiWordNet 3.0, the term sentiment scores generated by domain-dependent lexicon were more consistent with the review's sentiment polarity as shown in Table 2. Notably, it did not mean the sentiment inclination of term determined by domain-dependent was correct in a review, even it was in accordance with the polarity label of the review. For instance, the item 'fat (0.4, -0.46)' mean the term 'fat' is negative in R1 for domain-dependent lexicon since its score is -0.46. In fact, this review is about a body fat measurer. Thus the term 'fat' did not express any sentiment in this review. But for the health domain, 'fat' was not expected to be, thus it should express a negative sentiment. It was similar to the other domains. Moreover, some terms expressed special sentiment according to the domain-dependent lexicons, although they did not express any sentiment without context, which maybe because of the writing styles for different type reviews.

Again, the domain-dependent sentiment lexicon was more effective than the general one, since it considered a term could express different sentiment for different domains. And it could be applied on many applications including sentiment classification, opinion mining, opinion summarization, opinion retrieval, opinion question answering and so on.

CONCLUSION

As an important source, sentiment lexicon has seen a great deal of attention in recent years. In this study, the problem of constructing domain-dependent sentiment lexicon automatically was explored, which is valuable and crucial to many Web applications. To capture the term's sentiment inclination correctly, its relationship with sentiment labels was quantified by mutual information. The proposed approach was able to identify not only the unigram's sentiment but also the bigram's. And the sentiment of the later could not be determined by all general sentiment lexicons. Multiple sentiment classification tasks were performed to valid the effectiveness of the constructed lexicons. The experimental results showed the feature weighting based on the constructed lexicons achieved the best performance compared with the state of art methods.

ACKNOWLEDGMENTS

This study was supported by National Natural Science Foundation of China (No. 61063039), Guangxi Natural Science Foundation (No. 2012GXNSFBA053171) and Guangxi Key Laboratory of Trusted Software Foundation (No. kx201203).

REFERENCES

Baccianella, S., A. Esuli and F. Sebastiani, 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. Proceedings of the 7th Conference on International Language Resources and Evaluation, May 17-23, 2010, European Language Resources Association, Valletta, Malta, pp: 2200-2204.

Blitzer, J., M. Dredze and F. Pereira, 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, June 23-30, 2007, Prague, Czech Republic, pp: 440-447.

Hu, M. and B. Liu, 2004. Mining and summarizing customer reviews. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data, August 22-25, 2004, ACM Press, Washington, USA., pp: 168-177.

Leopold, E. and J. Kindermann, 2002. Text categorization with support vector machines. How to represent texts in input space. Mach. Learn., 46: 423-444.

Lu,Y., M. Castellanos, U. Dayal and C.X. Zhai, 2011. Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach. Proceedings of the 20th international conference on World wide web, March 28-April 1, 2011, Hyderabad, India, pp: 347-356.

Matsumoto, S., H. Takamra and M. Okumura, 2005. Sentiment classification using word sub-sequences and dependency sub-trees. Proceedings of the 9th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, May 18-20, 2005, Springer-Verlag, pp 301-311.

Ounis, I., M. de Rijke, C. Macdonald, G. Mishne and I. Soboroff, 2006. Overview of the trec 2006 blog track. Proceedings of the 15th Text Retrieval Conference, November 14-17, 2006, Gaithersburg, USA, pp: 93-101.

Paltoglou, G. and M. Thelwall, 2010. A study of information retrieval weighting schemes for sentiment analysis. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden, pp: 1386-1395.

Pang, B., L. Lee and S. Vaithyanathan, 2002. Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the Conference on Empirical Methods on Natural Language Processing, July 6-7, Association for Computational Linguistics Press, Philadelphia, USA., pp: 79-86.

Popescu, A.M. and M. Pennacchiotti, 2010. Detecting controversial event from twitter. Proceedings of the 19th ACM International Conference on Information and Knowledge Management, October 25-29, 2010, ACM, Toronto, Canada, pp: 1873-1876.

Turney, P.D., 2002. Thumbs up or thumbs down? Sentiment orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, July 11-12, 2002, Philadelphia, USA., pp: 417-424.

Xia, Y., L. Wang, K.F. Wong and M. Xu, 2008. Sentiment vector space model for lyric-based song sentiment classification. Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, June 16-7, 2008, Association for Computational Linguistics Stroudsburg, PA, USA., 133-136.