Better Document-level Sentiment Analysis from RST Discourse Parsing*

Parminder Bhatia and **Yangfeng Ji** and **Jacob Eisenstein**
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA 30308
parminder.bhatia243@gmail.com, {jiyfeng,jacobe}@gatech.edu

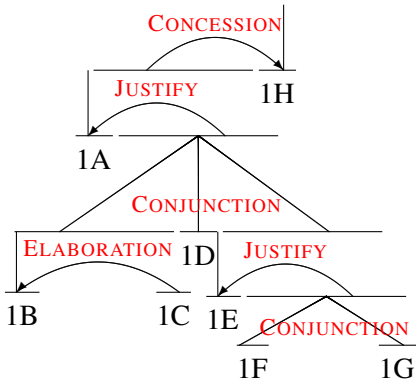
Abstract

Discourse structure is the hidden link between surface features and document-level properties, such as sentiment polarity. We show that the discourse analyses produced by Rhetorical Structure Theory (RST) parsers can improve document-level sentiment analysis, via composition of local information up the discourse tree. First, we show that reweighting discourse units according to their position in a dependency representation of the rhetorical structure can yield substantial improvements on lexicon-based sentiment analysis. Next, we present a recursive neural network over the RST structure, which offers significant improvements over classification-based methods.

1 Introduction

Sentiment analysis and opinion mining are among the most widely-used applications of language technology, impacting both industry and a variety of other academic disciplines (Feldman, 2013; Liu, 2012; Pang and Lee, 2008). Yet sentiment analysis is still dominated by bag-of-words approaches, and attempts to include additional linguistic context typically stop at the sentence level (Socher et al., 2013). Since document-level opinion mining inherently involves multi-sentence texts, it seems that analysis of document-level structure should have a role to play.

A classic example of the potential relevance of discourse to sentiment analysis is shown in Figure 1. In this review of the film *The Last Samurai*, the positive sentiment words far outnumber the negative sentiment words. But the discourse structure — indicated here with Rhetorical Structure Theory (RST; Mann and Thompson, 1988) — clearly favors the final sentence, whose polarity is negative. This example is illustrative in more than one way: it was originally identified by Voll and Taboada (2007), who found that *manually-annotated* RST parse trees improved lexicon-based sentiment analysis, but that automatically-generated parses from the SPADE parser (Soricut and Marcu, 2003), which was then state-of-the-art, did not.



[It could have been a great movie]^{1A} [It does have beautiful scenery,]^{1B} [some of the best since Lord of the Rings.]^{1C} [The acting is well done,]^{1D} [and I really liked the son of the leader of the Samurai.]^{1E} [He was a likable chap,]^{1F} [and I hated to see him die.]^{1G} [But, other than all that, this movie is nothing more than hidden rip-offs.]^{1H}

Figure 1: Example adapted from Voll and Taboada (2007).

Since this time, RST discourse parsing has improved considerably, with the best systems now yielding 5-10% greater raw accuracy than SPADE, depending on the metric. The time is therefore right to reconsider the effectiveness of RST for document-level sentiment analysis. In this paper, we present two different ways of combining RST discourse parses with sentiment analysis. The methods are both relatively simple, and

*Code is available at https://github.com/parry2403/R2N2

Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 2212–2218,
Lisbon, Portugal, 17-21 September 2015.

can be used in combination with an "off the shelf" discourse parser. We consider the following two architectures:

- Reweighting the contribution of each discourse unit, based on its position in a dependency-like representation of the discourse structure. Such weights can be defined using a simple function, or learned from a small of data.

- Recursively propagating sentiment up through the RST parse, in an architecture inspired by recursive neural networks (Smolensky, 1990; Socher et al., 2011).

Both architectures can be used in combination with either a lexicon-based sentiment analyzer, or a trained classifier. Indeed, for users whose starting point is a lexicon-based approach, a simple RST-based reweighting function can offer significant improvements. For those who are willing to train a sentiment classifier, the recursive model yields further gains.

2 Background

2.1 Rhetorical Structure Theory

RST is a compositional model of discourse structure, in which elementary discourse units (EDUs) are combined intro progressively larger discourse units, ultimately covering the entire document. Discourse relations may involve a *nucleus* and a *satellite*, or they may be *multinuclear*. In the example in Figure 1, the unit $1C$ is the satellite of a relationship with its nucleus $1B$; together they form a larger discourse unit, which is involved in a multinuclear CONJUNCTION relation.

The nuclearity structure of RST trees suggests a natural approach to evaluating the importance of segments of text: satellites tend to be less important, and nucleii tend to be more important (Marcu, 1999). This idea has been leveraged extensively in document summarization (Gerani et al., 2014; Uzêda et al., 2010; Yoshida et al., 2014), and was the inspiration for Voll and Taboada (2007), who examined intra-sentential relations, eliminating all words except those in the top-most nucleus within each sentence. More recent work focuses on reweighting each discourse unit depending on the relations in which it participates (Heerschop et al., 2011; Hogenboom et al., 2015). We consider such an approach, and compare it with a compositional method, in which sentiment polarity is propagated up the discourse tree.

Marcu (1997) provides the seminal work on automatic RST parsing, but there has been a recent spike of interest in this task, with contemporary approaches employing discriminative learning (Hernault et al., 2010), rich features (Feng and Hirst, 2012), structured prediction (Joty et al., 2015), and representation learning (Ji and Eisenstein, 2014; Li et al., 2014). With many strong systems to choose from, we employ the publicly-available DPLP parser (Ji and Eisenstein, 2014),[1]. To our knowledge, this system currently gives the best F-measure on relation identification, the most difficult subtask of RST parsing. DPLP is a shift-reduce parser (Sagae, 2009), and its time complexity is linear in the length of the document.

2.2 Sentiment analysis

There is a huge literature on sentiment analysis (Pang and Lee, 2008; Liu, 2012), with particular interest in determining the overall sentiment polarity (positive or negative) of a document. Bag-of-words models are widely used for this task, as they offer accuracy that is often very competitive with more complex approaches. Given labeled data, supervised learning can be applied to obtain sentiment weights for each word. However, the effectiveness of supervised sentiment analysis depends on having training data in the same domain as the target, and this is not always possible. Moreover, in social science applications, the desired labels may not correspond directly to positive or negative sentiment, but may focus on other categories, such as politeness (Danescu-Niculescu-Mizil et al., 2013), narrative frames (Jurafsky et al., 2014), or a multidimensional spectrum of emotions (Kim et al., 2012). In these cases, labeled documents may not be available, so users often employ a simpler method: counting matches against lists of words associated with each category. Such lists may be built manually from introspection, as in LIWC (Tausczik and Pennebaker, 2010) and the General Inquirer (Stone, 1966). Alternatively, they may be induced by bootstrapping from a seed set of words (Hatzivassiloglou and McKeown, 1997; Taboada et al., 2011). While lexicon-based methods may be less accurate than supervised classifiers, they are easier to apply to

[1] https://github.com/jiyfeng/DPLP

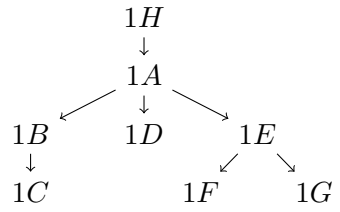


Figure 2: Dependency-based discourse tree representation of the discourse in Figure 1

new domains and problem settings. Our proposed approach can be used in combination with either method for sentiment analysis, and in principle, could be directly applied to other document-level categories, such as politeness.

2.3 Datasets

We evaluate on two review datasets. In both cases, the goal is to correctly classify the opinion polarity as positive or negative. The first dataset is comprised of 2000 movie reviews, gathered by Pang and Lee (2004). We perform ten-fold cross-validation on this data. The second dataset is larger, consisting of 50,000 movie reviews, gathered by Socher et al. (2013), with a predefined 50/50 split into training and test sets. Documents are scored on a 1-10 scale, and we treat scores ≤ 4 as negative, ≥ 7 as positive, and ignore scores of 5-6 as neutral — although in principle nothing prevents extension of our approaches to more than two sentiment classes.

3 Discourse depth reweighting

Our first approach to incorporating discourse information into sentiment analysis is based on quantifying the importance of each unit of text in terms of its discourse depth. To do this, we employ the *dependency-based discourse tree (DEP-DT)* formulation from prior work on summarization (Hirao et al., 2013). The DEP-DT formalism converts the constituent-like RST tree into a directed graph over elementary discourse units (EDUs), in a process that is a close analogue of the transformation of a headed syntactic constituent parse to a syntactic dependency graph (Kübler et al., 2009).

The DEP-DT representation of the discourse in Figure 1 in shown in Figure 2. The graph is constructed by propagating "head" information up the RST tree; if the elementary discourse unit e_i is the satellite in a discourse relation headed by e_j, then there is an edge from e_j to e_i. Thus, the "depth" of each EDU is the number of times in which it is embedded in the satellite of a discourse relation. The exact algorithm for constructing DEP-DTs is given by Hirao et al. (2013).

Given this representation, we construct a simple linear function for weighting the contribution of the EDU at depth d_i:

$$\lambda_i = \max(0.5, 1 - d_i/6). \quad (1)$$

Thus, at $d_i = 0$, we have $\lambda_i = 1$, and at $d_i \geq 3$, we have $\lambda_i = 0.5$. Now assume each elementary discourse unit contributes a prediction $\psi_i = \boldsymbol{\theta}^\top \boldsymbol{w}_i$, where $\boldsymbol{w}_i$ is the bag-of-words vector, and θ is a vector of weights, which may be either learned or specified by a sentiment lexicon. Then the overall prediction for a document is given by,

$$\Psi = \sum_i \lambda_i(\boldsymbol{\theta}^\top \boldsymbol{w}_i) = \boldsymbol{\theta}^\top(\sum_i \lambda_i \boldsymbol{w}_i). \quad (2)$$

Evaluation We apply this approach in combination with both lexicon-based and classification-based sentiment analysis. We use the lexicon of Wilson et al. (2005), and set $\theta_j = 1$ for words marked "positive", and $\theta_j = -1$ for words marked negative. For classification-based analysis, we set $\boldsymbol{\theta}$ equal to the weights obtained by training a logistic regression classifier, tuning the regularization coefficient on held-out data.

Results are shown in Table 1. As seen in the comparison between lines B1 and D1, discourse depth weighting offers substantial improvements over the bag-of-words approach for lexicon-based sentiment analysis, with raw improvements of $4-5\%$. Given the simplicity of this approach — which requires only a sentiment lexicon and a discourse parser — we strongly recommend the application of discourse depth weighting for lexicon-based sentiment analysis at the document level. However, the improvements for the classification-based models are considerably smaller, less than 1% in both datasets.

4 Rhetorical Recursive Neural Networks

Discourse-depth reweighting offers significant improvements for lexicon-based sentiment analysis, but the improvements over the more accurate classification-based method are meager. We therefore turn to a data-driven approach for combining sentiment analysis with rhetorical structure theory, based on recursive neural networks (Socher et al.,

	Pang & Lee	Socher et al.
Baselines		
B1. Lexicon	68.3	74.9
B2. Classifier	82.4	81.5
Discourse depth weighting		
D1. Lexicon	72.6	78.9
D2. Classifier	82.9	82.0
Rhetorical recursive neural network		
R1. No relations	83.4	85.5
R2. With relations	84.1	85.6

Table 1: Sentiment classification accuracies on two movie review datasets (Pang and Lee, 2004; Socher et al., 2013), described in Section 2.3.

2011). The idea is simple: recursively propagate sentiment scores up the RST tree, until the root of the document is reached. For nucleus-satellite discourse relations, we have:

$$\Psi_i = \tanh(K_n^{(r_i)}\Psi_{n(i)} + K_s^{(r_i)}\Psi_{s(i)}), \quad (3)$$

where i indexes a discourse unit composed from relation r_i, $n(i)$ indicates its nucleus, and $s(i)$ indicates its satellite. Returning to the example in Figure 1, the sentiment score for the discourse unit obtained by combining $1B$ and $1C$ is obtained from $\tanh(K_n^{(\text{elaboration})}\Psi_{1B} + K_s^{(\text{elaboration})}\Psi_{1C})$. Similarly, for multinuclear relations, we have,

$$\Psi_i = \tanh(\sum_{j \in n(i)} K_n^{(r_i)}\Psi_j). \quad (4)$$

In the base case, each elementary discourse unit's sentiment is constructed from its bag-of-words, $\Psi_i = \boldsymbol{\theta}^\top \boldsymbol{w}_i$. Because the structure of each document is different, the network architecture varies in each example; nonetheless, the parameters can be reused across all instances.

This approach, which we call a Rhetorical Recursive Neural Network (R2N2), is reminiscent of the compositional model proposed by Socher et al. (2013), where composition is over the constituents of the syntactic parse of a sentence, rather than the units of a discourse. However, a crucial difference is that in R2N2s, the elements Ψ and K are *scalars*: we do not attempt to learn a latent distributed representation of the sub-document units. This is because discourse units typically comprise multiple words, so that accurate analysis of the sentiment for elementary discourse units is not so difficult as accurate analysis of individual words. The scores for individual discourse units can be computed from a bag-of-words classifier, or, in future work, from a more complex model such as a recursive or recurrent neural network.

While this neural network structure captures the idea of compositionality over the RST tree, the most deeply embedded discourse units can be heavily down-weighted by the recursive composition (assuming $K_s < K_n$): in the most extreme case of a right-branching or left-branching structure, the recursive operator may be applied N times to the most deeply embedded EDU. In contrast, discourse depth reweighting applies a uniform weight of 0.5 to all discourse units with depth ≥ 3. In the spirit of this approach, we add an additional component to the network architecture, capturing the bag-of-words for the entire document. Thus, at the root node we have:

$$\Psi_{\text{doc}} = \gamma\boldsymbol{\theta}^\top(\sum_i \boldsymbol{w}_i) + \Psi_{\text{rst-root}}, \quad (5)$$

with $\Psi_{\text{rst-root}}$ defined recursively from Equations 3 and 4, $\boldsymbol{\theta}$ indicating the vector of per-word weights, and the scalar γ controlling the tradeoff between these two components.

Learning R2N2 is trained by backpropagating from a hinge loss objective; assuming $y_t \in \{-1, 1\}$ for each document t, we have the loss $L_t = (1 - y_t\Psi_{\text{doc},t})_+$. From this loss, we use backpropagation through structure to obtain gradients on the parameters (Goller and Kuchler, 1996). Training is performed using stochastic gradient descent. For simplicity, we follow Zirn et al. (2011) and focus on the distinction between contrastive and non-contrastive relations. The set of contrastive relations includes CONTRAST, COMPARISON, ANTITHESIS, ANTITHESIS-E, CONSEQUENCE-S, CONCESSION, and PROBLEM-SOLUTION.

Evaluation Results for this approach are shown in lines R1 and R2 of Table 1. Even without distinguishing between discourse relations, we get an improvement of more than 3% accuracy on the Stanford data, and 0.5% on the smaller Pang & Lee dataset. Adding sensitivity to discourse relations (distinguishing $K^{(r)}$ for contrastive and non-contrastive relations) offers further improvements on the Pang & Lee data, outperforming the baseline classifier (D2) by 1.3%.

The accuracy of discourse relation detection is only 60% for even the best systems (Ji and Eisen-

stein, 2014), which may help to explain why relations do not offer a more substantial boost. An anonymous reviewer recommended evaluating on gold RST parse trees to determine the extent to which improvements in RST parsing might transfer to downstream document analysis. Such an evaluation would seem to require a large corpus of texts with both gold RST parse trees and sentiment polarity labels; the SFU Review Corpus (Taboada, 2008) of 30 review texts offers a starting point, but is probably too small to train a competitive sentiment analysis system.

5 Related Work

Section 2 mentions some especially relevant prior work. Other efforts to incorporate RST into sentiment analysis have often focused on intra-sentential discourse relations (Heerschop et al., 2011; Zhou et al., 2011; Chenlo et al., 2014), rather than relations over the entire document. Wang et al. (2012) address sentiment analysis in Chinese. Lacking a discourse parser, they focus on explicit connectives, using a strategy that is related to our discourse depth reweighting. Wang and Wu (2013) use manually-annotated discourse parses in combination with a sentiment lexicon, which is automatically updated based on the discourse structure. Zirn et al. (2011) use an RST parser in a Markov Logic Network, with the goal of making polarity predictions at the sub-sentence level, rather than improving document-level prediction. None of the prior work considers the sort of recurrent compositional model presented here.

An alternative to RST is to incorporate "shallow" discourse structure, such as the relations from the Penn Discourse Treebank (PDTB). PDTB relations were shown to improve sentence-level sentiment analysis by Somasundaran et al. (2009), and were incorporated in a model of sentiment flow by Wachsmuth et al. (2014). PDTB relations are often signaled with explicit discourse connectives, and these may be used as a feature (Trivedi and Eisenstein, 2013; Lazaridou et al., 2013) or as posterior constraints (Yang and Cardie, 2014). This prior work on discourse relations within sentences and between adjacent sentences can be viewed as complementary to our focus on higher-level discourse relations across the entire document.

There are unfortunately few possibilities for direct comparison of our approach against prior work. Heerschop et al. (2011) and Wachsmuth et al. (2014) also employ the Pang and Lee (2004) dataset, but neither of their results are directly comparable: Heerschop et al. (2011) exclude documents that SPADE fails to parse, and Wachsmuth et al. (2014) evaluates only on individual sentences rather than entire documents. The only possible direct comparison is with very recent work from Hogenboom et al. (2015), who employ a weighting scheme that is similar to the approach described in Section 3. They evaluate on the Pang and Lee data, and consider only lexicon-based sentiment analysis, obtaining document-level accuracies between 65% (for the baseline) and 72% (for their best discourse-augmented system). Table 1 shows that fully supervised methods give much stronger performance on this dataset, with accuracies more than 10% higher.

6 Conclusion

Sentiment polarity analysis has typically relied on a "preponderance of evidence" strategy, hoping that the words or sentences representing the overall polarity will outweigh those representing counterpoints or rhetorical concessions. However, with the availability of off-the-shelf RST discourse parsers, it is now easy to include document-level structure in sentiment analysis. We show that a simple reweighting approach offers robust advantages in lexicon-based sentiment analysis, and that a recursive neural network can substantially outperform a bag-of-words classifier. Future work will focus on combining models of discourse structure with richer models at the sentence level.

Acknowledgments Thanks to the anonymous reviewers for their helpful suggestions on how to improve the paper. The following members of the Georgia Tech Computational Linguistics Laboratory offered feedback throughout the research process: Naman Goyal, Vinodh Krishan, Umashanthi Pavalanathan, Ana Smith, Yijie Wang, and Yi Yang. Several class projects in Georgia Tech CS 4650/7650 took alternative approaches to the application of discourse parsing to sentiment analysis, which was informative to this work; thanks particularly to Julia Cochran, Rohit Pathak, Pavan Kumar Ramnath, and Bharadwaj Tanikella. This research was supported by a Google Faculty Research Award, by the National Institutes of Health under award number R01GM112697-01, and by the Air Force Office of Scientific Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of these sponsors.

References

José M Chenlo, Alexander Hogenboom, and David E Losada. 2014. Rhetorical structure theory for polarity estimation: An experimental study. *Data & Knowledge Engineering*.

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 250–259, Sophia, Bulgaria.

Ronen Feldman. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4):82–89.

Vanessa Wei Feng and Graeme Hirst. 2012. Text-level Discourse Parsing with Rich Linguistic Features. In *Proceedings of the Association for Computational Linguistics (ACL)*, Jeju, Korea.

Shima Gerani, Yashar Mehdad, Giuseppe Carenini, Raymond T Ng, and Bita Nejat. 2014. Abstractive summarization of product reviews using discourse structure. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD.

Christoph Goller and Andreas Kuchler. 1996. Learning task-dependent distributed representations by backpropagation through structure. In *Neural Networks, IEEE International Conference on*, pages 347–352. IEEE.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 174–181, Madrid, Spain.

Bas Heerschop, Frank Goossen, Alexander Hogenboom, Flavius Frasincar, Uzay Kaymak, and Franciska de Jong. 2011. Polarity analysis of texts using discourse structure. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1061–1070. ACM.

Hugo Hernault, Helmut Prendinger, David A. duVerle, and Mitsuru Ishizuka. 2010. HILDA: A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse*, 1(3):1–33.

Tsutomu Hirao, Yasuhisa Yoshida, Masaaki Nishino, Norihito Yasuda, and Masaaki Nagata. 2013. Single-document summarization as a tree knapsack problem. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pages 1515–1520.

Alexander Hogenboom, Flavius Frasincar, Franciska de Jong, and Uzay Kaymak. 2015. Using rhetorical structure in sentiment analysis. *Communications of the ACM*, 58(7):69–77.

Yangfeng Ji and Jacob Eisenstein. 2014. Representation learning for text-level discourse parsing. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD.

Shafiq Joty, Giuseppe Carenini, and Raymond Ng. 2015. CODRA: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*, 41(3).

Dan Jurafsky, Victor Chahuneau, Bryan R Routledge, and Noah A Smith. 2014. Narrative framing of consumer sentiment in online restaurant reviews. *First Monday*, 19(4).

Suin Kim, JinYeong Bak, and Alice Haeyun Oh. 2012. Do you feel what i feel? social aspects of emotions in twitter conversations. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*, Menlo Park, California. AAAI Publications.

Sandra Kübler, Ryan McDonald, and Joakim Nivre. 2009. Dependency parsing. *Synthesis Lectures on Human Language Technologies*, 1(1):1–127.

Angeliki Lazaridou, Ivan Titov, and Caroline Sporleder. 2013. A bayesian model for joint unsupervised induction of sentiment, aspect and discourse representations. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 1630–1639, Sophia, Bulgaria.

Jiwei Li, Rumeng Li, and Eduard Hovy. 2014. Recursive deep models for discourse parsing. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

William Mann. 1984. Discourse structures for text generation. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics*, pages 367–375. Association for Computational Linguistics.

Daniel Marcu. 1997. The rhetorical parsing of natural language texts. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 96–103.

Daniel Marcu. 1999. The automatic construction of large-scale corpora for summarization research. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 137–144. ACM.

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

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.

Kenji Sagae. 2009. Analysis of Discourse Structure with Syntactic Dependencies and Data-Driven Shift-Reduce Parsing. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, pages 81–84, Paris, France, October. Association for Computational Linguistics.

Paul Smolensky. 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, 46(1):159–216.

Richard Socher, Cliff C. Lin, Andrew Y. Ng, and Christopher D. Manning. 2011. Parsing Natural Scenes and Natural Language with Recursive Neural Networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, Seattle, WA.

Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Swapna Somasundaran, Galileo Namata, Janyce Wiebe, and Lise Getoor. 2009. Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, Singapore.

Radu Soricut and Daniel Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 149–156.

Philip J. Stone. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

Maite Taboada. 2008. SFU review corpus. `http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html`.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Rakshit Trivedi and Jacob Eisenstein. 2013. Discourse connectors for latent subjectivity in sentiment analysis. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 808–813, Stroudsburg, Pennsylvania. Association for Computational Linguistics.

Vinícius Rodrigues Uzêda, Thiago Alexandre Salgueiro Pardo, and Maria Das Graças Volpe Nunes. 2010. A comprehensive comparative evaluation of rst-based summarization methods. *ACM Transactions on Speech and Language Processing (TSLP)*, 6(4):4.

Kimberly Voll and Maite Taboada. 2007. Not all words are created equal: Extracting semantic orientation as a function of adjective relevance. In *Proceedings of Australian Conference on Artificial Intelligence*.

Henning Wachsmuth, Martin Trenkmann, Benno Stein, and Gregor Engels. 2014. Modeling review argumentation for robust sentiment analysis. In *Proceedings of the International Conference on Computational Linguistics (COLING)*.

Fei Wang and Yunfang Wu. 2013. Exploiting hierarchical discourse structure for review sentiment analysis. In *Proceedings of the Conference on Asian Language Processing (IALP)*, pages 121–124.

Fei Wang, Yunfang Wu, and Likun Qiu. 2012. Exploiting discourse relations for sentiment analysis. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 1311–1320, Mumbai, India.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pages 347–354.

Bishan Yang and Claire Cardie. 2014. Context-aware learning for sentence-level sentiment analysis with posterior regularization. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD.

Yasuhisa Yoshida, Jun Suzuki, Tsutomu Hirao, and Masaaki Nagata. 2014. Dependency-based Discourse Parser for Single-Document Summarization. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*.

Lanjun Zhou, Binyang Li, Wei Gao, Zhongyu Wei, and Kam-Fai Wong. 2011. Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pages 162–171.

Cäcilia Zirn, Mathias Niepert, Heiner Stuckenschmidt, and Michael Strube. 2011. Fine-grained sentiment analysis with structural features. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, pages 336–344, Chiang Mai, Thailand.