Handling Negation for Sentiment Analysis: A Case Study Using Dependency Parse Tree on Amazon Reviews of Kindle

Smita Suresh Daniel[1], Ani Thomas[2], Neelam Sahu[3]
[1,3]Dept. of Computer Application and IT, Dr. C.V. Raman University, Bilaspur (C.G.) India.
[2]Dept. of Computer Applications, Bhilai Institute of Technology, Durg (C.G.) India.

ABSTRACT
The process of sentiment analysis is a task of detecting, extracting and classifying sentiments expressed in texts. It includes the understanding of the meaning of words within the text through natural language processing rules using dependency based parse trees, using grammatical relations among words to model a sentence, and hence to determine words that are affected by negation. This paper presents a framework for identifying. Calculating and representing the presence of negation in textual data using dependency parsers. It includes a list of rules for negative polarity identification and calculation. These negation rules are designed to improve sentiment analysis. This paper is a demonstration of an approach for identifying the scope of negation in a review and its calculation for the Amazon product- Kindle dataset.

Keywords: Negation Identification, Negation Calculation, Sentiment Analysis, Dependency Parsing

I INTRODUCTION

The aim of sentiment analysis is to find out the positive and negative feelings written in a text, but it contain negations that are very frequently used in text that completely change the polarity of words. Negation identification and detecting its scope within a sentence (text) are necessary in finding out the sentiments from a piece of text. Proper addressing of negation identification is an important aspect of sentiment analysis.. Negation identification is a difficult task and its complexity increases, since negation words such as not, nor etc., (syntactic negation) are not the only criterion for negation calculation. The linguistic patterns - prefixes (e.g.,n-, dis-, etc.) or suffixes (e.g., -less) also introduce the context of negation in textual data . Similarly, word intensifiers and diminishers (contextual valence shifter) also change the polarity of sentiments. These valence shifters do not only flip the polarity but also increase or decrease the degree to which a sentimental term is positive or negative [2].

On the other hand, negation does not mean to handle only 'not'. There are words and clauses like; no, not, n't, no way, without, nowhere, never, no longer, by no means, no more, by no means, at no time, etc. [2] which also inverts the meaning of a sentence.. This paper is an effort towards finding a method to handle the syntactic negation for sentiment analysis by not only using the sentiment and its intensity for words but also using the dependencies of these words and their relation within the sentences and sentence structure.

The negation in a text is assessed with the help of diminishes,, intensifiers and negation terms during the process of sentiment analysis.

The paper is structured as follows: Section II presents the related work in the area of sentiment analysis. Section III. Describes the proposed framework for sentiment analysis, and the existing resources used to generate dependencies. Section IV presents an application for negation handling in sentiment analysis. It also explains the basic techniques used in this framework for handling negation. Section V involves an analysis of the technique.

II RELATED WORK

Most researchers in the field of opinion mining have used the lexicons and lists of words, with word as basic unit of expression of emotions in any language. Lexicon based negation i.e., negation introduced by suffix and/or prefix is easily handled with the help of a good lexical resource, i.e., dictionary, ontology, database etc.

However, more emphasis on opinion analysis should be on how these words are joined and correlated with other words to give specific meanings in any language. This interrelationship of words makes up sentences, which is why it is important to emphasis on finding the scope of negation, diminshers or intensifiers. Syntactic and semantic differences make it difficult to interpret the intensity of polarity. When calculating the value of intensity of any sentence, there are always modifiers, which not only change the polarity of other words in the sentence but also affect its intensity. It is also difficult to identify which part of a clause a negation is changing in a sentence. The different methods used for negation identification and how they affect sentiment analysis of text are discussed below.

(a) Bag of Words- Bag of words (BOW) is a technique where each word in a document is represented by a separate variable numeric value or weight. It is the most widely used technique for sentiment analysis where negation in a sentence reverses the meanings of the sentence. Words like "not", "never", "no", etc., serve to reverse sentence meaning. Limitation of this method is that it is based on the list of words, and lists in any language can are very vast.

(b) Contextual Valence Shifter-Contextual Valence Shifters or modifiers are words, which changes, boosts, enhances and diminishes the meaning. Many researchers have shifted their research on sentiment analysis from BOW to Parts of Speech (POS) especially Verbs, Adjectives and Adverbs. Polarity is associated with every word. However, lots of modifiers are still needed to change or modify the valance associated with words. Negatives, intensifiers or diminishers are examples of contextual shifter. For example :

Negatives: The battery is good versus the battery is not good.

Intensifier: The charger is working well versus the charger is working very well.
Diminishers: It starts versus it hardly starts.

There is a need for relationship finder to define the scope of negation terms .Researchers have tried to define the scope by defining lists of verbs, adjectives and adverbs and defining their relationships for sentiment analysis [6]. Lists of positive and negative terms and a set of lists for modifiers define the scope of these modifiers as n- terms before and after positive or negative terms, although this n remained a constant.

(c) Semantic Relations- Semantic relations refer to the relationship between concepts or meanings for example antonym, synonym, homonym etc. It is evident from existing research that semantic relationship is also used for negation identification. It is clear that atomic words, which can provide a misleading polarity for sentences as words can be modified (weakened, strengthened, or reversed) based on lexicons. The use of linguistic structure of sentence for sentiment analysis was proposed in [9], where the polarity of a sentence is dependent upon the polarities of its parts: noun phrases (NP), verb phrases (VP) and individual parts of speech. Negation is handled by defining different intensities of negation words. In other words, the negation of words can change the polarity of an entire sentence or only parts of it [7].

Dependency Analysis on the semantic verb frames of each sentence, and apply a set of rules to each dependency relation to calculate the contextual valence of the whole sentence. A two phase process was proposed in [3] as another way of compositional semantics. In the first phrase they identify the polarity of words where all the words are classified on the basis of the level of their strength in terms of the scope in the sentence and in the second phase inference rules are used, which identify the polarity modification feature.

(d) Relations and Dependency Based- The grammatical relationships between the words within a sentence and syntactic dependencies help in extraction of textual relations. For context aware approach for sentiment analysis where the sentiment is evaluated towards a target entity or an event. The scope of words is defined by the clauses or phrases (noun phrase, verb phrase) in the sentence. The sentiment of sentences is understood by the heuristic rules defined to join the clauses. Simple tree based rules can identifying the dependent terms and later use some parts of speech based tools to understand the sentimental behavior of negation [2].

(e) Analysis of Negation- For the sentence level sentiment analysis clauses and phrases are required to be understood. They are further divided into sentences and into different types of sentences (simple, complex and compound). The sentence is made more complicated by adding declarative, interrogative, exclamatory and imperative sentences. In order to further complicate the problem as the comparison, contradiction, negation and irony and sarcasm might also be introduced in the sentences. Negation needs to identify its scope. Negation can be local (e.g., not good), or it can involve far distance dependencies (e.g., does not look very good) or the negation of the subject (e.g., no one thinks that it's good). It can also change its roles i.e., instead of negating and it can also intensify (e.g., not only good but amazing). In order to find out the scope of the negation, the sequence of words in the sentence should also be identified. On the whole, it is not only the negation of a word but the negation of the sentence.

The expression of negation within a sentence can be a verb, adverb, suffix or prefix. It might also occur more than once in a sentence and rather than cancelling each other it can give negative meaning, the following section explains the proposed framework.

III FRAMEWORK FOR SENTIMENT ANALYSIS

This section introduces a framework for sentiment analysis and explains how it is handling negation identification, scope of negation and calculation of sentiment on sentence level. The framework uses a combined approach to lexical and syntactic resources for sentiment analysis.Our system use the dependency parsing by the Stanford Parser, described in[9].The output consists in a sentiment score in [1, 0 , -1] for positive, neutral and negative reviews. In order to enhance the accuracy of sentiment evaluation, the text analysis process includes several pre- processing phases like tokenization, Parts of Speech tagging, lemmatization and reduction.

Its main components are briefly described in Sections (a) through (c).

The framework is presented in Figure 1 as shown below.

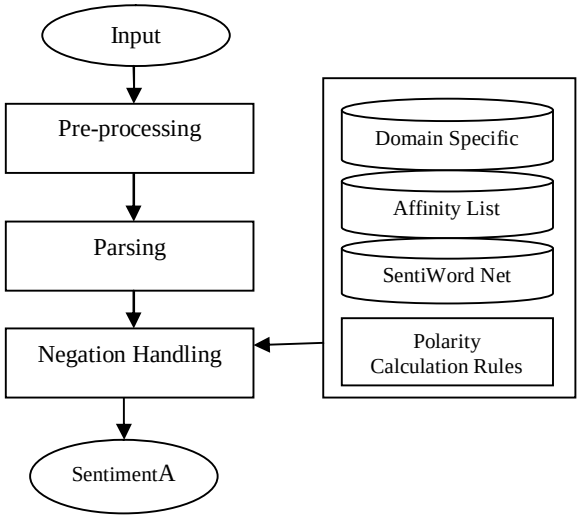

Fig.1 Framework for Sentiment Analysis

Its main components are briefly described in Sections (a) through (c).

(a) Pre-processor- The pre-processing phase of the system takes text as input and arranges all the data in required format. It splits data into sentences. Filter all sentences not containing negation triggers and extracts sentences containing negation triggers into a file and provides them to the syntactic parser.

(b) Syntactic Parser- The Parts of Speech tagger (POS) tags each word in the sentence. The name entities and phrases involved in a sentence are also identified in syntactic parsing. The Stanford dependency Parser identifies how different words are interacting within a sentence and identifies the syntactic dependencies/relationship within a sentence. The syntactic parser parses each sentence iteratively with all the identified information classifying the sentence as a question, an assertion or a comparison, using the rule of sentence type identification. It is parsed and stored in a file in CoNLLx format which is sent to the negation algorithm.

(c) Negation Algorithm- Here we exploit the dependency based parse tree to define a Negation algorithm. An approach used here is not only to invert the polarity of the term followed by the negation word but when a negation occurs in a sentence it is necessary to detect its scope that is the number of following words affected by it, which states that there is no fixed negation window. It depends on the structure of the sentence .The algorithm uses Stanford dependency parser to analyse the grammatical structure of the sentence. A dependency based parse tree is built for each sentence in the review having negation triggers .It explores it to find negation word using Depth First Search rule. If a negation word is found in a tree node, the algorithm inverts the polarity of the following child node and its sub trees. The sentiment score of the negation word is set to 0 as it does not have any sentiment values by itself.

The Negation Algorithm

Let S be an extracted Sentence from a split review having negation triggers.
Let T be the parse tree of Converted to CoNLL format.
Get_Neg_Scope(T) :Set rules to determine the scope of negation using maximum spanning tree and return range for the scope in a sentence.

If POS of trigger is RB/ DT/ JJ/ CC then tree from its immediate governor is used;
If POS of trigger is VB/IN/NN then tree from itself is used;
Span towards right or span left relatively to the position of trigger word till SUB arc, span nothing if there's no SUB arc or right part. (This rule is only applied to root node)

Let Polarity Score () calculate the sentiment score of each term within the scope of negation.

Invert the sentiment score if negation is true

IV USAGE OF FRAMEWORK FOR NEGATION

All the sentences having negation terms and clauses are forwarded from the pre-processor to the syntactic parser with other sentences. However, syntactic parser identifies the negation and POS that are involved in negation with the help of Parser during the syntactic parsing phase and takes care of the negation word and the scope associated with it. . In the negation identification process, the kind of negation i.e.," no one is interested in this new feature", where 'no' is used to determine the interest of one, is also identified. This process also takes care of the negation in conjunction sentences. The Dependency parsers identify the scope of the negation, if it is a single word or a phrase or a clause within a sentence.

These pharses and clause which is within this scope is extracted for polarity calculation by the sentiment analyzer. The polarity is inverted as per table no 1 . The following section explains how sentiment analyzer uses the extracted part of the sentence from the negation algorithm for polarity calculation.

(a) Sentiment Analyser- The sentiment analyser uses resources like Sentiwordnet [10] to extract the sentiment oriented words by using dependency relationships within the text. It identifies the semantics involved in a sentence, their meaning, polarity of each term and their intensity. It identifies the positive, negative and neutral words and assigns a value to each term and hence their intensities can also be calculated.

(b) Polarity Calculator- The words in a sentence, their meanings, alternative words, polarity of each word and intensity associated with each word are basic elements used by sentiment analyzer for sentiment identification. The polarity of sentence is usually based on the meaning of words. However, the negation changes the meaning of the words and polarity of the sentence. In order to calculate the polarity of a sentence, some rules are defined in Table 1.. Most negation words are classified as adverbs, suffix, prefix or verbs.

The scope of negation will be identified by the dependency tree, which indicates how negation is interacting with other words in the sentence.

This dependency parser will identify the scope of the negation - whether it is a single word or a phrase / clause within a sentence. The negation is handled in each phrase accordingly. The intensity of polarity will not exceed (+/-) 1, where + is for positive and – is for negative polarity. The polarity of a sentence is calculated using sentiwordnet.:

The Resulting Intensity = Total positives –total negatives

The positive/negative value of words in the Equation 1 is extracted from the SentiWordNet in order to calculate the polarity of a sentence. The extracted value from the SentiWordNet is reversed during this process if negation is 'True' as presented in Table 1.

Algorithm to Calculate Polarity

```
Function Calculate_Polarity  Returns Polarity
{
 polarity = 0
For Each  Extraction_Of_Sentence
{
get SentiWordNet value of all Adjectives and adverbs in the phrase
If (Sentence is Marked NEGATION by Syntax Parser) {
Reverse the SentiWordNet values of related
Adverbs /Adjectives }
                } }
Return polarity
}
```

Table 1
Rules specifying negation

Negation Word/phrase/ Clause	Associated Word/Phrase /Clause	Negation	Result
Negative	Positive	True	Positive
Negative	Positive	False	Negative
Negative	Negative	True	Negative
Negative	Negative	False	Positive

V ANALYSIS

The sentence polarity is calculated on the basis of the parts of a sentence. A sentence may contain either simple POS (Verb, Adverb, Adjectives, etc.) or complex parts of speech (Noun Phrase [Pronoun, Noun] or Verb Phrase [Verb, Noun Phrase], relations of possession, determiner, etc.). The following hierarchy is an example of POS in a complete sentence.

```
(Sentence
(Noun Phrase (Pronoun, Noun))
(Adverbial Phrase (Adverb))
(Verb Phrase (Verb)
 (Sentence
  (Verb Phrase (Verb)
    (Noun Phrase (Noun))) ) ) )
```

Sentiment polarity identification and calculation is a nested process. This process calculates the sentiment of the most inner most level first and then it calculates along with the next higher level, If there is a negation word the polarity will be calculated accordingly. If a negation word is found in a tree node, the algorithm inverts the polarity of its sub trees as they belong to the same clause.

For Example "Charger is never successful at charging." .

The dependency tree structure is as follows.

```
        (Sentence
          (Noun Phrase (charger))
```

(Verb Phrase (is)
(Adverbial Phrase (never))
(Adjectival Phrase (successful)
(Prepositional Phrase (at)
(Noun Phrase (charging))))))

Stanford Parser output:-

(ROOT
(S
(NP (DT
The) (NN
charger))
(VP (VBZ is)
(ADVP (RB never))
(ADJP (JJ successful)
(PP (IN at)
(S
(VP (VBG charging))))))
(. .)))

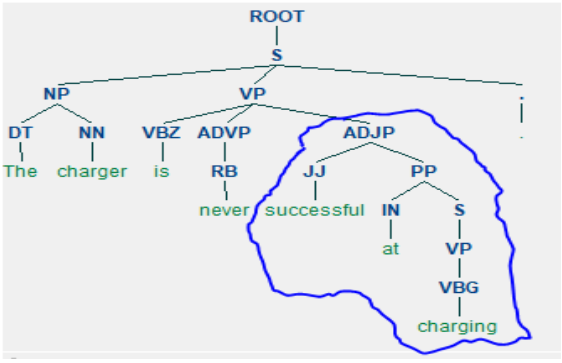

Fig. 2 Dependency Tree of the sentence

The resultant sentence structure is as shown below:

The charger is <SCOPE><NEG>never</NEG> successful at charging .</SCOPE>

Negation'never' is for the scope ('successful at charging') which is positive. This negation of positive phrase is a simple negation, which is presented in Figure no 2.

The confusion matrix for our experiment is shown in Table no 2. There is a 6% increase in accuracy after implementation of the above negation algorithm for our model for predicting sentiments for amazon reviews for the product kindle.

Table 2
Confusion Matrix for the Dataset used

	Predicted Negative	Predicted Neutral	Predicted Positive
Actual Negative	297	0	62
Actual Neutral	0	65	0
Actual Positive	27	0	349

We can calculate the Error estimation for n as Mean Squared Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - t_i)^2 = 0.08$$

where N is the number of review ie 800, x_i is the estimated sentiment score of the review i and t_i is the real score derived from human evaluation .

VI CONCLUSION

Negation has received little attention and its implication on the semantic understanding of sentences. This paper presents an approach for negation identification and calculation using a developed framework for sentiment analysis. These negation rules are designed in order to improve the sentiment text analysis. While, there are still a number of challenges to be addressed in the field of negation in sentiment analysis, the developed rules for negation calculation has good improvements in classification accuracy and helps to find the correct polarity.

REFERENCES

[1] Horn, L. R. and Kato, Y. (2000) Introduction: Negation and Polarity. at the Millennium.

[2] Jia , L., Yu, C. and Meng, W. (2009) The effect of negation on sentiment analysis and retrieval effectiveness. In: The 18th ACM conference on Information and knowledge management. Hong Kong, China. ACM, 1827-1830

[3] Choi, Y. and Cardie, C. (2008) Learning with compositional semantics as structural inference for subsentential sentiment analysis. In: The conference on Empirical Methods in Natural Language Processing. Honolulu, Hawaii. Association for Computational Linguistics, pp. 793-801.

[4] Kennedy, A. and Inkpen, D. (2005) Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. In: FINEXIN 2005. Ottawa.

[5] Kennedy, A. and Inkpen, D. (2006) Sentiment Classification of Movie Reviews Using Contextual Valence Shifters Intelligence. , Vol. 22 2, pp. 110-125

[6] Subrahmanian, V. S. and Reforgiato, D. (2008) AVA: Adjective- Verb-Adverb Combinations for sentiment Analysis. IEEE Intelligent Systems, Vol. 23, 4, pp. 43-50.

[7] Wiegand, M., et al. (2010) A survey on the role of negation in sentiment analysis. In: The Workshop on Negation and Speculation In Natural Language Processing. Uppsala, Sweden. Association for Computational Linguistics, 60-68

[8] Amna Asmi, Tanko Ishaya , IMMM 2012 . The second International Conference on Advances in Information Mining and Management .

[9] Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430

[10] Miller , G.A(1995) WordNet, a lexical database for English Communications of ACM , Vol. 38,11, pp. 39-41