

Scope of Negation Detection in Sentiment Analysis

Maral Dadvar
Human Media Interaction Group
University of Twente
Enschede, Netherlands
m.dadvar@ewi.utwente.nl

Claudia Hauff
Human Media Interaction Group
University of Twente
Enschede, Netherlands
c.hauff@ewi.utwente.nl

Franciska de Jong
Human Media Interaction Group
University of Twente
Enschede, Netherlands
f.m.g.dejong@ewi.utwente.nl

ABSTRACT

An important part of information-gathering behaviour has always been to find out what other people think and whether they have favourable (positive) or unfavourable (negative) opinions about the subject. This survey studies the role of negation in an opinion-oriented information-seeking system. We investigate the problem of determining the polarity of sentiments in movie reviews when negation words, such as *not* and *hardly* occur in the sentences. We examine how different negation scopes (window sizes) affect the classification accuracy. We used term frequencies to evaluate the discrimination capacity of our system with different window sizes. The results show that there is no significant difference in classification accuracy when different window sizes have been applied. However, negation detection helped to identify more opinion or sentiment carrying expressions. We conclude that traditional negation detection methods are inadequate for the task of sentiment analysis in this domain and that progress is to be made by exploiting information about how opinions are expressed implicitly.

Categories and Subject Descriptors

H.3.1 [Information Systems]: Content Analysis and Indexing – Linguistic processing.

General Terms

Reliability, Experimentation, Languages, Human Factors, Information Systems

Keywords

Scope Modelling, Movie Review Analysis, Opinion Mining

1. INTRODUCTION

With the rapid expansion of e-commerce, more products are being sold online. Industry or manufacturing companies that produce these products want to know how their customers feel about them. This information can be acquired by studying opinions from review portals (for example, Amazon and ConsumerReports). At the same time, users or consumers want to know which product to buy or which movie to watch, so they also read reviews and try to

make their decisions accordingly. However, gathering all this online information manually is time consuming. Therefore automatic sentiment analysis is important. Sentiment analysis is defined here as the task of identifying the opinions expressed in text and classifying texts accordingly. To do so, the main task is to extract the opinions, facts and sentiments expressed in these reviews. Example applications are, classifying products or reviews into ‘recommended’ or ‘not recommended’ [1, 2], opinion summarization [3] and subjectivity classification [1, 4] which is the task of determining whether a sentence or a paragraph contains the opinion of the writer. There are also other applications for sentiment analysis, for example, comparison of products, or general opinions on public policy. Sentiment analysis aims at classifying the sentiment of the opinions into polarity types (the common types are positive and negative). This text classification task is also referred to as polarity classification.

Negation is one of the most common linguistic means that can change text polarity. Therefore in sentiment analysis negation has to be taken into account [5, 6]. The scope size of a negation expression determines which sequence of words in the sentence is affected by negation words, such as, *no*, *not*, *never* [6]. Negation terms affect the contextual polarity of words but the presence of a negation word in a sentence does not mean that all of the words conveying sentiments will be inverted [7]. That is why we also have to determine the scope of negation in each sentence. One of the most noticeable works done on examining the affect of different scope models for negation is [7]. Jia et al. have used some linguistic rules to identify the scope of each negation term. The impact of scope modelling for negation applied for sentiment analysis has not been studied a lot compared to domains such as biomedical studies [8-10].

Linguistic negation is a complex topic and there are several forms to express a negative opinion. Negation can be morphological where it is either denoted by a prefix (“dis-”, “non-”) or a suffix (“-less”) [11]. It can be implicit, as in *with this act, it will be his first and last movie*. Although this sentence carries a negative opinion, no negative words are used. Negation can also be explicit, *this is not good*. This last type of negation will be the focus of our experiments. In this paper we studied the effect of scope modelling for negation by comparing the effect of different scope sizes (or window sizes) in the context of sentiment analysis, particularly with respect to sentiments expressed in movie reviews. Scope in negation detection is defined here as the window in which a negation word may affect the other elements of the sentence. We studied how opinions were expressed in each category of reviews and how adjectives and adverbs were used.

This paper is organized as follows; in section 2 the related work on scope detection for negation is introduced. Sections 3 and 4 explain the method and experimental setup. The results and evaluation of the model is presented in section 5 and we round off the paper with the discussion and conclusion in sections 6 and 7.

2. RELATED WORK

Recently [6] did a review on negation and its scope in sentiment analysis. This work presents various computational approaches to modelling negation in sentiment analysis. The focus of this paper is particularly on the scope of negation. It also discusses limits and challenges of negation modelling. For example, recognition of polar expressions (sentences which carry sentiments) is still a challenging task. The authors also discussed that the effectiveness of negation models can change in different corpora because of the specific construction of language in different contexts.

On the effect of negation on sentiment analysis, [7] introduces the concept of the scope of a negation term. The authors employ a decision tree to determine the polarity of the documents. The proposed scope detection method, considers static delimiters (unambiguous words) such as, *because*, dynamic delimiters (ambiguous words) such as, *like*, and heuristic rules which focus on polar expressions. For negation detection they have tried three window sizes; 3, 4 and 5. Their experimental results show that their method outperforms other methods in accuracy of sentiment analysis and the retrieval effectiveness of polarity classification in opinion retrieval. [12] suggests that the scope of negation should be the adjectives close to the negation word. Authors have suggested that the scope of a negation term to be its next 5 words.

In [1] the scope of a negation term is assumed to be the words between the negation term and the first punctuation mark following it. The accuracy of this work is 0.69 based on the previous version (Ver. 0.9) of movie review data. [13] introduces the concept of contextual valence shifters which consist of negation, intensifier and diminisher. Intensifiers and diminishers are terms that change the degree of the expressed sentiments. The sentence, *this movie is very good*, is more positive than *this movie is good*. In the sentence, *this movie is barely any good*, the term *barely* is a diminisher, which makes this statement less positive. They have used a term-counting method, a machine learning method and a combination of both methods on the same data collection as was used in our experiment. They found that combining the two systems slightly improved the results compared to machine learning or term-counting methods alone.

There are other studies on determining the scope of negation mostly in biomedical texts, using machine learning techniques. In recent work by Morante et al. [15], a metalearning approach to processing the scope of negation signals is studied, involving two classification tasks: identifying negation signals and finding the scope, using supervised machine learning methods. They achieved an error reduction of 32.07 %.

3. METHODS

The experiment to determine the sentiments expressed in movie reviews is based on term frequencies. We count the number of occurrences of positive words and negative words in each document. These numbers are compared with each other and the documents are classified accordingly as positive or negative. If the numbers of positive and negative words are equal the document is neutral.

When an explicit negation word occurs in a sentence, it is important to determine the range of words that are affected by this term. The scope may be only the next word after the negation word, for example, *the movie was not interesting* (window size = 1), or a wider range, for instance, *I do not call this film a comedy movie* (window size = 5). In the second sentence the effect of *not* is until the end of the sentence and not only the word following it. A negation does not negate every subsequent word in the sentence. There is no fixed window size. The window can be affected by different combinations of textual features such as adjectives, adverbs, nouns and verbs. When a positive or negative word falls inside the scope of a negation, its original meaning shifts to the opposite one and it is counted as the opposite polarity.

For extracting the opinion words we use the two wordlists. We do not use part of speech tags in our experiment. Considering the word senses given by WordNet¹, it was verified that almost all of the words in the wordlists are adjectives. Few of them belong to other categories (verbs or adverbs) which again only occur in one form, for example verbs such as “adore” and “detest”.

Negation terms are not restricted to *not*. The set of negation terms that we have used in this paper also includes *no*, *not*, *rather*, *hardly* and all the verbs that the word *not* can be concatenated to in the form of *n't*.²

4. EXPERIMENTAL SETUP

We used the Movie Review data set prepared by [14]. This data set contains 2000 movie reviews: 1000 positive and 1000 negative. These reviews were originally collected from the Internet Movie Database (IMDb) archive³. Their classification as positive or negative was automatically extracted from the ratings and will be used as ground truth.

In order to identify the positive and negative terms in the documents we use two wordlists. The positive wordlist⁴ consists of 136 words which are used to express positive opinions. For example, “good” is one of the positive words along with its synonyms such as, “fascinating” and “absorbing” which were also added to the list. The negative wordlist⁵ contains 109 negative words which are used to express negative opinions (for example “boring” and its synonyms “awful”, “dull” and “tedious”). These lists are derived from online dictionaries such as synonyms.com and the words proposed in [1]. Following [1] we also use “?” and “!” as negative words in the wordlist.

Of the total number of the words in the positive list, 20 never occurred in any of the reviews and the rest of the words occurred on average 44 times in the whole corpus. From the negative list, 18 have never occurred in any of the documents and the rest of the words were used 75 times on average in the corpus. Figures 1 and 2 illustrate the frequency of the 30 most repeated positive and negative words in the corpus.

¹ <http://wordnet.princeton.edu/> [Accessed 24 October 2010]

² List of the negation words is accessible online: <http://wwwhome.ewi.utwente.nl/~dadvarm/dir2011/negation.txt>

³ <http://www.cs.cornell.edu/people/pabo/movie-review-data/> [Accessed 24 October 2010]

⁴ The positive words list is accessible online: <http://wwwhome.ewi.utwente.nl/~dadvarm/dir2011/positive.txt>

⁵ The negative words list is accessible online: <http://wwwhome.ewi.utwente.nl/~dadvarm/dir2011/negative.txt>

Our aim in this work is to examine whether negation detection affects sentiment analysis and improves the classification. Moreover, we evaluated the effect of different window sizes (scope) in negation detection. We started our experiment by classifying the movie reviews, without considering the negation (step 1). In each document, the numbers of occurrences of the wordlists' words were counted. Accordingly the reviews were classified as positive, negative or neutral.

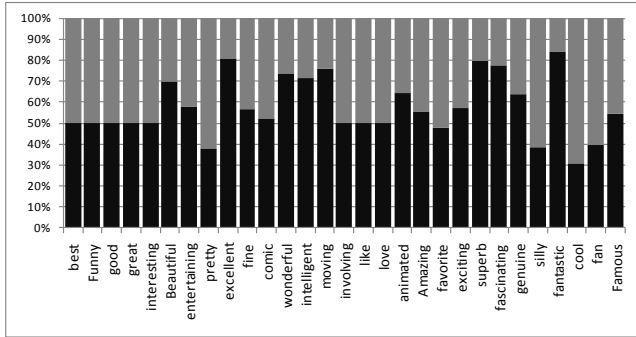


Figure 1. Frequency comparison (%) of the 30 most repeated positive words in positive (black) and negative (grey) documents.

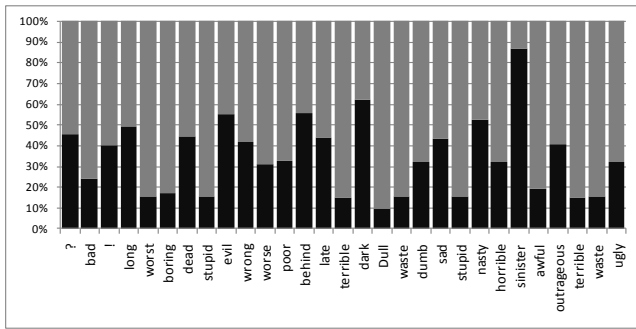


Figure 2. Frequency comparison (%) of the 30 most repeated negative words in positive (black) and negative (grey) documents.

In the second step we employed negation detection, considering only the term *not* as the negation word. We checked how results would change in different window sizes. Then (step 3) we extended our negation words by adding the words *no*, *rather*, *hardly*. The verbs which were negated with *n't* were then added to the negation word lists in step 4. We repeated our experiment with different window sizes from 1 up to and including 5.

We used Accuracy as our evaluation metric. We evaluated the classification of our system by comparing it with the Naive baseline which all the documents are classified as positive, i.e., precision 0.5, recall 1, and accuracy 0.5.

5. RESULTS

We repeated the [1] experiment using same word lists and corpus to evaluate our system. Pang et al. have used more limited wordlists (Negatives = 7, Positives = 7). Our results of sentiment analysis (without negation detection) with accuracy of 0.70 comply with the results of [1] with an accuracy of 69%. The overall accuracy of the first step of our experiment (sentiment analysis without negation) was 65%, true positive rate (recall) was

84% and precision was 62%. Table 1 shows the accuracy results after applying negation detection using only *not* as negation word (step 2). Accuracy results of step 3 and step 4 are shown in tables 2. There are no significant differences in the results with different negation words and window sizes. Negation detection in window sizes 4 and 2, and using *no*, *not*, *rather*, *hardly* as the negation words, resulted in more accurate classification. Recall was always higher than precision in all experiments which suggests poor negative review classification.

We also counted the number of adjectives and adverbs in the dataset. There were more adjectives and adverbs in positive documents compared to negative documents. (Table 3)

Table 1. Accuracy results of sentiment analysis (SA) before and after applying negation detection (ND) using only *not* as the negation word in different window sizes (WS)

Experiment	Recall	Precision	Accuracy
SA without ND	0.83	0.62	0.65
WS 5	0.83	0.62	0.65
WS 4	0.83	0.62	0.65
WS 3	0.83	0.62	0.65
WS 2	0.83	0.61	0.65
WS 1	0.83	0.61	0.65

Table 2. Accuracy results after applying negation detection using *no*, *not*, *rather*, *hardly*, and the verbs which were negated with *n't* as the negation words in different window sizes (WS)

Negation words	WS 5	WS 4	WS 3	WS 2	WS 1
<i>not</i>	0.65	0.65	0.65	0.65	0.65
<i>no</i> , <i>not</i> , <i>rather</i> , <i>hardly</i>	0.66	0.70	0.66	0.70	0.65
<i>no</i> , <i>not</i> , <i>rather</i> , <i>hardly</i> and the verbs which were negated with <i>n't</i> .	0.67	0.66	0.66	0.66	0.66

Table 3. Mean number of adjectives and adverbs in each review

Dataset	Type	Mean	Std. Dev.
Positives	Adjectives	66.0	31.6
	Adverbs	47.8	26.2
Negatives	Adjectives	57.5	24.6
	Adverbs	44.9	22.6

6. DISCUSSION

We studied the impact of negation detection in sentiment analysis in movie reviews. We tested different negation scopes to investigate how it would affect the polarity identification of the sentences. We hypothesized to observe significant improvements on the classification of the documents after applying negation detection. In our experiment we assumed that opinions are mostly expressed by the use of adjectives and adverbs. Therefore, we classified the reviews as negative or positive according to the

number of occurrences of these types of words. After failure analysis, we realized that most of the sentiments and opinions are expressed implicitly, for example, “ ... *I have a problem even regarding it as a film, it's more of a show*”.

The negation words that we have used in our experiment, according to grammatical rules should usually be followed by either an adjective or an adverb. Therefore, in our case (adjective and adverbs are not commonly used to express the opinions), negation did not have much influence on the outcome. The majority of reviewers have used sarcasm sentences, comparison and metaphor, for instance the sentences;

“now , I saw this scene coming from a mile away , but I said to myself , " that is impossible . there's no way they'll do that . . . oh god ! " they did do it . it's there . ”

“Now what didn't work in this movie? would be the rest of it”.

Although we extended the word lists compare to [1], the result of classification did not improve significantly. This can support our claim that since the opinions are mostly expressed indirectly, the number of adjectives does not have much effect on the outcome.

As it is illustrated in the figures 1 and 2, there are also words which are considered to be positive but are equally or even more occurred in the negative documents than the positive ones and vice versa. For example, the word *cool*, which is one of the words from the positive word list, it is more frequently occurred in negative documents than the positive documents. This can also be another reason for misclassifications. A pre-enhancement of the wordlists, considering the language used in the dataset, may also improve the classifications.

Many emotions and opinions are expressed in the form of question or surprise. The results show that “?” and “!” are the most repeated ones in the documents, ! in negative documents = 527, in positive documents = 352 and ? in negative documents = 1092, in positive documents = 913. As it was mentioned in [1], negative sentiments are most likely to be expressed by – at least – one of these punctuation marks.

Our results also illustrate higher recall than precision which implies a better discrimination capacity in positive documents (in step 1, TP = 794 vs. TN = 431). A possible reason for higher misclassifications in negative documents can also be the number of adjectives and adverbs. In the positive documents more opinions are conveyed by explicit use of adjectives or adverbs in comparison to the negative documents (Table 3).

More investigation on falsely classified documents revealed that in some cases the negation word appears after the words which convey sentiments. For example, “*sounds great huh? well it's not*”, where the adjective *great* is located four words before the negation word *not*.

7. CONCLUSION

We conclude that traditional negation detection methods are inadequate for sentiment analysis in this domain. In addition to the explicit elements, there are other indirect elements that affect the polarity of sentences, either positively or negatively. In some cases the opinion words are used before the negation word, therefore, it might be wise to also take them into account while setting the negation scope. It is important to study which lexical features are mainly used to express the sentiments implicitly. Sarcasm and metaphor detection may also improve the

classifications accuracy. We also would like to extend our research by performing more detailed analysis using machine learning approaches.

ACKNOWLEDGMENTS

We thank Thijs Verschoor and Dolf Trieschnigg for their useful advice on programming. Thanks also to Alessandro Valitutti for his helpful comments and technical supports. The language editing was kindly done by Lynn Packwood. We are grateful to four anonymous reviewers for their valuable comments and suggestions. This research is funded by the EU project PuppyIR <http://www.puppyir.eu> (EU FP7 231507).

REFERENCES

- [1] Pang, B., L. Lee, and S. Vaithyanathan. *Thumbs up? Sentiment classification using machinelearning techniques*. in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*. 2002.
- [2] Turney, P.D. *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*. in *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. 2002. Philadelphia, USA.
- [3] Ku, L., et al. *Major topic detection and its application to opinion summarization*. in *SIGIR '05 Proceedings of the 28th annual international ACM conference on Research and development in information retrieval*. 2005. Salvador, Brazil.
- [4] Yu, H. and V. Hatzivassiloglou. *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences*. in *Conference on Empirical Methods in Natural Language Processing*. 2003: Association for Computational Linguistics.
- [5] Jia, L., C. Yu, and W. Meng. *The effect of negation on sentiment analysis and retrieval effectiveness*. 2009: ACM.
- [6] Wiegand, M., et al. *A survey on the role of negation in sentiment analysis*. in *'10 Proceedings of the Workshop on Negation and Speculation in Natural Language Processing 2010*: Association for Computational Linguistics.
- [7] Jia, L., C. Yu, and W. Meng. *The effect of negation on sentiment analysis and retrieval effectiveness*. in *8th International Conference on Information and Knowledge Management*. 2009. Hong Kong.
- [8] Chapman, W., et al., *A simple algorithm for identifying negated findings and diseases in discharge summaries*. *Journal of biomedical informatics*, 2001. **34**(5): p. 301-310.
- [9] Goldin, I. and W. Chapman. *Learning to detect negation with 'not' in medical texts*. in *Workshop at the 26th ACM SIGIR Conference*. 2003.
- [10] Morante, R., A. Liekens, and W. Daelemans. *Learning the scope of negation in biomedical texts*. in *EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing 2008*.
- [11] Councill, I., R. McDonald, and L. Velikovich. *What's great and what's not: learning to classify the scope of negation for improved sentiment analysis*. in *'10 Proceedings of the Workshop on Negation and Speculation in Natural Language Processing 2010*.
- [12] Hu, M. and B. Liu. *Mining and summarizing customer reviews*. in *Tenth ACM International Conference on Knowledge Discovery and Data Mining*. 2004. Seattle, WA.
- [13] Kennedy, A. and D. Inkpen, *Sentiment classification of movie reviews using contextual valence shifters*. *Computational Intelligence*, 2006. **22**(2): p. 110-125.
- [14] Pang, B. and L. Lee. *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*. in *Proceedings of the ACL*. 2004: Association for Computational Linguistics.
- [15] Morante, R. and W. Daelemans. *A metalearning approach to processing the scope of negation*. in *Proceedings of the CoNLL*. 2009: Association for Computational Linguistics.