Accounting for Negation in Sentiment Analysis

Bas Heerschop
basheerschop@gmail.com

Paul van Iterson
paulvaniterson@gmail.com

Alexander Hogenboom
hogenboom@ese.eur.nl

Flavius Frasincar
frasincar@ese.eur.nl

Uzay Kaymak
kaymak@ese.eur.nl

Econometric Institute
Erasmus University Rotterdam
PO Box 1738, NL-3000 DR
Rotterdam, the Netherlands

ABSTRACT

Automated ways of analyzing sentiment in Web data are becoming more and more urgent as virtual utterances of opinions or sentiment are becoming increasingly abundant on the Web. The role of negation in sentiment analysis has been explored only to a limited extent. In this paper, we investigate the impact of accounting for negation in sentiment analysis. To this end, we utilize a basic sentiment analysis framework – consisting of a wordbank creation part and a document scoring part – taking into account negation. Our experimental results show that by accounting for negation, precision relative to human ratings increases with 1.17%. On a subset of selected documents containing negated words, precision increases with 2.23%.

1. INTRODUCTION

In recent years, utterances of opinions or sentiment have become increasingly abundant on the Web through messages on Twitter, on-line customer reviews, etcetera. The information contained in this ever-growing data source is invaluable to key decision makers, e.g., those making decisions related to reputation management or marketing. An understanding of what is going on in their particular markets is crucial for decision makers, yet the analysis of sentiment in an overwhelming amount of data is far from trivial.

Sentiment analysis aims to determine the attitude, evaluation, or emotions of the author with respect to the subject of a text. This may involve word sentiment scoring (i.e., learning the sentiment scores of single words), subject/aspect relevance filtering (i.e., determining the subject and/or aspect a sentiment carrying word is relevant to), subjectivity analysis (i.e., determining whether a sentence is subjective or objective), or sentiment amplification and negation (i.e., modifying sentiment strength on amplifying words and reversing sentiment scores on negated words). The impact of taking into account negation in sentiment analysis has not been demonstrated yet. Therefore, we present our first steps towards insight in the impact of negation on sentiment analysis. A more elaborate analysis may be found in an extended version of this work [2].

DIR 2011 February 4, 2011, Amsterdam, The Netherlands.

2. SENTIMENT ANALYSIS

Most approaches to sentiment analysis (i.e., classification) of documents essentially adhere to more or less similar frameworks consisting of creating a list of words and their associated sentiment from a training corpus and a subsequent method for scoring documents. An example of such a framework is the basic framework proposed by Ceserano et al. [1], who provide two word scoring algorithms based on supervised learning and three sentence-level document scoring algorithms with topic relevance filtering. Despite adhering to similar frameworks, document sentiment analysis approaches have several characteristic features distinguishing them from one another.

Sentiment may be scored on document level, sentence level, or window level. In this process, most approaches rely on a wordbank, typically containing per-word sentiment scores. Creation methods include supervised learning on a set of manually rated documents, learning through related word expansion, completely manual creation, or a combination of these methods. In matching words in a text with words in a wordbank, some approaches as lemmatization are designed to cope with syntactical variations. Part-of-speech tagging is also considered to be helpful in sentiment analysis, as it may help algorithms to, for example, distinguish sentiment-carrying words like adjectives or adverbs. Additionally, some algorithms attempt to identify subjective phrases or phrases relevant to the topic considered in order to boost sentiment analysis performance. Other helpful techniques include taking into account amplification or negation of sentiment carrying words. The role of negations has however been explored only to a limited extent. Therefore, we propose to shed some light onto the impact of accounting for negation in sentiment analysis.

3. SENTIMENT NEGATION

In order to assess the impact of sentiment negation, we propose a very simple sentiment analysis framework, consisting of wordbank creation and subsequent lexicon-based document scoring. Both parts have optional support for sentiment negation. We classify a document as either positive (1), neutral (0), or negative (-1). The score range of individual words is [-1, 1]. We focus on adjectives.

The first part of our framework facilitates wordbank creation, involving scoring sentiment of individual words (adjectives) w in a training corpus D_{train}. Our word scoring function is based on a pseudo-expected value function [1]. The sentiment score of any adjective w, $\text{score}(w)$, is based on its total relative influence on the sentiment over all documents $d \in D_w$, where $D_w \subseteq D$, with each document containing w:

$$\text{score}(w) = \frac{\sum_{d \in D_w} \text{score}(d) \times \inf(w, d, neg)}{|D_w|}, \quad (1)$$

where $\text{score}(d)$ is a document d's manually assigned score, $|D_w|$ is the number of documents in D_w, and $\inf(w,d,neg)$ is the relative influence of an adjective w in document d, with neg indicating whether to account for negation or not. This influence is calculated as the count $\text{freq}(w,d,neg)$ of w in d in terms of the total frequency $\sum_{w' \in d} \text{freq}(w',d,neg)$ of all sentiment carrying words w' in d:

$$\inf(w,d,neg) = \frac{\text{freq}(w,d,neg)}{\sum_{w' \in d} \text{freq}(w',d,neg)}. \quad (2)$$

In order to support negation in our framework, we use a variation of Hu and Liu's method [3] of negation. We first focus on a one-word scope for negation words in an attempt to tease out the effects of accounting for even the simplest forms of negation, as opposed to not accounting for negation at all. We only handle negation words that precede a sentiment word, as larger distances might cause noise in our results due to erroneously negated words. Support for negation is considered in the frequency computations by subtracting the number of negated occurrences of word w or w' in d from the number of non-negated occurrences of w or w' in d.

In the second part of our framework, the score $\text{eval}(d)$ of a document d containing n adjectives $\{w_1, w_2, \ldots, w_n\}$ is simply computed as the sum of the scores of the individual adjectives (the same adjective can appear multiple times), as determined using (1) and (2). In case negation is accounted for, we propose to use a document scoring function:

$$\text{eval}(d) = \sum_{w_i \in d} (-1)^{\text{negated}(w_i,d)} \times \text{score}(w_i), \quad (3)$$

where $\text{negated}(w_i,d)$ is a Boolean indicating whether the ith adjective in w is negated in d (1) or not (0). Using (3), the classification $\text{class}(d)$ of a document d can finally be determined as follows:

$$\text{class}(d) = \begin{cases} 1 & \text{if } \text{eval}(d) > 0.002, \\ 0 & \text{if } -0.021 \leq \text{eval}(d) \leq 0.002, \\ -1 & \text{if } \text{eval}(d) < -0.021, \end{cases} \quad (4)$$

where the thresholds have been optimized through hill-climbing.

4. EVALUATION

We have implemented our framework in C#, combined with a Microsoft SQL Server database. We have used a corpus of 13,628 human-rated Dutch documents on 40 different topics. Sentiment in these documents is classified as positive, negative, or neutral. In order to be able to asses the impact of negation, we have implemented two versions of our framework. The first version has no support for negation, whereas the second version supports negation both in the wordbank creation and in the document scoring part. Our framework only handles adjectives for sentiment analysis and uses the Teezir part-of-speech tagger (based on OpenNLP and trained on Dutch corpora) to identify adjectives in the corpus.

We have used 60% of our documents for training and 40% for testing. The training set was used to create wordbanks and to determine the best threshold level for document classification. Our software first retrieves all adjectives from the training corpus, where multiple occurrences of an adjective are not allowed. The list of adjectives thus extracted is subsequently used for creating a wordbank, by scoring all adjectives in the training set with word scoring function (1). Our software then scores documents in accordance with document scoring functions (3) and (4).

In order to evaluate the human judgements, we took a random sample of 224 documents and rated these for sentiment. We observed 56% strong agreement and 99% weak agreement between our judgement and the human annotations, where strong agreement means an exact match and weak agreement means that one rating is positive or negative, whereas the other is neutral. Most discrepancies between ratings can be explained by interpretation differences. It is for instance difficult for humans to pick up on subtle cases of sentiment, which can be expressed in irony and tone. The interpretation of such subtle uses of sentiment can differ from person to person. The two observed cases of strong disagreement are due to misinterpretation of the text.

We have evaluated the performance of our framework against human ratings in two set-ups: one with support for negation and one without support for negation. Precision improves with 1.17% from 70.41% without taking into account negation to 71.23% when accounting for negation. This improvement is even more evident when our framework is applied to a subset of the corpus, where each document contains negated words (not necessarily adjectives). On this subset of the corpus, precision increases with 2.23% from 69.44% without accounting for negation to 70.98% when taking into account negation. These results are notable given that only 0.85% of the sentences in the original corpus contain negations.

5. CONCLUSIONS AND FUTURE WORK

The main contribution of this paper lies in our reported endeavors of shedding some light onto the impact of accounting for negation in sentiment analysis. Our experiments with a basic sentiment analysis framework show that a relatively straightforward approach to accounting for negation already helps to increase precision with 1.17%. On a subset of selected documents containing negated words, precision increases with 2.23%; a notable result if we consider the fact that negation is sparsely used in our data set.

Nevertheless, it appears to be worthwhile to investigate the effects of optimizing the scope of influence of negation words in order to obtain more detailed insights in the impact of negation in sentiment analysis. We would also like to experiment with other types of words in our wordbank (e.g., adverbs, possibly combined with adjectives). Finally, we plan on taking into account degrees of negation.

6. ACKNOWLEDGEMENTS

We would like to thank Teezir (`http://www.teezir.com`) for their technical support, fruitful discussions, and for supplying us with data for this research.

7. REFERENCES

[1] Cesarano, C., Dorr, B., Picariello, A., Reforgiato, D., Sagoff, A., Subrahmanian, V.: OASYS: An Opinion Analysis System. In: AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs (CAAW 2006). pp. 21–26, AAAI Press (2006)

[2] Heerschop, B., van Iterson, P., Hogenboom, A., Frasincar, F., Kaymak, K.: Analyzing Sentiment in a Large Set of Web Data while Accounting for Negation. In: 7th Atlantic Web Intelligence Conference (AWIC 2011), Springer (Forthcoming 2011)

[3] Hu, M., Liu, B.: Mining and Summarizing Customer Reviews. In: 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004), pp. 168–177. ACM (2004)