

# Opinion Mining and Sentiment Analysis: A Survey

Zhang Wei

Department of Computer Science  
School of Computing  
National University of Singapore  
z-wei@comp.nus.edu.sg

## Abstract

With the increasing availability of opinion-rich resources such as online review sites and personal blogs, many research activities have focused on Opinion Mining. Meanwhile the application of Opinion Mining is becoming more and more widespread (e.g., movie review recommender system and product review). This survey is an overview of the work on Opinion Mining, covering the new challenges appear in Opinion Mining compared to traditional fact-based analysis, some available resources and evaluation datasets, and the techniques and approaches in this task (i.e. sentiment classification, subjectivity classification, feature/topic-based sentiment analysis, sentiment analysis of comparative sentences, opinion spam and utility of opinions). A discussion of future work is also provided.

## 1 Introduction

Web resources such as discussion forums, review sites and blogs have now given a public piece of information to every consumer. Recent surveys (e.g., (Horrigan, 2008)) have showed that a massive number of internet users gather information for product and services from internet, guiding their own choices and decisions by the opinions that other consumers have publicly expressed. With the increasing availability of opinion-rich resources, Sentiment Analysis (also called Opinion Mining) has recently received a lot of attention in the Natural Language Processing (NLP) community.

Meanwhile the application of Opinion Mining is becoming more and more widespread. Most of us are interested in other's opinions when pur-

chasing a product (Hu and Liu, 2004), using a service (Pang et al., 2002), finding opinions on political topics (Greene, 2007; Thomas et al., 2006), and many other decision making tasks. The field of Opinion Mining is also well-suited to various types of intelligence applications (e.g. businesses and organizations).

The work of Opinion Mining is that which deals with the computational treatment of opinion, sentiment, and subjectivity in text. Such work has come to be known as opinion mining, sentiment analysis, and subjectivity analysis (Pang and Lee, 2008). Wiebe (1994) called sentences that take a character's psychological point of view subjective, in contrast to sentences that objectively narrate events or describe the fictional world. In the paper by Dave et al. (2003), the appearance of Opinion Mining strongly associated with Web search or information retrieval.

So far, the demand, application and terminology for Opinion Mining are briefed. New challenges appear in Opinion Mining compared to traditional fact-based analysis are introduced in next section. Section 3 will introduce some available resources and evaluation datasets. As is well known, there are many research directions in this research area, e.g., sentiment classification (classifying an opinion document as positive or negative), subjectivity classification (determining whether a sentence is subjective or objective, and its associated opinion), feature/topic-based sentiment analysis (assigning positive or negative sentiments to topics or product features), sentiment analysis of comparative sentences (comparing the object with some other similar objects), opinion spam and utility of opinions (detecting opinion spam refers to fake or bogus opinion) (Liu, 2009). The techniques and approaches in these different directions will be covered. A discussion of future work is also provided at last.

## 2 Challenges

There are two main types (i.e., facts and opinions) of information on the web. However, current search engines (e.g., google) are all for facts expressed with topic keywords. There are many factors that make Opinion Mining difficult compared to traditional fact-based text analysis. The following are some of the factors.

### 2.1 Expression

Compared to topic, opinions are hard to express with keywords. While a topic is more likely to be emphasized by frequent occurrences of certain keywords, overall sentiment may not usually be highlighted through repeated use of the same terms. Sentiment can often be expressed in a more subtle manner making it difficult to be identified by any of a sentence or document's term when considered in isolation. An early study (Pang et al., 2002) on movie reviews suggest that coming up with the right set of keywords might be less trivial than one might initially think.

### 2.2 Negations

"I like this car" and "I do not like this car" are considered to be very similar by most similarity measures in fact-based text analysis. It is possible to deal with negations by attaching "NOT" to words occurring close to negation terms such as "no" or "not", so that in the bag-of-words representations of "I do not like this car", the token "like" is converted to a new token "like-NOT" (Das and Chen, 2001).

However, in sentence "No wonder this is considered one of the best.", the new token "best-NOT" is incorrect. Na et al. (2004) attempt to model negation more accurately by specific part-of-speech tag patterns and achieve about 3% improvement in accuracy. Another difficulty with modeling negation is that negation can be expressed in rather subtle ways. Sarcasm and irony can be quite difficult to detect (Pang and Lee, 2008).

### 2.3 Domain Considerations

The accuracy of sentiment classification can be influenced by the domain of the items to which it is applied (Aue and Gamon, 2005; Blitzer et al., 2007; Engstrom, 2004). The reason is that the same phrase can indicate different sentiment in different domains: "go read the book" most likely

indicates positive sentiment for book reviews, but negative sentiment for movie reviews. Difference in vocabularies across different domains also adds to the difficulty when applying classifiers trained on labeled data in one domain to test data in another (Pang and Lee, 2008). Engstrom (2004) studies how the accuracy of sentiment classification can be influenced by topic. Owsley et al. (2006) show the importance of building a domain-specific classifier. (Aue and Gamon, 2005) explore different approaches to customizing a sentiment classification system to a new target domain in the absence of large amounts of labeled data. Some research activities (Blitzer et al., 2007; Yang et al., 2006) focused on domain transfer problem and achieved a great improvement.

## 3 Available Resources

### 3.1 Annotated Datasets

One source of opinion, sentiment, and subjectivity labels is manual annotation (Ku et al., 2007; Wiebe et al., 2005). Current researchers in opinion mining and sentiment analysis have taken advantage of Rotten Tomatoes, Epinions, Amazon, and other sites where users furnish ratings along with their reviews. However, there is the question of whether one can trust the label in these sites. Liu et al. (2007) essentially re-labeled their review-quality Amazon data. The following list is some available datasets.

*Blog06*<sup>1</sup>. The University of Glasgow distributes this 25GB TREC test collection consisting of blog posts over a range of topics.

*Congressional floor-debate transcripts*<sup>2</sup>. This dataset includes speeches as individual documents together with labels for whether the speaker supported or opposed the legislation.

*Cornell movie-review datasets*<sup>3</sup>. These corpora, first introduced in the papers (Pang and Lee, 2004; Pang and Lee, 2005).

*Customer review datasets*<sup>4</sup> This dataset introduced in Hu and Liu (2004), consists of reviews of five electronics products download from Amazon and Cnet.

*Economining*<sup>5</sup>. This site hosted by the Stern School at New York University.

<sup>1</sup>[http://ir.dcs.gla.ac.uk/test\\_collections/access\\_to\\_data.html](http://ir.dcs.gla.ac.uk/test_collections/access_to_data.html)

<sup>2</sup><http://www.cs.cornell.edu/home/llee/data/convote.html>

<sup>3</sup><http://www.cs.cornell.edu/people/pabo/movie-review-data/>

<sup>4</sup><http://www.cs.uic.edu/~liub/FBS/CustomerReviewData.zip>

<sup>5</sup><http://economining.stern.nyu.edu/datasets.html>

The information for other available datasets such as *French sentences*, *French sentences*, *MPQA Corpus*, *Multiple-aspect restaurant reviews*, *Multi-Domain Sentiment Dataset*, *NTCIR multilingual corpus* and *Review-search results sets* can be found in the survey by Pang (2008).

### 3.2 Evaluation

Performance of the Opinion Mining System usually is measured using the following metrics: Accuracy, Recall and F-measure (e.g. (Somasingh and Wiebe, 2008)).

*TREC 2006 Blog Track.* TREC 2006 involved a Blog track: participating systems had to retrieve blog posts expressing an opinion about specified topic.

*TREC 2007 Blog Track.* The TREC 2007 Blog track retained the opinion retrieval task and instituted determining the sentiment status (positive, negative, or mixed) of the retrieved opinions as a subtask.

*TREC 2008 Blog Track.* In the TREC 2008 Blog track, the polarity identification problem was reposed as one of ranking of positive polarity retrieved documents by degree of positivity, and, similarly, ranking of negative-polarity retrieved documents by degree of negativity.

*NTCIR-6 opinion analysis pilot task.* The four constituent tasks were as follows: (1) Detection of opinionated sentences. (2) Detection of opinion holders. (3) (optional) Polarity labeling of opinionated sentences as positive, negative, or neutral. (4) (optional) Detection of sentences relevant to a given topic.

*NTCIR-7 Multilingual opinion analysis task (MOAT), 2008.* This exercise involves six subtasks: (1) Detection of opinionated sentences and opinion fragments within opinionated sentences. (2) Polarity labeling of opinion fragments as positive, negative or neutral. (3) (optional) Strength labeling of opinion fragments as very weak, average, or very strong. (4) (optional) Detection of opinion holders. (5) (optional) Detection of opinion targets. (6) (optional) Detection of sentences that are relevant to a given topic.

## 4 Sentiment and Subjectivity Classification

This is the area that has been researched the most in academia. It treats sentiment analysis as a text classification problem. Two sub-topics are: (1)

classifying an opinionated document as expressing a positive or negative opinion, and (2) classifying a sentence or a clause of the sentence as subjective or objective, and for a subjective sentence or clause classifying it as expressing a positive, negative or neutral opinion. (Liu, 2009)

### 4.1 Sentiment Classification

The task is to determine a document expresses a positive or negative opinion. The task is also known as document-level sentiment classification. The existing research assumes that the document is known to be opinionated.

#### 4.1.1 Supervised Method

As the two class labels are pre-defined (positive or negative), Sentiment classification can obviously be formulated as a supervised learning problem. Training and testing data used in existing research are listed in section 3. Existing supervised learning methods can be readily applied to sentiment classification, e.g., naive Bayesian, and support vector machines (SVM), etc.

Pang et al. (2002) used supervised machine learning to classify movie reviews. They extract different features from the review and use machine learning algorithms naive Bayesian, Maximum Entropy (ME) and Support Vector Machine (SVM) to classify the reviews. They achieved accuracies between 78.8% and 82.9%. Neutral reviews were not used in this work, which made the problem easier.

Current research activities use many features in machine learning. Some of the example features used in research and possibly in practice are listed below.

*Terms and their frequency:* It is similar with the traditional information retrieval task, to represent a text as a vector wherein the entries correspond to the terms or n-gram. Pang et al. (2002) obtained better performance using presence rather than frequency. The position information of a token within a textual unit can also have important effects (e.g. (Kim and Hovy, 2006; Pang et al., 2002)).

*Part of speech tags:* Adjectives have been employed as features by a number of researchers (e.g. (Mullen and Collier, 2004; Whitelaw et al., 2005)). Rather than focusing on isolated adjectives, Turney (2002) proposed to detect document sentiment based on selected phrases, where the phrases are chosen via a number of pre-specified

part-of-speech patterns, most including an adjective or an adverb. Besides adjectives and adverbs, Nasukawa and Yi (2003) also consider verbs as clue of sentiment. However, other word types may also affect text sentiment. For example, In sentence “Although it is good, the book is too expensive”, “Although” is important, due to expensive must be assigned higher weight.

*Opinion words and phrases:* Opinion words are words that are commonly used to express positive negative sentiments. Most of the work use part-of-speech (POS) tagging to extract them (Turney, 2002). The main approaches to identify the semantic orientation of a opinion word are statistical-based or lexicon-based. Some researchers (Hu and Liu, 2004) use WordNet to determine whether the extracted adjective has a positive or negative polarity.

*Syntactic:* Syntactic relations play an important role in Opinion Mining task. Collocations and syntactic patterns have been found to be useful for subjectivity detection. (Riloff and Wiebe, 2003)

#### 4.1.2 Unsupervised Method

It is not hard to identify the sentiment words and their orientation. Thus, using unsupervised method based on sentiment words would be quite natural. The algorithm in one paper (Turney, 2002) is as below:

Step 1: The first step of the algorithm is to extract phrases containing adjectives or adverbs from reviews. After applying a part of speech tagger, two consecutive words are extracted from the review if their tags conform to any of the pre-defined patterns(e.g., first word “JJ”, second word “NN” or “NNs”).

Step 2: The second step is to estimate the semantic orientation of the extracted phrases, using the PMI-IR algorithm. This algorithm uses mutual information as a measure of the strength of semantic association between two words.

Step 3: The third step is to calculate the average semantic orientation of the phrases in the given review and classify the review as recommended if the average is positive and otherwise not recommended.

As mentioned in this paper, the limitations of this work include the time required for queries and, for some applications, the level of accuracy that was achieved.

Besides this method, there are many other unsupervised methods. For example, Somasundaran

and Wiebe (2008) proposed an unsupervised approach to classifying the stance of a post in a dual-topic debate. For this, they first use a web corpus to learn preferences that are likely to be associated with a side. These learned preferences are then employed in conjunction with discourse constraints to identify the side for a given post.

#### 4.1.3 Other Issues

*Rating inference:* The more general problem of rating inference, where one must determine the authors evaluation with respect to a multi-point scale (e.g., one to five stars for a review) can be viewed simply as a multi-class text categorization problem. But in contrast to many topic-based multi-class classification problems, sentiment-related multi-class classification can also be naturally formulated as a regression problem because ratings are ordinal.

*Other non-factual information in text:* Researchers have considered various affect types, such as the six universal emotions: anger, disgust, fear, happiness, sadness, and surprise (Alm et al., 2005; Liu et al., 2003; Subasic and Huettner, 2001). An interesting application is in human-computer interaction: if a system determines that a user is upset or annoyed, for instance, it could switch to a different mode of interaction. (Pang and Lee, 2008)

### 4.2 Subjectivity Classification

Subjectivity classification (also known as sentence-level subjectivity classification) is to determine whether a sentence is subjective or objective, and its associated opinion. Thus, there are two subtask: 1. Subjectivity classification: Determine whether the sentence is a subjective sentence or an objective sentence, 2. Sentence-level sentiment classification: If the sentence is subjective, determine whether it expresses a positive or negative opinion.

Both problems are classification problems. Thus, traditional supervised learning methods are applicable. Wiebe et al. (1999) performed subjectivity classification using the naive Bayesian classifier. Subsequent research also used other learning algorithms (e.g. (Wilson et al., 2004; Yu and Hatzivassiloglou, 2003)).

One of the limitation in applying supervised learning is the manual effort involved in annotating a large number of training examples. To save the manual labeling effort, Hu and Liu (2004) pro-

posed an method using bootstrapping technique and the WordNet to determine customer reviews orientation.

For their work is focused on the overall opinion for a particular product feature, the sentiment classification for each individual opinion sentence comes to use as they need to group the opinion sentences for the same feature to find out the overall opinion towards that product feature. After opinion features have been identified, they determine the semantic orientation (i.e., positive or negative) of each opinion sentence. This consists of two steps: (1) for each opinion word in the opinion word list, they identify its semantic orientation using a bootstrapping technique and the WordNet and (2) They then decide the opinion orientation of each sentence based on the dominant orientation of the opinion words in the sentence. Their sentence orientation accuracy is 84.2%.

For the first subtask, early work by Hatzivasiloglou and Wiebe (2000) examined the effects of adjective orientation and gradability on sentence subjectivity. The goal was to tell whether a given sentence is subjective or not judging from the adjectives appearing in that sentence.

## 5 Feature-Based Sentiment Analysis

Although classifying opinionated texts at the document level or at the sentence level is useful in many cases, they do not provide the necessary detail needed for some other applications. A positive opinionated document on a particular object does not mean that the author has positive opinions on all aspects or features of the object. Likewise, a negative opinionated document does not mean that the author dislikes everything. In a typical opinionated text, the author writes both positive and negative aspects of the object, although the general sentiment on the object may be positive or negative. Document-level and sentence-level classification does not provide such information. (Liu, 2009) Thus, feature-based sentiment analysis is necessary. There are two main tasks:

1. Identify object features that have been commented on. For instance, in the sentence, The picture quality of this camera is amazing, the object feature is picture quality.

2. Determine whether the opinions on the features are positive, negative or neutral. In the above sentence, the opinion on the feature picture quality is positive.

Identification of product features is in some sense a standard information extraction task with little to distinguish it from other non-sentiment-related problems. However, Hu and Liu (2004) show that one can benefit from light sentiment analysis even for this sub-task, as described shortly. Thus, it is necessary to introduce the techniques in the feature extraction task.

Current online product reviews have two main format:

- 1, Pros, Cons and detailed review: The reviewer is asked to describe Pros and Cons separately and also write a detailed review. Epinions.com uses this format.

- 2, free format: The reviewer can write freely, i.e., no separation of Pros and Cons. Amazon.com uses this format. (Liu et al., 2005)

In the paper by Liu et al. (2005), they proposed a technique to identify product features from Pros and Cons in format 1. They use supervised rule discovery to perform this task. Firstly, they prepare a training dataset by manually labeling (or tagging) a large number of reviews. Then rules are generated. After post-processing, the result patterns are used to match and identify candidate features from new reviews. In the final step, they correct some mistakes made during extraction. They achieved around 90% accuracy.

In the paper by Hu and Liu (2004), they summarize their techniques for mining product features from format 2. To identify product features on which many people have expressed their opinions, association mining is used. They run the association miner CBA which is based the Apriori algorithm on the transaction set of noun/noun phrases produced in the previous step. They define an itemset as frequent if it appear in more than 1%. This method checks features that contain at least two words, which they call feature phrases, and remove those that are likely to be meaningless. To describe the meaning of redundant features, they use the concept of p-support (pure support). P-support of feature ftr is the number of sentences that ftr appears in as a noun or noun phrase, and these sentences must contain no feature phrase that is a superset of ftr. They use a minimum p-support value to prune those redundant features. If a feature has a p-support lower than the minimum p-support (in their system, they set it to 3) and the feature is a subset of another feature phrase (which suggests that the feature alone may not be inter-

esting), it is pruned. For instance, life by itself is not a useful feature while battery life is a meaningful feature phrase. They achieved around 70% accuracy.

## 6 Sentiment Analysis of Comparative Sentences

In general, a comparative sentence expresses a relation based on similarities or differences of more than one object. The comparison is usually conveyed using the comparative or superlative form of an adjective or adverb. Comparison mining consists of two tasks: 1. Identify comparative sentences, and classify the identified comparative sentences into different types or classes. 2. Extract comparative opinions from the identified sentences. Little research has been done on classifying whether a comparative sentence is opinionated or not. A approach bears some resemblance to the lexicon-based approach was proposed by (Ganapathibhotla and Liu., 2008) to identifying the preferred objects.

## 7 Opinion Spam and Utility of Opinions

To detect opinion spam refers to fake or bogus opinion, the idea is to determine how helpful a review is to a user. In fact, many review aggregation sites have been practicing this for years. They obtain the helpfulness or utility score of each review by asking reader to provide helpfulness feedbacks to each review.

There are generally three types of spam reviews:

Type 1 (untruthful opinions): These are reviews that deliberately mislead readers or opinion mining systems by giving undeserving positive reviews to some target objects in order to promote the objects and/or by giving unjust or malicious negative reviews to some other objects in order to damage their reputation. Untruthful reviews are also commonly known as fake reviews or bogus reviews.

Type 2 (opinions on brands only): These are reviews that do not comment on the specific products that they are supposed to review, but only comment on the brands, the manufacturers or the sellers of the products. Although they may be useful, they are considered as spam because they are not targeted at the specific products and are often biased. For example, in a review for a HP printer, the reviewer only wrote “I hate HP. I never buy any of their products”

Type 3 (non-opinions): These are not reviews or opinionated although they appear as reviews. There are two main sub-types: (1) advertisements, and (2) other irrelevant texts containing no opinions (e.g., questions, answers, and random texts). (Liu, 2009)

## 8 Discussion and Conclusion

### 8.1 Reference Resolution

Reference resolution may be useful for the Opinion Mining task, especially for the topic/feature-based sentiment analysis. When more than two objective appearing in the document, using reference resolution may improve the accuracy.

### 8.2 Temporal Relations

The time of reviews expressed by the reviewer may be important for sentiment analysis. The reviewer may think “windows 2000 OS” of Microsoft is good in 2000, but the reviewer may have negative opinion now due to the appearance of new OS “windows vista”. Thus, the reviews are ranked by temporal relations and then assigned different weight which may improve the performance of Opinion Mining system.

### 8.3 Review Holder

Due to the age, gender and other difference among the reviewers, different reviewers may be assigned different weight. If someone always provides spam refers to fake or bogus opinion, his review weight will be zero.

### 8.4 Conclusion

This survey has been to cover the main directions in the Opinion Mining area. The various approaches and techniques in these directions are compared and evaluated. A discussion of future work is also provided. The objective is to give a overview of the work on Opinion Mining.

## References

- C. O. Alm, D. Roth, and R. Sproat, 2005. *Emotions from text: Machine learning for text-based emotion prediction*, in Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP).
- A. Aue and M. Gamon. 2005. *Customizing sentiment classifiers to new domains: A case study*, in Proceedings of Recent Advances in Natural Language Processing (RANLP).

- J. Blitzer, M. Dredze, and F. Pereira, 2003. *Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification*, in Proceedings of the Association for Computational Linguistics (ACL).
- S. Das and M. Chen, 2001. *Yahoo! for Amazon: Extracting market sentiment from stock message boards*, in Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).
- K. Dave, S. Lawrence, and D. M. Pennock. 2003. *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews*, in Proceedings of WWW, pp. 519–528.
- C. Engstrom, 2003. *Topic Dependence in Sentiment Classification*, Masters thesis, University of Cambridge.
- G. Ganapathibhotla and B. Liu. 2008. *Identifying Preferred Entities in Comparative Sentences*, Proceedings of the International Conference on Computational Linguistics.
- S. Greene. 2007. *Lexical Semantics, Transitivity, and the Identification of Implicit Sentiment*, PhD thesis, University of Maryland.
- V. Hatzivassiloglou and J. Wiebe. 2000. *Effects of adjective orientation and gradability on sentence subjectivity*, in Proceedings of the International Conference on Computational Linguistics (COLING).
- J. A. Horrigan. 2008. *Online shopping*, Pew Internet and American Life Project Report.
- Minqing Hu and Bing Liu 2004. *Mining and summarizing customer reviews*, Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining. Aug. 22–25, 2004, Seattle, WA, USA.
- S. M. Kim and E. Hovy 2006. *Automatic identification of pro and con reasons in online reviews*, in Proceedings of the COLING/ACL Main Conference Poster Sessions, pp. 483–490.
- L.-W. Ku, Y.-S. Lo, and H.-H. Chen. 2007. *Test collection selection and gold standard generation for a multiply-annotated opinion corpus*, in Proceedings of the ACL Demo and Poster Sessions, pp. 89C92
- B. Liu, M. Hu, and J. Cheng 2005. *Opinion observer: Analyzing and comparing opinions on the web*, Proceedings of WWW, 2005.
- B. Liu. 2009. *Sentiment Analysis and Subjectivity*, To appear in Handbook of Natural Language Processing, Second Edition, (editors: N. Indurkha and F. J. Damerou), 2009 or 2010
- H. Liu, H. Lieberman, and T. Selker, 2003. *A model of textual affect sensing using real-world knowledge*, in Proceedings of Intelligent User Interfaces (IUI), pp. 125–132
- JJ. Liu, Y. Cao, C.-Y. Lin, Y. Huang, and M. Zhou 2007. *Low-quality product review detection in opinion summarization*, in Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp. 334–342.
- T. Mullen and N. Collier, 2004. *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.
- J.-C. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou, 2004. *Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews*, in Conference of the International Society for Knowledge Organization (ISKO), pp. 49–54.
- Tetsuya Nasukawa and Jeonghee Yi. 2003. *Tetsuya Nasukawa and Jeonghee Yi*, Proceedings of the 2nd international conference on Knowledge capture. Oct. 23–25, 2003, Sanibel Island, FL, USA.
- S. Owsley, S. Sood, and K. J. Hammond. 2006. *Domain specific affective classification of documents*, in AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW), pp. 181–183.
- Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. 2002. *Thumbs up? Sentiment classification using machine learning techniques*, Proceedings of the 7th Conference on Empirical Methods in Natural Language Processing (EMNLP-02).
- B. Pang and L. Lee, 2004. *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*, in Proceedings of the Association for Computational Linguistics (ACL), pp. 271–278.
- B. Pang and L. Lee, 2005. *Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales*, in Proceedings of the Association for Computational Linguistics (ACL), pp. 115–124,
- Bo Pang and Lillian Lee. 2008. *Opinion Mining and Sentiment Analysis*, Now Publishers Inc.
- E. Riloff and J. Wiebe, 2003. *Learning extraction patterns for subjective expressions*, in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).
- Swapna Somasundaran and Janyce Wiebe 2008. *Recognizing Stances in Online Debates*, in Proceedings of the Association for Computational Linguistics (ACL).
- P. Subasic and A. Huettner 2001. *Affect analysis of text using fuzzy semantic typing*, IEEE Transactions on Fuzzy Systems, vol. 9, pp. 483–496

- M. Thomas, B. Pang, and L. Lee. 2006. *Get out the vote: Determining support or opposition from congressional floor-debate transcripts*, in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 327–335.
- P. Turney, 2002. *Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews*, in Proceedings of the Association for Computational Linguistics (ACL), pp. 417–424.
- C. Whitelaw, N. Garg, and S. Argamon. 2005. *Using appraisal groups for sentiment analysis*, in Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM), pp. 625–631, ACM.
- J. M. Wiebe. 1994. *Tracking point of view in narrative*, Computational Linguistics, vol. 20, pp. 233–287.
- J. Wiebe, R. F. Bruce, and T. P. O Hara, 1999. *Development and use of a gold standard data set for subjectivity classifications*, Proceedings of the Association for Computational Linguistics (ACL), pp. 246–253
- J. Wiebe, T. Wilson, and C. Cardie, 2005. *Annotating expressions of opinions and emotions in language*, Language Resources and Evaluation (formerly Computers and the Humanities), vol. 39, pp. 164–210.
- T. Wilson, J. Wiebe, and R. Hwa 2004. *Just how mad are you? Finding strong and weak opinion clauses*, Proceedings of AAAI, pp. 761–769.
- H. Yang, L. Si, and J. Callan. 2006. *Knowledge transfer and opinion detection in the TREC2006 blog track*, in Proceedings of TREC.
- H. Yu and V. Hatzivassiloglou 2003. *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences*, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).