Evaluating Sentiment Analysis Methods and Identifying Scope of Negation in Newspaper Articles

S Padmaja Dept. of CSE, UCE Osmania University Hyderabad Prof. S Sameen Fatima Dept. of CSE, UCE Osmania University Hyderabad Sasidhar Bandu Dept. of English Salman Bin Abdul-Aziz University Saudi Arabia

Abstract-Automatic detection of linguistic negation in free text is a demanding need for many text processing applications including Sentiment Analysis. Our system uses online news archives from two different resources namely NDTV and The Hindu. While dealing with news articles, we performed three subtasks namely identifying the target; separation of good and bad news content from the good and bad sentiment expressed on the target and analysis of clearly marked opinion that is expressed explicitly, not needing interpretation or the use of world knowledge. In this paper, our main focus was on evaluating and comparing three sentiment analysis methods (two machine learning based and one lexical based) and also identifying the scope of negation in news articles for two political parties namely BJP and UPA by using three existing methodologies. They were Rest of the Sentence (RoS), Fixed Window Length (FWL) and Dependency Analysis (DA). Among the sentiment methods the best F-measure was SVM with the values 0.688 and 0.657 for BJP and UPA respectively. On the other hand, the F measures for RoS, FWL and DA were 0.58, 0.69 and 0.75 respectively. We observed that DA was performing better than the other two. Among 1675 sentences in the corpus, according to annotator I, 1,137 were positive and 538 were negative whereas according to annotator II, 1,130 were positive and 545 were negative. Further we also identified the score of each sentence and calculated the accuracy on the basis of average score of both the annotators.

Keywords—Sentiment Analysis; Negation Identification; News Articles

I. INTRODUCTION

Access to popular communication platforms have given a way for the public to generate emotions, opinions, sentiments, evaluations, appraisals and attitudes towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. The volume of such information created every day is massive. Through such Online Social Networks (OSN) we share this information as part of our everyday lives for better understanding and integrating into our surrounding reality. With the effect of such networks on a daily basis, we start recoiling our valuable decisions and actions with certain predefined notions created by others through their reviews. Thus, it is both interesting and challenging to see what and how the trends in such OSNs change from time to time. Comprehensive investigations can lead to appropriate predictions. This gives an edge to Text Mining and explores the area of Sentiment Analysis (Opinion Mining) which is designated as an automatic processing of texts to detect opinions.

Authors of such OSNs express their opinions freely, but, this is not the case with news articles. The major difference between news and product reviews is that the target of the sentiment is less concrete and is expressed much less explicitly. Another major difference is Newspaper articles is that they give an impression of objectivity to refrain from using clearly positive or negative vocabulary. They resort to other means to express their opinion, like embedding statements in a more complex discourse, omitting facts that highlight some important people. For this reason sentiment analysis on news text is rather difficult compared to others. Moreover, the automatic detection of the scope of linguistic negation is a problem encountered in wide variety of documents like understanding tasks, medical data mining, general fact or relation extraction, question answering, sentiment analysis and many more.

The goal of this work is to evaluate and compare three sentiment analysis methods (two machine learning based and one lexical based) and also to identify the scope of negation in news articles for two political parties namely BJP and UPA by using three existing methodologies namely Rest of the Sentence (RoS), Fixed Window Length (FWL) and Dependency Analysis (DA) with respect to sentiment expressed in online news archives.

II. RELATED WORK

Sentiment Analysis of Natural Language texts is a broad and expanding field. A text may contain both Subjective and Objective sentiments. Wiebe (1994) [2] defines Subjective text as the "linguistic expression of somebody's opinions, sentiments, emotions, evaluations, beliefs and speculations". In her definition, the author was inspired by the work of the linguist Ann Ban field (1982) [3], who defines subjective as a sentence that takes a character's point of view and that present private states (that are not open to objective observation or verification), defined by Quirk (1985) [4], of an experiencer, holding an attitude, optionally towards an object. Bing Liu (2010) [5] defines Objective text as the facts that are expressed about entities, events and their properties. Esuli and Sebastiani (2006) [6] define Sentiment Analysis as a recent discipline at the crossroads of Information Retrieval and Computational Linguistics which is concerned not with the topic a document is about, but with the opinion it expresses.

While a wide range of human moods can be captured through Sentiment Analysis Hannak (2012) [7] says majority of studies focus on identifying the polarity of a given text—

that is to automatically identify if a message about a certain topic is positive or negative.

Polarity analysis has numerous applications especially on news articles using several methods. Pang and Lee (2002) [8] broadly classifies Sentiment Analysis methods into machinelearning-based and lexical-based. Machine learning methods often rely on supervised classification approaches, where sentiment detection is framed as a binary (i.e., positive or negative). This approach requires labeled data to train classifiers [8]. While one advantage of learning-based methods is their ability to adapt and create trained models for specific purposes and contexts, their drawback is the availability of labeled data and hence the low applicability of the method on new data. This is because labeling data might be costly or even prohibitive for some tasks. On the other hand, lexical-based methods make use of a predefined list of words, where each word is associated with a specific sentiment. The lexical methods vary according to the context in which they were created [8].

Sentiment Analysis work has been handled heavily in subjective text types where the target is clearly defined and unique across the text as the case in movie or product reviews. But when applying Sentiment Analysis to the News domain, Alexandra Balahur (2009) [9] says it is necessary to clearly define the scope of the tasks in three levels. They are definition of the target; separation of good and bad news content from the good and bad sentiment expressed on the target and analysis of clearly marked opinion that is expressed explicitly, not needing interpretation or the use of world knowledge. Thus, on the same lines we built a corpus of 689 political instances from three different news databases namely, The Hindu, Times of India and Economic Times for three distinct Indian parties namely United Progressive Alliance (UPA), Telugu Desam Party (TDP) and Telangana Rashtra Samithi (TRS) to analyze the choice of certain words used in political texts to influence the sentiments of public in polls. Following is the glimpse of work done using news articles by various authors either by adopting machine learning based (MLB) or Lexical based (LB) methods with their respective accuracies.

Also, Negation and its scope in the context of sentiment analysis has been studied in the past [10]. However, others have studied various forms of negation within the domain of sentiment analysis, including work on content negators, which typically are verbs such as "hampered", "lacked", "denied", etc. [10] [11]. A recent study by Danescu- Niculescu-Mizil et al. looked at the problem of finding downward call for operators that include a wider range of lexical items, involving soft negators such as adverbs "rarely" and "hardly" [13]. With the absence of a general purpose corpus annotating the precise scope of negation in sentiment corpora, many studies incorporate negation terms through heuristics or softconstraints in statistical models. In the work of Wilson et al., a supervised polarity classifier is trained with a set of negation features derived from a list of cue words and a small window around them in the text [12]. Choi and Cardie et al. combine different kinds of negators with lexical polarity items through various compositional semantic models, both heuristic and machine learned, to improve phrasal sentiment analysis [11]. In that work [11] the scope of negation was either left undefined or determined through surface level syntactic patterns similar to the syntactic patterns from Moilanen and Pulman [10].

III. DATA SETS

The work described in this paper was part of a larger research to improve the accuracy of sentiment analysis in the daily political news present in online news archives. For our study a sample of 513 Political news opinions, dating from January 1, 2014 to March 31, 2014, were obtained from two online Indian news archives namely The Hindu ¹ and NDTV ². This sample of 513 political news opinions contained 1675 sentences in total.

We extracted only political news pertaining to 2014 General Elections for two leading parties explicitly "UPA" and "BJP". Each sentence was manually annotated with the scope of negation by two annotators, after achieving inter-annotator agreement of 91% with a second annotator on a smaller subset of 20 sentences containing negation.

Among 1675 sentences in the corpus according to annotator I, 1,137 were positive and 538 were negative whereas according to annotator II, 1,130 were positive and 545 were negative sentences. Inter-annotator agreement was calculated using strict exact span criteria where both the existence and the left/right boundaries of a negation span were required to match.

The obtained political news corpus was annotated with a general principle to consider minimal span of a negation covering only the portion of the text being negated semantically and according to the following instructions:

A. Negation words

Words like "never", "no", or "not" in its various forms are not included in negation scope. For example, in the sentence, "It was not XYZ", only "XYZ" is annotated as the negation span.

B. Noun phrases

Typically entire noun phrases are annotated as within the scope of negation if a noun within the phrase is negated. For example, in the sentence, "The consequence of the act was not due to Modi" the string "due to Modi" is annotated. This is also true for more complex noun phrases, e.g., "People did not expect Sonia to act in such a way" should be annotated with the span "expect Sonia to act in such a way".

C. Adjectives in noun phrases

Do not annotate an entire noun phrase if an adjective is to be negated - consider the negation of each term separately. For instance, "Not top-drawer political party, but still wins. "top drawer" is negated, but "political party" may not, since it is still party, just not "top-drawer".

D. Adverbs/Adjective phrases

a) Case 1: Adverbial comparatives like "very," "really," "less," "more", etc., annotate the entire adjective phrase, e.g., "It was not very good" should be annotated with the span "very good".

b) Case 2: If only the adverb is directly negated, only

¹http://www.thehindu.com/opinion/

²http://www.ndtv.com/article/list/opinion/

annotate the adverb itself. e.g., "Not only was it great", or "Not quite as great": in both cases the subject still is "great", so just "only" and "quite" should be annotated, respectively.

However, there are cases where the intended scope of adverbial negation is greater, e.g., the adverb phrase "just a small part" in "Modi was on stage for the entire speech. It was not just a small part".

c) Case 3: "as good as X". Try to identify the intended scope, but typically the entire phrase should be annotated, e.g., "It was not as good as I remember".

Note that Case 2 and 3 can be intermixed, e.g., "Not quite as good as I remember", in this case follow 2 and just annotate the adverb "quite", since it was still partly "as good as I remember", just not entirely.

E. Verb Phrases

If a verb is directly negated, annotate the entire verb phrase as negated, e.g., "appear to be fair" would be marked in "He did not appear to be fair". For the case of verbs (or adverbs), we made no special instructions on how to handle verbs that are content negators. For example, for the sentence "I can't deny it was good", the entire verb phrase "deny it was good" would be marked as the scope of "can't". Ideally annotators would also mark the scope of the verb "deny", effectively canceling the scope of negation entirely over the adjective "good". As mentioned previously, there are a wide variety of verbs and adverbs that play such a role and recent studies have investigated methods for identifying them [3] [4]. We leave the identification of the scope of such lexical items as future work.

One of the freely available resources for evaluating negation detection performance is the Bio-Scope corpus [3], which consists of annotated clinical radiology reports, biological full papers, and biological abstracts. Annotations in Bio-Scope consist of labeled negation and speculation cues along with the boundary of their associated text scopes. Each cue is associated with exactly one scope, and the cue itself is considered to be part of its own scope.

Traditionally, negation detection systems have encountered difficulty in parsing the full papers subcorpus, which contains nine papers and a total of 2670 sentences.

IV. EXTRACTING FEATURES

A. Bag-of-Words Features

Here each feature indicates the number of occurrences of a word in the document. The news for a given day is represented by a normalized unit length vector of counts, excluding common stop words and features that occur fewer than 20 times in our corpus [16].

B. Entity Features

As shown by Wiebe et al., it is important to know not only what is being said but about whom it is said [15]. The term "victorious" by itself is meaningless when discussing an election – meaning comes from the subject.

Similarly, the word "scandal" is bad for a candidate but good for the opponent. Subjects can often be determined by proximity. If the word "scandal" and "UPA" are mentioned in the same sentence, this is likely to be bad for "Sonia Gandhi". A small set of entities relevant to the party can be defined priori to give context to features. For example, the entities "Sonia Gandhi," "Rahul Gandhi", "Dr. Manmohan Singh", "UPA" and "Congress party" were known to be relevant before the general election. News is filtered for sentences that mention exactly one of these entities. Such sentences are likely about that entity, and the extracted features are conjunctions of the word and the entity. For example, the sentence "Sonia Gandhi is facing another scandal" produces the feature "Sonia Gandhi-scandal" instead of just "scandal." Two, Context disambiguation comes at a high cost: about 70% of all sentences do not contain any predefined entities and about 7% contain more than one entity [17].

These likely relevant sentences are unfortunately discarded, although future work could reduce the number of discarded sentences using co reference resolution.

C. Dependency Features

While entity features are helpful they cannot process multiple entity sentences. These sentences may be the most helpful since they indicate entity interactions [2]. Consider the following three example sentences:

- Narendra Modi defeated Rahul Gandhi in the debate.
- Rahul Gandhi defeated Narendra Modi in the debate.
- Narendra Modi, the president of BJP, defeated Sonia Gandhi in the last night's debate.

Obviously, the first two sentences have very different meanings for each candidate's campaign. However, representations considered so far do not differentiate between these sentences, nor would any heuristic using proximity to an entity. Three effective features rely on the proper identification of the subject and object of "defeated." Longer n-grams, which would be very sparse, would succeed for the first two sentences but not the third.

To capture these interactions, sentences were part of speech tagged, parsed with a dependency parser. The resulting parses encode dependencies for each sentence, where word relationships are expressed as parent-child links. The parse for the third sentence above indicates that "Narendra Modi" is the subject of "defeated," and "Sonia Gandhi" is the object. Features are extracted from parse trees containing the predefined entities (as mentioned in subsection 4.2). Note that they capture events and not opinions.

V. IMPLEMENTATION

In the pre-processing stage, the data is cleaned to hold only what is essential for the analysis. Steps like tokenization, stop word removal, lemmatization and pos tagging were performed using NLTK and Stanford POS tagger.

To bring a comparison between lexical based and machine learning methods, we implemented SentiWordNet (SWN) [10] which is based on English lexical dictionary called WordNet [5] and two machine learning based algorithms namely Naive Bayes (NB) and Support Vector Machine (SVM) using WEKA with 10 fold cross validation [7] respectively. The scope of negation detection is limited to explicit rather than implied negations within a single sentence.

A. Dictionary Tagging

POS tagged sentences were given as an input to the Dictionary tagger. Dictionary tagger then tags each token of every sentence with tags like positive, negative, negation (inv). SentiWordNet values were taken to tag the tokens.

B. Negation Scope Determination

The scope of negation detection is limited to explicit rather than implied negations within a single sentence. A lexicon of negations was created to identify the presence of negation in the sentence. Using a statistics driven approach, Klima et al. was the first to identify negation words by analyzing word cooccurrence with n-grams that are cues for the presence of negation [6]. Klima's lexicon served as a starting point for the present work and was further refined through the manual inclusion of selected negation cues from the corpus. The final list of cues used for the evaluation is presented in Table 1.

TABLE I. LEXICON OF EXPLICIT NEGATIONS

Hardly	Lack	Neither	Nor
Never	No	Nobody	None
Nothing	Nowhere	Not	n't
cannot	Without	Bad	Uninspired
Expensive	Disappoint	Ditch	Misunderstand

The above list of lexicon serves as a reliable signal to detect the presence of explicit negations. It does not provide any means of inferring the scope of negation. To detect the scope of negation in the sentence three different approaches were implemented.

The first approach was Rest of the Sentence (RoS). In this method, all the words that follow the negation keyword are reversed in a sentence. For this, we used the negation tags given by the dictionary tagger. For every token containing the tag, if the previous tokens contains negation tag, we reverse the polarity of the current token, append negation token to the current token so that it gets propagated to the last word of the sentence and then add it to the score value.

Algorithm for Polarity Calculation

For a sentence score If the negation tag is identified in a sentence Return{reverse the values of the next tokens from the negation tag and take the sum of all scores in the sentence} Else Return {sum of all scores in the sentence}

The second was the Fixed Window Length (FWL) approach in which we considered a fixed length of 4 words followed by a negation keyword. Every word in a sentence was tagged as positive, negative or negation by the dictionary tagger as discussed in subsection 5.1. If the tagged sentence

contains negation, then we started a counter equal to the window size to reverse the polarity of the tokens next to negation till the size is attained and then the resultant was added to the score value.

Algorithm for Polarity Calculation

For a sentence score

If the negation tag is identified in a sentence Return {reverse the value for four consecutive scores from negation tag and then add the total scores in the sentence} Else

Return {sum of all scores in the sentence}

The third approach was Dependency Analysis (DA). Only unigram features were employed, but each unigram feature vector is expanded to include bigram and trigram representations derived from the current token in conjunction with the prior and subsequent tokens. The distance measures can be explained as follows. Token-wise distance is simply the number of tokens from one token to another, in the order they appear in a sentence. Dependency distance was more involved, and was calculated as the minimum number of edges that must be traversed in a dependency tree to move from one node (or token) to another. Each edge was considered to be bidirectional. The number 0 implies that a token was, or was part of, an explicit negation cue. The numbers 1-4 encode stepwise distance from a negation cue, and the number 5 was used to jointly encode the concept as "not applicable". To get the parse tree of the sentences, we used Stanford parser. The reason for that was in the negation identification process, the kind of negation i.e. "No one likes his behavior", where 'no' is used to determine the behavior of one, is also identified. This rocess also takes care of the negation in conjunction sentences.

The output of which was given to the nltk parse function to get the Tree object of nltk, so that traversing through the parse tree was made possible.

Having determined the scope of negation cues, the sentiment scores associated with the words in the negation keywords' scope can be inverted. To this end, we introduce unigram sentiment modifiers, which are initialized at a value of 1, indicating that the sentiment score retrieved from the sentiment lexicon is considered to be the true sentiment score associated with that word in the considered context. In case a word is negated, the sentiment modifier may be multiplied with an inversion factor. Initially, we assume this factor to be equal to -1. Finally, when all word scores have been determined while accounting for negation, sentences can be classified as either positive or negative. To this end, we use a sentence scoring function. If the sum of word-level sentiment scores in a sentence produces a number smaller than 0, the sentence is classified as negative, else, the sentence is classified as a positive sentence. We ignored those sentences whose score is 0 as we are considering only two class problem.

Algorithm for Polarity Calculation

For a sentence score

If the negation word is identified in the sentence Return {reverse the polarities of its parent nodes and then add the total scores in the sentence} Else

Return {sum of all scores in the sentence}

VI. RESULTS

In order to understand the advantages, disadvantages, and limitations of the various sentiment analysis methods and analyze the choice of words from news articles in winning a particular party, we present comparison results among them which are as shown below -

A. Prediction Performance

We illustrate a comparative performance evaluation of each method in terms of correctly predicted polarity. Here we depict the results for precision, recall, accuracy, and F-measure for the three previously described methods. Table 3 shows the performance of the results obtained for each labeled dataset. For the F-measure, a score of 1 is ideal and 0 is the worst possible. Among the methods the best F-measure was SVM with the values 0.688 and 0.657 for UPA and TRS respectively and NB with the value 0.723 for TDP. On par, the accuracy of SWN, NB and SVM with the values 0.742, 0.725 and 0.666 are the highest scores in UPA, TDP and TRS respectively.

TABLE II. PREDICTED PERFORMANCE FOR LABELED DATASET

Metric	BJP			UPA		
	SVM	NB	SWN	SVM	NB	SWN
Recall	0.79	0.62	0.58	0.70	0.75	0.50
Precision	0.60	0.75	0.86	0.69	0.71	0.54
F-Measure	0.68	0.65	0.56	0.67	0.73	0.51
Accuracy	0.72	0.62	0.74	0.70	0.75	0.61

The evaluation metric for three different methodologies was calculated separately for both the parties. The results are as shown below:

 TABLE III.
 RESULTS TO DETECT COMPARISON OF NEGATION BETWEEN THREE EXISTING METHODS

Metric	UPA			BJP		
	FWL	RoS	DA	FWL	RoS	DA
Recall	0.69	0.53	0.72	0.68	0.56	0.77
Precision	0.61	0.51	0.69	0.64	0.59	0.79
F-Measure	0.66	0.54	0.71	0.72	0.61	0.78

From the results we can observe that DA methodology was outperforming when compared to RoS and FWL.

TABLE IV.	OBSERVATIONS DRAWN FROM THE RESULTS AFTER APPLYING
	DA SCOPE OF NEGATION DETECTION

Negation	First word/ Phrase/ clause	Second word/ Phrase/clause	Result
Present	Positive	Positive	Negative
Absent	Positive	Positive	Positive
Present	Positive	Negative	Positive
Absent	Positive	Negative	Negative
Present	Negative	Positive	Positive
Absent	Negative	Positive	Negative
Present	Negative	Negative	Negative
Absent	Negative	Negative	Positive

VII. CONCLUSION

In this paper, we began the comparison of three representative sentiment analysis methods like Support Vector Machine, Naïve Bayes and SentiWordNet.

Our comparison study focused on detecting the polarity of content (i.e., positive and negative affects) from good or bad news for two different Indian political parties. Thus by extracting the average predicted performance we observed that the choice of certain words used in political text was influencing the Sentiments in favor of BJP which might be one of the causes for them be the winners in Elections 2014.

We also study the concept of scope of negation (t) identification which is precisely the sequence of words affected by t. Three sets of experiments were performed to bring a comparison in identifying the scope of negation in news articles for two political parties. Experimental results show that DA method outperforms better than other two.

ACKNOWLEDGMENT

I am indebted to the EMR Division under HRD Group of Council of Scientific & Industrial Research (CSIR) for providing me CSIR Research Fellowship (SRF).

REFERENCES

- [1] Wiberg, M., "The interaction Society: Theories Practice and Supportive Technologies." Idea Group Inc. 2004.
- [2] Wiebe, J., "Tracking point of view in narrative." Computational Linguistics, 20, 1994.
- [3] Banfield, A. "Unspeakable sentences: Narration and Representation in the Language of Fiction." Routledge and Kegan Paul, 1982.
- [4] Quirk, R., " A Comprehensive Grammar of the English Language", Longman Publishing House, 1985.
- [5] Bing Liu, "Sentiment Analysis and Subjectivity", Handbook of Natural Language Processing, Second Edition, 2010.

- [6] Esuli, A. and F. Sebastiani, "SentiWordNet: A Publicly Available Resource for Opinion Mining." In Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2006, Italy, 2006.
- [7] A. Hannak, E. Anderson, L. F. Barrett, S. Lehmann, A. Mislove, and M. Riedewald. Tweetin'in the rain: Exploring societal-scale effects of weather on mood. In Int'l AAAI Conference on Weblogs and Social Media (ICWSM), 2012.
- [8] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine Learning techniques. In ACL Conference on Empirical Methods in Natural Language Processing, pages 79–86, 2002.
- [9] Alexandra Balahur, Ralf Steinberger, "Rethinking Sentiment Analysis in the News: from Theory to Practice and back", European Commission-Joint Research Centre, University of Alicante, Department of Software and Computing Systems. WOMSA, pp.1-12, 2009.
- [10] Karo Moilanen and Stephen Pulman 2007. Sentiment Composition. Proceedings of the Recent Advances in Natural Language Processing International Conference Borovets, Bulgaria.
- [11] Yejin Choi and Claire Cardie. 2008. Learning with Compositional Semantics as Structural Inference For Subsentential Sentiment Analysis. Proceedings of the Conference on Empirical Methods on Natural Language Processing. ACL, Honolulu, HI.

- [12] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann, 2005, Recognizing contextual polarity in phrase level sentiment analysis. Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing Vancouver, Canada.
- [13] Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Richard Ducott, 2008, Without a 'doubt'? Unsupervised discovery of downwardentailing operators. Proceedings of The 10th Annual Conference of the North American Chapter of the Association for Computational Linguistics. ACL, Boulder, CO.
- [14] Kevin Lerman, Ari Gilder, Mark Dredze, 2008 Reading the Markets: Forecasting Public Opinion of Political Candidates by News Analysis.
- [15] Gunnel Tottie. 1991. Negation in English Speech and Writing: A Study in Variation Academic, San Diego, CA.
- [16] Penn_Treebank, 1992, The Penn Treebank Project [WWW]. Available from: http://www.cis.upenn.edu/~treebank/ [Accessed April 21, 2012]
- [17] Insaac G. Councill, Ryan McDonald, 2010, What's Great and What's Not: Learning to Classify the Scope of Negation for Improved Sentiment Analysis, Negation and Speculation in Natural Language Processing (NeSp-NLP 2010), Proceedings of the Workshop, Uppnsala, Sweden, 10 July 2010.
- [18] User Manual http://iasri.res.in/ebook/win_school_aa/notes/WEKA.pdf
- [19] NITK 3.0 documentation, http://nltk.org/