Learning Text Patterns to Detect Opinion Targets

Filipa Peleja and João Magalhães
NOVA LINCS, Departamento de Informática, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa, Lisboa, Portugal

Keywords: Sentiment Analysis, Opinion Mining, Opinion Targets.

Abstract: Exploiting sentiment relations to capture opinion targets has recently caught the interest of many researchers. In many cases target entities are themselves part of the sentiment lexicon creating a loop from which it is difficult to infer the overall sentiment to the target entities. In the present work we propose to detect opinion targets by extracting syntactic patterns from short-texts. Experiments show that our method was able to successfully extract 1,879 opinion targets from a total of 2,052 opinion targets. Furthermore, the proposed method obtains comparable results to SemEval 2015 opinion target models in which we observed the syntactic structure relation that exists between sentiment words and their target.

1 INTRODUCTION

Sentiment analysis is a research area that has been quite active in the last decade. From the first techniques of review analysis (Pang et al., 2002), to more recent approaches of tweet analysis (Bollen, 2010; Diakopoulos and Shamma, 2010), the field has progressed much. This is intrinsically related to the popularity of the Web which led to changes in peoples habits and as a consequence, we have observed an amount of opinionated text data that previously to these changes did not exist. Sentiment analysis approaches can be divided into three levels of granularity: document level (Ghorbel and Jacot, 2010; Pang and Lee, 2004), sentence level (Riloff and Wiebe, 2003) and aspect level. Aspect level sentiment analysis provides a finer-grain analysis as it aims to identify different opinion components. Hence it enables one to identify likes/dislikes that target specific product features.

The analysis of opinionated text also known as subjective text involves the detection of words, phrases or sentences that express a sentiment. Although this area has been researched in academia, the problem is still far from being solved (Liu, 2012). One of the main challenges is that opinionated language varies over a broad range of discourse, and a system with a fixed vocabulary will not be enough to represent users' opinion. Another challenge is to identify relevant mentions to opinion targets which are accompanied by related sentiment words. From an algorithmic perspective, the challenge is to analyse how these sentiment words affect the public image of opinion targets. Previous work (Hu and Liu, 2004; Liu, 2012) has introduced significant advances in detecting product aspects or features. It is reasonable to apply similar methods to detect sentiment words influence in entities' reputation. However, unlike products opinions that target specific entities are not structured around a fixed set of aspects or features (Albornoz et al., 2012). Users comments in the Twitter social network is limited to a maximum of 140 characters and each tweet is usually composed of a single sentence. Hence, we take the assumption that the sentiment expressed in a tweet is composed within that single sentence (Pak and Paroubek, 2010). In tweets the opinion targets is quite diverse since there is a large range of different topics: named entities and noun phrases that are the object of a sentiment.

Opinion target detection is an important task, in particular, to evaluate how impacts the reputation of a product that is targeted by an opinion. Opinion target is usually the entity that the opinion is about (Kim and Hovy, 2006). For example, the sentiment word "fantastic" and the opinion target "camera" in "A fantastic camera on Pinterest". In this study, to identify opinion targets, we investigate a syntactic sentence parsing method. As, we argue that there is a fixed pattern structure that is indicative of the existence of an opinion target. The overall task is structured in the following steps:

- Sentiment words detection.
- Subjective classification.

Peleja, F. and Magalhães, J..
Learning Text Patterns to Detect Opinion Targets.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 1: KDIR*, pages 337-343
ISBN: 978-989-758-158-8

- Expand existing sentiment words lexicon.
- For each tweet obtain the triple
  <polarity, sentiment word, opinion target >.

The main contribution of this paper is the method that we propose to automatically identify opinion targets. To detect opinion targets the method studies the syntactic structure of subjective sentences.

2 RELATED WORK

Sentiment analysis enfolds various techniques to detect words that express a positive and negative feeling or emotion. These words are commonly known as sentiment words or opinion words. Beyond words, n-grams (contiguous sequence of n words) and idiomatic expressions are commonly used as sentiment words. For example, the word “terrible”, the n-gram “quite wonderful” and the idiomatic expression “break a leg”. At document or sentence level sentiment words can be used to predict sentiment classes for users opinions (Liu, 2012). Unlike sentiment analysis at document and sentence level, the entity (or aspect level) allows a finer-grain analysis. Entity or aspect level captures specific product features that users dislike and like (Hu and Liu, 2004). Turney (2002) proposed a document level approach to evaluate reviews polarity in which an unsupervised learning algorithm is used to evaluate reviews polarity. For each review is observed the average polarity of its constituent words or phrases. Others (Pang et al., 2002; Heerschop et al., 2011) have also proposed to solve a sentiment analysis problem using a document level approach. A common use of sentence level sentiment analysis is to capture subjective sentences (Wiebe et al., 1999). In a subjectivity classification the goal is to distinguish between sentences that express factual information (objective) and sentences that express an opinion (subjective) (Hatzivassiloglou and Wiebe, 2000). To perform an aspect-based sentiment analysis task an initial step is required: distinguish between objective from subjective sentences. Several different methods have been proposed to perform subjective classification in social media platforms like Twitter, where users comment on a large collection of different subjects (Go et al., 2009; Wiebe et al., 1999). For this task supervised and unsupervised algorithms have been applied. According to (Liu, 2012) the supervised classification is mostly adopted by researchers. Hence, in our framework to capture subjective sentences we will apply a supervised subjective classification.

The task of detecting overall sentiment, opinion holders and targets implies several steps (Liu, 2012). In a sentence level sentiment analysis approach Meena and Prabhakar (2007) showed that rules based on atomic sentiments of individual phrases can be helpful to decided the overall sentiment of a sentence. However, only adjectives and verbs were considered as features, which implies that only those can be related to the opinion target (Meena and Prabhakar, 2007). For instance, in another work (Wilson et al., 2009) show that other words families (e.g. nouns) may share dependency relations with opinion targets (also referred as aspects) which might be indicative of the sentiment expressed towards those terms. Previous work has also introduced a system based on statistical classifiers to identify semantic relationships (Gildea and Jurafsky, 2002). Their system analyses the prior probabilities of various combinations of semantic roles (predicate verb, noun, or adjective) to automatically label domain-specific semantic roles such as Agent, Patient, Speaker or Topic. Similarly to Gildea and Jurafsky (2002) semantic roles detection we propose to analyze sentences lexical and syntactic relations to automatically label opinion targets.

Generally, aspect-based sentiment classification is split into two main approaches: supervised learning and lexicon-based (Liu, 2010). Regarding the supervised learning approach several well-known machine learning algorithms have been adapted to a sentiment analysis evaluation (Pang et al., 2002; Pang and Lee, 2005). However a supervised learning method depends more on the coverage of dataset than a lexicon-based approach which demands a greater effort to scale up to different domains. A lexicon-based approach is typically an unsupervised evaluation and handles more easily domain issues. Additionally, in lexicon-based approaches sentiment lexicons (e.g. SentiWordNet (Esuli and Sebastiani, 2006)) are commonly used to discover new opinion word, also semantic lexicons can also be used as seed to capture new opinion words.

3 TEXT-PATTERNS FOR OPINION TARGETS

The analysis of subjective text involves the detection of words, phrases or sentences that express a sentiment. However, one of the main challenges is to identify opinion targets. Within subjective text, opinion targets tend to be accompanied by sentiment words. For example, tripod and beautiful in *“We have here a very beautiful tripod”*. To this aim, we propose to explore the syntax structure of subjective sentences correlation with opinion targets. Given a set of labeled data, in which it is available the opinion tar-

Table 1: Text-patterns for opinion targets.

Rule (1-5)	Rule (6-10)	Rule (11-15)	Rule (16-20)
n_v#A	v_in#B	v_v#A	prp_v#B
n_n#A	a_n#A	v_n#A	v_n#B
n_in#B	n_in#A	n_a#A	v_#A
n_n#B	n_v#B	in_dt#B	v_a#A
n_#A	a_n#B	v_dt#B	n_r#A

gets labels, the proposed method identifies a set of syntactic patterns that correlate with the opinion targets. Table 1 presents 20 of the 35 extracted rules. In this table #A refers to after and #B before. The word-families tags are represented by personal pronoun (prp), preposition or subordinating conjunction (in), determiner (dt), noun (n), adjective (a), verb (v) and adverb (r). This syntactic patterns are extracted from the analysis of a Twitter dataset that contains the annotation of one or more opinion target per each tweet (Twitter dataset will be further described in Section 4).

3.1 Processing Textual Data

The first step of the proposed method aims to translate the text into a representation that resolves writing typos and the usage of internet slang. To this end the textual data is split in sentence level and the tokens are mapped according to patterns of repeated letters, internet slang words and emotion expressions (i.e. =) to represent a smile). Also, tokens that express a sentiment are identified and mapped to its corresponding sentiment weight in sentiment lexicons.

The scope of the sentiment expressed is determined by the identified sentiment tokens. It is weighted the sentiment expressed in conditional expressions and sentiment shifters. These correspond to tokens that neutralize the sentiment weight or invert the polarity of sentiment tokens respectively.

3.2 Subjective Textual Data

Subjectivity in natural language refers to certain combinations of the language used to express an opinion (Liu, 2010). Early work (Wiebe, 1994) defines subjectivity classification as an algorithm that evaluates in a sentence or document the linguist elements that express a sentiment. Since for this task the goal is to create a classifier that can distinguish subjective from objective sentences, we perform this task by creating a classification model that uses subjective, and objective, labeled data to train the model. Hence, this method will allow to detect the existing subjective and objective vocabulary differences. For this task we have chosen Vowpal Wabbit (VW) linear sigmoid function.

3.3 Sentiment Words Lexicon

One of the most important indicators in the analysis of subjective text are sentiment words. Researchers have examined the viability of building such lexicons (Baccianella et al., 2010; Rao and Ravichandran, 2009). Obtaining a sentiment lexicon is an important but complex step with many unsolved questions (Liu, 2012). Depending on the domain, sentiment words may hold opposite directions and come with different sentiment weights. To this end, we propose a corpus-based approach to detect sentiment words. In the proposed method a seed list of generic sentiment words is used to classify sentence polarity. These words are later used to learn additional sentiment words.

We follow a statistical approach to detect and weight relevant sentiment sentiment words. The sentiment weight for a given unigram and bigram is computed with the Chi-square (χ^2) probabilistic model:

$$\chi^2 = \frac{N(AD-CB)^2}{(A+B)(B+D)(A+B)(C+D)}, \tag{1}$$

where w is an unigram or bigram, N the number of positive and negative sentences, A the number o occurrences of w in positive sentences, B the number of occurrences of w in negative sentences, C the number of occurrences of positive sentences in which w did not occur, D the number of negative sentences in which w did not occur.

4 EXPERIMENTS

4.1 Datasets

The proposed framework is split in four main tasks: subjective evaluation, expanded sentiment words lexicon, sentence polarity evaluation and identification of the opinion targets. For this tasks the following datasets were used:

- Subjective (Pang and Lee, 2005): This dataset is used for the subjective classifier. Contains 5,000 subjective and 5,000 objective sentences from Rotten Tomatoes movie reviews and the respective IMDb movies' plot summaries. (Pang and Lee, 2005) marked Rotten Tomatoes snippets as subjective sentences, and IMDb plot summaries as objective sentences.
- IMDb-Extracted: This dataset is used to expand the sentiment-lexicon. A total of 7,443,722 sentences were collected from IMDb. The IMDb

Table 2: Detailed information of IMDb-Extracted dataset.

Split	#Sentences	Description
A	3,890,540	Train polarity classifier.
B	3,553,182	Test polarity classifier.

reviews are rated in a range from 1 to 10 stars. Following previous work (Pang et al., 2002; Bespalov et al., 2011; Moshfeghi et al., 2011; Qu et al., 2010) reviews rated above 6 were labeled as positive, otherwise negative. Also, if a sentence belongs to a positive review is labeled as positive, otherwise negative. This dataset contains 4,705,351 positive and 2,738,371 negative sentences. Table 2 presents the detained information on this dataset.

- Twitter: This dataset contains a total of 4,341 tweets in which 2,815 were manually annotated with related concepts (e.g. PER/Obama; ORG/NASA [1]. For the present work the annotated tweets were used to train and evaluate the opinion target detection approach.
- Restaurants: SemEval 2015 Task 12 (Aspect Based Sentiment Analysis) released a opinion target dataset [2]. This dataset contains 1,850 sentences in which enclose a total of 2,187 opinion targets.

The evaluation of the algorithms is given by the standard evaluation measures of precision (p), recall (r) and F-score, which is the harmonic mean between p and r,

$$Fscore = \frac{2 \cdot p \cdot r}{p + r} \quad (2)$$

4.2 Extracted Sentiment Words

In this step, Freeling natural language analyzer is used to perform grammatical tagging. Also, in each tweet jargon is identified and evaluated (e.g. emotions and internet slang). The following expressions were used as sentiment shifters “not”, “however”, “rather”, “never”, “nothing” and “scarcely”; and “if”, “though”, “without” and “despite” as conditional expressions. To build the sentiment lexicon different sources were used to identify and score the intensity of an opinion word: Twitrratr (Go et al., 2009), SentiWordNet (Esuli and Sebastiani, 2006), PMI-IR (Turney, 2002; Turney, 2001), emotions smiles, and an acronyms list of internet slang.

Twitrratr evaluates the sentiment in humans generated tweets and contains a list of 174 positive and

[1]http://oak.dcs.shef.ac.uk/msm2013/challenge.html

[2]http://alt.qcri.org/semeval2015/task12/

Table 3: Examples of acronyms and smiles used to express emotions.

Acronyms	YTB - You are The Best	Positive
	BF4L - Best Friends For Life	Positive
Smiles	=)	Positive
	=(	Negative

Table 4: Detailed information of IMDb-Extracted dataset.

split	#sentences		#total
	#positive	#negative	
A	2,462,991	1,427,549	3,890,540
B	2,242,360	1,310,822	3,553,182

185 negative words. SentiWordNet (Esuli and Sebastiani, 2006) is a popular linguistic dictionary that contains a lexicon created semi-automatically by means of linguistic classifiers and human annotation. Regarding this lexicon, 154,745 opinion words were considered. Finally, an acronyms list that contains 352 internet slang acronyms and an emotion smiles list that contain 85 labeled emotions (Table 3). For the IMDb-Extracted dataset, pointwise mutual information ((Turney, 2002; Turney, 2001)) was applied. Pointwise mutual information (PMI) observes the probability of two words co-occurring together, and individually by measuring the degree of statistical dependence between two words. For this task it was used as references words “excellent” and “poor”.

$$PMI = \frac{hits(word, ”excellent”) \cdot hits(”poor”)}{hits(word, ”poor”) \cdot hits(”excellent”} \quad (3)$$

In Equation 3 $hits(word)$ and $hits(word, excellent)$ are obtained by the number of hits a search engine returns using these keywords as search queries. Using PMI we obtained 63,771 opinion words. Furthermore, applying the method proposed in Section 3.3, we captured and scored a total of 2,643,317 opinion words. These opinion words were extracted from the 3,890,400 sentences of the IMDb-Extracted split A dataset (Table 4).

4.3 Subjective Classification

To conduct the subjective classification experiments IMDb-Extracted dataset is split into two disjoints subsets for evaluation purposes (see Table 4). Each sentence in the IMDb-Extracted dataset has on average 19 words and 114 characters. For this task the sentences were evenly split into two subsets: train and test. The train split contains 3,333 subjective and objective sentences respectively and the test split contains 1,667 subjective and objective sentences respectively. For this task we achieve an F-score of 67

Table 5: Sentences from the IMDb-Extracted dataset classified as objective (OBJ) and subjective (SUBJ).

OBJ	"It was the first film made by Thomas Edison on his motion picture, camera." "In this film Melies designs, bullet that resembles a rocket."
SUBJ	"The movie's director knew how to arouse people's imagination's and thought this project would work and it did!" "Despite this viewers in the 21st century can still relate to these, themes and enjoy the story."

Table 6: Polarity classification of subjective sentences.

	VW	Nave-Bayes(I)	Nave-Bayes(II)
Precision	0.77	0.66	0.62
Recall	0.67	0.35	0.88
F1-score	0.72	0.45	0.73

Table 5 illustrates an example taken from IMDb-Extracted dataset. The sentences were evaluated with the subjective classifier model in which two were classified as objective (OBJ), as the other two were classified as subjective (SUBJ).

4.4 Evaluation: Polarity Classification

To evaluate the extracted sentiment words the following classifications were performed: (i) Linear classifier (VW [3]): each sentence is represented by the respective opinion words frequencies, (ii) Naive-Bayes(I): each sentence is represented by its sentiment word sentiment lexicon score and (iii) Naive-Bayes(II): it is only observed sentiment words that occur in the extracted sentiment words. In (i) the model is built with a train and test split from IMDb-Extracted and (ii) and (iii) have no training phase.

Table 6 presents the polarity classification in which VW outperforms the Naive-Bayes implementations. These results illustrate the discriminative nature of using opinion words in a bag-of-words sentiment classification. VW algorithm can better define the boundaries between positive and negative sentences, however with this classifier we lose the syntax of the sentence which allows our method to identify opinion targets. For this task we used the extracted sentiment words (Section 3.3) as available lexicons.

[3]Vowpal Wabbit available at `https://github.com/JohnLangford/vowpal_wabbit/wiki`.

Table 7: Results of proposed method and SemEval 2015 systems opinion target detection for the Restaurants dataset.

	Precision
Our method	0.71
Elixa	0.69
NLANGP	0.71
IHS-RD-Belarus	0.68

4.5 Evaluation: Opinion Target Detection

For the Twitter and Restaurant datasets Figure 1 illustrates the results obtained by using a maximum of 3 and 6 opinion target patterns from the existing 35 opinion target patterns (see Table 1). In this task each sentence is evaluated with 35 opinion target patterns and as a consequence, we obtain several opinion target candidates. The opinion targets relevance are ranked by using the scores from the sentiment word lexicon. For example, in the sentence "2,000 fetuses found hidden at Thai Buddhist temple _URL_ via _Mention" that is labeled with the opinion target "LOC[4]/Thai Buddhist temple" we obtained 10 opinion targets in which the ones ranked highest are "thai buddhist"/20.9, "url via"/7.2 and "found fetuses"/6.7.

For the Restaurants dataset Table 7 shows the obtained results with the proposed method and (Pontiki et al., 2015) reported results. Elixa team achieved the best results in the SemEval submissions. For the Elixa experiments the authors chose the best combination of features using 5-fold cross validation. The features can be a token and token shape in a 2 range window, 4 characters in a prefix or suffix, bigrams and trigrams. Furthermore the authors have induced three types of word representations Brown (Brown et al., 1992), Clark (Clark, 2003) and Word2Vec (Mikolov et al., 2013).The other teams NLANGP and IHS-RD are described with more detail in (Pontiki et al., 2015). The obtained results show that our method is able to achieve competitive results. Unlike the other methods our approach takes into account the syntactic parsing of each sentence to detect opinion targets patterns.

4.5.1 Detected Opinion Targets

In Figure 1 the predicted opinion target is evaluated as correct if one or more words of the predicted opinion target is within the opinion target label. In addition, it was evaluated the performance of the opinion targets patterns when it is evaluated as correct only for an exact match. For this task we obtained a fairly lower

[4]Location

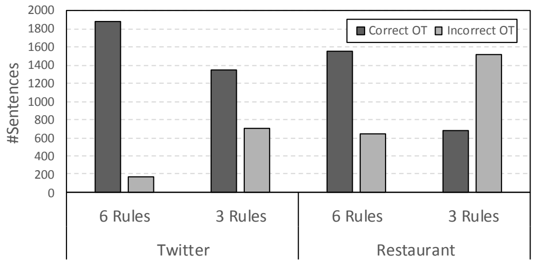

Figure 1: Opinion target prediction with 3 and 6 rules for the Twitter and Restaurant datasets.

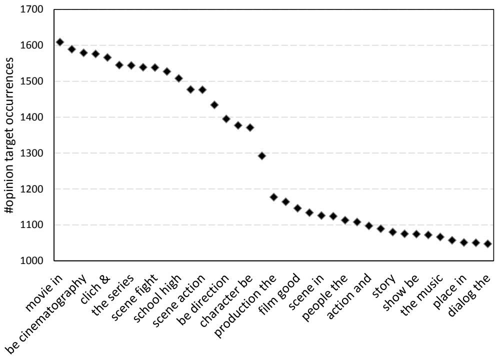

Figure 2: Opinion targets detection.

performance. Observing the results for 6 and 3 opinion target patterns it was correctly predicted 161 and 76 sentences, respectively. Examples of incorrect predictions are “canada de”, “julian win” and “url iran” where the correct opinion targets are “canada”, “julian assange” and “iran”.

Figure 2 illustrates the most frequent opinion targets obtained with the IMDb dataset. For 2,029,121 sentences were extracted 1,740,006 opinion targets. These are promising results since the most frequent opinion targets are clearly related to the domain, and commonly correspond to opinion targets used by users in movie reviews.

5 CONCLUSIONS

In this paper we propose a method to automatically capture opinion targets from humans opinionated sentences. To tackle this problem it was performed an analysis of subjective short-texts (i.e. tweets). To this end we detected text patterns that tend to co-occur previous and after an opinion target. The captured opinion targets are ranked according to its sentiment relevance. With the proposed method we have correctly identified 1,879 from 2,053 opinion targets. Our results show that we are able to extend existing sentiment lexicons. For a sentiment classification task our sentiment lexicon was able to achieve an F1 score of 0.73 which represents an improvement of 28% over baseline sentiment lexicons.

ACKNOWLEDGEMENTS

This work was supported by MULTISENSOR project, partially funded by the European Commission, under the contract number FP7-610411.

REFERENCES

Albornoz, J. C., Chugur, I., and Amigó, E. (2012). Using an Emotion-based Model and Sentiment Analysis Techniques to Classify Polarity for Reputation. In Forner, P., Karlgren, J., and Womser-Hacker, C., editors, *Conference and Labs of the Evaluation Forum, Online Working Notes (CLEF)*, volume 1178.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC)*, 25:2200–2204.

Bespalov, D., Bai, B., Shokoufandeh, A., and Qi, Y. (2011). Sentiment Classification Based on Supervised Latent n-gram Analysis. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 375–382.

Bollen, J. (2010). Determining the public mood state by analysis of microblogging posts. *Alife XII Conf. MIT Press*, page 667.

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479.

Clark, A. (2003). Combining distributional and morphological information for part of speech induction. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1*, pages 59–66. Association for Computational Linguistics.

Diakopoulos, N. and Shamma, D. (2010). Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1195–1198.

Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC)*, 6:417–422.

Ghorbel, H. and Jacot, D. (2010). Sentiment Analysis of French Movie Reviews. *Advances in Distributed Agent-Based Retrieval Tools, 4th International Workshop on Distributed Agent-based Retrieval Tools (DART), Springer Heidelberg*, pages 97–108.

Gildea, D. and Jurafsky, D. (2002). Automatic Labeling of Semantic Roles. *Computational Linguistics, MIT Press Linguistics*, 28(3):245–288.

Go, A., Bhayani, R., and Huang, L. (2009). Twitter Sentiment Classification using Distant Supervision. *CS224N Project Technical report, Stanford*, pages 1–12.

Hatzivassiloglou, V. and Wiebe, J. M. (2000). Effects of adjective orientation and gradability on sentence subjectivity. *Proceedings of the 18th Conference on Computational Linguistics (COLING)*, 1:299–305.

Heerschop, B., Goossen, F., Hogenboom, A., Frasincar, F., Kaymak, U., and De Jong, F. (2011). Polarity analysis of texts using discourse structure. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*, pages 1061–1070.

Hu, M. and Liu, B. (2004). Mining opinion features in customer reviews. *Proceedings of the Association for the Advancement of Artificial Intelligence 19th International Conference on Artifical Intelligence (AAAI)*, pages 755–760.

Kim, S.-M. and Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 1–8.

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing, CRC Press, Taylor and Francis Group*.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies, Morgan and Claypool Publishers*, pages 1–167.

Meena, A. and Prabhakar, T. V. (2007). Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis. In Amati, G., Carpineto, C., and Romano, G., editors, *Proceedings of the 29th European Conference on Advances in Information Retrieval (ECIR)*, volume 4425, pages 573–580.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Moshfeghi, Y., Piwowarski, B., and Jose, J. M. (2011). Handling data sparsity in collaborative filtering using emotion and semantic based features. *Proceedings of the 34th international ACM conference on Research and development in Information Retrieval (SIGIR)*, pages 625–634.

Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 10:1320–1326.

Pang, B. and Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the Association of Computational Linguistics (ACL)*, pages 271–278.

Pang, B. and Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 43(1):115–124.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 10:79–86.

Pontiki, M., Galanis, D., Papageogiou, H., Manandhar, S., and Androutsopoulos, I. (2015). Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), Denver, Colorado*.

Qu, L., Ifrim, G., and Weikum, G. (2010). The bag-of-opinions method for review rating prediction from sparse text patterns. *Proceedings of the 23rd International Conference on Computational Linguistics (COLING)*, pages 913–921.

Rao, D. and Ravichandran, D. (2009). Semi-supervised Polarity Lexicon Induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (COLING)*, EACL '09, pages 675–682, Stroudsburg, PA, USA. Association for Computational Linguistics.

Riloff, E. and Wiebe, J. (2003). Learning extraction patterns for subjective expressions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 105–112.

Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*, pages 417–424.

Turney, P. D. (2001). Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. *Proceedings of the 12th European Conference on Machine Learning (EMCL)*, 2167:491–502.

Wiebe, J. M. (1994). Tracking point of view in narrative. *Journal of Computational Linguistics, MIT Press Cambridge*, 20(2):233–287.

Wiebe, J. M., Bruce, R. F., and O'Hara, T. P. (1999). Development and use of a gold-standard data set for subjectivity classifications. *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics (ACL)*, pages 246–253.

Wilson, T., Wiebe, J., and Hoffmann, P. (2009). Recognizing Contextual Polarity: An Exploration of Features for Phrase-level Sentiment Analysis. *Journal Computational Linguistics*, 35(3):399–433.