Optimal Feature Selection for Sentiment Analysis

Basant Agarwal and Namita Mittal

Malaviya National Institute of Technology, Jaipur, India
thebasant@gmail.com, nmittal@mnit.ac.in

Abstract. Sentiment Analysis (SA) research has increased tremendously in recent times. Sentiment analysis deals with the methods that automatically process the text contents and extract the opinion of the users. In this paper, *unigram* and *bi-grams* are extracted from the text, and composite features are created using them. Part of Speech (POS) based features adjectives and adverbs are also extracted. Information Gain (IG) and Minimum Redundancy Maximum Relevancy (mRMR) feature selection methods are used to extract prominent features. Further, effect of various feature sets for sentiment classification is investigated using machine learning methods. Effects of different categories of features are investigated on four standard datasets i.e. Movie review, product (book, DVD and electronics) review dataset. Experimental results show that composite features created from prominent features of *unigram* and *bi-gram* perform better than other features for sentiment classification. mRMR is better feature selection method as compared to IG for sentiment classification. Boolean Multinomial Naïve Bayes (BMNB) algorithm performs better than Support Vector Machine (SVM) classifier for sentiment analysis in terms of accuracy and execution time.

Keywords: Sentiment Analysis, feature selection methods, machine learning, Information Gain, Minimum Redundancy Maximum Relevancy (mRMR), composite features.

1 Introduction

Sentiment Analysis (SA) is a task that finds the opinion (e.g. positive or negative) from the text documents like product reviews /movie reviews [1], [9]. As user generated data is increasing day by day on the web, it is needed to analyze those contents to know the opinion of the users, and hence it increases the demand of sentiment analysis research. People express their opinion about movies and products etc. on the web blogs, social networking websites, content sharing sites and discussion forums etc. These reviews are beneficial for users and companies. Users can know about various features of products that can help in taking decision of purchasing items. Companies can improve their products and services based on the reviews. Sentiment analysis is very important for e-Commerce companies to know the online trends about the products and services. Example of sentiment analysis includes identifying movie popularity from online reviews; which model of a camera is liked by most of the users and which music is liked the most by people etc.

A. Gelbukh (Ed.): CICLing 2013, Part II, LNCS 7817, pp. 13–24, 2013.

Sentiment classification is to assign a document into categories (positive, negative and neutral) by its subjective information. The challenge in movie review polarity classification is that the generally real facts are also mixed with actual review data. It is difficult to extract opinion from reviews when there is a discussion of the plot of the movie, discussion of the good qualities of actors of the movie but in the end overall movie is disliked. One of the biggest challenges of this task is to handle negated opinion. Product review domain considerably differs from movie review dataset. In product reviews, reviewer generally writes both positives and negative opinion, because some features of the product are liked and some are disliked. It is difficult to classify that review into positive and negative class. Also, some feature specific comments are written in the review, for example like battery life of the laptop is less, but overall performance is good. To identify overall sentiment of these type of reviews are difficult. Generally, product review dataset contains more comparative sentences than movie review dataset, which is difficult to classify [6].

Machine learning methods have been extensively used for sentiment classification [1], [2], [9]. The Bag of Words representation is commonly used for sentiment classification, resulting very high dimensionality of the feature space. Machine learning algorithm can handle this high-dimensional feature space by using feature selection methods which eliminate the noisy and irrelevant features [17].

In proposed approach, *unigram* and *bi-grams* feature set are extracted from text, and various composite feature sets are created. Effect of various feature sets are investigated for sentiment classification using Boolean Multinomial Naïve Bayes (BMNB) [18] and Support Vector Machine (SVM) [11] classifiers. Information Gain (IG) and Minimum Redundancy Maximum Relevancy (mRMR) feature selection techniques are used to extract prominent features.

Contributions of this paper are as follows.

1. Different composite feature set are created using *unigram* and *bi*-gram that perform better than other features.
2. Used mRMR feature selection method for sentiment analysis, and compared its performance with the IG.
3. Compared the performance of BMNB and SVM for sentiment analysis, and found that BMNB classifier performs better than state of art SVM classifier.
4. Proposed method is evaluated on four standard datasets on varied domain reviews.

The paper is organized as follows: A brief discussion of the earlier research work is given in Section 2. Feature selection methods used for reducing the feature vector size are discussed in Section 3. Section 4 describes the machine learning algorithm used in the experiments. Dataset, Experimental setup and results are discussed in Section 5. Finally, Section 6 describes conclusions.

2 Related Work

A lot of work has been done for feature selection for sentiment classification [9],[10], [13], [16], [17] using machine learning methods [1], [2], [5], [13]. Pang and Lee [2] used *unigrams*, *bi*-grams and adjectives for creating feature vector. Authors used different machine learning algorithms like NB, SVM, and Maximum-Entropy (ME) for sentiment analysis of movie review dataset. Further, they investigated that presence or absence of a term in the feature vector gives better classification results than using term frequency, and concluded that SVM performs best amongst classifiers. Sentiment classification using machine learning methods face problem of dealing high dimension of the feature vector [1], [13]. Many researchers worked on reducing feature vector size with different feature selection methods. The performance comparison of standard machine learning techniques with different feature selection methods have been discussed [1], [5], [9], [13]. Pang and Lee [4] used minimum cut method for sentiment polarity detection. Authors eliminated the objective sentences from the documents. In [3], Categorical Probability Proportion Difference (CPPD) feature selection method is proposed, which is capable of selecting the features which are relevant and capable of discriminating the class.

O' keefe et al. [15] compared three feature selection methods and feature weighting scheme for sentiment classification. Wang et al. [14] proposed a new Fisher's discriminant ratio based feature selection method for text sentiment classification. Abbasi et al. [17] found that information gain or genetic algorithm improves the accuracy of sentiment classification. They also proposed Entropy Weighted Genetic Algorithm (EWGA) by combining the two, which produces high accuracy. S. Tan [13], discussed four feature selection methods Mutual Information (MI), IG, Chi square (CHI), and Document Frequency (DF) for sentiment classification on Chinese documents, using five machine learning algorithms i.e. K- nearest neighbour, Centroid classifier, Winnow classifier, NB and SVM. Authors observed that IG performs best among all the feature selection methods and SVM gives best results among machine learning algorithms.

Verma et al. [6] used semantic score for initial pruning of semantically less important terms, further by using information gain feature selection technique important features are extracted, for better classification accuracy. Part-of-speech (POS) information is commonly used in sentiment analysis and opinion mining [5], [9]. There are several comparisons of efficiency of adjectives, adverbs, verbs and other POS [1], [9], [20]. Turney [7] proposed a sentiment classification method using phrases based on POS patterns, mostly including adjective and adverbs.

3 Feature Selection Method

Feature selection methods select important features by eliminating irrelevant features. Reduced feature vector comprising relevant features improves the computation speed and increases the accuracy of machine learning methods [10], [17].

3.1 Minimum Redundancy Maximum Relevance (mRMR)

The Minimum Redundancy Maximum Relevance (mRMR) feature selection method [12] is used to identify the discriminant features of a class. mRMR method selects features those have high dependency to class (maximum relevancy) and minimum dependency among features (minimum redundancy). Sometimes relevant features with maximum relevancy with the class may have redundancy among features. When two features have redundancy then if one feature is eliminated, there is not much difference in class discrimination [12].

Mutual information is used for calculating the correlation/dependency between features and class attribute, and among features. mRMR feature selection technique selects features which have high mutual information (maximum relevant) with the class attribute and eliminate features which have high mutual information (highly correlated) among themselves (minimum redundant).

3.2 Information Gain (IG)

Information gain (IG) is one of the important feature selection techniques for sentiment classification. IG is used to select important features with respect to class attribute. It is measured by the reduction in the uncertainty in identifying the class attribute when the value of the feature is known. The top ranked (important) features are selected for reducing the feature vector size in turn better classification results. Information gain of a term can be calculated by using equation 1 [11].

$$\begin{aligned} IG(t) = -\sum_{J=1}^{K} P(C_J) \log(P(C_J)) + \\ P(w)\sum_{J=1}^{K} P(C_J \mid w) \log(P(C_J \mid w)) + P(\overline{w})\sum_{J=1}^{K} P(C_J \mid \overline{w}) \log P(C_J \mid \overline{w}) \end{aligned} \tag{1}$$

Here, $P(C_J)$ is the fraction of number of documents that belongs to class C_j out of total documents and $P(w)$ is fraction of documents in which term w occurs. $P(C_j|w)$ is computed as fraction of documents from class C_j that have term w.

4 Machine Learning Algorithms

4.1 Multinomial Naïve Bayes

Naive Bayes [11] is frequently used for text classification problems. It is computationally very efficient and easy to use. The Naïve Bayes assumption is that features are conditionally independent of one another, given the class [11]. A Multinomial Naive Bayes classifier [18] with Term Frequency is a probability based learning method, which constructs a model by using term frequency of a feature/word to represent documents.

In Boolean Multinomial Naïve Bayes (BMNB) [18], TF of a word in a document is counted as 1 if that term is present else it is counted as zero.

4.2 Support Vector Machine (SVM)

SVM is a supervised learning method [9], [11]. SVM finds a hyperplane that divides the training documents in such a way that both the class data points are maximum separable. SVM has shown to be superior in comparison to other machine learning algorithms, in case of limited but sufficient training samples. SVM has been widely used for text classification [11], [18] and sentiment analysis [1], [2], [9].

5 Dataset, Experimental Setup and Results

5.1 Dataset Used

To evaluate the prominent features, feature selection method and best machine learning algorithm, one of the most popular publically available movie review dataset [4] is used. This standard dataset, known as Cornell Movie Review Dataset is consisting of 2000 reviews that contain 1000 positive and 1000 negative reviews collected from Internet Movie Database. To make experiment scientifically more stable, we used product review dataset consisting amazon products reviews provided by Blitzer et al. [22]. Reviews are available for different domains. We used product reviews of books, DVD and electronics for experiments. Each domain has 1000 positive and 1000 negative labelled reviews. An average number of words per document are larger in Movie review dataset as compared to product review dataset.

5.2 Features Extraction and Selection in Proposed Approach

In proposed approach, each review is pre-processed in such a way that machine learning algorithm is applied. Negation word (no, not, never, didn't, don't, can't) reverses the polarity of the sentence that is important to handle for sentiment classification. It is done by concatenating first word after the negation word that should not be a stop word. For example, "this is not a good movie", polarity of word "good" is reversed by "not", and it becomes "notgood" after negation handling [6]. Boolean weighting scheme is used for representing text document.

Features are categorized on the basis of the way we have extracted them from the text. The categories are (i) words occurring in the document i.e. *unigrams*, *bi*-gram. (ii) POS based words, i.e. adjectives, adverbs.

(i) In the first category, firstly negation handling is performed. Then document is tokenized, and stop words are removed. Each word is stemmed according to Porter's algorithm [8]. In the pre-processing phase, Document Frequency (DF) is used for initial pruning of unimportant features by eliminating features occurring in less number of documents. Firstly feature set using *unigram* (*F1* feature set) and *bi*-gram (*F2* feature set) features are generated. *Bi-gram* based features (*F2*) are capable of handling negation words in the context of the text [2] that is why there is no need of negation handling explicitly in this case.

Further, prominent feature sets and composite feature sets are created from *unigram* and *bi-gram* features. Prominent features are extracted from *unigrams* with IG and mRMR, we call it as *PIGF1* (***P**rominent **IG F**eatures **1**-gram*) and *PmRMRF1* (***P**rominent **mRMR F**eatures **1**-gram*) respectively. Similarly, optimal features are extracted from *bi-grams* with IG and mRMR, those are *PIGF2* (***P**rominent **IG F**eatures **2**-gram*) and *PmRMRF2* (***P**rominent **mRMR F**eatures **2**-gram*) respectively. Further, by combining *unigrams* and *bi-grams*, ***Com**posite **F**eature set* (*ComF*) is created. Then, by combining ***p**rominent unigram* and *bi*-gram ***IG f**eatures* (*PIGF1* and *PIGF2*), ***P**rominent **Com**posite* IG features (*ComPIG*) are created. Similarly, by using Prominent *unigram* and *bi-gram* mRMR features (*PmRMRF1*, *PmRMRF2*), ***P**rominent **Com**posite **mRMR*** features *ComPmRMR* feature set is created.

(ii) In the second category, Stanford POS tagging software[1] is used for tagging each term according to Part of Speech. Stop word removal and stemming is not performed in this method for extracting features, as the same word can occur with different POS. For example, die as Noun is different than die as Verb. Adjective and Adverbs are extracted because these are considered as important features for sentiment classification [1, 5, 9]. Feature sets namely *P1* and *P2* are generated using adjectives and adverbs respectively. Further, **Com**posite **F**eature set (*ComP*) is also created by combining POS features (*P1 and P2*).

5.3 Evaluation Metrics

Precision, Recall, Accuracy and F- measure are used for evaluating performance of sentiment classification [11]. Precision for a class C is the fraction of total number of documents that are correctly classified and total number of documents that classified to the class C (sum of True Positives (TP) and False Positives (FP)). Recall is the fraction of total number of correctly classified documents to the total number of documents that belongs to class C (sum of True Positives and False Negative (FN)). F –measure is the combination of both precision and recall, is given by

$$F - Measure = 2 * (precision * recall) / (precision + recall) \tag{2}$$

F-measure is used to report the performance of classifiers for the sentiment classification.

5.4 Results and Discussions

Different feature vector generated after pre-processing are further used for the classification. Among different machine learning algorithms Support Vector Machine (SVM) classifiers are the mostly used for sentiment classification [2], [5], [6], [10], [13], [17]. In our Experiments, BMNB and SVM are used for classifying review documents into positive or negative sentiment polarity, since BMNB can perform better than SVM in case some appropriate feature selection method is used. Evaluation of classification is done by 10 fold cross validation [21]. Linear SVM and Naïve Bayes Multinomial are used for all the experiments with default setting in WEKA [19].

[1] `http://nlp.stanford.edu/software/`

Determination of Prominent Feature and Classifiers

The performance of different feature sets are compared with respect to F-measure values using BMNB and SVM classifiers. F-measure values for all the features with BMNB and SVM classifiers for four datasets are shown in Table 1.

For unigram features (*F1*), BMNB is performing better than SVM for all the dataset except movie review dataset. This is because SVM performs better with large feature vector as number of unique terms in movie review dataset is larger over product review datasets (refer Table 2). When we consider *PIGF1* and *PmRMRF1* features, performance of BMNB increased significantly compare to their unigram features because BMNB is very sensitive with the noisy features. If noisy and irrelevant features are removed from the feature vector, BMNB can perform better. Also, performance of SVM is increased compared to its performance with unigram features (refer Table 1). Performance is increased due to IG and mRMR methods removed noisy and irrelevant features from the feature vector which deteriorate the performance of a classifier.

It can be observed from Table 1 that *bi-gram* feature set individually doesn't give better performance as compared to unigram features. However, when prominent *bi-grams* are extracted in *PIGF2* and *PmRMRF2* with IG and mRMR, F-measure values are increased due to the fact that feature selection methods (IG and mRMR) reduce the noisy and irrelevant features.

Further, when composite feature vector *ComF* (combining *unigram* and *bi-gram*) is considered, performance of both the classifier (SVM and BMNB) improves but at the cost of execution overhead as given in Table 1. As Feature vector size of *ComF* features is large, so it is required to filter the irrelevant and noisy features for better classification results. That is done by creating feature vector by combining only prominent features of both unigram and bigrams denoted as *ComPIG, ComPmRMR*. *ComPIG* and *ComPmRMR* features produce significantly good results with small feature vector size. Performance (in terms of F-measure) of *ComPmRMR* presents greater than *ComPIG*. F-measure for BMNB classifier is 82.7% with *unigram (F1)* features, while with the same classifier *ComPmRMR* gives 91.1% (+10.15%) with movie review dataset. Similarly, for other datasets, *ComPmRMR* outperforms other feature selection methods.

mRMR feature selection method performs better than IG as IG selects relevant features based on reduction in uncertainty in identifying the class after knowing the value of the feature. It does not eliminate redundant features. However, mRMR discards redundant features which are highly correlated among features, and retain relevant features having minimum correlation. It is intuitive that when *unigram* and *bi-gram* features are combined, redundancy remains there. So, in case of composite features more information is included but at the cost of redundancy, which is removed with the use of mRMR feature selection method. Since, IG only considers relevancy of the feature with the class, it only includes important features of both *unigram* and *bi-gram* but not considering the effect of redundancy. In case of mRMR method, it includes prominent features of both unigram and bi-gram, with eliminating the redundant features.

Table 1. F-measure (%) for different features sets and feature selection methods

	Movie		***Book***		***DVD***		***Electronics***	
	BMNB	SVM	BMNB	SVM	BMNB	SVM	BMNB	SVM
F1	82.7	84.2	80.9	76.2	78.9	77.3	80.8	76.5
PIGF1	89.2	85.8	89.3	84.2	89.1	84.5	86.4	84.6
PmRMRF1	90.2	87.1	90.1	84.1	90.1	85.3	87.2	84.9
F2	79.2	78.8	68.6	66.8	67.1	68.0	72.6	70.4
PIGF2	81.1	80.4	80.4	75.4	74.8	77.1	79.2	74.9
PmRMRF2	80.1	81.4	81.1	76.0	76.1	75.5	80.2	76.0
ComF	87.0	86.7	82.6	79.5	79.9	79.3	85.2	80.8
ComPIG	90.6	89.2	92.1	87.1	90.4	87.3	91.3	88.1
ComPmRMR	91.1	90.2	92.5	88.3	91.5	88.0	91.8	89.0
P1	80.8	81.1	79.4	77.9	74.0	74.6	78.6	77.5
P2	70.4	68.2	72.5	71.2	68.0	67.9	68.2	66.4
ComP	82.1	82.4	81.4	80.9	77.8	79.0	79.0	81.2

When only adjectives are considered to generate feature vector, it is observed that performance is degraded as compared to *unigrams* features. Adverbs individually are performing worse as compared to adjectives and *unigram* features. Combining Adjectives and adverbs gives performance near to base *unigram* features (refer Table1). Composite features (*ComP*) perform better as compared to the features considered independently with respect to F- measure value.

Both mRMR and IG perform considerably better with optimal features for classifying instances compare to results reported in previous literature. We observed during experiments that mRMR and IG selects approximately 65-70% features in common for all the dataset considered. However, remaining 30-35% features in IG features set were those features, which were correlated with other features. mRMR feature selection method was able to remove those redundant features to included more relevant features which IG method was unable to do. mRMR discards unwanted noisy features and retains only relevant feature with minimum correlation among features. That is why mRMR feature selection method performed better as compared to IG.

Dependency among attributes inevitably decrease the power of NB classifier [11]. mRMR selects the prominent features out of complete feature set those are not correlated among features. It is observed from the experiments that performance of BMNB increased significantly after removing the irrelevant and noisy features. This is due to the fact that prominent features are less likely to be depended among themselves. BMNB after mRMR feature selection method performs best because mRMR feature selection technique is capable of removing the correlation among the features. In addition, BMNB is significantly faster than SVM.

Effect of Feature Vector Size on Classification Performance

For deciding, in what ratio prominent features should be selected from unigrams and bigrams? We empirically experimented with different combination of prominent features vector sizes. It is observed that unigrams are more important than bigram that is also resembles with the results of Table1. So, we decided to include *unigram* and *bi-gram* in 60:40 percent ratio. For example, to create *ComPIG* feature vector size of 1000, top 600 features are selected from *PIGF1* and top 400 features are selected from *PIGF2*.

Table 2. Feature vector size for all the features for different datasets

S.No	Features	Movie Review	Book	DVD	Electronics
1	*F1*	9045	5391	5955	4270
2	*PIGF1 and PmRMRF1*	600	480	720	480
3	*F2*	6050	6484	8888	5513
4	*PIGF2 and PmRMRF2*	400	320	480	320
5	*ComF*	15095	11875	14843	9783
6	*ComPIG and ComPmRMR*	1000	800	1200	800
7	*P1*	1330	1120	1280	980
8	*P2*	377	350	400	310
9	*ComP*	1707	1470	1680	1290

Effect of feature vector size is also experimented with feature selection technique on performance of classifier. Understanding the limitation of space, we report the performance of IG and mRMR for composite features i.e. *ComPmRMR* and *ComPIG* for BMNB classifier since composite features performed best among all the features and BMNB to be better than SVM. Feature vector size for all the features is shown in Table 2. Effect of different feature vector size with IG and mRMR on the performance of BMNB classifiers on different dataset is shown in Figure 1-2.

How many features should be selected for classification is taken based on these observations? For taking this decision, it is observed from Figure 1-2 that if feature size is not reduced much, F-measure value is varying in a narrow range, and that is approximately 10-15% of total features. Therefore, with empirically experimenting, we selected very less number of features for creating feature vector. Feature vector sizes used for our experiments are shown in Table 2.

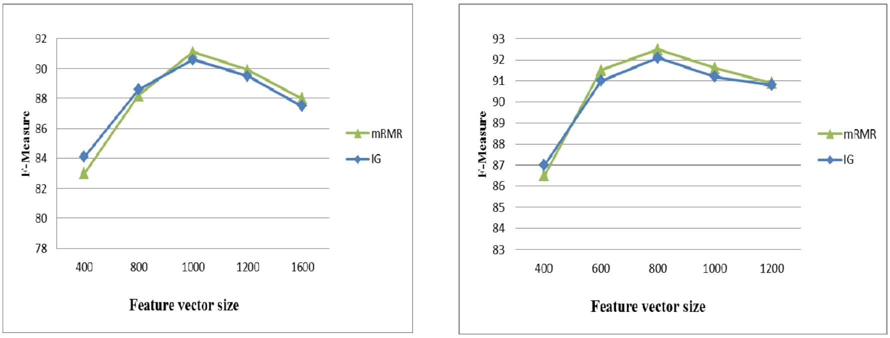

Fig. 1. Effect of feature size for *ComPmRMR* feature with BMNB classifier on Movie Review and book dataset respectively

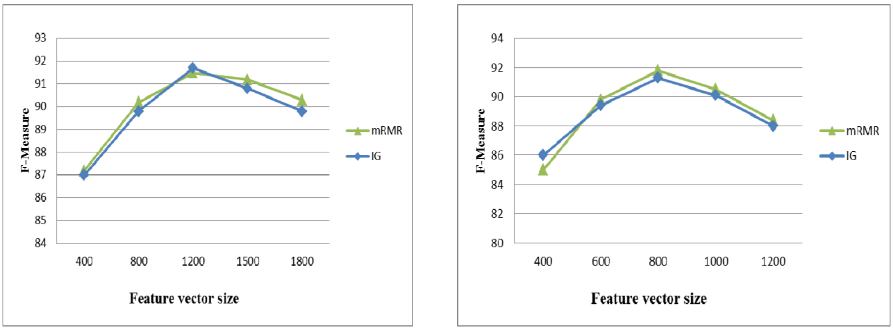

Fig. 2. Effect of feature size for *ComPmRMR* feature with BMNB classifier on DVD and electronics dataset respectively

6 Conclusion

In this paper, different features like unigrams, bigrams, adjectives, adverbs were extracted and composite features were created. Effect of various categories of features was investigated on four different standard dataset of different domains. Composite feature of prominent features of *unigram* and *bi-gram* gives better performance as compared to unigrams, bigrams, adjectives, adverbs individually with respect to F-measure. IG and mRMR feature selection methods are used for extracting predominant features. Comparative performance of IG and mRMR is investigated for sentiment classification, and it is observed that mRMR performs better than IG. It is due to the fact that mRMR feature selection method is capable of selecting relevant features as well as it can eliminate redundant features unlike IG which can only compute importance of the feature. SVM and BMNB classifiers are used for sentiment classification. Performance of BMNB is better as compared to SVM in terms of performance, and significantly better than SVM in terms of execution time. The advantage of using *unigrams* and *bi-grams* over other POS based features are that they are easy to extract, while POS based features require tagger to extract the features, and POS tagging is very slow process. BMNB perfomed best with prominent mRMR composite features (*ComPmRMR*) in terms of execution time and accuracy for sentiment classification. We wish to compare the performance of these features on more datasets of different domain, and also study the affect of proposed method on non-english documents.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), 1–135 (2008)
2. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 79–86 (2002)
3. Agarwal, B., Mittal, N.: Categorical Probability Proportion Difference (CPPD): A Feature Selection Method for Sentiment Classification. In: Proceedings of the 2nd Workshop on Sentiment Analysis where AI Meets Psychology (SAAIP 2012), COLING 2012, Mumbai, pp. 17–26 (2012)
4. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the Association for Computational Linguistics (ACL), pp. 271–278 (2004)
5. Mullen, T., Collier, N.: Sentiment analysis using support vector machines with diverse information sources. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 412–418 (2004)
6. Verma, S., Bhattacharyya, P.: Incorporating semantic knowledge for sentiment analysis. In: Proceedings of ICON 2009, Hyderabad, India (2009)
7. Turney, P.: Thumbs up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: ACL 2002, pp. 417–424 (2002)
8. Porter, M.F.: An Algorithm for Suffix Stripping. Program 14(3), 130–137 (1980)
9. Liu, B.: Sentiment Analysis and Subjectivity. In: Indurkhya, N., Damerau, F.J. (eds.) Handbook of Natural Language Processing, 2nd edn., pp. 627–666. Chapman & Hall (2010)
10. Abbasi, A., France, S., Zhang, Z., Chen, H.: Selecting Attributes for Sentiment Classification Using Feature Relation Networks. IEEE Transactions on Knowledge and Data Engineering 23, 447–462 (2011)

11. Witten, I.H., Frank, E.: Data mining: Practical Machine Learning Tools and techniques. Morgan Kaufmann, San Francisco (2005)
12. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Analysis and Machine Intelligence 27(8), 1226–1238 (2005)
13. Tan, S., Zhang, J.: An empirical study of sentiment analysis for chinese documents. Expert Systems with Applications 34, 2622–2629 (2008)
14. Wang, S., Li, D., Wei, Y., Li, H.: A feature selection method based on fisher's discriminant ratio for text sentiment classification. In: Liu, W., Luo, X., Wang, F.L., Lei, J. (eds.) WISM 2009. LNCS, vol. 5854, pp. 88–97. Springer, Heidelberg (2009)
15. O'Keefe, T., Koprinska, I.: Feature Selection and Weighting Methods in Sentiment Analysis. In: Proceedings of the 14th Australasian Document Computing Symposium (2009)
16. Nicholls, C., Song, F.: Comparison of feature selection methods for sentiment analysis. In: Farzindar, A., Kešelj, V. (eds.) Canadian AI 2010. LNCS, vol. 6085, pp. 286–289. Springer, Heidelberg (2010)
17. Abbasi, A., Chen, H.C., Salem, A.: Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. ACM Transactions on Information Systems 26(3), Article no:12 (2008)
18. Manning, C.D., Raghvan, P., Schutze, H.: Introduction to information retrieval. Cambridge University Press, Cambridge (2008)
19. WEKA. Open Source Machine Learning Software Weka, http://www.cs.waikato.ac.nz/ml/weka/
20. Benamara, F., Cesarano, C., Picariello, A., Reforgiato, D., Subrahmanian, V.S.: Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM) (2007)
21. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI 1995 Proceedings of the 14th International Joint Conference on Artificial Intelligence, vol. 2, pp. 1137–1143 (1995)
22. Blitzer, J., Dredze, M., Pereira, F.: Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: Proceedings of the Association for Computational Linguistics (ACL), pp. 440–447 (2007)