# Literature Review: Discourse and Subjectivity in the Analysis of Movie Reviews

**Christopher Hidey**
Department of Computer Science
Columbia University
New York, NY 10027
chidey@cs.columbia.edu

## 1 Introduction

I propose a project to analyze sentiment at the document level using discourse features. I plan to build on existing work using linguistic connectives that may identify similar or contrasting sentiment. Previous work has focused on using explicit markers that identify discourse relations. Discourse relations indicate how two text spans are logically connected. However, many discourse relations are not indicated explicitly but rely on implicit structure to indicate the discourse relation. Identification of these implicit relations is helped by the use of word pairs-cross products of words that are separated by a discourse marker. I hypothesize that these word pairs, when trained on data marked with polarity, will help with the identification of document level sentiment.

I plan to use the corpus created from reviews on the Internet Movie Database (IMDB) (Maas et al., 2011). This corpus contains movie reviews ranging in length from one sentence to a few paragraphs. The dataset as previously used is split into positive and negative reviews, where a positive review has a rating between 7 and 10 inclusive and a negative review has a rating between 1 and 4 inclusive.

## 2 Motivation

Research in sentiment analysis has often been concerned with how the polarity of the words may appear to change but in fact reflect the expression of the author in an alternative manner (Pang and Lee, 2008). Simple lexical approaches fail to take into account the context in which the sentiment is expressed. Sarcasm, for example, may be used to express negative sentiment, but a lexical approach would identify mostly positive words. Negation is another aspect of sentiment where it is often difficult to identify the scope of where the negation applies. Lastly, the author of a text may express their opinion relatively to something that was previously written. This may take the form of expanding a new idea, juxtapose a good example with a negative one, or providing an explanation. Discourse structure indicates relations between clauses or sentences such as comparison and contrast. These relations are used by authors to help with coherence, the way a text is structured for readability. Identifying these relations can help a sentiment classifier to learn when the context has changed and how to represent sentiment differently.

## 3 Related Work

Lately, there has been increased interest in taking advantage of properties of discourse to help identify sentiment. Zhou focuses on predicting sentiment analysis at the discourse level (Zhou, 2013). Lazaridou et al. have also created a model for unsupervised joint inference for discourse and sentiment using Bayesian networks (Lazaridou et al., 2013).

Early work in sentiment analysis recognized the use of common discourse markers to identify adjectives with similar or contrasting polarity (Hatzivassiloglou and McKeown, 1997). Hatzivassiloglou and McKeown created a classifier that takes advantage of the linguistic intuition that adjectives that are conjoined with “and” have similar sentiment and adjectives conjoined with “but” have opposite sentiment. The authors created an initial set of adjectives manually annotated with positive or negative polarity from the Wall Street Journal Corpus. They used annotated pairs to train and test two classifiers: a logistic regression classifier and a rule-based classifier. Then they used the classifier to cluster the adjectives into one of two classes. Because these clusters have no sense of positive or negative, they assigned the positive class to the one with the highest average frequency. They found that it is possible to identify the polarity of these adjectives with high accuracy.

Recent work on sentiment analysis for Twitter has looked at the use of "lightweight" discourse features (Mukherjee and Bhattacharyya, 2012). Researchers focused on models for unstructured, noisy text because many lexical models are trained on structured text and perform poorly out of domain. They identify a list of discourse connectives and semantic operators which may affect the polarity of a clause and create an algorithm to harness this information and weight the polarity according to the discourse information. They used a lexicon based system and train a support vector machine (SVM) in a supervised framework. They also tested their model on structured text (travel reviews) and found that it performs well.

Some researchers created a model for document level sentiment using latent sentence subjectivity (Yessenalina et al., 2010). They state that when using only annotator rationales generated by human judges to support the document level sentiment that it is possible to obtain much higher accuracy than using the full document. Their claim is that this is analogous to only using subjective sentences. However, documents marked with sentiment that also have individual sentences annotated with subjectivity are difficult to obtain so they model the subjectivity as a latent variable using a latent structured SVM. At training time, they attempt to find the weights that maximize:

$$\hat{w} = arg \max_{w} \sum_{t} \max_{h} w^T f(y_t, x_t, h)$$

The feature vector $f$ consists of polarity and subjectivity features $\psi_{pol}$ and $\psi_{subj}$ and they design these features to be orthogonal such that $\psi_{pol}^T \psi_{subj} = 0$. They also include transition features for subjectivity in $\psi_{subj}$.

Other researchers furthered this work by including discourse features (Trivedi and Eisenstein, 2013). They use explicit discourse connectives to identify when there is a change in polarity. For their model, they use a latent structured SVM to train a classifier on movie reviews, using sentence subjectivity as a latent variable. Similarly, the feature vector $f$ is composed of several subsets of features: polarity and subjectivity features based on a bag-of-words model and they also include transition features based on whether certain discourse connectives are present.

This research is closest to what I propose to do. I plan to introduce features for identifying implicit discourse relations as well. Much of the work using discourse relations has focused on exploiting structure when an explicit marker is present. Implicit discourse relations are much more difficult to identify than the explicit relations (Pitler et al, 2009). However, the performance on identifying implicit relations has improved by making use of features other than lexical ones, using syntax, semantics, or features derived using distributional methods. For this project, I plan to research the use of one of these distributional methods: word pairs, which have been used extensively in discourse analysis.

Although implicit discourse relations are more difficult to identify, they are not much less prominent. According to the theory of discourse in the Penn Discourse Tree Bank (PDTB) (Prasad et al, 2008), these discourse relations can be marked explicitly or conveyed implicitly. In the Penn Discourse Treebank corpus, there are 18,459 explicit relations but 16,053 implicit relations. One class of relation that should be relevant to sentiment is *Comparison*. For this class, there are 5,471 explicit relations and 2,441 implicit ones, which is a significant number to be missing.

## 4 Word Pair Features for Implicit Discourse Relation Disambiguation

One approach to identifying discourse relations involves the use of word pairs created from the cross product of words that span a known discourse connective. Early work derived these word pairs from training data (Marcu, 2001). Each word pair was used as a separate feature in a classifier and improvements in identifying implicit relations were gained.

However, this approach results in very sparse vectors used as features. The lack of adequate data for all possible pairs of words requires the model to make inferences it cannot.

Later work looked at using aggregated word pairs as features (Biran and McKeown, 2013). Instead of using word pairs derived from a training set, researchers used the Gigaword corpus to create counts of pairs of words across each of the 102 explicit discourse markers listed by the PDTB and normalized the counts using TF-IDF. Then during training, when an explicit marker is not present, word pairs are created from the cross product of all subsequent words and the cosine similarity between this new vector and each of the 102 word pairs is used as a feature.

I plan to research the polarity of words in context, spanning a discourse marker. Initially, I will do some analysis to determine which discourse markers are most likely to indicate transitions in polarity to and from positive, negative, or neutral polarity. There are 102 explicit markers in the PDTB which all have some likelihood of represent changes in sentiment. This part of the project will be done with lexical analysis of polarity, perhaps using OpinionFinder (Wilson et al., 2005). I plan to use a large corpus marked with sentiment, such as set-aside data from the IMDB corpus. I will experiment with how to weight these markers according to their likelihood of indicating a transition in polarity. Previous methods were not entirely data-driven, instead using linguistic observations to categorize these markers even though some markers are ambiguous.

Next, I plan to create word pairs from another set aside data set used for product reviews, which will be used to calculate cosine similarity between sentences as a feature. When an explicit discourse marker is not present, it is still useful to attempt to determine if there is an implicit discourse relation present. I will experiment with an overall word pair model and separate word pairs from positive and negative reviews, which may change the context in which certain word pairs appear.

Still, even with the aggregated approach, it is not possible to observe every possible word pair without substantially more data. One possible direction is to research word embeddings, latent vector representations of a word that allow for comparisons of words in a latent space, such as word2vec (Mikolov et al., 2013). Using these word embeddings, it may be possible to train separate models to predict whether two words would appear as a word pair for an explicit discourse marker. For example, even though we may not have a frequency estimate for the words "good" and "terrible" when we see the connective "but," we have probably seen "good" and "bad", and "terrible" is similar to "bad" so we can approximate this word pair.

It should also be noted that although many argument spans for discourse relations occur entirely within a single sentence (intra-sentence), previous work mostly focused on relations between sentences (inter-sentence).

## 5 Additional Features

Finally, because I am interested in selecting the subset of relevant subjective sentences, I plan to include some methods from text summarization. Text summarization methods are often concerned with the selection of sentences that most represent the text as a whole. Thus, structural features such as the relative index of the sentence in the document or paragraph are useful. Similarly, I might expect that the subjective sentences would tend to occur in similar locations. A reviewer might start off by saying "This movie was terrible" or they might start with a summary by stating "This movie was about ..." before reviewing the movie in a later paragraph.

For the same reason, I will examine some other global document features. I could create a spline model of the documents to determine local and global maxima and minima. Some aspects of language that might indicate their subjectivity can be determined using the dictionary of affect: pleasantness, activation, and imagery (Whissell, 1989). By calculating the slope and using that as a feature it will give an idea of the proximity to these critical points.

Another possibility is to use sentence similarity (Guo and Diab, 2012) to compare a sentence with the first or last sentence in the text. This would achieve two things: it would give some sense of redundant information for the current sentence and if the first sentence is subjective or subjective it would provide a measure of comparison.

## References


Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1-2), pp. 1135.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. *Predicting the semantic orientation of adjectives*. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, 174-181.

Angeliki Lazaridou, Ivan Titov, and Caroline Sporleder. 2013. *A Bayesian Model for Joint Unsupervised Induction*. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages 1630-1639.

Rakshit Trivedi and Jacob Eisenstein. 2013. *Discourse Connectors for Latent Subjectivity in Sentiment Analysis*. Proceedings of NAACL-HLT 2013, pages 808-813.

Yudong Zhou. 2013. *Fine-grained Sentiment Analysis with Discourse Structure*. Master Thesis, Saarland University.

Maas, Andrew L. and Daly, Raymond E. and Pham, Peter T. and Huang, Dan and Ng, Andrew Y. and Potts, Christopher. 2011. *Learning Word Vectors for Sentiment Analysis*. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 142-150.

Weiwei Guo and Mona Diab. 2012. *Modeling Sentences in the Latent Space*. In Proceedings of ACL.

Subhabrata Mukherjee and Pushpak Bhattacharyya. 2012. *Sentiment Analysis in Twitter with Lightweight Discourse Analysis*. Proceedings of COLING 2012: Technical Papers, pages 18471864.

Ainur Yessenalina, Yisong Yue, and Claire Cardie. 2010. *Multi-Level structured models for Document Level Sentiment Classification*. In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP).

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi and Bonnie Webber. 2008. *The Penn Discourse Treebank 2.0*. In Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC). Marrakech, Morocco.

The PDTB Research Group 2008. *The PDTB 2.0. Annotation Manual*. Technical Report IRCS-08-01. Institute for Research in Cognitive Science, University of Pennsylvania.

Emily Pitler and Ani Nenkova. 2009. *Using Syntax to Disambiguate Explicit Discourse Connectives in Text*. Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Suntec, Singapore, 4 August 2009.

Emily Pitler, Annie Louis, and Ani Nenkova. 2009. *Automatic sense prediction for implicit discourse relations in text* Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, Suntec, Singapore, 4 August 2009.

Or Biran and Kathleen McKeown. 2013. *Aggregated Word Pair Features for Implicit Discourse Relation Disambiguation*. In proceedings of ACL 2013, Sofia, Bulgaria.

Daniel Marcu and Abdessamad Echihabi 2001. *An Unsupervised Approach to Recognizing Discourse Relations*. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.

Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005. *Opinionfinder: A system for subjectivity analysis*. In Proceedings of HLT-EMNLP: Interactive Demonstrations.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. *Distributed Representations of Words and Phrases and their Compositionality*. In Proceedings of NIPS.

Cynthia Whissell. 1989. The dictionary of affect in language. Emotion: Theory, research, and experience, 4:113-131. Academic Press, London.