A Survey on the Role of Negation in Sentiment Analysis

Michael Wiegand
Saarland University
Saarbrücken, Germany
michael.wiegand@lsv.uni-saarland.de

Alexandra Balahur
University of Alicante
Alicante, Spain
abalahur@dlsi.ua.es

Benjamin Roth and **Dietrich Klakow**
Saarland University
Saarbrücken, Germany
benjamin.roth@lsv.uni-saarland.de
dietrich.klakow@lsv.uni-saarland.de

Andrés Montoyo
University of Alicante
Alicante, Spain
montoyo@dlsi.ua.es

Abstract

This paper presents a survey on the role of *negation* in sentiment analysis. Negation is a very common linguistic construction that affects polarity and, therefore, needs to be taken into consideration in sentiment analysis.
We will present various computational approaches modeling negation in sentiment analysis. We will, in particular, focus on aspects, such as level of representation used for sentiment analysis, negation word detection and scope of negation. We will also discuss limits and challenges of negation modeling on that task.

1 Introduction

Sentiment analysis is the task dealing with the automatic detection and classification of opinions expressed in text written in natural language.
Subjectivity is defined as the linguistic expression of somebody's opinions, sentiments, emotions, evaluations, beliefs and speculations (Wiebe, 1994). Subjectivity is opposed to objectivity, which is the expression of facts. It is important to make the distinction between subjectivity detection and sentiment analysis, as they are two separate tasks in natural language processing. Sentiment analysis can be dependently or independently done from subjectivity detection, although Pang and Lee (2004) state that subjectivity detection performed prior to the sentiment analysis leads to better results in the latter.
Although research in this area has started only recently, the substantial growth in subjective information on the world wide web in the past years has made sentiment analysis a task on which constantly growing efforts have been concentrated. The body of research published on sentiment analysis has shown that the task is difficult, not only due to the syntactic and semantic variability of language, but also because it involves the extraction of indirect or implicit assessments of objects, by means of emotions or attitudes. Being a part of subjective language, the expression of opinions involves the use of nuances and intricate surface realizations. That is why the automatic study of opinions requires fine-grained linguistic analysis techniques and substantial efforts to extract features for machine learning or rule-based systems, in which subtle phenomena as *negation* can be appropriately incorporated.
Sentiment analysis is considered as a subsequent task to subjectivity detection, which should ideally be performed to extract content that is not factual in nature. Subsequently, 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*.
This paper presents a survey on the role of *negation* in sentiment analysis. Negation is a very common linguistic construction that affects polarity and, therefore, needs to be taken into consideration in sentiment analysis. Before we describe the computational approaches that have been devised to account for this phenomenon in sentiment analysis, we will motivate the problem.

2 Motivation

Since subjectivity and sentiment are related to expressions of personal attitudes, the way in which this is realized at the surface level influences the manner in which an opinion is extracted and its polarity is computed. As we have seen, sentiment analysis goes a step beyond subjectivity detection,

including polarity classification. So, in this task, correctly determining the valence of a text span (whether it conveys a positive or negative opinion) is equivalent to the success or failure of the automatic processing.
It is easy to see that Sentence 1 expresses a positive opinion.

1. I *like*$^{+}$ this new Nokia model.

The polarity is conveyed by *like* which is a *polar expression*. Polar expressions, such as *like* or *horrible*, are words containing a prior polarity. The negation of Sentence 1, i.e. Sentence 2, using the negation word *not*, expresses a negative opinion.

2. I do [*not* like^{+}]$^{-}$ this new Nokia model.

In this example, it is straightforward to notice the impact of negation on the polarity of the opinion expressed. However, it is not always that easy to spot positive and negative opinions in text. A negation word can also be used in other expressions without constituting a negation of the proposition expressed as exemplified in Sentence 3.

3. *Not only* is this phone expensive *but* it is *also* heavy and difficult to use.

In this context, *not* does not invert the polarity of the opinion expressed which remains negative.
Moreover, the presence of an actual negation word in a sentence does not mean that all its polar opinions are inverted. In Sentence 4, for example, the negation does not modify the second polar expression *intriguing* since the negation and *intriguing* are in separate clauses.

4. [I do [*not* like^{+}]$^{-}$ the design of new Nokia model] but [it contains some intriguing^{+} new functions].

Therefore, when treating negation, one must be able to correctly determine the scope that it has (i.e. determine what part of the meaning expressed is modified by the presence of the negation).
Finally, the surface realization of a negation is highly variable, depending on various factors, such as the impact the author wants to make on the general text meaning, the context, the textual genre etc. Most of the times, its expression is far from being simple (as in the first two examples), and does not only contain obvious negation words, such as *not*, *neither* or *nor*. Research in the field has shown that there are many other words that invert the polarity of an opinion expressed, such as *diminishers*/*valence shifters* (Sentence 5), *connectives* (Sentence 6), or even *modals* (Sentence 7).

5. I find the functionality of the new phone *less* practical.

6. Perhaps it is a great phone, *but* I fail to see why.

7. In theory, the phone *should* have worked even under water.

As can be seen from these examples, modeling negation is a difficult yet important aspect of sentiment analysis.

3 The Survey

In this survey, we focus on work that has presented novel aspects for negation modeling in sentiment analysis and we describe them chronologically.

3.1 Negation and Bag of Words in Supervised Machine Learning

Several research efforts in polarity classification employ supervised machine-learning algorithms, like Support Vector Machines, Naïve Bayes Classifiers or Maximum Entropy Classifiers. For these algorithms, already a low-level representation using bag of words is fairly effective (Pang et al., 2002). Using a bag-of-words representation, the supervised classifier has to figure out by itself which words in the dataset, or more precisely feature set, are polar and which are not. One either considers all words occurring in a dataset or, as in the case of Pang et al. (2002), one carries out a simple feature selection, such as removing infrequent words. Thus, the standard bag-of-words representation does not contain any explicit knowledge of polar expressions. As a consequence of this simple level of representation, the reversal of the polarity type of polar expressions as it is caused by a negation cannot be explicitly modeled. The usual way to incorporate negation modeling into this representation is to add artificial words: i.e. if a word *x* is preceded by a negation word, then rather than considering this as an occurrence of the feature *x*, a new feature *NOT_x* is created. The scope of negation cannot be properly modeled with this representation either. Pang et al. (2002), for example, consider every word until the next punctuation mark. Sentence 2 would, therefore, result in the following representation:

8. I do not NOT_like NOT_this NOT_new NOT_Nokia NOT_model.

The advantage of this feature design is that a plain occurrence and a negated occurrence of a word are

reflected by two separate features. The disadvantage, however, is that these two contexts treat the same word as two completely different entities.
Since the words to be considered are unrestricted, any word – no matter whether it is an actual polar expression or not – is subjected to this negation modification. This is not only linguistically inaccurate but also increases the feature space with more sparse features (since the majority of words will only be negated once or twice in a corpus).
Considering these shortcomings, it comes to no surprise that the impact of negation modeling on this level of representation is limited. Pang et al. (2002) report only a *negligible* improvement by adding the artificial features compared to plain bag of words in which negation is not considered.
Despite the lack of linguistic plausibility, supervised polarity classifiers using bag of words (in particular, if training and testing are done on the same domain) offer fairly good performance. This is, in particular, the case on coarse-grained classification, such as on document level. The success of these methods can be explained by the fact that larger texts contain redundant information, e.g. it does not matter whether a classifier cannot model a negation if the text to be classified contains twenty polar opinions and only one or two contain a negation. Another advantage of these machine learning approaches on coarse-grained classification is their usage of higher order *n-grams*. Imagine a labeled training set of documents contains frequent bigrams, such as *not appealing* or *less entertaining*. Then a feature set using higher order n-grams implicitly contains negation modeling. This also partially explains the effectiveness of bigrams and trigrams for this task as stated in (Ng et al., 2006).
The dataset used for the experiments in (Pang et al., 2002; Ng et al., 2006) has been established as a popular benchmark dataset for sentiment analysis and is publicly available[1].

3.2 Incorporating Negation in Models that Include Knowledge of Polar Expressions - Early Works

The previous subsection suggested that appropriate negation modeling for sentiment analysis requires the awareness of polar expressions. One way of obtaining such expressions is by using a polarity lexicon which contains a list of polar expressions and for each expression the corresponding polarity type. A simple rule-based polarity classifier derived from this knowledge typically counts the number of positive and negative polar expressions in a text and assigns it the polarity type with the majority of polar expressions. The counts of polar expressions can also be used as features in a supervised classifier. Negation is typically incorporated in those features, e.g. by considering negated polar expressions as unnegated polar expressions with the opposite polarity type.

3.2.1 Contextual Valence Shifters

The first computational model that accounts for negation in a model that includes knowledge of polar expressions is (Polanyi and Zaenen, 2004). The different types of negations are modeled via *contextual valence shifting*. The model assigns scores to polar expressions, i.e. positive scores to positive polar expressions and negative scores to negative polar expressions, respectively. If a polar expression is negated, its polarity score is simply inverted (see Example 1).

$$\text{clever } (+2) \rightarrow \text{not clever } (-2) \tag{1}$$

In a similar fashion, diminishers are taken into consideration. The difference is, however, that the score is only reduced rather than shifted to the other polarity type (see Example 2).

$$\text{efficient } (+2) \rightarrow \text{rather efficient } (+1) \tag{2}$$

Beyond that the model also accounts for modals, presuppositional items and even discourse-based valence shifting. Unfortunately, this model is not implemented and, therefore, one can only speculate about its real effectiveness.

Kennedy and Inkpen (2005) evaluate a negation model which is fairly identical to the one proposed by Polanyi and Zaenen (2004) (as far as simple negation words and diminishers are concerned) in document-level polarity classification. A simple scope for negation is chosen. A polar expression is thought to be negated if the negation word immediately precedes it. In an extension of this work (Kennedy and Inkpen, 2006) a parser is considered for scope computation. Unfortunately, no precise description of how the parse is used for scope modeling is given in that work. Neither is there a comparison of these two scope models measuring their respective impacts.

[1] `http://www.cs.cornell.edu/people/pabo/movie-review-data`

Final results show that modeling negation is important and relevant, even in the case of such simple methods. The consideration of negation words is more important than that of diminishers.

3.2.2 Features for Negation Modeling

Wilson et al. (2005) carry out more advanced negation modeling on expression-level polarity classification. The work uses supervised machine learning where negation modeling is mostly encoded as features using polar expressions. The features for negation modeling are organized in three groups:

- negation features
- shifter features
- polarity modification features

Negation features directly relate to negation expressions negating a polar expression. One feature checks whether a negation expression occurs in a fixed window of four words preceding the polar expression. The other feature accounts for a polar predicate having a negated subject. This frequent long-range relationship is illustrated in Sentence 9.

9. [No politically prudent Israeli]$_{subject}$ could support$_{polar\ pred}$ either of them.

All negation expressions are additionally disambiguated as some negation words do not function as a negation word in certain contexts, e.g. *not to mention* or *not just*.

Shifter features are binary features checking the presence of different types of *polarity shifters*. Polarity shifters, such as *little*, are weaker than ordinary negation expressions. They can be grouped into three categories, general polarity shifters, positive polarity shifters, and negative polarity shifters. General polarity shifters reverse polarity like negations. The latter two types only reverse a particular polarity type, e.g. the positive shifter *abate* only modifies negative polar expressions as in *abate the damage*. Thus, the presence of a positive shifter may indicate positive polarity. The set of words that are denoted by these three features can be approximately equated with diminishers.

Finally, *polarity modification features* describe polar expressions of a particular type modifying or being modified by other polar expressions. Though these features do not explicitly contain negations, language constructions which are similar to negation may be captured. In the phrase [*disappointed*$^{-}$ *hope*$^{+}$]$^{-}$, for instance, a negative polar expression modifies a positive polar expression which results in an overall negative phrase.

Adding these three feature groups to a feature set comprising bag of words and features counting polar expressions results in a significant improvement. In (Wilson et al., 2009), the experiments of Wilson et al. (2005) are extended by a detailed analysis on the individual effectiveness of the three feature groups mentioned above. The results averaged over four different supervised learning algorithms suggest that the actual negation features are most effective whereas the binary polarity shifters have the smallest impact. This is consistent with Kennedy and Inkpen (2005) given the similarity of polarity shifters and diminishers.

Considering the amount of improvement that is achieved by negation modeling, the improvement seems to be larger in (Wilson et al., 2005). There might be two explanations for this. Firstly, the negation modeling in (Wilson et al., 2005) is considerably more complex and, secondly, Wilson et al. (2005) evaluate on a more fine-grained level (i.e. expression level) than Kennedy and Inkpen (2005) (they evaluate on document level). As already pointed out in §3.1, document-level polarity classification contains more redundant information than sentence-level or expression-level polarity classification, therefore complex negation modeling on these levels might be more effective since the correct contextual interpretation of an individual polar expression is far more important[2].

The fine-grained opinion corpus used in (Wilson et al., 2005; Wilson et al., 2009) and all the resources necessary to replicate the features used in these experiments are also publicly available[3].

3.3 Other Approaches

The approaches presented in the previous section (Polanyi and Zaenen, 2004; Kennedy and Inkpen, 2005; Wilson et al., 2005) can be considered as the works pioneering negation modeling in sentiment analysis. We now present some more recent work on that topic. All these approaches, however, are heavily related to these early works.

[2]This should also explain why most subsequent works (see §3.3) have been evaluated on fine-grained levels.

[3]The corpus is available under: `http://www.cs.pitt.edu/mpqa/databaserelease` and the resources for the features are part of OpinionFinder: `http://www.cs.pitt.edu/mpqa/opinionfinderrelease`

3.3.1 Semantic Composition

In (Moilanen and Pulman, 2007), a method to compute the polarity of headlines and complex noun phrases using compositional semantics is presented. The paper argues that the principles of this linguistic modeling paradigm can be successfully applied to determine the subsentential polarity of the sentiment expressed, demonstrating it through its application to contexts involving sentiment propagation, polarity reversal (e.g. through the use of negation following Polanyi and Zaenen (2004) and Kennedy and Inkpen (2005)) or polarity conflict resolution. The goal is achieved through the use of syntactic representations of sentences, on which rules for composition are defined, accounting for negation (incrementally applied to constituents depending on the scope) using negation words, shifters and negative polar expressions. The latter are subdivided into different categories, such that special words are defined, whose negative intensity is strong enough that they have the power to change the polarity of the entire text spans or constituents they are part of.

A similar approach is presented by Shaikh et al. (2007). The main difference to Moilanen and Pulman (2007) lies in the representation format on which the compositional model is applied. While Moilanen and Pulman (2007) use syntactic phrase structure trees, Shaikh et al. (2007) consider a more abstract level of representation being verb frames. The advantage of a more abstract level of representation is that it more accurately represents the meaning of the text it describes. Apart from that, Shaikh et al. (2007) design a model for sentence-level classification rather than for headlines or complex noun phrases.

The approach by Moilanen and Pulman (2007) is not compared against another established classification method whereas the approach by Shaikh et al. (2007) is evaluated against a non-compositional rule-based system which it outperforms.

3.3.2 Shallow Semantic Composition

Choi and Cardie (2008) present a more lightweight approach using compositional semantics towards classifying the polarity of expressions. Their working assumption is that the polarity of a phrase can be computed in two steps:

- the assessment of polarity of the constituents
- the subsequent application of a set of previously-defined inference rules

An example rule, such as:

$$Polarity([\text{NP1}]^{-}\ [\text{IN}]\ [\text{NP2}]^{-}) = + \quad (3)$$

may be applied to expressions, such as $[\textit{lack}]^{-}_{\text{NP1}}\ [\textit{of}]_{\text{IN}}\ \ [\textit{crime}]^{-}_{\text{NP2}}$ *in rural areas*. The advantage of these rules is that they restrict the scope of negation to specific constituents rather than using the scope of the entire target expression.

Such inference rules are very reminiscent of *polarity modification features* (Wilson et al., 2005), as a negative polar expression is modified by positive polar expression. The rules presented by Choi and Cardie (2008) are, however, much more specific, as they define syntactic contexts of the polar expressions. Moreover, from each context a direct polarity for the entire expression can be derived. In (Wilson et al., 2005), this decision is left to the classifier. The rules are also similar to the syntactic rules from Moilanen and Pulman (2007). However, they involve less linguistic processing and are easier to comprehend[4]. The effectiveness of these rules are both evaluated in rule-based methods and a machine learning based method where they are anchored as constraints in the objective function. The results of their evaluation show that the compositional methods outperform methods using simpler scopes for negation, such as considering the scope of the entire target expression. The learning method incorporating the rules also slightly outperforms the (plain) rule-based method.

3.3.3 Scope Modeling

In sentiment analysis, the most prominent work examining the impact of different scope models for negation is (Jia et al., 2009). The scope detection method that is proposed considers:

- static delimiters
- dynamic delimiters
- heuristic rules focused on polar expressions

Static delimiters are unambiguous words, such as *because* or *unless* marking the beginning of another clause. *Dynamic delimiters* are, however,

[4] It is probably due to the latter, that these rules have been successfully re-used in subsequent works, most prominently Klenner et al. (2009).

ambiguous, e.g. *like* and *for*, and require disambiguation rules, using contextual information such as their pertaining part-of-speech tag. These delimiters suitably account for various complex sentence types so that only the clause containing the negation is considered.

The *heuristic rules* focus on cases in which polar expressions in specific syntactic configurations are directly preceded by negation words which results in the polar expression becoming a delimiter itself. Unlike Choi and Cardie (2008), these rules require a proper parse and reflect grammatical relationships between different constituents.

The complexity of the scope model proposed by Jia et al. (2009) is similar to the ones of the compositional models (Moilanen and Pulman, 2007; Shaikh et al., 2007; Choi and Cardie, 2008) where scope modeling is exclusively incorporated in the compositional rules.

Apart from scope modeling, Jia et al. (2009) also employ a complex negation term disambiguation considering not only phrases in which potential negation expressions do not have an actual negating function (as already used in (Wilson et al., 2005)), but also *negative rhetorical questions* and *restricted comparative sentences*.

On sentence-level polarity classification, their scope model is compared with

- a simple negation scope using a fixed window size (similar to the negation feature in (Wilson et al., 2005))
- the text span until the first occurrence of a polar expression following the negation word
- the entire sentence

The proposed method consistently outperforms the simpler methods proving that the incorporation of linguistic insights into negation modeling is meaningful. Even on polarity document retrieval, i.e. a more coarse-grained classification task where contextual disambiguation usually results in a less significant improvement, the proposed method also outperforms the other scopes examined.

There have only been few research efforts in sentiment analysis examining the impact of scope modeling for negation in contrast to other research areas, such as the biomedical domain (Huang and Lowe, 2007; Morante et al., 2008; Morante and Daelemans, 2009). This is presumably due to the fact that only for the biomedical domain, publicly available corpora containing annotation for the scope of negation exist (Szarvas et al., 2008). The usability of those corpora for sentiment analysis has not been tested.

3.4 Negation within Words

So far, negation has only be considered as a phenomenon that affects entire words or phrases. The word expressing a negation and the words or phrases being negated are disjoint. There are, however, cases in which both negation and the negated content which can also be opinionated are part of the same word. In case, these words are lexicalized, such as *flaw-less*, and are consequently to be found a polarity lexicon, this phenomenon does not need to be accounted for in sentiment analysis. However, since this process is (at least theoretically) productive, fairly uncommon words, such as *not-so-nice*, *anti-war* or *offensive-less* which are not necessarily contained in lexical resources, may emerge as a result of this process. Therefore, a polarity classifier should also be able to decompose words and carry out negation modeling within words.

There are only few works addressing this particular aspect (Moilanen and Pulman, 2008; Ku et al., 2009) so it is not clear how much impact this type of negation has on an overall polarity classification and what complexity of morphological analysis is really necessary. We argue, however, that in synthetic languages where negation may regularly be realized as an affix rather than an individual word, such an analysis is much more important.

3.5 Negation in Various Languages

Current research in sentiment analysis mainly focuses on English texts. Since there are significant structural differences among the different languages, some particular methods may only capture the idiosyncratic properties of the English language. This may also affect negation modeling.

The previous section already stated that the need for morphological analyses may differ across the different languages.

Moreover, the complexity of scope modeling may also be language dependent. In English, for example, modeling the scope of a negation as a fixed window size of words following the occurrence of a negation expression already yields a reasonable performance (Kennedy and Inkpen, 2005). However, in other languages, for example German, more complex processing is required as the negated expression may either precede (Sen-

tence 10) or follow (Sentence 11) the negation expression. Syntactic properties of the negated noun phrase (i.e. the fact whether the negated polar expression is a verb or an adjective) determine the particular negation construction.

10. *Peter mag den Kuchen nicht.*
    Peter likes the cake not.
    'Peter does not like the cake.'

11. *Der Kuchen ist nicht köstlich.*
    The cake is not delicious.
    'The cake is not delicious.'

These items show that, clearly, some more extensive cross-lingual examination is required in order to be able to make statements of the general applicability of specific negation models.

3.6 *Bad* and *Not Good* are Not the Same

The standard approach of negation modeling suggests to consider a negated polar expression, such as *not bad*, as an unnegated polar expression with the opposite polarity, such as *good*. Liu and Seneff (2009) claim, however, that this is an oversimplification of language. *Not bad* and *good* may have the same polarity but they differ in their respective polar strength, i.e. *not bad* is less positive than *good*. That is why, Liu and Seneff (2009) suggest a compositional model in which for individual adjectives and adverbs (the latter include negations) a prior rating score encoding their intensity and polarity is estimated from pros and cons of on-line reviews. Moreover, compositional rules for polar phrases, such as *adverb-adjective* or *negation-adverb-adjective* are defined exclusively using the scores of the individual words. Thus, adverbs function like universal quantifiers scaling either up or down the polar strength of the specific polar adjectives they modify. The model independently learns what negations are, i.e. a subset of adverbs having stronger negative scores than other adverbs. In short, the proposed model provides a unifying account for intensifiers (e.g. *very*), diminishers, polarity shifters and negation words. Its advantage is that polarity is treated compositionally and is interpreted as a continuum rather than a binary classification. This approach reflects its meaning in a more suitable manner.

3.7 Using Negations in Lexicon Induction

Many classification approaches illustrated above depend on the knowledge of which natural language expressions are polar. The process of acquiring such lexical resources is called lexicon induction. The observation that negations co-occur with polar expressions has been used for inducing polarity lexicons on Chinese in an unsupervised manner (Zagibalov and Carroll, 2008). One advantage of negation is that though the induction starts with just positive polar seeds, the method also accomplishes to extract negative polar expressions since negated mentions of the positive polar seeds co-occur with negative polar expressions. Moreover, and more importantly, the distribution of the co-occurrence between polar expressions and negations can be exploited for the selection of those seed lexical items. The model presented by Zagibalov and Carroll (2008) relies on the observation that a polar expression can be negated but it occurs more frequently without the negation. The distributional behaviour of an expression, i.e. significantly often co-occurring with a negation word but significantly more often occurring without a negation word makes up a property of a polar expression. The data used for these experiments are publicly available[5].

3.8 Irony – The Big Challenge

Irony is a rhetorical process of intentionally using words or expressions for uttering meaning that is different from the one they have when used literally (Carvalho et al., 2009). Thus, we consider that the use of irony can reflect an implicit negation of what is conveyed through the literal use of the words. Moreover, due to its nature irony is mostly used to express a polar opinion.

Carvalho et al. (2009) confirm the relevance of (verbal) irony for sentiment analysis by an error analysis of their present classifier stating that a large proportion of misclassifications derive from their system's inability to account for irony.

They present predictive features for detecting irony in positive sentences (which are actually meant to have a negative meaning). Their findings are that the use of emoticons or expressions of gestures and the use of quotation marks within a context in which no reported speech is included are a good signal of irony in written text. Although the use of these clues in the defined patterns helps to detect some situations in which irony is present, they do not fully represent the phenomenon.

[5] `http://www.informatics.sussex.ac.uk/users/tz21/coling08.zip`

A data-driven approach for irony detection on product-reviews is presented in (Tsur et al., 2010). In the first stage, a considerably large list of simple surface patterns of ironic expressions are induced from a small set of labeled seed sentences. A pattern is a generalized word sequence in which content words are replaced by a generic *CW* symbol. In the second stage, the seed sentences are used to collect more examples from the web, relying on the assumption that sentences next to ironic ones are also ironic. In addition to these patterns, some punctuation-based features are derived from the labeled sentences. The acquired patterns are used as features along the punctuation-based features within a k nearest neighbour classifier. On an in-domain test set the classifier achieves a reasonable performance. Unfortunately, these experiments only elicit few additional insights into the general nature of irony. As there is no cross-domain evaluation of the system, it is unclear in how far this approach generalizes to other domains.

4 Limits of Negation Modeling in Sentiment Analysis

So far, this paper has not only outlined the importance of negation modeling in sentiment analysis but it has also shown different ways to account for this linguistic phenomenon. In this section, we present the limits of negation modeling in sentiment analysis.

Earlier in this paper, we stated that negation modeling depends on the knowledge of polar expressions. However, the recognition of genuine polar expressions is still fairly brittle. Many polar expressions, such as *disease* are ambiguous, i.e. they have a polar meaning in one context (Sentence 12) but do not have one in another (Sentence 13).

12. He is a *disease* to every team he has gone to.

13. Early symptoms of the *disease* are headaches, fevers, cold chills and body pain.

In a pilot study (Akkaya et al., 2009), it has already been shown that applying *subjectivity word sense disambiguation* in addition to the feature-based negation modeling approach of Wilson et al. (2005) results in an improvement of performance in polarity classification.

Another problem is that some polar opinions are not lexicalized. Sentence 14 is a negative *pragmatic opinion* (Somasundaran and Wiebe, 2009) which can only be detected with the help of external world knowledge.

14. The next time I hear this song on the radio, I'll throw my radio out of the window.

Moreover, the effectiveness of specific negation models can only be proven with the help of corpora containing those constructions or the type of language behaviour that is reflected in the models to be evaluated. This presumably explains why rare constructions, such as negations using connectives (Sentence 6 in §2), modals (Sentence 7 in §2) or other phenomena presented in the conceptual model of Polanyi and Zaenen (2004), have not yet been dealt with.

5 Conclusion

In this paper, we have presented a survey on the role of negation in sentiment analysis. The plethora of work presented on the topic proves that this common linguistic construction is highly relevant for sentiment analysis.

An effective negation model for sentiment analysis usually requires the knowledge of polar expressions. Negation is not only conveyed by common negation words but also other lexical units, such as diminishers. Negation expressions are ambiguous, i.e. in some contexts do not function as a negation and, therefore, need to be disambiguated. A negation does not negate every word in a sentence, therefore, using syntactic knowledge to model the scope of negation expressions is useful.

Despite the existence of several approaches to negation modeling for sentiment analysis, in order to make general statements about the effectiveness of specific methods systematic comparative analyses examining the impact of different negation models (varying in complexity) with regard to classification type, text granularity, target domain, language etc. still need to be carried out.

Finally, negation modeling is only one aspect that needs to be taken into consideration in sentiment analysis. In order to fully master this task, other aspects, such as a more reliable identification of genuine polar expressions in specific contexts, are at least as important as negation modeling.

Acknowledgements

Michael Wiegand was funded by the BMBF project NL-Search under contract number 01IS08020B. Alexandra Balahur was funded by Ministerio de Ciencia e Innovación - Spanish Government (grant no. TIN2009-13391-C04-01), and Conselleria d'Educación-Generalitat Valenciana (grant no. PROMETEO/2009/119 and ACOMP/2010/286).

References

C. Akkaya, J. Wiebe, and R. Mihalcea. 2009. Subjectivity Word Sense Disambiguation. In *Proceedings of EMNLP*.

P. Carvalho, L. Sarmento, M. J. Silva, and E. de Oliveira. 2009. Clues for Detecting Irony in User-Generated Contents: Oh...!! It's "so easy" ;-). In *Proceedings of CIKM-Workshop TSA*.

Y. Choi and C. Cardie. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of EMNLP*.

Y. Huang and H. J. Lowe. 2007. A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports. *JAMIA*, 14.

L. Jia, C. Yu, and W. Meng. 2009. The Effect of Negation on Sentiment Analysis and Retrieval Effectiveness. In *Proceedings of CIKM*.

A. Kennedy and D. Inkpen. 2005. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. In *Proceedings of FINEXIN*.

A. Kennedy and D. Inkpen. 2006. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. *Computational Intelligence*, 22.

M. Klenner, S. Petrakis, and A. Fahrni. 2009. Robust Compositional Polarity Classification. In *Proceedings of RANLP*.

L. Ku, T. Huang, and H. Chen. 2009. Using Morphological and Syntactic Structures for Chinese Opinion Analysis. In *Proceedings ACL/IJCNLP*.

J. Liu and S. Seneff. 2009. Review Sentiment Scoring via a Parse-and-Paraphrase Paradigm. In *Proceedings of EMNLP*.

K. Moilanen and S. Pulman. 2007. Sentiment Construction. In *Proceedings of RANLP*.

K. Moilanen and S. Pulman. 2008. The Good, the Bad, and the Unknown. In *Proceedings of ACL/HLT*.

R. Morante and W. Daelemans. 2009. A Metalearning Approach to Processing the Scope of Negation. In *Proceedings of CoNLL*.

R. Morante, A. Liekens, and W. Daelemans. 2008. Learning the Scope of Negation in Biomedical Texts. In *Proceedings of EMNLP*.

V. Ng, S. Dasgupta, and S. M. Niaz Arifin. 2006. Examining the Role of Linguistic Knowledge Sources in the Automatic Identification and Classification of Reviews. In *Proceedings of COLING/ACL*.

B. Pang and L. Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. In *Proceedings of ACL*.

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proceedings of EMNLP*.

L. Polanyi and A. Zaenen. 2004. Context Valence Shifters. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*.

M. A. M. Shaikh, H. Prendinger, and M. Ishizuka. 2007. Assessing Sentiment of Text by Semantic Dependency and Contextual Valence Analysis. In *Proceedings of ACII*.

S. Somasundaran and J. Wiebe. 2009. Recognizing Stances in Online Debates. In *Proceedings of ACL/IJCNLP*.

G. Szarvas, V. Vincze, R. Farkas, and J. Csirik. 2008. The BioScope Corpus: Annotation for Negation, Uncertainty and Their Scope in Biomedical Texts. In *Proceedings of BioNLP*.

O. Tsur, D. Davidov, and A. Rappoport. 2010. ICWSM - A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews. In *Proceeding of ICWSM*.

J. Wiebe. 1994. Tracking Point of View in Narrative. *Computational Linguistics*, 20.

T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. In *Proceedings of HLT/EMNLP*.

T. Wilson, J. Wiebe, and P. Hoffmann. 2009. Recognizing Contextual Polarity: An Exploration for Phrase-level Analysis. *Computational Linguistics*, 35:3.

T. Zagibalov and J. Carroll. 2008. Automatic Seed Word Selection for Unsupervised Sentiment Classification of Chinese Text. In *Proceedings of COLING*.