Domain-Based Lexicon Enhancement for Sentiment Analysis

Aminu Muhammad, Nirmalie Wiratunga, Robert Lothian and Richard Glassey

IDEAS Research Institute, Robert Gordon University, Aberdeen UK
{a.b.muhammad1, n.wiratunga, r.m.lothian, r.j.glassey}@rgu.ac.uk

Abstract. General knowledge sentiment lexicons have the advantage of wider term coverage. However, such lexicons typically have inferior performance for sentiment classification compared to using domain focused lexicons or machine learning classifiers. Such poor performance can be attributed to the fact that some domain-specific sentiment-bearing terms may not be available from a general knowledge lexicon. Similarly, there is difference in usage of the same term between domain and general knowledge lexicons in some cases. In this paper, we propose a technique that uses distant-supervision to learn a domain focused sentiment lexicon. The technique further combines general knowledge lexicon with the domain focused lexicon for sentiment analysis. Implementation and evaluation of the technique on Twitter text show that sentiment analysis benefits from the combination of the two knowledge sources. The technique also performs better than state-of-the-art machine learning classifiers trained with distant-supervision dataset.

1 Introduction

Sentiment analysis concerns the study of opinions expressed in text. Typically, an opinion comprises of its polarity (positive or negative), the target (and aspects) to which the opinion was expressed and the time at which the opinion was expressed [14]. Sentiment analysis has a wide range of applications for businesses, organisations, governments and individuals. For instance, a business would want to know customer's opinion about its products/services and that of its competitors. Likewise, governments would want to know how their policies and decisions are received by the people. Similarly, individuals would want make use of other people's opinion (reviews or comments) to make decisions [14]. Also, applications of sentiment analysis have been established in the areas of politics [3], stock markets [1], economic systems [15] and security concerns [13] among others.

Typically, sentiment analysis is performed using machine learning or lexicon-based methods; or a combination of the two (hybrid). With machine learning, an algorithm is trained with sentiment labelled data and the learnt model is used to classify new documents. This method requires labelled data typically generated through labour-intensive human annotation. An alternative approach to generating labelled data called distant-supervision has been proposed [9, 23]. This approach relies on the appearance of certain emoticons that are deemed to signify positive (or negative) sentiment to tentatively labelled documents as positive (or negative). Although, training data generated through distant-supervision have been shown to do well in sentiment classification [9], it is hard to integrate into a machine learning algorithm, knowledge which is not available from its training data. Similarly, it is hard to explain the actual evidence on which a machine learning algorithm based its decision.

The lexicon-based, on the other hand, involves the extraction and aggregation of terms' sentiment scores offered by a lexicon (i.e prior polarities) to make sentiment prediction. Sentiment lexicons are language resources that associate terms with sentiment polarity (positive, negative or neutral) usually by means of numerical score that indicate sentiment dimension and strength. Although sentiment lexicon is necessary for lexicon-based sentiment analysis, it is far from enough to achieve good results [14]. This is because the polarity with which a sentiment-bearing term appears in text (i.e. contextual polarity) could be different from its prior polarity. For example in the text "the movie sucks", although the term 'sucks' seems highly sentiment-bearing, this may not be reflected by a sentiment lexicon. Another problem with sentiment lexicons is that they do not contain domain-specific, sentiment-bearing terms. This is especially more common when a lexicon generated from standard formal text is applied in sentiment analysis of informal text.

In this paper, we introduce lexicon enhancement technique (LET) to address the the afore-mentioned problems of lexicon-based sentiment analysis. LET leverages the success of distant-supervision to mine sentiment knowledge from a target domain and further combines such knowledge with the one obtained from a generic lexicon. Evaluation of the technique on sentiment classification of Twitter text shows performance gain over using either of the knowledge sources in isolation. Similarly, the techniques performs better than three standard machine learning algorithms namely Support Vector Machine, Naive Bayes and Logistic Regression. The main contribution of this paper is two-fold. First, we introduce a new fully automated approach of generating social media focused sentiment lexicon. Second, we propose a strategy to effectively combine the developed lexicon with a general knowledge lexicon for sentiment classification.

The remainder of this paper is organised as follows. Section 2 describes related work. The proposed technique is presented in Section 3. Evaluation and discussions appear in Section 4, followed by conclusions and future work in Section 5.

2 Related Work

Typically, three methods have been employed for sentiment analysis namely machine learning, lexicon based and hybrid. For machine learning, supervised classifiers are trained with sentiment labelled data commonly generated through labour-intensive human annotation. The trained classifiers are then used to classify new documents for sentiment. Prior work using machine learning include the work of Pang et al [20], where three classifiers namely, Naïve Bayes (NB), Maximum Entropy (ME) and Support Vector Machines (SVMs) were used for the task. Their results show that, like topic-based text classification, SVMs perform better than NB and ME. However, performance of all the three classifiers in sentiment classification is lower than in topic-based text classification. Document representation for machine learning is an unordered list of terms that appear in the documents (i.e. bag-of-words). A binary representation based on term presence or absence attained up to 87.2% accuracy on a movie review dataset [18]. The addition of phrases that are used to express sentiment (i.e. appraisal groups) as additional features in the binary representation resulted in further improvement of 90.6% [32] while best result of 96.9% was achieved using term-frequency/inverse-document-frequency (tf/idf) weighting [19]. Further sentiment analysis research using machine learning attempt to improve classification accuracy with feature selection mechanisms. An approach for selecting bi-gram features was introduced in [16]. Similarly, feature space reduction based on subsumption hierarchy was introduced in [24]. The aforementioned works concentrate on sentiment analysis of reviews, therefore, they used star-rating supplied with reviews to label training and test data instead of hand-labelling. This is typical with reviews, however, with other forms of social media (e.g. discussion forums, blogs, tweets e.t.c.), star-rating is typically unavailable. Distant-supervision has been employed to generate training data for sentiment classification of tweets [9, 23]. Here, emoticons supplied by authors of the tweets were used as noisy sentiment labels. Evaluation results on NB, ME and SVMs trained with distant-supervision data but tested on hand-labelled data show the approach to be effective with ME attaining the highest accuracy of 83.0% on a combination of unigram and bigram features. The limitation of machine learning for sentiment analysis is that it is difficult to integrate into a classifier, general knowledge which may not be acquired from training data. Furthermore, learnt models often have poor adaptability between domains or different text genres because they often rely on domain specific features from their training data. Also, with the dynamic nature of social media, language evolves rapidly which may render a previous learning less useful.

The lexicon based method excludes the need for labelled training data but requires sentiment lexicon which several are readily available. Sentiment lexicons are dictionaries that associate terms with sentiment values. Such lexicons are either manually generated or semi-automatically generated from generic knowledge sources. With manually generated lexicons such as General Inquirer [25] and Opinion Lexicon [12], sentiment polarity values are assigned purely by humans and typically have limited coverage. As for the semi-automatically generated lexicons, two methods are common, *corpus-based* and *dictionary-based*. Both methods begin with a small set of seed terms. For example, a positive seed set such as ‘good’, ‘nice’ and ‘excellent’ and a negative seed set could contain terms such as ‘bad’, ‘awful’ and ‘horrible’. The methods leverage on language resources and exploit relationships between terms to expand the sets. The two methods differ in that corpus-based uses collection of documents while the dictionary-based uses machine-readable dictionaries as the lexical resource. Corpus-based was used to generate sentiment lexicon [11]. Here, 657 and 679 adjectives were manually annotated as positive and negative seed sets respectively. Thereafter, the sets were expanded to conjoining adjectives in a document collection based on the connectives ‘and’ and ‘but’ where ‘and’ indicates similar and ‘but’ indicates contrasting polarities between the conjoining adjectives. Similarly, a sentiment lexicon for phrases generated using the web as a corpus was introduced in [29, 30]. Dictionary-based was used to generate sentiment lexicon in [2, 31]. Here, relationships between terms in WordNet [8] were explored to expand positive and negative seed sets. Both corpus-based and dictionary-based lexicons seem to rely on standard spelling and/or grammar which are often not preserved in social media [27].

Lexicon-based sentiment analysis begins with the creation of a sentiment lexicon or the adoption of an existing one, from which sentiment scores of terms are extracted and aggregated to predict sentiment of a given piece of text. Term-counting approach has been employed for the aggregation. Here, terms contained in the text to be classified are categorised as positive or negative and the text is classified as the class with highest number of terms [30]. This approach does not account for varying sentiment intensities between terms. An alternative approach is the aggregate-and-average strategy [26]. This classifies a piece of text as the class with highest average sentiment of terms. As lexicon-based sentiment analysis often rely on generic knowledge sources, it tends to perform poorly compared to machine learning.

Hybrid method, in which some elements from machine learning and lexicon based are combined, has been used in sentiment analysis. For instance, sentiment polarities of terms obtained from lexicon were used as additional features to train machine learning classifiers [5, 17]. Similarly, improvement was observed when multiple classifiers formed from different methods are used to classify a document [22]. Also, machine learning was employed to optimize sentiment scores in a lexicon [28]. Here, initial score for terms, assigned manually are increased or decreased based on observed classification accuracies.

3 Lexicon Enhancement Technique

Lexicon enhancement technique (LET) addresses the semantic gap between generic and domain knowledge sources. As illustrated in Fig. 1, the technique involves obtaining scores from a *generic lexicon*, automated domain *data labelling using distant-supervision*, *domain lexicon generation* and *aggregation strategy for classification*. Details of these components is presented in the following sub sections.

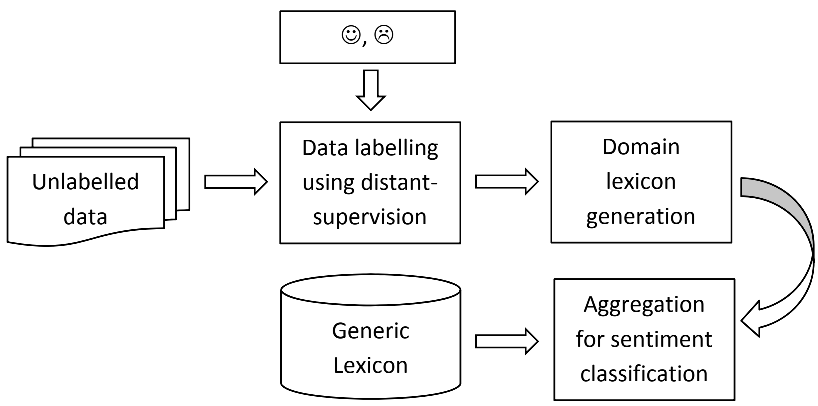

Fig. 1. Diagram showing the architectural components of the proposed technique (LET)

3.1 Generic Lexicon

We use SentiWordNet [7] as the source of generic sentiment scores for terms. SentiWordNet is a general knowledge lexicon generated from WordNet [8]. Each synset (i.e. a group of synonymous terms based on meaning) in WordNet is associated with three numerical scores indicating the degree of association of the synset with positive, negative and objective text. In generating the lexicon, seed (positive and negative) synsets were expanded by exploiting *synonymy* and *antonymy* relations in WordNet, whereby synonymy preserves while antonymy reverses the polarity with a given synset. As there is no direct synonym relation between synsets in WordNet, the relations: see_also, similar_to, pertains_to, derived_from and attribute were used to represent synonymy relation while direct antonym relation was used for the antonymy. Glosses (i.e. textual definitions) of the expanded sets of synsets along with that of another set assumed to be composed of objective synsets were used to train eight ternary classifiers. The classifiers are used to classify every synset and the proportion of classification for each class (positive, negative and objective) were deemed as initial scores for the synsets. The scores were optimised by a random walk using the PageRank [4] approach. This starts with manually selected synsets and then propagates sentiment polarity (positive or negative) to a target synset by assessing the synsets that connect to the target synset through the appearance of their terms in the gloss of the target synset. SentiWordNet can be seen to have a tree structure as shown in Fig. 2. The root node of the tree is a term whose child nodes are the four basic PoS tags in WordNet (i.e. noun, verb, adjective and adverb). Each PoS can have multiple word senses as child nodes. Sentiment scores illustrated by a point within the triangular space in the diagram are attached to word-senses. Subjectivity increases (while objectivity decreases) from lower to upper, and positivity increases (while negativity decreases) from right to the left part of the triangle.

We extract scores from SentiWordNet as follows. First, input text is broken into unit tokens (tokenization) and each token is assigned a lemma (i.e. corresponding dictionary entry) and PoS using Stanford CoreNLP library[1]. Although in SentiWordNet scores are associated with word-senses, disambiguation is usually not performed as it does not seem to yield better results than using either the average score across all senses of a term-PoS or the score attached to the most frequent sense of the term (e.g. in [21], [17], [6]). In this work, we use average positive (or negative) score at PoS level as the positive (or negative) for terms as shown in Equation 1.

$$gs(t)_{dim} = \frac{\sum_{i=1}^{|senses(t,PoS)|} ScoreSense_i(t,PoS)_{dim}}{|senses(t,PoS)|} \tag{1}$$

Where $gs(t)_{dim}$ is the score of term t (given its part-of-speech, *PoS*) in the sentiment dimension of dim (dim is either positive or negative). $ScoreSense_i(t,PoS)_{dim}$ is the sentiment score of the term t for the part-of-speech (*PoS*) at sense i. Finally, $|senses(t,PoS)|$ is number of word senses for the part-of-speech (*PoS*) of term t.

[1] nlp.stanford.edu/software/corenlp.shtml

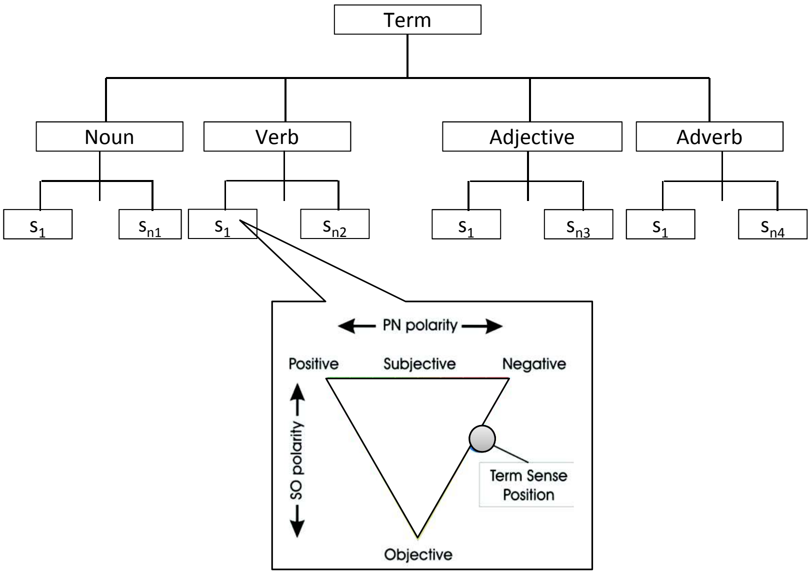

Fig. 2. Diagram showing the structure of SentiWordNet

3.2 Data Labelling Using Distant-Supervision

Distant-supervision offers an automated approach to assigning sentiment class labels to documents. It uses emoticons as noisy labels for documents. It is imperative to have as many data as possible at this stage as this affects the reliability of scores to be generated at the subsequent stage. Considering that our domain of focus is social media, we assume there will be many documents containing such emoticons and, therefore, large dataset can be formed using the approach. Specifically, in this work we use Twitter as a case study. We use a publicly available distant-supervision dataset for this stage [9][2]. This dataset contains 1,600,000 tweets balanced for positive and negative sentiment classes. We selected first 10,000 tweets from each class for this work. This is because the full dataset is too big to conveniently work with. For instance, building a single machine learning model on the full dataset took several days on a machine with 8GB RAM, 3.2GHZ Processor and 64bit Operating System. However, we aim to employ "big data" handling techniques to experiment with larger datasets in the future. The dataset is preprocessed to reduce feature space using the approach introduced in [9]. That is, all user names (i.e. words that starts with the @ symbol) are replaced with the token 'USERNAME'. Similarly all URLs (e.g. "http://tinyurl.com/cvvg9a") are replaced with the token 'URL'. Finally, words consisting of sequence of three or more repeated character (e.g. "haaaaapy") are normalised to contain only two of such repeated character in sequence.

[2]The dataset available from Sentiment140.com

3.3 Domain Lexicon Generation

Domain sentiment lexicon is generated at this stage. Each term from the distant-supervision dataset is associated with positive and negative scores. Positive (or negative) score for a term is determined as the proportion of the term's appearance in positive (or negative) documents given by equation 2. Separate scores for positive and negative classes are maintain in order to suit integration with the scores obtained from the generic lexicon (SentiWordNet). Table 1 shows example terms extracted from the dataset and their associated positive and negative scores.

$$ds(t)_{dim} = \frac{\sum\limits_{dim} \mathrm{tf}(t)}{\sum\limits_{Alldocuments} \mathrm{tf}(t)} \tag{2}$$

Where $ds(t)_{dim}$ is the sentiment score of term t for the polarity dimension dim (positive or negative) and $\mathrm{tf}(t)$ is document term frequency of t.

Table 1. Some terms from the domain lexicon

Term	Sentiment Scores	
	Positive	Negative
ugh	0.077	0.923
sucks	0.132	0.868
hehe	0.896	0.104
damn	0.241	0.759
argh	0.069	0.931
thx	1	0
luv	0.958	0.042
xoxo	0.792	0.208

3.4 Aggregation Strategy for Sentiment Classification

At this stage, scores from generic and domain lexicons for each term t are combined for sentiment prediction. The scores are combined so as to complement each other according to the following strategy.

$$Score(t)_{dim} = \begin{cases} 0, & \text{if } gs(t)_{dim} = 0 \text{ and } ds(t)_{dim} = 0 \\ gs(t)_{dim}, & \text{if } ds(t)_{dim} = 0 \text{ and } gs(t) > 0 \\ ds(t)_{dim}, & \text{if } gs(t)_{dim} = 0 \text{ and } ds(t) > 0 \\ \alpha \times gs(t)_{dim} + (1-\alpha) \times ds(t)_{dim}, & \text{if } gs(t)_{dim} > 0 \text{ and } ds(t)_{dim} > 0 \end{cases}$$

The parameter, α, controls a weighted average of generic and domain scores for t when both scores are non-zero. In this work we set α to 0.5 thereby giving equal weights to both scores. However, we aim to investigate an optimal setting for the parameter in the future. Finally, sentiment class for a document is determined using aggregate-and-average method as outlined in Algorithm 1.

Algorithm 1 Sentiment Classification

1: **INPUT:** *Document*
2: **OUTPUT:** *class* ▷ document sentiment class
3: Initialise: *posScore*, *negScore*
4: **for all** $t \in Document$ **do**
5:     **if** $Score(t)_{pos} > 0$ **then**
6:         $posScore \leftarrow posScore + Score(t)_{pos}$
7:         $nPos \leftarrow nPos + 1$ ▷ increment number of positive terms
8:     **end if**
9:     **if** $Score(t)_{neg} > 0$ **then**
10:         $negScore \leftarrow negScore + Score(t)_{neg}$
11:         $nNeg \leftarrow nNeg + 1$ ▷ increment number of negative terms
12:     **end if**
13: **end for**
14: **if** $posScore/nPos > negScore/nNeg$ **then return** *positive*
15: **else** **return** *negative*
16: **end if**

4 Evaluation

We conduct a comparative study to evaluate the proposed technique (LET). The aim of the study is three fold, first, to investigate whether or not combining the two knowledge sources (i.e. LET) is better than using each source alone. Second, to investigate performance of LET compared to that of machine learning algorithms trained with distant-supervision data since that is the state-of-the-art use of distant-supervision for sentiment analysis. Lastly, to study the behaviour of LET on varying dataset sizes. We use hand-labelled Twitter dataset, introduced in [9][3] for the evaluation. The dataset consists of 182 positive and 177 negative tweets.

4.1 LET Against Individual Knowledge Sources

Here, the following settings are compared:

1. LET: The proposed technique (see Algorithm 1)
2. Generic: A setting that only utilises scores obtained from the generic lexicon (SentiWorNet). In Algorithm 1, $Score(t)_{pos}$ (line 5) and $Score(t)_{neg}$ (line 9) are replaced with $gs(t)_{pos}$ and $gs(t)_{neg}$ respectively.
3. Domain: A setting that only utilises scores obtained from the domain lexicon. In Algorithm 1, $Score(t)_{pos}$ (line 5) and $Score(t)_{neg}$ (line 9) are replaced with $ds(t)_{neg}$ and $ds(t)_{neg}$ respectively.

Table 2 shows result of the comparison. The LET approach performs better than Generic and Domain. This is not suprising since LET utilises generic knowledge which could have been omitted by Domain and also, domain knowledge which could have

[3]The dataset is available from Sentiment140.com

Table 2. Performance accuracy of individual knowledge sources and LET

Generic	Domain	LET
60.33	71.26	75.27

been omitted by Generic. Also the result shows that the generated domain lexicon (Domain) is more effective than the general knowledge lexicon (Generic) for sentiment analysis.

4.2 LET Against Machine Learning and Varying Dataset Sizes

Three machine learning classifiers namely Naïve Bayes (NB), Support Vector Machine (SVM) and Logistic Regression (LR) are trained with the distant-supervision dataset and then evaluated with the human-labelled test dataset. These classifiers are selected because they are the most commonly used for sentiment classification and typically perform better than other classifiers. We use presence and absence (i.e. binary) feature representation for documents and Weka [10] implementation for the classifiers. Furthermore, we use subsets of the distant-supervision dataset (16000, 12000, 8000 and 4000; also balanced for positive and negative classes) in order to test the effect of varying distant-supervision dataset sizes for LET (in domain lexicon generation, see Section 3.3) and the machine learning classifiers.

Table 3. LET compared to machine learning methods on varying data sizes

Classifier / Dataset size	NB	SVM	LR	LET
4,000	60.17	61.00	66.02	68.70
8,000	54.04	59.61	69.64	73.10
12,000	54.04	62.12	71.03	73.80
16,000	54.04	62.95	71.87	75.27
20,000	54.60	62.40	73.26	75.27

Table 3 shows result of the experiment. LET performs better than any of the machine learning classifiers. This can be attributed to the fact that LET utilises generic knowledge which the machine learning classifiers could not have acquired from the training dataset, especially, as the distant-supervision dataset may contain incorrect labels. As for the behaviour of LET and the classifiers on varying dataset sizes, they all tend to improve in performance with increased dataset size as depicted by Fig. 3, with the exception of SVM for which the performance drops. Interestingly however, the difference between the algorithms appeared to be maintained over the different dataset sizes. This shows that the domain lexicon generated in LET becomes more accurate with increased dataset size in a similar manner that a machine learning classifier becomes more accurate with increased training data.

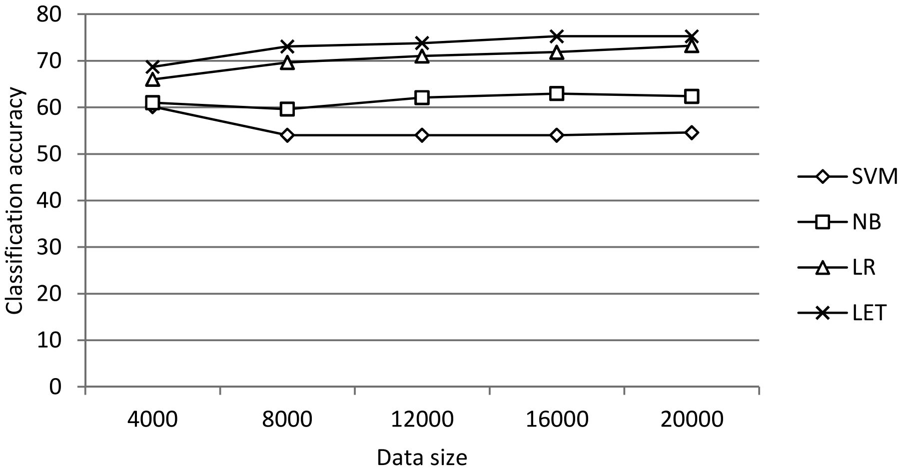

Fig. 3. LET compared to machine learning methods on varying data sizes

5 Conclusions and Future Work

In this paper, we presented a novel technique for enhancing generic sentiment lexicon with domain knowledge for sentiment classification. The major contributions of the paper are that we introduced a new approach of generating domain-focused lexicon which is devoid of human involvement. Also, we introduced a novel strategy to combine generic and domain lexicons for sentiment classification. Experimental evaluation shows that the technique is effective and better than state-of-the-art machine learning sentiment classification trained the same dataset from which our technique extracts domain knowledge (i.e. distant-supervision data).

As part of future work, we plan to conduct an extensive evaluation of the technique on other social media platforms (e.g. discussion forums) and also, to extend the technique for subjective/objective classification. Similarly, we intend perform experiment to find an optimal setting for α and improve the aggregation strategy presented.

References

[1] Arnold, I., Vrugt, E.: Fundamental uncertainty and stock market volatility. Applied Financial Economics 18(17), 1425–1440 (2008)
[2] Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Annual Conference on Language Resouces and Evaluation (2010)
[3] Baron, D.: Competing for the public through the news media. Journal of Economics and Management Strategy 14(2), 339–376 (2005)
[4] Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: Seventh International World-Wide Web Conference (WWW 1998) (1998)
[5] Dang, Y., Zhang, Y., Chen, H.: A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. IEEE Intelligent Systems 25, 46–53 (2010)

[6] Denecke, K.: Using sentiwordnet for multilingual sentiment analysis. In: ICDE Workshop (2008)
[7] Esuli, A., Baccianella, S., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC10) (2010)
[8] Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press (1998)
[9] Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Processing pp. 1–6 (2009)
[10] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. SIGKDD Explor. Newsl. 11(1), 10–18 (Nov 2009)
[11] Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: Proceedings of the 35th Annual Meeting of the ACL and the 8th Conference of the European Chapter of the ACL. pp. 174–181. New Brunswick, NJ (1997)
[12] Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 168–177 (2004)
[13] Karlgren, J., Sahlgren, M., Olsson, F., Espinoza, F., Hamfors, O.: Usefulness of sentiment analysis. In: 34th European Conference on Information Retrieval (2012)
[14] Liu, B.: Sentiment Analysis and Subjectivity, chap. Handbook of Natural Language Processing, pp. 627–666. Chapman and Francis, second edn. (2010)
[15] Ludvigson, S.: Consumer confidence and consumer spending. The Journal of Economic Perspectives 18(2), 29–50 (2004)
[16] Mukras, R., Wiratunga, N., Lothian, R.: Selecting bi-tags for sentiment analysis of text. In: Proceedings of the Twenty-seventh SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence (2007)
[17] Ohana, B., Tierney, B.: Sentiment classification of reviews using sentiwordnet. In: 9th IT&T Conference, Dublin, Ireland (2009)
[18] Pang, B., Lee, L.: Polarity dataset v2.0, 2004. online (2004), http://www.cs.cornell.edu/People/pabo/movie-review-data/.
[19] Pang, B., Lee, L.: Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1), 1–135 (2008)
[20] Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods on Natural Language Processing (2002)
[21] Pera, M., Qumsiyeh, R., Ng, Y.K.: An unsupervised sentiment classifier on summarized or full reviews. In: Proceedings of the 11th International Conference on Web Information Systems Engineering. pp. 142–156 (2010)
[22] Prabowo, R., Thelwall, M.: sentiment analysis: A combined approach. Journal of Informetrics 3(2), 143–157 (2009)
[23] Read, J.: Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL Student Research Workshop. pp. 43–48. ACLstudent '05, Association for Computational Linguistics, Stroudsburg, PA, USA (2005)
[24] Riloff, E., Patwardhan, S., Wiebe, J.: Feature subsumption for opinion analysis. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP-06) (2006)
[25] Stone, P.J., Dexter, D.C., Marshall, S.S., Daniel, O.M.: The General Inquirer: A Computer Approach to Content Analysis. MIT Press, Cambridge, MA (1966)
[26] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational Linguistics 37, 267–307 (2011)

[27] Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology 63(1), 163–173 (2012)

[28] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., Kappas, A.: Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology 61(12), 2444–2558 (2010)

[29] Turney, P., et al.: Mining the web for synonyms: Pmi-ir versus lsa on toefl. In: Proceedings of the twelfth european conference on machine learning (ecml-2001) (2001)

[30] Turney, P.D.: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics. pp. 417–424 (2002)

[31] Valitutti, R.: Wordnet-affect: an affective extension of wordnet. In: In Proceedings of the 4th International Conference on Language Resources and Evaluation. pp. 1083–1086 (2004)

[32] Whitelaw, C., Garg, N., Argamon., S.: Using appraisal groups for sentiment analysis. In: 14th ACM International Conference on Information and Knowledge Management (CIKM 2005). pp. 625–631 (2005)