

Domain Sentiment Matters: A Two Stage Sentiment Analyzer

Raksha Sharma and Pushpak Bhattacharyya

Dept. of Computer Science and Engineering

IIT Bombay, Mumbai, India

{raksha, pb}@cse.iitb.ac.in

Abstract

There are words that change its polarity from domain to domain. For example, the word *deadly* is of positive polarity in the cricket domain as in “Shane Warne is a ‘deadly’ leg spinner”. However, ‘I witnessed a deadly accident’ carries negative polarity and going by the sentiment in cricket domain will be misleading. In addition to this, there exist *domain-specific* words, which have the same polarity across domains, but are used very frequently in a particular domain. For example, *blockbuster*, is specific to the movie domain. We combine such words as *Domain Dedicated Polar Words (DDPW)*.

A concise feature set made up of principal polarity clues makes the classifier less expensive in terms of time complexity and enhances the accuracy of classification. In this paper, we show that DDPW make such a concise feature set for sentiment analysis in a domain. Use of domain-dedicated polar words as features beats the state of art accuracies achieved independently with unigrams, adjectives or Universal Sentiment Lexicon (USL).

1 Introduction

The general approach of Sentiment Analysis (SA) is to summarize the semantic polarity (i.e., positive or negative) of sentences/documents by analysis of the orientation of the individual words (Riloff and Wiebe, 2003; Pang and Lee, 2004; Danescu-Niculescu-Mizil et al., 2009; Kim and Hovy, 2004; Takamura et al., 2005). In the real world, most sentiment analysis applications are domain oriented. All business organizations are interested in sentiment information about the product they deal with. For instance, an automobile organiza-

tion is concerned only about recognizing the sentiment information received for automobiles only. Therefore, a list of Domain Dedicated Polar Words (DDPW) can be proved as the best lexical resource for domain oriented sentiment analysis.

Most sentiment analysis applications rely on the Universal Sentiment Lexicons (USL) as a key feature along with additional features (Riloff and Wiebe, 2003). There are many USL resources like senti-word-net¹, subjectivity lexicon² by Wiebe and a list of positive and negative opinion words³ by Liu. These lexicons contain only those words that are usual and have the same polarity across all the domains. These universal sentiment lexicons have the following problems:

- The words that have fluctuating polarity across domains, but have fixed polarity in a domain are strong candidate for the sentiment analysis in that domain. We call such words *chameleon words*. Consider the following example of *fluctuating polarity* phenomenon.
 1. *The cars steering was unpredictable while driving.* (-ve sentiment)
 2. *The story line of Palmetto was unpredictable.* (+ve sentiment)

The word *unpredictable* bears *negative* polarity in the automobile domain, but it is *positive* in the movie domain. Hence, *unpredictable* assigns negative polarity to the first sentence and positive polarity to the second sentence. Due to the absence of *chameleon* words like *unpredictable*, the USL based classifier fails to determine the correct polarity of the sentences that contain chameleon words.

- On the other hand, consistency in use of a polar word in a particular domain, makes it a

¹<http://sentiwordnetisti.cnr.it/>

²<http://mpqa.cs.pitt.edu/>

³<http://www.cs.uic.edu/liub/FBS/sentiment-analysis.html>

very strong candidate for sentiment analysis in that domain. Consider the following example of *fluctuating regularity* (frequency of usage) phenomenon.

1. *It's another summer **blockbuster** with plot points that are beyond unbelievable.* (+ve sentiment)
2. *The main Characters were **miscast*** (-ve sentiment)

The words *blockbuster* and *miscast* are used very frequently in the movie domain to express the opinion in comparison with other domains. The absenteeism of such strong polarity clues for a domain, makes the USL impoverished for sentiment analysis in a particular domain.

We combine the chameleon words and the domain specific regular words into a single unit: *Domain-Dedicated polar words (DDPW)*. The DDPW are missing from the USL, because of their fluctuating polarity and regularity across domains.

In this paper, we present sentiment analysis in a domain as a two stage process. Identification of domain-dedicated polar words prior to implementation of classification algorithms leads to less expensive and more efficient sentiment analysis system (Section 2). We examine the role of domain-dedicated polar words in three domains: movie, tourism and product. Our results show that use of domain-dedicated polar words as features either beats or equates the accuracy achieved independently with unigrams, adjectives or USL in all the three domains (Section 6). The two stage approach is depicted in the figure 1.

The first stage implements the Chi-Square⁴ test on the difference in the counts of the word in positive and negative documents to detect domain-dedicated polar words. The second stage accomplishes the sentiment analysis task using the output of the first stage as features. We experimented with three standard classification algorithms: Neural Network (NN) classification, Logistic Regression (LR) classification and Support Vector Machine (SVM) classification. Accuracy figures (Section 6) substantiate the effectiveness of the *two stage sentiment analysis in a domain* in comparison with the single stage sentiment analysis that relies on standard features like, universal sentiment lexicons, adjectives or unigrams.

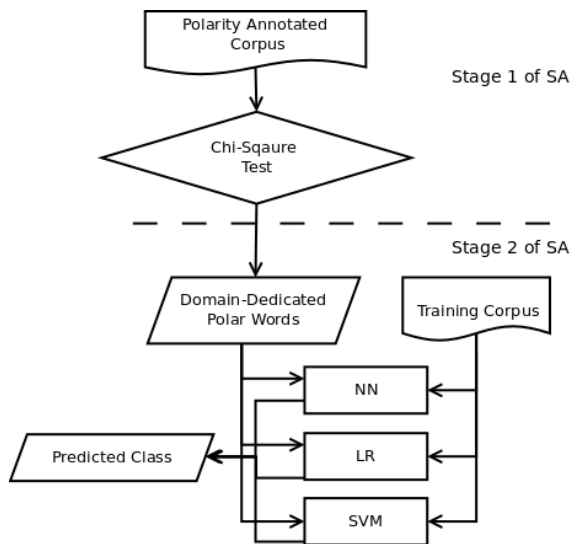


Figure 1: The two stage approach for SA in a domain

In this paper, we use subjectivity lexicon given by Wiebe as universal sentiment lexicon. Section 2 helps illustrate the reason behind improvement in accuracy with a very small feature set: domain-dedicated polar words. In section 3, we formulate the Chi-Square test to depict the generation of DDPW. Section 4 expands on the ML based classification algorithms that are used in stage-2. Section 5 and 6 illustrate the experimental set up and results respectively. Sections 7 and 8 discuss related work and conclusion.

2 Artifacts of Domain Significance in Sentiment Analysis

A feature is a piece of information that is potentially useful for prediction. Coming up with a big feature set increases the time complexity of the classifier. In addition to this, presence of irrelevant or redundant features misleads the classifier, hence, results in a poor accuracy. Pang et al. (2002) observed that the top 2633 unigrams are better features than unigrams or adjectives for sentiment classification of a document. They also proved that ‘term presence’ is more informative than ‘term frequency’ for sentiment prediction of a document.

A concise feature set made up of principal polarity clues makes the classifier less expensive in terms of time complexity and enhances the accuracy of classification. Our work falls in the same series, that is, identification of decisive components for sentiment analysis in a domain. As

⁴<http://math.hws.edu/javamath/ryan/ChiSquare.html> 131

most of the applications of SA are domain specific, therefore, we can restrict ourselves to a domain for SA. For domain oriented sentiment analysis, we can come up with more prominent features, such as domain dedicated polar words. The following examples⁵ from the movie domain help illustrate the problem we attempt to address.

juvenile, surreal, unpredictable, predictable, timeless, thrilling, well-made, well-written

The exemplified words are highly polar in the movie domain, but because of their fluctuating polarity and regularity in use are not included in most of the existing universal sentiment lexicons. We tested for the universal sentiment lexicon given by Wiebe and find 664 words that are not present in the USL, but are extracted as DDPW. The words shown in the example are a few of them. On the other hand, these features can be extracted as a subset of unigrams or adjectives from the input corpus, but at the cost of higher training time complexity and poor generalization in classification. Therefore, we define identification of domain-dedicated polar words as the foremost step (stage-1) for sentiment analysis in a domain.

In literature, Unigrams are considered as state-of-art features for sentiment classification, we are able to achieve the same level of accuracy with DDPW as features. However, instead of words, one uses word senses (synset ids in WordNets) as features, the accuracy improves dramatically. Balamurali et al. (2011) reported accuracy above 85% with (*sense + words*) as features. However, accuracy accomplishment is a function of investment in annotation. This improvement is not significant enough to justify the cost of annotation.

However, the criterion of domain-dedication does not equally exist with all the polar words. There are words that have uniform polarity and regularity (frequency of usage) across domains. This phenomenon is considered implicitly by our proposed approach of DDPW extraction. Consequently, we are able to extract deterministic polar words that have uniform polarity and regularity across domains. Consider the following example from the movie domain.

enjoyable, entertaining, magnificent, impressive, irritating, awful, annoying, weakest

⁵All the examples reported in the paper are part of the output obtained through the Chi-Square test in stage-1. 132

The significant occurrence of such deterministic polar words in a particular class (positive or negative) in the movie review corpus assures the satisfiability criterion of the Chi-Square test, hence, their extraction as DDPW in stage-1 of the proposed SA system.

3 Identification of Domain Dedicated Polar Words

The orientation of the polarity of a word and its frequency of usage vary from domain to domain. Such domain-dedicated polar words are the important clues for sentiment analysis in that domain. This section illustrates the stage-1 of the two-stage sentiment analysis approach that generates domain-dedicated polar words. We have performed an extensive evaluation of DDPW participation in sentiment analysis using three domains: movie, product and tourism.

3.1 Domain and Dataset

Providing polarity information about movie reviews is a very useful service (Turney, 2002). Its proof is the continuously growing popularity of the several film review websites⁶⁷. For movie domain, we use 1000 positive and 1000 negative reviews⁸. Product reviews directly affect the business of e-commerce organizations. For product domain, we use 1000 positive and 1000 negative reviews (music instruments) from Amazon, used by Blitzer et al. (2007). The third domain is the tourism domain, a more accurate sentiment analysis in tourism domain can suggest a more accurate place for visit. We use 700 positive and 700 negative tourism reviews, used by Khapra et al. (2010) to train a word sense disambiguation system. In this paper, we report domain-dedicated words for the movie, product and tourism domain and show that these words are better features than universal sentiment lexicons, unigrams and adjectives for sentiment analysis in the respective domain.

3.2 Chi-Square Test

The Chi-Square test is a statistical test to identify the class (positive/negative) of the encountered word. We use Chi-Square test to extract domain-dedicated words from the corpus in stage-1 of the proposed sentiment analyzer. As Chi-square test

⁶www.rottentomatoes.com

⁷www.imdb.com

⁸Available at: [www.cs.cornell.edu/people/pabo/movie-review-data/\(review corpus version 2.0\)](http://www.cs.cornell.edu/people/pabo/movie-review-data/(review%20corpus%20version%202.0)).

requires values of two parameters, that are, expected count and observed count, we consider the arithmetic mean of the count in positive and negative files as expected count of the word in positive and negative classes, which is also a null hypothesis. The alternative hypothesis states that there is a statistically significant difference between the observed count and the mean value.

On the basis of the deviation of the observed count from the mean value (expected count), Chi-square test decides the polarity of a word for a particular domain to which documents belong. The statistically significant deviation resulted from the Chi-Square test shows that the word appears in a particular class of documents more frequently. This appearance is not by chance, rather, there is some reason behind its occurrence in that class of documents. The reason behind this significant deviation is the polarity orientation of the word that makes it a part of positive or negative documents more frequently. For example, *thought-provoking*, *superb*, *thrilling*, *tremendous* have positive polarity in the movie domain, so they would occur more frequently in the positive reviews rather than negative reviews (Sharma and Bhattacharyya, 2013). The Chi-Square test is formulated as follows:

$$X^2(W) = ((C_p - \mu)^2 + (C_n - \mu)^2) / \mu \quad (1)$$

Here, C_p is the count of a word W in the positive documents and C_n is the count in negative documents. μ represents an average of the word's count in positive and negative documents. μ is the expected count or null hypothesis, while C_p and C_n are the observed count of W . If the Chi-square test results in a value that is greater than the threshold value, then there is a significant difference between the expected and the observed. Since there is an inverse relation⁹ between the Chi-Square value and the *probability of word given NULL hypothesis is true*, a high Chi-Square value indicates that the probability is very poor. Therefore, we reject NULL hypothesis, that is, we reject the uniform distribution of the word in positive and negative class. However, we assume that the word W belongs to a particular class (alternative hypothesis), either positive or negative.

To understand the identification of DDPW from the corpus, Consider the example of Chi-Square

⁹The Chi-square score and probability table is given at <http://faculty.southwest.tn.edu/jwilliams/probab2.gif>

test performed for the word *unpredictable* in the movie domain. The count of the word *unpredictable* in positive (22) and negative files (10) are taken from the considered movie review corpus.

$$X^2(Unpredictable) = ((22-16)^2 + (10-16)^2) / 16 \quad (2)$$

The word “*unpredictable*” results in a Chi-Square value of 4.5, that is greater than 3.84 (Standard Threshold Value in Statistics). This relation implies that $P(Data|NULL-Hypothesis\ is\ true)$ is less than 0.05 (5%) for the word *unpredictable*. Hence, reject NULL hypothesis and accept the alternative hypothesis, that is, the word *unpredictable* belongs to a particular class.

Bruce and Wiebe (1999) proved that ‘adjectives’ are the best candidate to adhere the polarity. They established a statistically significant correlation between sentence subjectivity and the presence of adjectives. At this stage, we also have considered adjectives¹⁰ only as domain-dedicated words, but we believe that domain-dedicated words are not limited to adjective only. The same approach can be applied to find domain-dedicated words from other part of speeches.

4 Sentiment Classification in stage-2

The final class of the document is predicted in stage-2 of the proposed sentiment analysis system. Our utmost goal is to examine the behavior of SA system using a concise feature set: *Domain Dedicated Polar Words (DDPW)*, which are extracted as output of stage-1. For this purpose, we experimented with three machine learning based classification algorithms, that are, *Neural Network*, *Logistic Regression* and *SVM*.

4.1 Neural Network (NN)

Neural networks are able to produce a complex non linear hypothesis function for classification. Nowadays, NN has become “state-of-the-art” technique for many applications because of the fast computers that can solve a big network, (Yanagimoto et al., 2013; Hui, 2011). In our case also, the classification accuracy attained by Neural Network surpasses SVM and LR.

¹⁰Bidirectional Stanford POS tagger is used to find the part of speech of the word.

4.2 Logistic Regression (LR)

LR classifier is a non linear classifier. Non linearity is achieved through sigmoid function (equation 3) that estimates the probability of the document belonging to a class. If LR results in a probability value higher than 0.5, it implies that the document has more than 50% chance of being positive, else the document bears negative polarity. The Logistic (Sigmoid) function simulated by LR is as follows.

$$\text{Hypothesis}(X) = 1/(1 + e^{(-\text{Theta}' * X)}) \quad (3)$$

Here, X is the input feature vector of a document (Section 4), Theta is also a vector that contains the weights assigned to X .

4.3 Support Vector Machine (SVM)

The support vector machine (SVM) has been proven to be highly effective at traditional text categorization and sentiment classification (Joachims, 1998; Pang et al., 2002). SVM predicts the class of the document on the basis of a linear function, that is, z 's score. LR takes a decision at z equals to zero, while SVM takes a decision at two boundaries: z equals to +1 and z equals to -1.

$$z = \text{Theta}' * X \quad (4)$$

5 Experimental Setup

For all the three observed domains: movie, product and tourism, we divide the data into five equal size folds, maintaining the balance of the negative and positive files in each fold. All the results reported in section 6 are the average five-fold cross validation results on the dataset.

In our work, we compare the performance of sentiment analysis system based on features, identified by Chi-Square test with the systems that are based on universal sentiment lexicon, Union of DDPW and USL, top-adjectives, adjectives and unigrams. Since training Logistic Regression is expensive with a large set of features, we considered only those unigrams that appear at least four times in the corpus. In the same way, top adjectives are chosen such that they appear at least five times in the corpus. These constants are chosen such that it satisfy both the conditions; it fetches maximum number of words which will be used as features and all those words occur frequently in the corpus.

Since all the three techniques, NN, LR, and SVM take input as a feature vector, each document is represented as a feature vector. Let $p_j(d)$ denotes the presence of feature j in the document d and n be the size of the feature set. Then, each document d is represented as a feature vector as shown here.

$$\vec{d} = (p_1(d), p_2(d), p_3(d), \dots, p_n(d))$$

6 Results and Discussion

The table 1 shows the classification accuracies for the movie domain obtained with various feature sets and techniques separately. Accuracy is calculated as a fraction of total input documents that are correctly classified by the classifiers. The accuracies resulting from using only DDPW as features are shown in row (1) of table 1.

In literature, unigrams are considered as state-of-art features (Ng et al., 2006; Pang et al., 2002) for sentiment analysis, we also experimented with unigrams. Domain-dedicated words as features perform better than unigrams with all the three classification algorithms. At the same time, domain-dedicated words speed up the classification process with a small feature set of size 920. Since adjectives have been crucial clues in sentiment prediction (Hatzivassiloglou and Wiebe, 2000), we experimented with all the adjectives and top adjectives. We find that both the feature sets are not as effective as DDPW.

We experimented with a universal sentiment lexicon provided by Wiebi to capture more context in general. Such sentiment lexicons are good source of polar words with a compact size, but are independent of any domain. The absence of polar words from USL that are crucial for movie domain (e.g., blockbuster) causes accuracy to decline by 5% to 8%. On the other hand, inclusion of DDPW with the USL leads to a big increment in accuracy in comparison with USL only. Yet, the results shown in row (5) of table 1 are relatively poor: the feature set, consisting of 920 domain-dedicated polar words provide more information than the intersection of DDPW and USL.

Figure 2 shows the maximum classification accuracy obtained in the three domains using six feature sets. From figure 2, we can observe that DDPW outperform the accuracy obtained with the other features in the movie domain. In case of tourism domain, DDPW equate the accuracy obtained with unigrams, which is the highest accu-

	Features	Number of features	NN	LR	SVM
(1)	DDPW	920	85.50	83.50	84.50
(2)	Unigrams	18345	83.50	80.00	82.50
(3)	Adjectives	11151	83.00	82.75	80.25
(4)	Top-Adjectives	2500	82.50	82.00	81.50
(5)	DDPW \cup USL	2220	81.50	81.75	81.75
(6)	USL	1946	76.50	75.50	76.00

Table 1: Average five-fold cross-validation accuracies, in percentage. Boldface: best performance achieved through NN, LR and SVM, for the given feature set

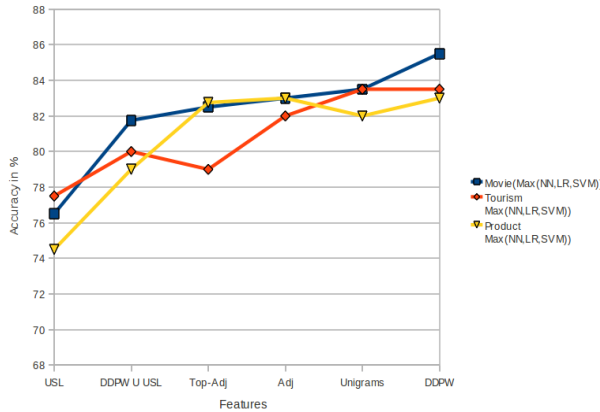


Figure 2: Sentiment classification accuracy in all three domains

racy for the tourism domain. For product domain, DDPW equate the accuracy obtained with adjectives, which is the highest accuracy for the product domain.

In terms of relative performance of classification techniques, Neural Network tends to perform the best, although the differences are not very large. As a whole, accuracy figures validate the prominence of identification of domain-dedicated polar words prior to the implementation of classification algorithm.

7 Related Work

Several works use the universal sentiment lexicons to decide whether a sentence expresses a sentiment (Riloff and Wiebe, 2003; Whitelaw et al., 2005; Mukherjee et al., 2012). Considering that the USL solely is not sufficient to achieve satisfactory performance, there are some more works that combine additional feature types for sentiment classification exist (Yu and Hatzivassiloglou, 2003; Kim and Hovy, 2004; McDonald et al., 2007; Melville et al., 2009; Ng et al., 2006).

Wiebe (2000), for the first time, worked in the

area of sentiment lexicon. She focused on the problem of identifying subjective adjectives with the help of the corpus. She proposed an approach to find subjective adjectives based on the distributional similarity from the Lin (1998) thesaurus. Her approach was seeded by manually provided strong subjectivity clues. She used this new set of adjectives to find subjectivity in sentences, just by the presence of an adjective from the new set. However, the approach was unable to predict sentiment orientations of newly found subjective adjectives and sentences. Moreover, they did not take into account the domain-dedicated polar words and domain-dedicated sentiment analysis.

Pang et al. (2002) for the first time applied machine learning techniques for sentiment classification. They implemented Naive Bayes, maximum entropy classification, and support vector machines. They used frequency or the presence of unigrams or bigrams as features. In addition to this, they used combinations of unigrams and bigrams, unigrams and Part of Speech, unigrams and their position as features. They got the highest accuracy of 82.9% with SVM using a feature vector of size 16165. Besides this, they showed that simply using the 2633 most frequent unigrams are better choice. The feature vector made up of 2633 most frequent unigrams yielded performance comparable to that of using all unigrams (16165) from corpus. Our approach based on Chi-Square test identifies key features from unigrams and enhances the performance.

There are a few researchers who have worked for domain oriented sentiment analysis. The work of Qiu et al. (2009) exploited the relationship between sentiment words and product features. Their method begins with a seed set, then they extract product features that are modified by some sentiment word in the domain dependent corpus.

For example *zoom, flash, resolution* in the camera domain can be modified by *high, poor, nice*, respectively. The process was executed iteratively. The extraction rules are defined based on the relationships described in the dependency trees. They proposed that a feature should receive the same polarity throughout the review and the words extracted by the features will receive the polarity of the feature. However, the reviewer may associate polarity towards a feature with time. To understand this fact consider the following scenario.

When I purchased this camera, the battery was good, but now it is disastrous.

The change in time changes the user's views for a feature in the same review.

8 Conclusion

In this paper, we propose that if we restrict the sentiment analysis task to a domain, then domain-dedicated polar words are the best features for sentiment prediction. For this purpose, we present the SA system as a two stage process, stage-1 identifies decisive words from a domain specific corpus for sentiment analysis in that domain. Stage-2 uses these words as features for classification task. Use of domain-dedicated polar words as features outperforms or equates the accuracies achieved independently with unigrams, adjectives, top-adjectives, Universal Sentiment Lexicon (USL) and union of USL and DDPW. Besides improvement in sentiment analysis, the research can be useful for creating writing aids for authors and in natural language generation.

References

- A. R. Balamurali, Aditya Joshi, and Pushpak Bhattacharyya. 2011. Harnessing wordnet senses for supervised sentiment classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '11*, pages 1081–1091, Stroudsburg, PA, USA. Association for Computational Linguistics.
- John Blitzer, Mark Dredze, and Fernando Pereira. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, volume 7, pages 440–447. Citeseer.
- Rebecca F. Bruce and Janyce M. Wiebe. 1999. Recognizing subjectivity: a case study in manual tagging. *Nat. Lang. Eng.*, 5(2):187–205, June.
- Cristian Danescu-Niculescu-Mizil, Gueorgi Kossinets, Jon Kleinberg, and Lillian Lee. 2009. How opinions are received by online communities: A case study on amazon.com helpfulness votes. In *Proceedings of the 18th International Conference on World Wide Web, WWW '09*, pages 141–150, New York, NY, USA. ACM.
- Vasileios Hatzivassiloglou and Janyce M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th Conference on Computational Linguistics - Volume 1, COLING '00*, pages 299–305, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Wei Hui. 2011. A neural dynamic model based on activation diffusion and a micro-explanation for cognitive operations. In *Cognitive Informatics Cognitive Computing (ICCI*CC), 2011 10th IEEE International Conference on*, pages 397–404.
- Thorsten Joachims. 1998. *Text categorization with support vector machines: Learning with many relevant features*. Springer.
- Mitesh M Khapra, Sapan Shah, Piyush Kedia, and Pushpak Bhattacharyya. 2010. Domain-specific word sense disambiguation combining corpus based and wordnet based parameters. In *In 5th International Conference on Global Wordnet (GWC2010)*. Citeseer.
- Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, page 1367. Association for Computational Linguistics.
- Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 17th international conference on Computational linguistics - Volume 2, COLING '98*, pages 768–774, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. 2007. Structured models for fine-to-coarse sentiment analysis. In *Annual Meeting-Association For Computational Linguistics*, page 432.
- Prem Melville, Wojciech Gryc, and Richard D Lawrence. 2009. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1275–1284. ACM.
- Subhabrata Mukherjee, Akshat Malu, Balamurali A.R., and Pushpak Bhattacharyya. 2012. Twisent: a multistage system for analyzing sentiment in twitter. In *Proceedings of the 21st ACM international conference on Information and knowledge management, CIKM '12*, pages 2531–2534, New York, NY, USA. ACM.

- Vincent Ng, Sajib Dasgupta, and S. M. Niaz Arifin. 2006. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In *Proceedings of the COLING/ACL on Main Conference Poster Sessions*, COLING-ACL '06, pages 611–618, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international joint conference on Artificial intelligence*, pages 1199–1204.
- Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 105–112, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Raksha Sharma and Pushpak Bhattacharyya. 2013. Detecting domain dedicated polar words. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 661–666, Nagoya, Japan, October. Asian Federation of Natural Language Processing.
- Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 133–140, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Peter D. Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 417–424, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Casey Whitelaw, Navendu Garg, and Shlomo Argamon. 2005. Using appraisal groups for sentiment analysis. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, CIKM '05, pages 625–631, New York, NY, USA. ACM.
- Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 735–740. AAAI Press.
- Hidekazu Yanagimoto, Mika Shimada, and Akane Yoshimura. 2013. Document similarity estimation for sentiment analysis using neural network. In *Computer and Information Science (ICIS), 2013 IEEE/ACIS 12th International Conference on*, pages 105–110.
- Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, EMNLP '03, pages 129–136, Stroudsburg, PA, USA. Association for Computational Linguistics.