NIC_LBA: Negations and Intensifier Classification of microblog data using Lexicon Based Approach

A.Angelpreethi[1], and Dr.S.Britto Ramesh Kumar[2].

1. Research Scholar, Department of Computer Science, St.Joseph's College, Trichy-2.
2. Assistant Professor, Department of Computer Science, St.Joseph's College, Trichy-2.

Abstract: Due to the vast generation of online information, there are huge number of opportunities and challenges created for both consumers and internet users. Opinion Mining is the digital acknowledgement of public feelings, emotions, opinions and services. It is also helps individual users and business peoples and organizations to monitor the other people comment about a particular items, services, issue and topics. The objective of the paper is to feat the affluence of user opinions collected from micro blogging sites to analyse the semantic orientation of reviews by categorizing them into positive, negative classes to identify and classify negations and intensifiers along with slang words and emoticons.

Key Words: Negations, Intensifiers, Slang Words, Emoticons, Opinion Mining.

1. Introduction

Opinion Mining or sentiment analysis is a challenging natural language processing and text mining problem for automatic extraction, classification and summarization of sentiments and feelings expressed through online text. It is also helps individual users and business peoples and organizations to monitor the other people comment about a particular items, services, issue and topics. Consequently, user opinions measured as important source of information in opinion mining applications mainly for decision-making.

Identifying the polarity of the review is essential task in opinion mining systems. Especially subjectivity identification is an essential process. Because subjective sentences expresses users opinion while the objective sentences expresses the fact that is no opinion. This subjectivity classification can be done in three levels. i.e., Document level, Sentence Level and Aspect level. Using any level of classification there is some opinion related to entity. This paper addresses the classification of negations and intensifiers along with slang words, emoticons and emotion terms from the microblogging sites.

This paper organized as section 2 describes the background of the research work, section 3 describes related research work, section 4 presents the motivation of the proposed work, section 5,6 presents the objective and proposed work. At last, section 7 is the conclusion.

2. Background:

Negations are computational linguistic because they inversely affect the polarity of the sentiment. No,not, nor, shouldn't, are the negation words. Whenever negation appear in a sentence or review it must have affected with sentiment and it is difficult to determine the sentiment polarity of the word in a sentence level sentiment analysis. Generally, negation terms may starts with negative affix like un-, in-, dis-, a-, an- , non-, im-, il-, ir-, less etc., Intensifiers are modifiers which enhance or reduce the polarity strength of the opinionated terms in

a sentence. For example very, amazingly, so, too are intensifiers. Those words enhance or reduce the strength of an opinion term. Inensifiers and negations play a vital role in sentiment classification.

3. Related Research

On the effect of negation on sentiment analysis, jia et al [1] introduces the concept of the scope of a negation term. The authors employ a decision tree to determine the polarity of the documents. The proposed scope detection method, considers static delimiters (unambiguous words) such as, because, dynamic delimiters (ambiguous words) such as, like, and heuristic rules which focus on polar expressions. For negation detection they have tried three window sizes; 3, 4 and 5. Their experimental results show that their method outperforms other methods in accuracy of sentiment analysis and the retrieval effectiveness of polarity classification in opinion retrieval.

Wiegand et al., [2] did a review on negation and its scope in sentiment analysis. This work presents various computational approaches to modelling negation in sentiment analysis. The focus of this paper is particularly on the scope of negation. It also discusses limits and challenges of negation modelling. For example, recognition of polar expressions (sentences which carry sentiments) is still a challenging task. The authors also discussed that the effectiveness of negation models can change in different corpora because of the specific construction of language in different contexts.

Kennedy et al., introduces the concept of contextual valence shifters which consist of negation, intensifier and diminisher. Intensifiers and diminishers are terms that change the degree of the expressed sentiments. The sentence, this movie is very good, is more positive than this movie is good. The film is scarcely good in the phrase; the word barely is a diminisher, making this declaration less positive. They have used a term-counting method, a machine learning method and a combination of both methods on the same data collection as was used in our experiment. They found that combining the two systems slightly improved the results compared to machine learning or term-counting methods alone.

Nicholas et al., [3] proposed the algorithm, which sums up the sentiment scores of the terms in the review considering negations and intensifiers; here positive score of a word is taken as one and negative score as -2. To manage negation and intensifier, the entire algorithm works well. However, in case of two negative sentences, both can have same score i.e. this is very bad or this is very wicked. In proposed methodology author will, determined second sentence is more negative than first one.

4. Motivation

Considering the existing works, the proposed work planned a method opine_neg, which can determine the semantic nature of the review. Most of the users expresses their experiences through microblogging sites. Therefore, this work is tested twitter micro blogging site.

The proposed opine_neg approach is rule-based classification supported by many dictionaries, such as Derivative Slang Dictionary, Acronym Slang Dictionary, Shortened Slang Dictionary, Emoticon Dictionary,

SWN Dictionary, Negation Dictionary and Intensifier Dictionary. The main motivation of the work is the lexicon based dictionary approach to classify the reviews based on rule-based method.

5. Objective

The main objective of this paper is to propose a novel technique for negations and intensifiers handling. The sub objectives are defined as follows:

- To propose an effective technique for sentence level opinion mining to reduce the data sparsity.
- To design a rule based approach to handle negations and intensifiers.
- To improve the accuracy of opinion mining and sentiment classification in sentence level sentiment analysis.

6. Proposed Work

In the proposed model there are eight types of opinions are handle for each review. Each review can uses trigram based feature selection. There are two types of sentences. Namely subjective and objective sentences. The sentence, which has opinions, are called subjective sentences. Objective sentences are factual sentences. Table 1 presents the detailed rules for subjective reviews. Each rule can process based on the four steps.

Step 1: Allocate the score using SWN based classifier and allocate Tag

Step 2: Compare the scores of positive and negative

Step 3: Handle negations and intensifiers using rules.

Step 4: summing up of all the scores

Table 1. Rules for Opine_neg approach

Rule No	Rule	Result
Rule 1	(Pve_wd, Pve_sc, ADJ)	+Pve_score
Rule 2	(In, In_sc, ADV) and (Pve_wd, Pve_sc, ADJ)	Pve_score + In_score
Rule 3	(Neg, -Neg_sc, ADJ) and (Pve_wd, Pve_sc, ADJ)	Neg_sc + Product(Pve_sc , (-1))
Rule 4	(Neg, -Neg_sc, ADV) and (In, In_sc, ADV)and (Pve_wd, Pve_sc, ADJ)	(-1) * sum(Neg_sc, In_sc, Pve_sc)

Rule 5	(Nve_wd, Nve_sc, ADJ)	- Nve_sc
Rule 6	(In, In_sc, ADV) and (Nve_wd, -Nve_sc, ADJ)	$(-1)*$ sum(In_sc, Nve_sc)
Rule 7	(Neg, -Neg_Nve_sc, ADV) and (Nve_wd, -Nve_sc, ADJ)	-Neg_Nve_sc + Product (-Nve_sc, (-1))
Rule 8	(Neg, Neg_Nve_sc, ADV) and (In, In_sc, ADV) and (Nve_wd, - Nve_sc, ADJ)	(-Nve_sc) + Product (In_sc, (-1))+Product(- Nve_sc, (-1))

where In➔ Intensifier, In_sc ➔ Score(or) polarity value of the Intensifier, Pve_Wd ➔ Positive sentiment word, Pve_sc ➔ Polarity value of the positive sentiment, ADJ➔ Adjective, ADV➔ Adverb, Neg➔ Negation, Neg_sc➔ Polarity value of the negation, Neg_Nve_sc ➔ Negative Score of the negation, Nve_Wd ➔Negative sentiment word.

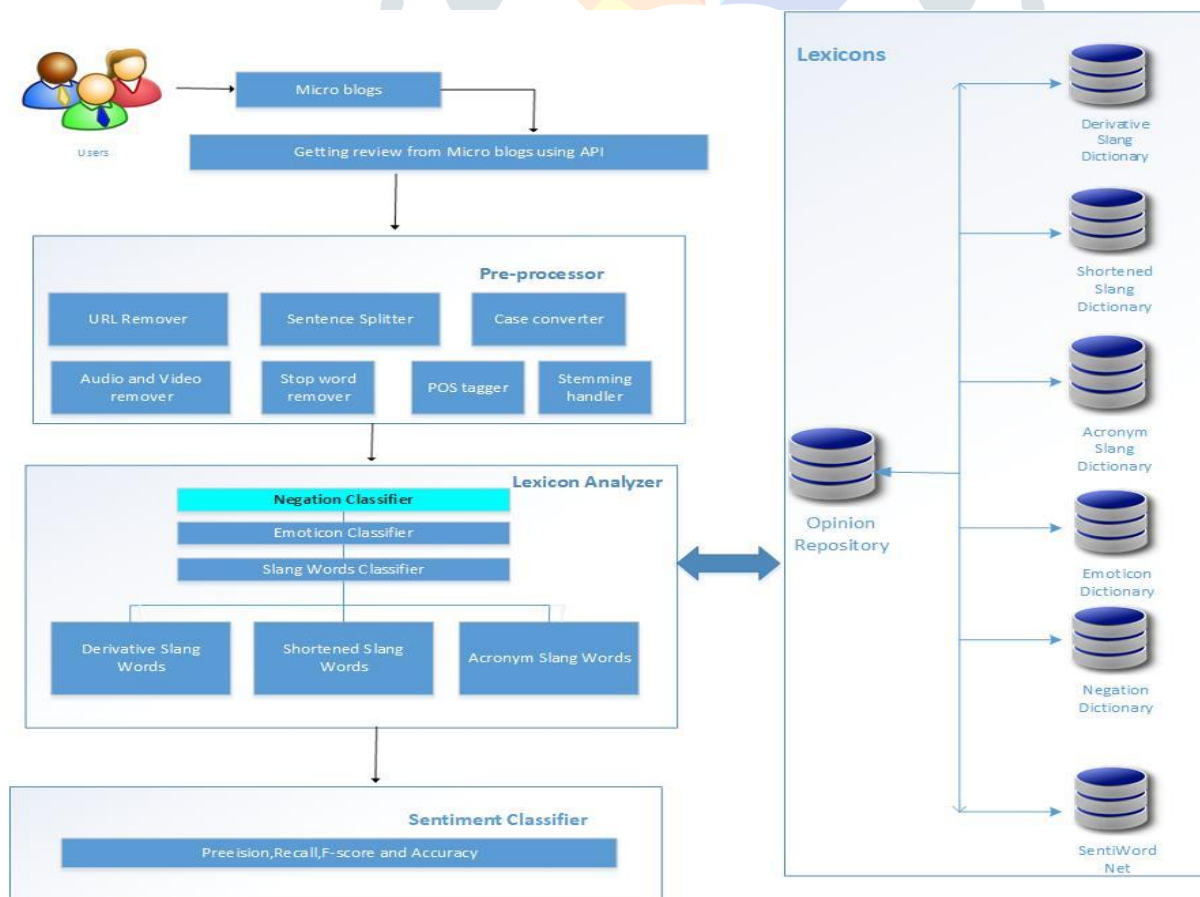The proposed methodological diagram is presented in Figure1.

Fig 1. Methodological diagram of the opine_neg approach

Tweets are collected from the twitter microblogging site. Tweets are store it in a HDFS. These tweets are further separated using (.) separator. After separating each reviews, it will be pre-processed. The pre-processed reviews are given as an input to the opine_neg approach. Opine_neg approach performs and focuses on classification method for the slang words, acronyms and emoticons with negation and intensifiers. The pre-processed data have been implied into the Opine_neg algorithm, each tweet is considered as T. Basically unigram based feature selection, which will be split the word in single, and bigram based feature selection split the words in couple. So, Trigram based feature selection is applied on tweets, which consider sentiments, intensifiers and negation terms. Then tweets will be pre-processed and store it in a document. Each record consists of sentiment, positive score, negative score and its POS tag. For each record, the comparison is performed whether the positive score of the term is greater than or equal to negative score of the term. If the positive score is greater the next step automatically take the positive score otherwise negative score is taken for the next step. After allocation of scores slang, words and emoticons was handled using opine_lexi approach[4]. Whenever negations appear in a sentence the rule based opine_neg approach is performed.

Intensifiers are identified using Part-of-speech tagging method. In a new record, intensifiers are found in a sentence it will be swapped by the sign of intensifier score as sign of next adjective. Similarly, Negations are found in record, it will swap the sign of intensifier and adjective scores as sign of negative scores. At the last, all the values will be summed up to find the intensity of the sentiment. If the intensity of the sentiment is greater than 0 it will be considered as positive sentiment. Intensity of the sentiment is less than 0 it will be considered as negative. Here intensifiers and negations are handled using the rules which was depicted in table 2 and 3. For negation will carefully check whether the previous word is negation or not. If the previous word is negation, then it will automatically revert the value that is multiplied by minus one. So, the sentiment orientation will be changed positive score changed into negative score and vice versa.

Table 2 Rules for Negation and Intensifiers

Orientation of Emotion / Sentiment word	Maximum score of the sentiment word	Presence of Negation	Presence of Negation Score	Presence of Intensifier	Presence of Intensifier Score	Result
Positive	Positive	False	False	False	False	Positive
Positive	Positive	False	False	True	True	Positive
Positive	Positive	True	True	False	False	Negative
Positive	Positive	True	True	True	True	Negative

Negative	Negative	False	False	False	False	Negative
Negative	Negative	False	False	True	True	Negative
Negative	Negative	True	True	False	False	Positive
Negative	Negative	True	True	True	True	Positive

Table 3. Rules for Negations

Orientation of Sentiment word	**Preceding word is Negation**	**Result**
Positive	True	Positive
Positive	False	Positive
Positive	True	Negative
Positive	False	Negative
Negative	True	Negative
Negative	False	Negative
Negative	True	Positive
Negative	False	Positive

Conclusion:

In recent times, opinion mining is popular among every for better decision making. In this paper, a novel technique was proposed to handle negations and intensifiers. Opine_neg approach is proposed to classify the user reviews for finding the semantic orientation. In addition, this approach can reduce the data sparsity. Proposed work is tested with twitter data set with improved results. Thus the proposed technique increases the accuracy of the opinion mining in the sentence level sentiment classification.

References.

1. Jianqiang, Z, "Pre-processing boosting Twitter sentiment analysis", International Conference on Smart City/SocialCom/SustainCom (SmartCity), IEEE, 2015, pp. 748-753.

2. Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M., "Learning subjective language. Computational linguistics", vol. 30, no. 3, pp. 277 308, 2004.

3. Nicholas Pröllochs, Stefan Feuerriegel, and Dirk Neumann. "Detecting negation scopes for financial news sentiment using reinforcement learning", System Sciences (HICSS), 49th Hawaii International Conference on .IEEE, 2016, pp. 1164-1173.

4. A.AngelPreethi, and Dr.S.Britto Ramesh Kumar, "A Dictionary-based Approach for improving the accuracy of opinion mining on big data", International Journal of Research and Analytical Reviews (IJRAR), Vol. 5, Issue 4, Oct-Dec 2018, pp- i836-844.

5. Pak A. and Paroubek P, "Twitter as A Corpus for Sentiment Analysis and Opinion Mining", Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), Valletta, Malta. European Language Resources Association (ELRA), Vol. 10, 2010, pp. 1920-1326.

6. Deepak Singh Tomar, Pankaj Sharma, A text polarity analysis using SentiWordNet based algorithm, IJCSIT 2016 ISSN:0975-9646

7. PeimanBarnaghi, John G. Breslin and Parsa Ghaffari, "Opinion Mining and Sentiment Polarity on Twitter and Correlation Between Events and Sentiment", Second International Conference on Data Computing Service and Applications (BigDataService), IEEE, 2016, pp. 52-57.

8. Rehab M Duwairi, and Mohammad A. Alshboul, "Negation-aware framework for sentiment analysis in Arabic reviews", Future Internet of Things and Cloud (FiCloud), 3rd International Conference on IEEE, 2015, pp. 731-735.

9. Shuigui Huang, Wenwen Han, Xirong Que and Wendong Wang, "Polarity Identification of Sentiment Words based on Emoticons", Ninth International Conference on Computational Intelligence and Security (ICCIS), IEEE, 2013, pp. 134-138.

10. A.AngelPreethi, and Dr.S.Britto Ramesh Kumar, " Dom_Classi: An Enhanced Weighting Mechanism for Domain Specific Words using Frequency based Probability", International Journal of Applied Engineering Research, Vol.14, Issue 1, 2019, pp 140-148.