Effect of Negation in Sentiment Analysis

Wareesa Sharif, Noor Azah Samsudin, Mustafa Mat Deris, Rashid Naseem, Muhammad Faheem Mushtaq
Faculty of Computer Science and Information Technology
University Tun Hussein Onn Malaysia (UTHM), 86400, Johor, Malaysia
wareesa786@gmail.com
{azah, mmustafa}@uthm.edu.my
{rnsqau, faheem.mushtaq88}@gmail.com

ABSTRACT: *Sentiment analysis is the process to study of people opinion, emotion and way of considering a matter and take the decision into different categorize like positive, negative and neutral in data mining. The sentiment analysis is used to find out negation within the text using Natural Language Processing rules. Our aim is to detect negation affect on consumer reviews which look like positive but exactly negative in sense. A number of different approaches have been used, but these approaches do not provide an efficient and appropriate way of calculating negation sense in sentiment analysis. The proposed modified negation approach presents a way of calculating negation identification and is helpful to improve classification accuracy. The main achievement of this approach is that it is helpful for calculating the negation in sentiment analysis without the words not, no, n't, never etc. This method produced a significant result for review classification by accuracy, precision, and recall.*

Keywords: Sentiment Analysis, Review mining, Data mining, Negation Identification

Received: 3 January 2017, Revised 22 February 2017, Accepted 27 February 2017

1. Introduction

Billions of web users are attracted to use web 2.0 as a medium of storing, distributing, broadcasting and retrieving data. Sentiment analysis is a process that computationally identifies and categorizes people attitudes, emotions, evaluation, and opinion in Natural Language Processing (NLP). The most useful textual data for storing and communication of information in organization exist on the web in the textual format [1, 2]. Text documents are considered as "bag of words" [3]. In WordNet, synonyms and synset of the word do exist. These words can be differentiated as part-of-speech (POS). Synset defines the

unique concept of natural language description and also describe the relation of synsets (antonymy, hyponymy, and meronymy) [4]. Each WordNet synset í assigned scores with the range of [1, 0] [5].

This analysis is important to find a critical idea of a sentence in a line or word. It is also helpful in social media to check people's opinions. The importance of sentiment analysis is increased due to a huge volume of opinionated data on the Web which is a quick source to collect people's opinions about anything, and sentiment analysis helps to determine whether a sentence is positive or negative with the help of the NLP technique. Most of the past studies [4, 6, 7] worked on sentiment analysis. They found the negation rules with the help of "not, no, none", but without these words a sentence can be negative naturally. Similarly, [8] conducted a review study in the context of opinion mining and argued that some of the words look like positive but naturally they are negative in their meanings. For Example, a person is too fat or the person is slim, both sentences are negative by nature. In contrast to that, if we use the mobile is slim then it will be positive by nature. The existing system cannot differentiate the sentence as positive or negative by nature. Therefore, our aim is to detect negation affect on consumer reviews which look like positive but exactly negative in sense.

1.1 Overview of Paper

The rest of paper is organized as follows. Section 2 defines literature review which includes the overview of negation that is used in our proposed framework. Section 3 explains the research methodology and results. Section 4 presents the conclusion of this research.

2. Literature Review

With the increase of information on various social media, forums, blogs etc., automatic sentiment analysis become necessary to handle this information [8]. Sentiment analysis is a technique that analyzes people's attitude, opinion, evaluation, emotion on entities such organization, events, product, services etc [9]. Several past studies [10, 11, 12, 13] has been done in the area of sentiments and opinions mining. Before the year of 2000, it was not an effective area and little research had been done [14]. Now it becomes an active research area due to ever increasing consumers' tendencies to post their post-consumption experience with products in the form of consumer reviews.

In linguistics, Negation is the process of turning an assertion that can change the text polarity [6, 15]. Therefore, negation has been taken into account [16]. The main scope of the negation expression is to determine the sequence of the words in a sentence affected by negation words, such as never, not, no [17]. Negation words affect the sentence polarity, if a sentence has negation words. Its mean that the sentence polarity will be inverted or vice versa. That's why the scope of each sentence has to calculate. To identify the negation, different linguistic rules have been applied to identify the negation scope [16].

Negation can increase sentence complexity and it is difficult to find. Negation words such as nor, not, and no etc are used for negation identification. The pattern suffixes (e.g –less) or prefix (e.g dis-, un- etc) are also introducing the negation context [18]. The number of researchers [8, 19, 7, 17] have worked on negation. They used no, n't, not, never, no more, no longer, nowhere, no way, etc., to identify negation, and the precision of negative word "not" is low at 63% [16]. Whereas, the sentiments analysis applies on three levels: document level, sentence level, and aspect level.

3. Research Methodology

In this section, architecture of the proposed system is described. As shown in Figure 1, there are five main components of the framework: (1) Pre-processing, (2) Syntactic Parser, (3) Dependency Table, (4) Rules for Negation (4) Classification.

3.1 Sentence Dependency

The dependency between words in the sentences helps to extract the relationship of textual data. Decision tree of review is presented in figure 2.

3.2 Analysis

The sentence level polarity calculated on the basis of POS (part of speech). A sentence may have verb, adjective, adverb, noun etc. Hierarchy of sentence is follows:

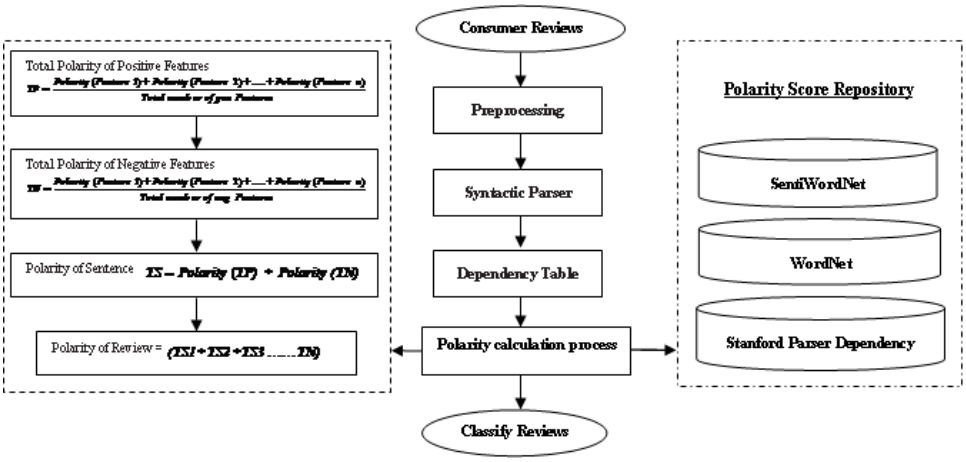


Figure 1. Framework for proposed Negation technique

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

A nested process is implicated in sentiment polarity calculation. It first computes the most inner level and then calculates the next higher level, it is also called sentiment propagation [20]. This nested process calculates sentiment polarity and intensity of the sentences.

3.3 Polarity Calculation

Sentiment analysis involves finding the semantics of sentences. The words have meaning, alternative words, polarity, intensity associated with words. Polarity score depends on the meaning of words. Negation affects the sentence meaning and changes the polarity of the whole sentence. Some sentences do not have words like not, never, no etc. but they have a negative sense. For calculating the negation of these sentences, rules are defined in the following sections.

3.3.1 Assigning Polarity

The positive and negative words are calculated for this equation,

$$\text{Polarity of words} = ((+/-)w_i) \quad \text{Where } 0 < i < n \tag{1}$$

The operator "+" indicate for positive words and "–" indicate the negative words, and the value of i range from '0' to 'n' words. The Stanford parser dependency helps to identify the negation scope. The negation is calculated as follows:

All noun subjects of all sentences of a paragraph are used to find the nature of the sentence. Matches nsubj(w1,w2) in all the sentences and find the polarity score of each matches words and then apply following formula on each word:

$$\text{Negation intensity} = -(w1+w2) \tag{2}$$

w1 indicate the first word and w2 indicate the second word in noun subject dependency. The same procedure will be used for all sentences and finally, the total result is multiplied of sentences score.

3.3.2 Evaluating Polarity

The final polarity score is calculated after evaluating the total sentence. The result of equation 1 and equation 2 are used for evaluating the total sentence polarity score. Both equation results give the final result. Following example illustrates how to calculate polarity score.

Example

"*I purchased a HP pro-book 4510 seven months ago. The video quality is good. However, my Brother thinks laptop is too heavy for him*".

After reviewing above paragraph, the polarity of opinion is calculated. The nature of opinion is mentioned as follows: (1) positive opinion expressed about video quality, (2) positive opinion about battery life, and (3) negative opinion about the weight of HP laptop.

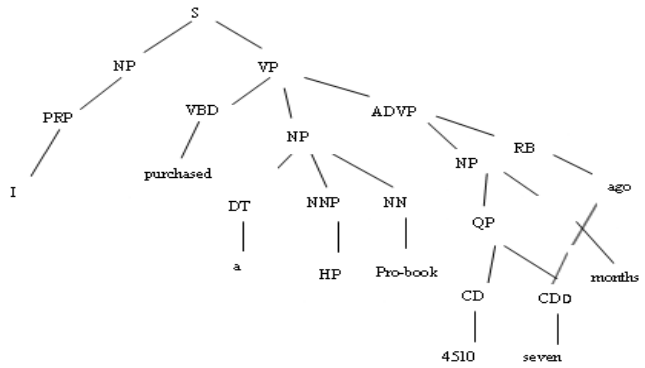
From this review, the following observations are considered both negative and positive about HP laptop. Sentences express a positive opinion about a laptop as a whole but Sentence 3 expresses the negative opinion about the weight of HP laptop. From this analysis, we can make the final observation.

Tokenization

```
(ROOT
 (S
  (NP (PRP I))
  (VP (VBD purchased)
   (NP (DT a) (JJ pro-book) (CD 4510))
   (ADVP
    (NP (CD seven) (NNS months))
    (RB ago)))
  (. .)))
(ROOT
 (S
  (NP (DT The) (JJ video) (NN quality))
  (VP (VBZ is)
   (ADJP (JJ good)))
  (. .)))
(ROOT
 (S
  (ADVP (RB However))
  (, ,)
  (NP (PRP$ my) (NN Brother))
  (VP (VBZ thinks)
   (SBAR
    (S
     (NP (NN laptop))
     (VP (VBZ is)
      (ADJP (RB too) (JJ heavy)
       (PP (IN for)
        (NP (PRP him))))))))
  (. .)))
```

The sentence tree that describes the structure of the sentence and POS hierarchy is depicted in Figure 2.

The sentence is "I purchased a HP Pro-book 4510 seven month ago".



Note: I/PRP purchased/VBD a/DT pro-book/ JJ 4510/CD seven/CD months/NNS ago/ RB./. The/DT video/JJ quality/NN is/VBZ good/JJ ./. However/RB ,/, my/ RP$ Brother/NN thinks/VBZ laptop/NN is/VBZ too/RB heavy/JJ for/IN him/PRP./.

Figure 2. Dependency Structure Tree

3.4 Universal Dependency

Universal dependency is collected using Stanford Parser for identifying the grammatical relationship in the sentences. We extract dependency relation of each sentence which does not provide by direct dependency tree.

nsubj(purchased-2, I-1)
root(ROOT-0, purchased-2)

det(4510-5, a-3)
amod(4510-5, pro-book-4)
dobj(purchased-2, 4510-5)
nummod(months-7, seven-6)
nmod:npmod(ago-8, months-7)
advmod(purchased-2, ago-8)

det(quality-3, The-1)
amod(quality-3, video-2)
nsubj(good-5, quality-3)
cop(good-5, is-4)
root(ROOT-0, good-5)

advmod(thinks-5, However-1)
nmod:poss(Brother-4, my-3)
nsubj(thinks-5, Brother-4)
root(ROOT-0, thinks-5)
nsubj(heavy-9, laptop-6)
cop(heavy-9, is-7)
advmod(heavy-9, too-8)
ccomp(thinks-5, heavy-9)
case(him-11, for-10)
nmod(heavy-9, him-11)

We work on nsubj() dependency, match each nsubj() with the given table values, and subsequently matched values are calculated. nsubj refers to a "*nominal subject which is a noun phrase and syntactic subject of a clause. The governor of this relation might not always be a verb: when the verb is a copular verb, the root of the clause is the complement of the copular verb, which can be an adjective or noun*" [21].

3.5 Dependency Table

The dependency table has nsubj, score and polarity of the sentence, as shown in Table 1. The dependency between words is nsubj(heavy-9, Laptop-6) nominal subject that shows the relationship between words such as laptop and heavyweight.

Nominal Subject Dependency Words	Score(W1+W2)	Polarity
Heavyweight, Laptop	-(WS1+WS2)=-R	Negative
Heavyweight, Mobile		
Heavyweight, Tab		
Heavyweight, phone		
Heavyweight, Laptop		
Heavyweight, Monitor		
Heavyweight, Keyboard		
Heavyweight, Mouse		
Heavyweight, LCD		
Heavy, LCD	-(0.25+0)=-0.25	Negative
Heavy, Mobile		
Heavy, Tab		
Heavy, Computer		
Heavy, Laptop		
Heavy, Monitor		
Heavy, Keyboard		
Heavy, Mouse		

Table 1. Dependency Table

3.6 Calculation of Polarity Score

The score calculated to Adverb, Adjective, Verb, and Noun and on the basis of dependencies the score is calculated in the above table. The table has 500 words nominal subject dependency. Dependency calculation is illustrated in the following example:

Sentence 1

purchased/VBD HP/NNP pro-book/NN
0 0 0

months/NNS ago/RB ./.
0 0

Sentence 2

video/JJ quality/NN is/VBZ good/JJ ./.
0 0.125 0 0.6

Sentence 3

However/RB Brother/NN thinks/VBZ
0.1 0 0 0 Laptop/NN is/VBZ heavy/JJ

0 -0.25

Positive polarity is calculated as:

Feature Positive polarity:

$$pos = \sum_{i=1}^{n} \frac{wi}{n}$$

Total Polarity of Positive Features:

$$TP = \frac{Polarity(F1) + Polarity(F2) + ... + Polarity(Fn)}{Total\ numb\ er\ of\ pos\ \ Features}$$

Negative polarity is calculated as:

Feature Negative polarity:

$$neg = -\sum_{i=1}^{n} \frac{wi}{n}$$

Total Polarity of Negative Features:

$$TN = \frac{Polarity(F1) + Polarity(F2) + ... + Polarity(Fn)}{Total\ numb\ er\ of\ neg\ \ Features}$$

Sentence 1: PolarityPos = (0+0+0+0+0)/5=0
PolarityNeg=(0)/0=0

Sentence2: PolarityPos = (0+0.125+0+0.6+0)/5 = 0.181,
PolarityNeg = (0)/0 = 0

Sentence3: PolarityPos = (0.1+0+0+0)/4=0.025
PolarityNeg = (-0.25/1) = -0.25

Algorithm 1: Algorithm of Negation Effect

Pos= Positive review in dataset
Neg = Negative review in dataset
FOR Extract Features and apply polarity from SentiWordNet
{

Feature Positive polarity $pos = \sum_{i=1}^{n} \frac{wi}{n}$

Feature Negative polarity $neg = -\sum_{i=1}^{n} \frac{wi}{n}$

If polarity(TP)!=0 and Polarity(TN)==0 then assign Polarity(TN)=1

Total Polarity of Positive Features $TP = \frac{Polarity(Feature\ 1) + Polarity(Feature\ 2) + ... + Polarity(Feature\ n)}{Total\ numb\ er\ of\ pos\ \ Features}$

Total Polarity of Negative Features $TN = \frac{Polarity(Feature\ 1) + Polarity(Feature\ 2) + ... + Polarity(Feature\ n)}{Total\ numb\ er\ of\ neg\ \ Features}$

Polarity of Sentence *TS = Polarity (TP) * Polarity (TN)*
Polarity of Review = *(TS1 * TS2 * TS3 .... TN)*
}

>-1	>-0.75	>-0.5	>-0.25	0	<0.3	<0.6	<0.75	<1.0
HighlyNeg	Moderatively Neg	Neg	WeakNeg	Pos	WeakPos	Pos	Meditatively Pos	Strongly Pos

Table 2. Polarity Measurements

With the help of this method, we can find hidden negativity of sentence. The negative score will decrease the value of polarity and help to find more correct polarity score. Sense of the first sentence is neutral and sense of the second sentence is positive, but the sense of third sentence is negative. Therefore, the whole polarity Score lies in a Weak Positive category.

3.7 Polarity Measurements

Polarity table has different categories Neutral, Positive, Moderate Positive, Strongly Positive, Negative, Moderate Negative, Strongly Negative. These polarity words help to define the text polarity. All categories have a range that will define the polarity category as shown in above Table 2.

3.8 Reliability and Validity

Sentiment analysis was performed on 500 reviews that were collected from Amazon and trustedreviews.com websites. The given technique is applied to check whether reviews have a positive sense or negative sense. Results are shown in Table 3.

Precision = tp / Tp+fp

Recall = tp / Tp+fn

Accuracy = tp+tn / Tp+Tn+fp+fn

Method Apply on	Polarity	Precision	Recall	Accuracy
Sentences	Without Negation	79%	79%	88%
	Negation	84.87%	84.81%	91.8%

Table 3. Results

The result in table showed that if negation is ignored in the review. The precision is 79% and leads to misclassification. With negation result improved as 84.87% precision. The review which is classifying in positive but actually belongs to a negative class. Our modified negation approach also improves the recall and accuracy as 84.81 and 91.8% than without negation classification.

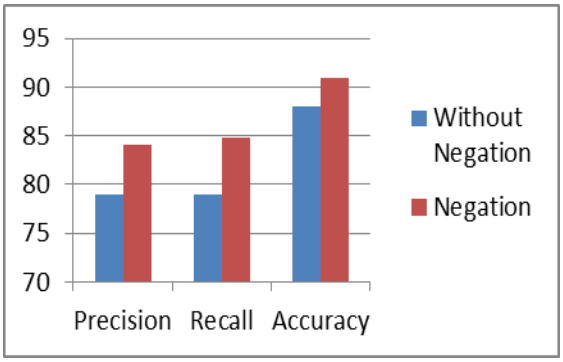


Figure 3. Result of Modified Negation Approach

4. Conclusion

In this paper, our result showed that impact of negation on customer reviews. We find correct polarity of reviews by calculating dependencies of negation words. We also investigate and experiment on customer reviews that how negation affects the polarity of positive reviews that actually belong to negative reviews. This paper presents a technique which calculates the polarity of negative review for mobile and laptop with novel polarity technique. The result presented the improved precision, recall and accuracy with negation words.

Acknowledgment

The authors would like to thank the Universiti Tun Hussein Onn Malaysia (UTHM) and Ministry of Higher Education Malaysia for supporting this research under Project Vot No. U497.

References

[1] Kosala, R., Blockeel, H. (2000). Web Mining Research: A Survey. *ACM SIGKDD Explorations Newsletter* 2 (1) 1-15.

[2] McKnight, W. (2005). Text Data Mining in Business Intelligence : Building Business Intelligence. Information Management, January, 2005.

[3] Grobelnik, M., Mladenic, D. (2004). Text-mining tutorial. *The Proceedings of Learning Methods for Text Understanding and Mining, Grenoble, France*.

[4] Dadvar, M., Hauff, C., De Jong, F. (2011). Scope of negation detection in sentiment analysis.

[5] Baccianella, S., Esuli, A., Sebastiani, F. (2010). *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In*: Paper presented at the LREC.

[6] Hogenboom, A., Van Iterson, P., Heerschop, B., Frasincar, F., Kaymak, U. (2011). *Determining negation scope and strength in sentiment analysis. In*: Paper presented at the Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on.

[7] Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S., Khan, K. H. (2016). Sentiment analysis and the complex natural language. *Complex Adaptive Systems Modeling,* 4 (1) 1.

[8] Banker, S., Patel, R. (2016). Abrief Review of Sentiment analysis methods.

[9] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies,* 5 (1) 1-167.

[10] Karkare, V. Y., Gupta, S. R. (2013). A Survey on Product Evaluation using Opinion Mining. *International Journal Of Computer Science And Applications,* 6 (2).

[11] Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T. (2002). Mining product reputations on the web. *In*: Paper presented at the Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.

[12] Tripathy, A., Agrawal, A., Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications,* 57, 117-126.

[13] Yadav, S. K. (2015). Sentiment analysis and classification: A survey. *International Journal of Advance Research in Computer Science and Management Studies,* 3 (3).

[14] Shaikh, A., Shaikh, R., Gupta, T. Sentiment Analysis to Understand User Context in Web Document.

[15] Asmi, A., Ishaya, T. (2012). Negation identification and calculation in sentiment analysis. *In*: *The Second International Conference on Advances in Information Mining and Management* (p. 1-7).

[16] Jia, L., Yu, C., Meng, W. (2009). *The effect of negation on sentiment analysis and retrieval effectiveness. In*: Paper presented at the Proceedings of the 18th ACM conference on Information and knowledge management.

[17] Wiegand, M., Balahur, A., Roth, B., Klakow, D., Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. *In*: Paper presented at the Proceedings of the workshop on negation and speculation in natural language processing.

[18] Councill, I. G., McDonald, R., Velikovich, L. (2010). *What's great and what's not: learning to classify the scope of negation for improved sentiment analysis.* Paper presented at the Proceedings of the workshop on negation and speculation in natural

language processing.

[19] Blanco, E., Moldovan, D. I. (2011). *Some Issues on Detecting Negation from Text.* Paper presented at the FLAIRS Conference.

[20] Heerschop, B., Hogenboom, A., Frasincar, F. (2011). *Sentiment lexicon creation from lexical resources, In:* Paper presented at the Business Information Systems.

[21] De Marneffe, M.-C., Manning, C. D. (2010). Stanford typed dependencies manual (2008). *URL: http://nlp. stanford. edu/ software/dependencies_manual. pdf* [access date: 30h].