What's Great and What's Not: Learning to Classify the Scope of Negation for Improved Sentiment Analysis

Isaac G. Councill
Google, Inc.
76 Ninth Avenue
New York, NY 10011
icouncill@google.com

Ryan McDonald
Google, Inc.
76 Ninth Avenue
New York, NY 10011
ryanmcd@google.com

Leonid Velikovich
Google, Inc.
76 Ninth Avenue
New York, NY 10011
leonidv@google.com

Abstract

Automatic detection of linguistic negation in free text is a critical need for many text processing applications, including sentiment analysis. This paper presents a negation detection system based on a conditional random field modeled using features from an English dependency parser. The scope of negation detection is limited to explicit rather than implied negations within single sentences. A new negation corpus is presented that was constructed for the domain of English product reviews obtained from the open web, and the proposed negation extraction system is evaluated against the reviews corpus as well as the standard BioScope negation corpus, achieving 80.0% and 75.5% F1 scores, respectively. The impact of accurate negation detection on a state-of-the-art sentiment analysis system is also reported.

1 Introduction

The automatic detection of the scope of linguistic negation is a problem encountered in wide variety of document understanding tasks, including but not limited to medical data mining, general fact or relation extraction, question answering, and sentiment analysis. This paper describes an approach to negation scope detection in the context of sentiment analysis, particularly with respect to sentiment expressed in online reviews. The canonical need for proper negation detection in sentiment analysis can be expressed as the fundamental difference in semantics inherent in the phrases, "this is great," versus, "this is not great." Unfortunately, expressions of negation are not always so syntactically simple.

Linguistic negation is a complex topic: there are many forms of negation, ranging from the use of explicit cues such as "no" or "not" to much more subtle linguistic patterns. At the highest structural level, negations may occur in two forms (Givón, 1993): morphological negations, where word roots are modified with a negating prefix (e.g., "dis-", "non-", or "un-") or suffix (e.g., "-less"), and syntactic negation, where clauses are negated using explicitly negating words or other syntactic patterns that imply negative semantics. For the purposes of negation scope detection, only syntactic negations are of interest, since the scope of any morphological negation is restricted to an individual word. Morphological negations are very important when constructing lexicons, which is a separate but related research topic.

Tottie (1991) presents a comprehensive taxonomy of clausal English negations, where each form represents unique challenges for a negation scope detection system. The top-level negation categories – denials, rejections, imperatives, questions, supports, and repetitions – can be described as follows:

- **Denials** are the most common form and are typically unambiguous negations of a particular clause, such as, "There is no question that the service at this restaurant is excellent," or, "The audio system on this television is not very good, but the picture is amazing."

- **Rejections** often occur in discourse, where one participant rejects an offer or suggestion of another, e.g., "Can I get you anything else? No." However, rejections may appear in expository text where a writer explicitly rejects a previous supposition or expectation, for instance, "Given the poor reputation of the manufacturer, I expected to be disappointed with the device. This was not the case."

- **Imperatives** involve directing an audience

Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, pages 51–59,
Uppsala, July 2010.

away from a particular action, e.g., "Do not neglect to order their delicious garlic bread."

- **Questions**, rhetorical or otherwise, can indicate negations often in the context of surprise or bewilderment. For example, a reviewer of a desk phone may write, "Why couldn't they include a decent speaker in this phone?", implying that the phone being reviewed does not have a decent speaker.

- **Supports** and **Repetitions** are used to express agreement and add emphasis or clarity, respectively, and each involve multiple expressions of negation. For the purpose of negation scope detection, each instance of negation in a support or repetition can be isolated and treated as an independent denial or imperative.

Tottie also distinguishes between intersentential and sentential negation. In the case of intersentential negation, the language used in one sentence may explicitly negate a proposition or implication found in another sentence. Rejections and supports are common examples of intersentential negation. Sentential negation, or negations within the scope of a single sentence, are much more frequent; thus sentential denials, imperatives, and questions are the primary focus of the work presented here.

The goal of the present work is to develop a system that is robust to differences in the intended scope of negation introduced by the syntactic and lexical features in each negation category. In particular, as the larger context of this research involves sentiment analysis, it is desirable to construct a negation system that can correctly identify the presence or absence of negation in spans of text that are expressions of sentiment. It so follows that in developing a solution for the specific case of the negation of sentiment, the proposed system is also effective at solving the general case of negation scope identification.

This rest of this paper is organized as follows. §2 presents related work on the topic of automatic detection of the scope of linguistic negations. The annotated corpora used to evaluate the proposed negation scope identification method are presented in §3, including a new data set developed for the purpose of identifying negation scopes in the context of online reviews. §4 describes the proposed negation scope detection system. The novel system is evaluated in §5 in terms of raw results on the annotated negation corpora as well as the performance improvement on sentiment classification achieved by incorporating the negation system in a state-of-the-art sentiment analysis pipeline. Lessons learned and future directions are discussed in §6.

2 Related work

Negation and its scope in the context of sentiment analysis has been studied in the past (Moilanen and Pulman, 2007). In this work we focus on explicit negation mentions, also called functional negation by Choi and Cardie (2008). However, others have studied various forms of negation within the domain of sentiment analysis, including work on content negators, which typically are verbs such as "hampered", "lacked", "denied", etc. (Moilanen and Pulman, 2007; Choi and Cardie, 2008). A recent study by Danescu-Niculescu-Mizil et al. (2009) looked at the problem of finding downward-entailing operators that include a wider range of lexical items, including soft negators such as the adverbs "rarely" and "hardly".

With the absence of a general purpose corpus annotating the precise scope of negation in sentiment corpora, many studies incorporate negation terms through heuristics or soft-constraints in statistical models. In the work of Wilson et al. (2005), a supervised polarity classifier is trained with a set of negation features derived from a list of cue words and a small window around them in the text. Choi and Cardie (2008) combine different kinds of negators with lexical polarity items through various compositional semantic models, both heuristic and machine learned, to improve phrasal sentiment analysis. In that work the scope of negation was either left undefined or determined through surface level syntactic patterns similar to the syntactic patterns from Moilanen and Pulman (2007). A recent study by Nakagawa et al. (2010) developed an semi-supervised model for sub-sentential sentiment analysis that predicts polarity based on the interactions between nodes in dependency graphs, which potentially can induce the scope of negation.

As mentioned earlier, the goal of this work is to define a system that can identify exactly the scope of negation in free text, which requires a robustness to the wide variation of negation expression,

both syntactic and lexical. Thus, this work is complimentary to those mentioned above in that we are measuring not only whether negation detection is useful for sentiment, but to what extent we can determine its exact scope in the text. Towards this end in we describe both an annotated negation span corpus as well as a negation span detector that is trained on the corpus. The span detector is based on conditional random fields (CRFs) (Lafferty, McCallum, and Pereira, 2001), which is a structured prediction learning framework common in sub-sentential natural language processing tasks, including sentiment analysis (Choi and Cardie, 2007; McDonald et al., 2007)

The approach presented here resembles work by Morante and Daelemans (2009), who used IGTree to predict negation cues and a CRF metalearner that combined input from k-nearest neighbor classification, a support vector machine, and another underlying CRF to predict the scope of negations within the BioScope corpus. However, our work represents a simplified approach that replaces machine-learned cue prediction with a lexicon of explicit negation cues, and uses only a single CRF to predict negation scopes, with a more comprehensive model that includes features from a dependency parser.

3 Data sets

One of the only freely available resources for evaluating negation detection performance is the BioScope corpus (Vincze et al., 2008), which consists of annotated clinical radiology reports, biological full papers, and biological abstracts. Annotations in BioScope consist of labeled negation and speculation cues along with the boundary of their associated text scopes. Each cue is associated with exactly one scope, and the cue itself is considered to be part of its own scope. Traditionally, negation detection systems have encountered the most difficulty in parsing the full papers subcorpus, which contains nine papers and a total of 2670 sentences, and so the BioScope full papers were held out as a benchmark for the methods presented here.

The work described in this paper was part of a larger research effort to improve the accuracy of sentiment analysis in online reviews, and it was determined that the intended domain of application would likely contain language patterns that are significantly distinct from patterns common in the text of professional biomedical writings. Correct analysis of reviews generated by web users requires robustness in the face of ungrammatical sentences and misspelling, which are both exceedingly rare in BioScope. Therefore, a novel corpus was developed containing the text of entire reviews, annotated according to spans of negated text.

A sample of 268 product reviews were obtained by randomly sampling reviews from Google Product Search[1] and checking for the presence of negation. The annotated corpus contains 2111 sentences in total, with 679 sentences determined to contain negation. Each review was manually annotated with the scope of negation by a single person, after achieving inter-annotator agreement of 91% with a second person on a smaller subset of 20 reviews containing negation. Inter-annotator agreement was calculated using a strict exact span criteria where both the existence *and* the left/right boundaries of a negation span were required to match. Hereafter the reviews data set will be referred to as the Product Reviews corpus.

The Product Reviews corpus was annotated according to the following instructions:

1. **Negation cues:** Negation cues (e.g., the words "never", "no", or "not" in it's various forms) are not included the negation scope. For example, in the sentence, "It was not X" only "X" is annotated as the negation span.

2. **General Principles:** Annotate the minimal span of a negation covering only the portion of the text being negated semantically. When in doubt, prefer simplicity.

3. **Noun phrases:** Typically entire noun phrases are annotated as within the scope of negation if a noun within the phrase is negated. For example, in the sentence, "This was not a review" the string "a review" is annotated. This is also true for more complex noun phrases, e.g., "This was not a review of a movie that I watched" should be annotated with the span "a review of a movie that I watched".

4. **Adjectives in noun phrases:** Do not annotate an entire noun phrase if an adjective is all that is being negated - consider the negation of each term separately. For instance, "Not

[1] http://www.google.com/products/

top-drawer cinema, but still good...": "top-drawer" is negated, but "cinema" is not, since it is still cinema, just not "top-drawer".

5. **Adverbs/Adjective phrases:**

   (a) Case 1: Adverbial comparatives like "very," "really," "less," "more", etc., annotate the entire adjective phrase, e.g., "It was not very good" should be annotated with the span "very good".

   (b) Case 2: If only the adverb is directly negated, only annotate the adverb itself. E.g., "Not only was it great", or "Not quite as great": in both cases the subject still "is great", so just "only" and "quite" should be annotated, respectively. However, there are cases where the intended scope of adverbial negation is greater, e.g., the adverb phrase "just a small part" in "Tony was on stage for the entire play. It was not just a small part".

   (c) Case 3: "as good as X". Try to identify the intended scope, but typically the entire phrase should be annotated, e.g., "It was not as good as I remember". Note that Case 2 and 3 can be intermixed, e.g., "Not quite as good as I remember", in this case follow 2 and just annotate the adverb "quite", since it was still partly "as good as I remember", just not entirely.

6. **Verb Phrases:** If a verb is directly negated, annotate the entire verb phrase as negated, e.g., "appear to be red" would be marked in "It did not appear to be red".

For the case of verbs (or adverbs), we made no special instructions on how to handle verbs that are content negators. For example, for the sentence "I can't deny it was good", the entire verb phrase "deny it was good" would be marked as the scope of "can't". Ideally annotators would also mark the scope of the verb "deny", effectively canceling the scope of negation entirely over the adjective "good". As mentioned previously, there are a wide variety of verbs and adverbs that play such a role and recent studies have investigated methods for identifying them (Choi and Cardie, 2008; Danescu-Niculescu-Mizil et al., 2009). We leave the identification of the scope of such lexical items and their interaction with explicit negation as future work.

hardly	lack	lacking	lacks
neither	nor	never	no
nobody	none	nothing	nowhere
not	n't	aint	cant
cannot	darent	dont	doesnt
didnt	hadnt	hasnt	havnt
havent	isnt	mightnt	mustnt
neednt	oughtnt	shant	shouldnt
wasnt	wouldnt	without	

Table 1: Lexicon of explicit negation cues.

The Product Reviews corpus is different from BioScope in several ways. First, BioScope ignores direct adverb negation, such that neither the negation cue nor the negation scope in the the phrase, "not only," is annotated in BioScope. Second, BioScope annotations always include entire adjective phrases as negated, where our method distinguishes between the negation of adjectives and adjective targets. Third, BioScope includes negation cues within their negation scopes, whereas our corpus separates the two.

4 System description

As the present work focuses on explicit negations, the choice was made to develop a lexicon of explicit negation cues to serve as primary indicators of the presence of negation. Klima (1964) was the first to identify negation words using a statistics-driven approach, by analyzing word co-occurrence with n-grams that are cues for the presence of negation, such as "either" and "at all". Klima's lexicon served as a starting point for the present work, and was further refined through the inclusion of common misspellings of negation cues and the manual addition of select cues from the "Neg" and "Negate" tags of the General Inquirer (Stone et al., 1966). The final list of cues used for the evaluations in §5 is presented in Table 1. The lexicon serves as a reliable signal to detect the presence of explicit negations, but provides no means of inferring the scope of negation. For scope detection, additional signals derived from surface and dependency level syntactic structure are employed.

The negation scope detection system is built as an individual annotator within a larger annotation pipeline. The negation annotator relies on two dis-

tinct upstream annotators for 1) sentence boundary annotations, derived from a rule-based sentence boundary extractor and 2) token annotations from a dependency parser. The dependency parser is an implementation of the parsing systems described in Nivre and Scholz (2004) and Nivre et al. (2007). Each annotator marks the character offsets for the begin and end positions of individual annotation ranges within documents, and makes the annotations available to downstream processes.

The dependency annotator controls multiple lower-level NLP routines, including tokenization and part of speech (POS) tagging in addition to parsing sentence level dependency structure. The output that is kept for downstream use includes only POS and dependency relations for each token. The tokenization performed at this stage is recycled when learning to identify negation scopes.

The feature space of the learning problem adheres to the dimensions presented in Table 2, and negation scopes are modeled using a first order linear-chain conditional random field (CRF)[2], with a label set of size two indicating whether a token is within or outside of a negation span. The features include the lowercased token string, token POS, token-wise distance from explicit negation cues, POS information from dependency heads, and dependency distance from dependency heads to explicit negation cues. Only unigram features are employed, but each unigram feature vector is expanded to include bigram and trigram representations derived from the current token in conjunction with the prior and subsequent tokens.

The distance measures can be explained as follows. Token-wise distance is simply the number of tokens from one token to another, in the order they appear in a sentence. Dependency distance is more involved, and is calculated as the minimum number of edges that must be traversed in a dependency tree to move from one node (or token) to another. Each edge is considered to be bidirectional. The CRF implementation used in our system employs categorical features, so both integer distances are treated as encodings rather than continuous values. The number 0 implies that a token is, or is part of, an explicit negation cue. The numbers 1-4 encode step-wise distance from a negation cue, and the number 5 is used to jointly encode the concepts of "far away" and "not applicable". The maximum integer distance is 5, which was determined empirically.

[2]Implemented with CRF++: http://crfpp.sourceforge.net/

Feature	Description
Word	The lowercased token string.
POS	The part of speech of a token.
Right Dist.	The linear token-wise distance to the nearest explicit negation cue to the right of a token.
Left Dist.	The linear token-wise distance to the nearest explicit negation cue to the left of a token.
Dep1 POS	The part of speech of the the first order dependency of a token.
Dep1 Dist.	The minimum number of dependency relations that must be traversed to from the first order dependency head of a token to an explicit negation cue.
Dep2 POS	The part of speech of the the second order dependency of a token.
Dep2 Dist.	The minimum number of dependency relations that must be traversed to from the second order dependency head of a token to an explicit negation cue.

Table 2: Token features used in the conditional random field model for negation.

The negation annotator vectorizes the tokens generated in the dependency parser annotator and can be configured to write token vectors to an output stream (training mode) or load a previously learned conditional random field model and apply it by sending the token vectors directly to the CRF decoder (testing mode). The output annotations include document-level negation span ranges as well as sentence-level token ranges that include the CRF output probability vector, as well as the alpha and beta vectors.

5 Results

The negation scope detection system was evaluated against the data sets described in §3. The negation CRF model was trained and tested against the Product Reviews and BioScope biological full papers corpora. Subsequently, the practical effect of robust negation detection was measured in the context of a state-of-the-art sentiment analysis system.

Corpus	Prec.	Recall	F1	PCS
Reviews	81.9	78.2	80.0	39.8
BioScope	80.8	70.8	75.5	53.7

Table 3: Results of negation scope detection.

5.1 Negation Scope Detection

To measure scope detection performance, the automatically generated results were compared against each set of human-annotated negation corpora in a token-wise fashion. That is, precision and recall were calculated as a function of the predicted versus actual class of each text token. Tokens made up purely of punctuation were considered to be arbitrary artifacts of a particular tokenization scheme, and thus were excluded from the results. In keeping with the evaluation presented by Morante and Daelemans (2009), the number of perfectly identified negation scopes is measured separately as the percentage of correct scopes (PCS). The PCS metric is calculated as the number of correct spans divided by the number of true spans, making it a recall measure.

Only binary classification results were considered (whether a token is of class "negated" or "not negated") even though the probabilistic nature of conditional random fields makes it possible to express uncertainty in terms of soft classification scores in the range 0 to 1. Correct predictions of the absence of negation are excluded from the results, so the reported measurements only take into account correct prediction of negation and incorrect predictions of either class.

The negation scope detection results for both the Product Reviews and BioScope corpora are presented in Table 3. The results on the Product Reviews corpus are based on seven-fold cross validation, and the BioScope results are based on five-fold cross validation, since the BioScope data set is smaller. For each fold, the number of sentences with and without negation were balanced in both training and test sets.

The system was designed primarily to support the case of negation scope detection in the open web, and no special considerations were taken to improve performance on the BioScope corpus. In particular, the negation cue lexicon presented in Table 1 was not altered in any way, even though BioScope contains additional cues such as "rather than" and "instead of". This had a noticeable effect on on recall in BioScope, although in several cases the CRF was still able to learn the missing cues indirectly through lexical features.

Condition	Prec.	Recall	F1	PCS
BioScope, trained on Reviews	72.2	42.1	53.5	52.2
Reviews, trained on Bioscope	58.8	68.8	63.4	45.7

Table 4: Results for cross-trained negation models. This shows the results for BioScope with a model trained on the Product Reviews corpus, and the results for Product Reviews with a model trained on the BioScope corpus.

In general, the system performed significantly better on the Product Reviews corpus than on BioScope, although the performance on BioScope full papers is state-of-the-art. This can be accounted for at least partially by the differences in the negation cue lexicons. However, significantly more negation scopes were perfectly identified in BioScope, with a 23% improvement in the PCS metric over the Product Reviews corpus.

The best reported performance to date on the BioScope full papers corpus was presented by Morante and Daelemans (2009), who achieved an F1 score of 70.9 with predicted negation signals, and an F1 score of 84.7 by feeding the manually annotated negation cues to their scope finding system. The system presented here compares favorably to Morante and Daelemans' fully automatic results, achieving an F1 score of 75.5, which is a 15.8% reduction in error, although the results are significantly worse than what was achieved via perfect negation cue information.

5.2 Cross training

The degree to which models trained on each corpus generalized to each other was also measured. For this experiment, each of the two models trained using the methods described in §5.1 was evaluated against its non-corresponding corpus, such that the BioScope-trained corpus was evaluated against all of Product Reviews, and the model derived from Product Reviews was evaluated against all of BioScope.

The cross training results are presented in Table 4. Performance is generally much worse, as expected. Recall drops substantially in BioScope,

which is almost certainly due to the fact that not only are several of the BioScope negation cues missing from the cue lexicon, but the CRF model has not had the opportunity to learn from the lexical features in BioScope. The precision in BioScope remains fairly high, and the percentage of perfectly labeled scopes remains almost the same. For Product Reviews, an opposing trend can be seen: precision drops significantly but recall remains fairly high. This seems to indicate that the scope boundaries in the Product Reviews corpus are generally harder to predict. The percentage of perfectly labeled scopes actually increases for Product Reviews, which could also indicate that scope boundaries are less noisy in BioScope.

5.3 Effect on sentiment classification

In addition to measuring the raw performance of the negation scope detection system, an experiment was conducted to measure the effect of the final negation system within the context of a larger sentiment analysis system.

The negation system was built into a sentiment analysis pipeline consisting of the following stages:

1. Sentence boundary detection.
2. Sentiment detection.
3. Negation scope detection, applying the system described in §4.
4. Sentence sentiment scoring.

The sentiment detection system in stage 2 finds and scores mentions of n-grams found in a large lexicon of sentiment terms and phrases. The sentiment lexicon is based on recent work using label propagation over a very large distributional similarity graph derived from the web (Velikovich et al., 2010), and applies positive or negative scores to terms such as "good", "bad", or "just what the doctor ordered". The sentence scoring system in stage 4 then determines whether any scored sentiment terms fall within the scope of a negation, and flips the sign of the sentiment score for all negated sentiment terms. The scoring system then sums all sentiment scores within each sentence and computes overall sentence sentiment scores.

A sample of English-language online reviews was collected, containing a total of 1135 sentences. Human raters were presented with consecutive sentences and asked to classify each sentence as expressing one of the following types of sentiment: 1) positive, 2) negative, 3) neutral, or 4) mixed positive and negative. Each sentence was reviewed independently by five separate raters, and final sentence classification was determined by consensus. Of the original 1135 sentences 216, or 19%, were found to contain negations.

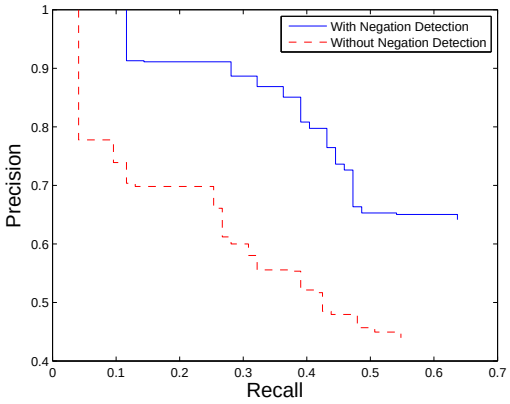

Figure 1: Precision-recall curve showing the effect of negation detection on positive sentiment prediction.

The effect of the negation system on sentiment classification was evaluated on the smaller subset of 216 sentences in order to more precisely measure the impact of negation detection. The smaller negation subset contained 73 sentences classified as positive, 114 classified as negative, 12 classified as neutral, and 17 classified as mixed. The number of sentences classified as neutral or mixed was too small for a useful performance measurement, so only sentences classified as positive or negative sentences were considered.

Figures 1 and 2 show the precision-recall curves for sentences predicted by the sentiment analysis system to be positive and negative, respectively. The curves indicate relatively low performance, which is consistent with the fact that sentiment polarity detection is notoriously difficult on sentences with negations. The solid lines show performance with the negation scope detection system in place, and the dashed lines show performance with no negation detection at all. From the figures, a significant improvement is immediately apparent at all recall levels. It can also be inferred from the figures that the sentiment analysis system is significantly biased towards positive predictions: even though there were significantly more sentences classified by human raters as neg-

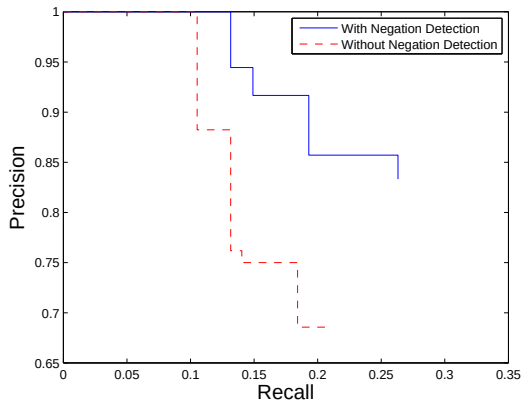

Figure 2: Precision-recall curve showing the effect of negation detection on negative sentiment prediction.

Metric	w/o Neg.	w/ Neg.	% Improv.
Positive Sentiment			
Prec.	44.0	64.1	35.9
Recall	54.8	63.7	20.0
F1	48.8	63.9	29.5
Negative Sentiment			
Prec.	68.6	83.3	46.8
Recall	21.1	26.3	6.6
F1	32.3	40.0	11.4

Table 5: Sentiment classification results, showing the percentage improvement obtained from including negation scope detection (w/ Neg.) over results obtained without including negation scope detection (w/o Neg.).

ative, the number of data points for positive predictions far exceeds the number of negative predictions, with or without negation detection.

The overall results are presented in Table 5, separated by positive and negative class predictions. As expected, performance is improved dramatically by introducing negation scope detection. The precision of positive sentiment predictions sees the largest improvement, largely due to the inherent bias in the sentiment scoring algorithm. F1 scores for positive and negative sentiment predictions improve by 29.5% and 11.4%, respectively.

6 Conclusions

This paper presents a system for identifying the scope of negation using shallow parsing, by means of a conditional random field model informed by a dependency parser. Results were presented on the standard BioScope corpus that compare favorably to the best results reported to date, using a software stack that is significantly simpler than the best-performing approach.

A new data set was presented that targets the domain of online product reviews. The product review corpus represents a departure from the standard BioScope corpus in two distinct dimensions: the reviews corpus contains diverse common and vernacular language patterns rather than professional prose, and also presents a divergent method for annotating negations in text. Cross-training by learning a model on one corpus and testing on another suggests that scope boundary detection in the product reviews corpus may be a more difficult learning problem, although the method used to annotate the reviews corpus may result in a more consistent representation of the problem.

Finally, the negation system was built into a state-of-the-art sentiment analysis system in order to measure the practical impact of accurate negation scope detection, with dramatic results. The negation system improved the precision of positive sentiment polarity detection by 35.9% and negative sentiment polarity detection by 46.8%. Error reduction on the recall measure was less dramatic, but still significant, showing improved recall for positive polarity of 20.0% and improved recall for negative polarity of 6.6%.

Future research will include treatment of implicit negation cues, ideally by learning to predict the presence of implicit negation using a probabilistic model that generates meaningful confidence scores. A related topic to be addressed is the automatic detection of sarcasm, which is an important problem for proper sentiment analysis, particularly in open web domains where language is vernacular. Additionally, we would like to tackle the problem of inter-sentential negations, which could involve a natural extension of negation scope detection through co-reference resolution, such that negated pronouns trigger negations in text surrounding their pronoun antecedents.

Acknowledgments

The authors would like to thank Andrew Hogue and Kerry Hannan for useful discussions regarding this work.

References

Yejin Choi and Claire Cardie. 2007. Structured Local Training and Biased Potential Functions for Conditional Random Fields with Application to Coreference Resolution. *Proceedings of The 9th Conference of the North American Chapter of the Association for Computational Linguistics,* ACL, Rochester, NY.

Yejin Choi and Claire Cardie. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. *Proceedings of the Conference on Empirical Methods on Natural Language Processing.* ACL, Honolulu, HI.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Richard Ducott. 2008. Without a 'doubt'? Unsupervised discovery of downward-entailing operators. *Proceedings of The 10th Annual Conference of the North American Chapter of the Association for Computational Linguistics.* ACL, Boulder, CO.

Talmy Givón. 1993. *English Grammer: A Function-Based Introduction.* Benjamins, Amsterdam, NL.

Edward S. Klima. 1964. Negation in English. *Readings in the Philosophy of Language.* Ed. J. A. Fodor and J. J. Katz. Prentice Hall, Englewood Cliffs, NJ: 246-323.

John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional random elds: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the International Conference on Machine Learning*. Morgan Kaufmann, Williamstown, MA.

Ryan McDonald, Kerry Hannan, Tyler Neylon, Mike Wells, and Jeff Reynar. 2007. Structured Models for Fine-to-Coarse Sentiment Analysis. *Proceedings of the Annual Meeting of the Association for Computational Linguistics.* Prague, Czech Republic.

Karo Moilanen and Stephen Pulman 2007. Sentiment Composition. *Proceedings of the Recent Advances in Natural Language Processing International Conference* Borovets, Bulgaria

Roser Morante and Walter Daelemans. 2009. A metalearning approach to processing the scope of negation. *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL).* ACM, Boulder, CO.

Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi. 2010. Dependency Tree-based Sentiment Classification using CRFs with Hidden Variables. *Proceedings of The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics* ACL, Los Angeles, CA.

Joakim Nivre and Mario Scholz. 2004. Deterministic Dependency Parsing of English Text. *Proceedings of the 20th International Conference on Computational Linguistics.* ACM, Geneva, Switzerland.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gulsen Eryigit Sandra Kubler, Svetoslav Marinov and Erwin Marsi. 2007. MaltParser: A language-independent system for data-driven dependency parsing *Natural Language Engineering* 13(02):95–135

Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis.* MIT Press, Cambridge, MA.

Gunnel Tottie. 1991. *Negation in English Speech and Writing: A Study in Variation* Academic, San Diego, CA.

Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. 2010. The viability of web-derived polarity lexicons. *Proceedings of The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics.* ACL, Los Angeles, CA.

Veronika Vincze, György Szarvas, Richárd Farkas, György Móra, and János Csirik. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(Suppl 11):S9.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing* Vancouver, Canada.