Journal of Theoretical and Applied Information Technology
31st March 2019. Vol.97. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

COMPREHENSIVE ANALYSIS OF SCOPE OF NEGATION FOR SENTIMENT ANALYSIS OVER SOCIAL MEDIA

[1]NIKHIL KUMAR SINGH, [2]DEEPAK SINGH TOMAR

[1]Reserach Scholar, MANIT, Department of Computer Science & Engineering, India

[2]Associate Professor. MANIT, Department of Computer Science & Engineering, India

E-mail: [1]Nikhilsinghmanit@gmail.com, [2]Deepaktomarmanit@gmail.com

ABSTRACT

Handle Negation is essential for effective sentiment analysis decision support system. Negation control comprised identification of negation cues, scope of negation and their influence within it. Negation can either invert, reduce or increase the polarity score. This paper present comprehensive assessment of recent research on negation control for sentiment analysis technique. Explore negation cue and scope detection techniques in collaboration with classification technique over social media data set. This assessment has included the evaluation of sentiment classification (Support vector machine, Navies Bayes, Linear Regression and Random Forest) and scope detection techniques (conjunction Analysis, Punctuation Mark and grammatical dependency tree) over presented prepossessing framework. This paper yield interesting result about collective response of negation scope detection and classification technique for sentiment analysis. Negative scope feature vector significantly increase the polarity classification accuracy of sentiment classification technique. Grammatical dependency tree in collaboration with SVM and Naves Bayes can detect negation with better accuracy as compere to conjunction and punctuation word scope detection technique.

Keywords: *Sentiment Analysis, Negation Cues, Scope Detection, Conjunction Analysis, Punctuation Mark, Grammatical Dependency Tree, Navies Bayes, Linear Regression, Random Forest, SVM.*

1. INTRODUCTION

Sentiment analysis (SA) is the computational examination of the opinion, attitudes, emotions of speaker/writer towards some topic and identification of non- trivial, subjective information from text repository. This field was known as opinion mining, point of view and subjectivity before the term sentiment analysis came into being [1].

Now these days, SA is rapidly growing field due to the rise of online massage sharing platform such as blogs, social media and commercial website. Billions of people share their experiences, knowledge and views on recent trend of politics, economics and other global- critical issue on daily basis. Recently Sentiment Analysis, subjectivity and Opinion mining engrossed momentous interest from both the research community and Marketing Agency [2]. The main goal of sentiment analysis is to classify the opinion according to its polarity level positive, negative, or neutral [3].

Sentiment analysis has extensive applications, ranging from analyzing product review [4] for enhancing sales and improving marketing strategies, forecast stock market fluctuations [5] , identifying ideological shifts in political issue [6], predict movie review [7],and Government Electronic-rule making [8] i.e. citizen opinions on a law before its approval. Even if the enormous amount of work has been carried out in field of sentiment analysis still there is some open challenges related to the multilingual SA , classify the sentence with slangs, symbol, misspelled word and Idioms , Sarcastic sentences SA and handle negation and identify polarity score in negative sentiments [3].

This paper summarized the effect of negation cues over sentiment analysis and present a comparative analysis of recent negation cues and scope detection technique. This paper present a framework securitizing and preprocessed social media data set and formulate the supervised classification technique for negative sentiment analysis. Evaluate comparative analysis of negation cues and scope detection technique with supervised sentiment classification approach and yield intersecting facts about the capabilities and deficiency of negation cues and scope detection technique.

The rest of the paper is organized as follows: Sect. 2 presents over view of Negative sentiment analysis; Sect. 3 covers negation cues and scope detection technique Sect. 4 covers related

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

work on negation handle mechanism for sentiment analysis and polarity detection over social media data set. Sects. 5 present a framework for securitizing and preprocessed social media data set and 5.1-5.3 explain how social media data are processed, step for prepossessing, negation cue and scope technique for efficient SA and experimental Contents for performance evaluation respectively. Section 6 describe the experimental setup for comparative evaluation of different scope detection technique with classification approach for sentiment analysis over social media and finally, Sect. 7 concludes the paper and outlines the founding and future work.

2. NEGATION SENTIMENT ANALYSIS

Negation is a linguistic phenomenon that acts as polarity influence that can changes the semantic of sentence and changes the polarity of sentence from positive to negative or swing the polarity strength. As a result, the essential treatments for negation in SA are required. Cambria and Hussain [9] state that negation is a complex phenomenon that studied under different disciplines. In NLP, negation treats as operator and scope is a principle feature of operators, i.e. negation influence the meaning of other phase of the sentence within their scope. Negations can not only inverting the meaning of single words or phase of words but also reduce the polarities of opinionated word. For example consider following sentences S_1, S_2 and S_3.

Sentence (S_1):- *"The battery life was not long but it's ok for me."*

Sentence (S_2):- *"this cell phone doesn't have a nice camera, voice quality."*

Sentence (S_3):- *"This cell phone is less relevant for selfie"*

Where in Sentence S_1, scope of negation "Not" is only limited to the next word after negation i.e. "Long". Where negation only invert the meaning of word "Long". Whereas in Sentence S_2, Scope of negation "not" is till the end of sentence. On other hand in sentence S_3 use diminisher "Less" to reduce the polarities of opinionated words instead of completely revers the polarities.

Method to handle negation in sentiment analysis is depend upon type of negative linguistic patterns and class negative word used in respective negative sentence as shown in figure 1 and table 1. Table 1 contain list of negations which serve as an indicator of the presence of a negation with different linguistic patterns.

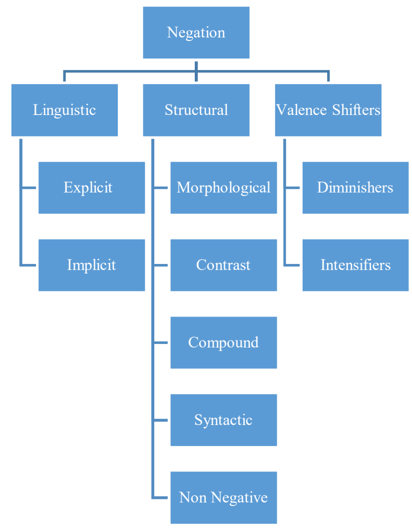

Figure 1: Classification of Negative Cues

Table 1: Negation Words

Negation Class	Negative Word
Explicit Negation	no, not, rather, never, none, nobody, nothing, neither, nor, nowhere (in all tense class)
Implicit Negation	scarcely, hardly, few, seldom, little, forget, fail, doubt, deny and etc.
Diminisher	hardly, few, , little, less
Syntactic	no, not, rather, never, none, nobody, nothing, neither, nor, nowhere (in all tense class)
Morphological	Prefixes: de-, dis-, il-, im-, in-, ir-, mis-, non-, un- , Suffix: -less
Intensifiers	Absolute, badly, biggest, epic, specially, eternally, exceptionally, extremely, freak in, fuckin, hella, huge, incredibly, major, massive, mighty, most, deadly, ever, really, ridiculous, significant, So, such, super, truly, ultimate, undoubtedly, very

Depending upon assertion linguistic patterns, negation in negative sentences may be occur explicitly (with explicit clues such as not, no etc.)

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and implicitly (with implicit clues such as scarcely, hardly, few, seldom, little, only and etc.).For expressing the negative opinion, if negation encoded opinionated word has been used then its implicit negation whereas if standard negation cues are used with opinionated word then it is explicit negation. The list of explicit and implicit negation cue are listed in table 1.For example consider following sentences S_4, S_5 and S_6.

Sentence (S_4):- *"This cell phone is not good"*

Sentence (S_5):- *"My personal experience to use this cell phone is horrible."*

Sentence (S_6):- *"Sound system of this cell phone is superb, I'm suffering from headache after enjoying the song!!!"*

Sentence (S_7):- *"This cell phone is irrelevant for selfie!!!"*

For instance Sentence S_4 have explicit negative sentiment about the cell phone whereas sentence S_5 use "horrible" as opinionated that encode negative sentiment about cell phone. On other hand sentence S_6 use irony to reflect its negative sentiment about the respective product. Whereas at structural level negative sentence may be appear with morphological, syntactic, contrast, compound and non-negative negations. In Morphological negation, negative meaning is carried out by modifying opinionated word either by prefix (e.g. ir-, non-, un- etc.) or suffix (e.g. -less). Whereas in Syntactic negations, explicit negation cues are used to revise the polarity of a single opinionated word or a sequence of words. For instance sentence S_7 use morphological negation to show its negative concern about the cell phone. Whereas S_1, S_2, S_4 and S_5 syntactic negative sentence.

In contrast negation, negative expression show contrast or manage opposition between opinionated terms. While compound negation express comparison or inequality between opinionated term. Whereas in non-negative negation that's used for interrogative and conditional sentences, negative cues and opinionated term may not contain any opinion or sentiment. For instance, sentence S_8, S_9 and S_{10} shows contrast, compound and non-negative negation respectively.

Contrast negation (S_8):- *"I brought this cell phone not for MP3 but for its camera resolution"*

Compound negation (S_9):- *"Camera resolution of cell phone is not better than other."*

Non-negative Negation (S_{10}):- *"Is camera quality of this cell phone is not good?"*

Intensifier and diminisher phase of word use as valance shifter in negation. Valance shifter usually degrade or upgrade the polarity strength instead of inverting the polarity of opinionated word. For instance, sentence S_{11} and S_{12} shows intensifier and diminisher based valance shifter in negation. Where the term "Very much" in sentence S_{11} degrade the negative polarity orientated by the term "not relevant" while the term "less" used in S_{12} shift positive polarity of front camera towards little bit negative.

Negative Intensifier (S_{11}):- *"This cell phone is not very much relevant for me."*

Negative Diminisher (S_{12}):- *"Front camera of this cell phone is less relevant for selife"*

3. NEGATIVE CUES & SCOPE DETECTION

Since negation can change or invert the polarity of opinionated term so it's very important to identification of Negative term for quantifying sentiment polarity. For example in the review 1 "it is nice phone", "nice" bears a positive polarity about phone but in other review such as review 2 "it is not a nice phone" bears a negative sentiment about phone. Christopher Potts from Stanford University, Linguistics Department [10] state that opinionated words behave contrarily when comes with negation and provided a list of negation term that change or invert the polarity of opinionated word. But it is not worth every time for example if negation word is a part of phrase such as "not only", "not wholly", "not all", "not just", "not quite", "not least", "no question", "not to mention" and "no wonder" then it is not to be contemplate as negation.

Handel Negation in sentiment analysis are comprises of two main sub task namely negative cues and scope detection. Negative cues detection is focus to identify the specific phase of words or term that responsible for negation in sentences. Generally sentiment analysis system identify the negation cues by keywords matching from pre-define list of negation words as shown in table 1 or any other subjective corpus. AL-Sharuee [11] employed a predefined list of negation and customized an antonym dictionary to exchange adjectives and adverbs comes after negation terms with their antonym. For example, in the review 2 after removing the negation term "not" the word "nice" is changed to "bad".

Whereas aim of Scope detection is to figure out the influence of linguistic coverage of cues in the sentence and examine how and how much negation cues affect the polarity of opinionated word that are within its scope. Possibly the effect of negation sentiment analysis was first analysis by Das & Chen (2001) [12] over financial domain for market analysis and developed one system for cleaning and normalization the negative finical comment by adding the term "NOT" in place of negation cues. Recently number of researcher has use this technique to handle negation for sentiment analysis in political issues, market analysis of new product [13, 14], stock market analysis [15, 16], online learning [17-19] and other.

In the same ways Pang et al. [20] handle the negation but consider the scope of negation from negation cues to next punctuation mark. Consequently, punctuation marks like quotation mark either single or double, colon, semi colon, comma, braces, exclamation and question mark has been used by number of researcher [21, 22] for negation scope detection. For example consider the sentence's S_{14} and S_{15} where scope of negation "not" in both the sentences limited to next punctuation mark and negation only invert the polarity of opinionated word superb and grown respectively. But there is one exceptional case with punctuation mark comma i.e. whenever comma is used to separate the phase of sentences in place of "and" and "or" then similar pattern of sentiment is flow in either side. For example consider the sentences S_{16} where scope of negation "don't" lies beyond the punctuation mark comma.

Sentence (S_{14}):- *"Sound system of this cell phone is not superb! But I am happy with its battery life."*

Sentence (S_{15}):- *"The production has not grown, but we are hopeful for the future."*

Sentence (S_{16}):- *"this cell phone don't having good battery life, nice sound system and touch screen."*

Apart from punctuation mark, conjunction word [23-26] is also being used for negation scope detection. Conjunction word generally used to change the sentiment orientation of the sentence before and after the contrary word like but, however, and, or, whereas unless and etc. for example "I don't like this cell phone but battery life is superb." where but clause significantly manage deviation of sentiment form first part of sentence to second part [27]. Whereas in some cases conjunction word `and' and `or' allow to extend the scope of negation appear in first phase beyond its occurrence. For example "I don't like power backup and sound system or music system of this cell phone" where "and" and "or" allow to flow the sentiment orientation of opinionated word like beyond its occurrence.

Recently number of researcher has used supervised syntactic rule based approach with static window size, grammatical dependency and semantic rule with POS and N-gram technique for scope detection. Morante [28] developed a bioscope corpus of bio medical text on behalf of handicraft rule based on grammatical dependency graph and keyword matching. Bioscope corpus provide a semantic relation between lists of negative cues with their respective scopes. Apart from that Prllochs et al. [29] use probabilistic Hidden marov model , Nakagawa et al. [30] developed a semi-supervised model for building handicraft rule for grammatical dependency graph. Whereas Jimenez-Zafra et al. [31] use iSOL lexicon for dependency tree formation.

4. RELATED WORK

Negation handling in sentiment analysis involving the identification the term indicate negation i.e. cue detection and identify linguistic coverage of the impact of the cue i.e. scope detection. Negation scope detection by cues grammatical structure based supervised syntactic rule use negative seed words for training purpose [32-35]. M. Ghiassi [36] applied supervised inference rules for polarity designation and feature are tagged with four different polarity indicator ie "XP" (extremely positive), "VP" (very positive), "SP" (somewhat positive), "SN" (somewhat negative), "VN" (very negative) or "XN" (extremely negative) by using information gain. Whereas Orestes Appel [21] has developed a fuzzy set theory using Naive Bayes and Maximum Entropy classification techniques that estimate sentiment polity and its intensity. This technique divide the Sentiment Oriented Word from the Mild to the Intensive, in five polarity parts, primarily Poorly slight , Moderate, very and Most intensive sentiment word.

Vicente García-Díaz [37] present probabilistic D-social Platform that highlight the negativity of expressions found in a text by using Naïve Bayes classifier. Ioannis Korkontzelos [22] has use part of speech (POS) technique to identify grammatical

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dependency relation between negation cue and opinionated word for analyzing adverse drug reaction. Claudia Diamantini [38] has constructs the dependency-based parse tree after analysis grammatical structure and clauses separation of sentence by using statistical dependency parser for identify of negation cues through a depth-first search (DFS) strategy. Tian Kang[39] has tackle negation cues and scope as sequence labeling problems and use Conditional Random Fields (CRF) for 'BIO' tagging to represent the boundaries of negation cues. Nicolas Pröllochs [29] has predict negation scopes by two-sided approach ie reinforcement learning and machine learning approaches for balanced classification by manually labeled dataset.

Polarity shift through negative cues affect sentiment classification accuracy. Recent research has been focus over Statistical techniques to distinguish explicit and implicit polarity shifts estimation. RuiXia [40] has employed rule-based method to detect polarity shifts in explicit negations and contrasts, and a statistical method to detect some implicit polarity shifts such as sentiment inconsistencies. Eric S. Telz [41] pointed out the disclaimer marker that can swift the polarity of the message. Namely, no, never, and without (without) According to the author, when one of them matches, the remaining disclaimer marker is abandoned. They have followed two types of rules; which depends on the input text. If the text is not parsed by the feeling, by using only information on the text, some rules (regular expressions) are applied to reject the rejection marker to the nearest word, the second approach is to take advantage of explanatory information. Uses a set of rules, about 50 rules were designed to consider the disclaimer markers. The disclaimer marker is linked to the prohibited marker with the closest word. Negation marker are attached to content words for example in sentence "The approach is neither nice nor reasonable", the negation marker "No" is attached to its two adjective and sentence become "The approach is not nice not reasonable".

M. Ghiassi [36] has worked on the valence shifter, mainly intensifiers, diminishers, negative and sarcasm. They have divided the negative feature set into four equivalence class using the b-gram technique i.e. horrible, neglect, addicted and deceived. Salud María Jiménez-Zafra [42] has validate the SFU review -NEG corpus for the supervised polarity classification system and add evaluation of effect of negative cues wither its increase or decrease the polarity opinioned word within its scope before final polarity classification in existing three phase polarity classification task. Asad Abdi [27] handle the negation and contrasts oriented word in multi document summarization using dictionary based approach. Murtadha Talib AL-Sharuee [11] handling intensifiers and negation using SentiWordNet and use antonym dictionary to replace adjectives and adverbs that follow negation terms with their opposite sentiment words. For example, after removing the negation term, the word "not like" is changed to "hate".

5. COMPARATIVE ANALYSIS

Comparative analysis of recent scope detection and classification technique in Negative Sentiment analysis are present interesting and useful facts regarding the state-of-the-art of sentiment analysis in presence of negation. This paper present a three tier framework for comparing the performance negative scope detection technique collaborated with supervised classifier followed by prepossessing of social media data set for negative sentiment analysis as shown in figure 2.

5.1 Preprocessing For Sentiment Analysis

Social media data set used for sentiment analysis is the collection of tweets or comment that has been posted by social media user. Social media user use domain specific slag language, emoticons, symbols, idioms and sarcastic sentences to post their tweets.

Presented framework explored the unique properties social media data and try to refine by sentence splitting, slag replacement, word normalization and negation control preprocessing step for better sentiment classification.

5.1.1 Sentence Splitting: - Sentence splitting preprocessing phase is used to magnified review document or comment into sentence level by splitting the review before and after the sentence delimiter (".", "?", "!") . For example consider the review of home theater system posted by person "P_1".

Home theater- User Review "P_1"

"I bought a 5.1 speaker home theater system of this company last year. D sound system is really awesome. I m loooooving it."

Sentence splitting phase split the review P_1 into three different sentence as sentence S_1, S_2 and S_3 .

S_1:- *"I bought a 5.1 speaker home theater system of this company last year."*

S_2:- *"D sound system is really awesome."*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

***S_3:-** "I m loooooving it"*

5.1.2 Slag Replacement: - Slag replacement replaced the slag word by keywords matching from corpus of pre-define list of frequently used semantic and emoticons. For example consider the unprocessed comment S_{16} and S_{17} where tokens "*D* " and "m" are compared to entries in slag corpus and return processed comment S_{18} and S_{19} with token "The" and "am".

***Unprocessed comment S_{16}:-** "D sound system is really awesome."*

***Unprocessed comment S_{17}:-** "I m loooooving it"*

***Processed comment S_{18}:-** "The sound system is really awesome."*

***Processed comment S_{19}:-** "I am loooooving it"*

5.1.3 Word Normalization: - For word normalization this framework used Rogets Thesaurus corpus and match the phase of words with entries in Rogets Thesaurus. If it's not matched, repeated letters are sub sequentially compact until it's not matched. For example consider the unprocessed comment S_{19} where the token "loooooving" are compared to entries in Rogets thesaurus and return refine one i.e. "loving" with processed comment S_{20}.

***Processed comment S_{20}:-** "I am loving it"*

5.2 Negation Control Mechanism

For handle the negation in sentiment analysis this framework use conjunction analysis, punctuation mark identification and grammatical dependency tree scope detection technique followed by negation cues detection.

5.2.1 Negation Cues Detection: - Negation cues are the term that indicate negation in opinion sentences. This framework identify the negation cues by keywords matching from pre-define list of negation words or corpus of negation words. Identified negation cues are replaced by token "NEGATION" and processed comment forwarded for scope detection. For example consider the sentences S_{21} & S_{22}.

***Sentence (S_{21}):-** "Sound system of this cell phone is not superb but I am happy with its battery life"*

***Sentence (S_{22}):-** "The production has not grown, we are hopeful for the future"*

5.2.2 Negation Scope detection: - Scope detection identify the linguistic coverage of the impact of negation cues in opinion sentences. This frame work use conjunction analysis, punctuation mark identification and grammatical dependency tree scope detection technique to identify the linguistic impact of "NEGATION" token provided by its successor negation cues detection phase.

A. **Conjunction Analysis: -** Conjunction words like "but", "expect", "however", "whereas"," although", "and", "or", "unless", "nevertheless" and etc. are limited the influence of opinionated word occur before and after its appearances in sentences. For example consider the sentence S_{21} where reviewer has different opinion about different aspect of cell phone. Reviewer have negative opinion about sound system but positive opinion about battery life. In sentence S_{21} conjunction clause "but" help to configure the scope of two opposite sentiment orientated opinionated word "not superb" and "happy" before and after its appearance.

B. **Punctuation Mark Identification: -** Punctuation Mark (",", "!", ";") is also being used for demarcation of scope of negation between a negation cue and the next punctuation mark. For example consider the sentence S_{22} where reviewer has negative sentiment about current production but he hopeful for future. Where comma "," is use to separate out these two different aspect about the production. Negation "Not" only influence the polarity of word "grown".

C. **Grammatical Dependency Tree:-** Grammatical dependency of among opinionated word and negation with respect to their order of appearances help to handle negation and word sense disambiguation [38].Negations cues are identified by examine the grammatical dependency relationship among negation words (e.g., no, not, any and less) and opinionated word. Whereas dependency parser is consider to determine scope of negation [43] with either POS [44, 45] or N-gram [35, 37, 39, 41, 42, 46]. POS and N-gram trigger provides grammatical marking (as noun, verb, adjective, adverb, coordinating conjunction etc.) and lowest level of grammatical syntactic relationship to determine scope of negation.

Journal of Theoretical and Applied Information Technology
31st March 2019. Vol.97. No 6
© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

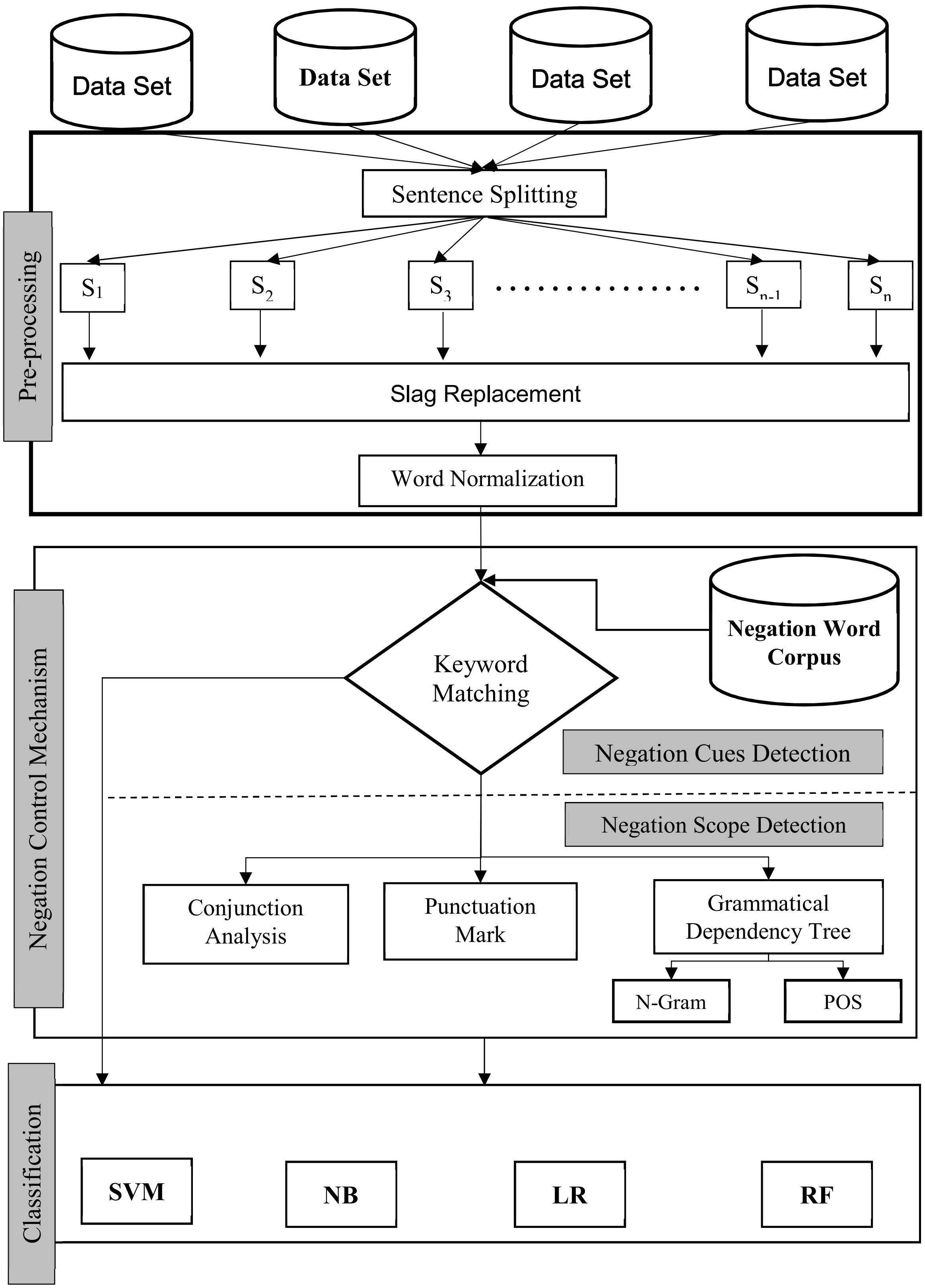

Figure 2:-Feature Extraction and Preprocessing

In grammatical semantic relation negation word (i) if modify polarity of adjective then describes it as and noun (ii) if modify the polarity of verb then describes it as noun and (iii) if modify the polarity of adverb then describe it as adjective or as verb.

5.3 CLASSIFICATION

After examine the scope of negation present in negative sentences classification technique classifies relevant data set into three different opinion class as positive , negative and Neural. This paper evaluate the performance of Classifiers SVM, Naives Bayes, Linear regression and random Forest with different scope detection based on conjunction word , punctuation mark and grammatical dependency tree

5.3.1 Support Vector Machine: - Support vector machine maximize the margin of separator hyper plane to classify the social media data set into positive and negative sentiment polarity classes [3]. Whereas for incorporating scope of negation as a relevant feature for negative sentence sentiment analysis, SVM treat all the token in scope as negative vector space as shown in equation.

$$W_{sv}^{n} = \{(w_i^n, w_{sc})n_t \in t_d\} \ldots 1$$

Where

- t_d is text data set,
- W_{sv}^n is negative vector space
- w^i_n is the word in negative scope and
- W_{sc} is the set of word in scope of negation
- n_t is negative token.

And finally create an optimal hyper plane that refine the marginal width of hyper plane as shown in equation 2, 3.

$$(W_{sv}^{P} + Positive) \geq P_{SV} \; \forall \; positive\ sentiment\ over\ the\ positive\ hyper\ plane \ldots \ldots ..2$$

$$(W_{sv}^{n} + negative) \geq N_{SV} \; \forall \; negative\ sentiment\ over\ the\ negative\ hyper\ plane \ldots \ldots ..3$$

5.3.2 Naive Bayes: - Naïve bayes return the sentiment polarity (positive, negative) of any sentence S on the basis of maximum posterior probability as shown in equation 4 and 5 [3].

$$S_P = \underset{p\epsilon(Positive,Negative)}{argmax} \quad P(p|s) \ldots \ldots \ldots \ldots \ldots ..4$$

$$P(p|s) = \frac{P(s|p)P(p)}{P(s)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots....5$$

Where P (p|s) is final posterior probability and P(s|p) is the probability of sentence S belong to polarity class ‘p’. Whereas P(p) and P(s) is the independent probability polarity class ‘p’ and sentence ‘s’.

Whereas for incorporating scope of negation as a relevant feature for negative sentence sentiment analysis, NB treat all the token in scope as independent probability entity as shown in equation 6.

$$P(n|\mathrm{scope}) = P(\mathrm{n}|\mathrm{sw}_1) * P(\mathrm{n}|\mathrm{sw}_2) * P(\mathrm{n}|\mathrm{sw}_3) * \ldots \ldots .* P(\mathrm{n}|\mathrm{sw}_n) \ldots \ldots \ldots \ldots ..6$$

Where P (n| scope) are independent given the polarity Class (P) and each word in scope of negation substitute their individual probability for exploring polarity classes.

5.3.3 Random forest

Random forest predict the sentiment polarity class of sentence S by building randomized regression trees $\{r_n(x, pc, ds), m \geq 1\}$ based relationship between polarity class and sentences as shown in equation 7.

$$\overline{r_n}(x, ds) = E_{pc}[r_n(x, pc, ds)] \ldots\ldots\ldots.7$$

Where E_{pc} is exception on polarity class (pc) classification with random parameter (r) on condition x and data set (ds). Whereas incorporation of scope of negation as conditional parameter X lead to minimized exception (Epc) on polarity class and increase classification rate.

5.3.4 Linear Regression: - Linear function separate positive and negative sentiment oriented sentences into two different classes by finding a decision boundary that linearly separate the data set as shown in equation 8. Where P passing the polarity function C*x through the threshold function as shown in equation 9.

$$P_C(x) = \begin{cases} +ve\ ; if\ C * x \geq 0\ (Polarity\ score) \\ -ve\ ; if\ C * x < 0 (Polarity\ Score) \end{cases} \ldots 8$$

$$P_c(x) = threshold\ C * x \ldots \ldots \ldots \ldots \ldots \ldots 9$$

6. ENVIRONMENT SETUP RESULT ANALYSIS

For Comparative analysis of recent negation scope detection and classification technique five different experimental campaigns over two different source of data set have been carried out.

Table 2:- Data Set Description

Reference	Platform	Data set Name	Total number of tweets/Review	Positive (tweets/Review)	Negative (tweets/Review)
47	Twitter	Stanford data set	1600000	80000	80000
48		Sanders Twitter Sentiment Corpus	1224	570	654
49	Amazon	Smartphone Review	17500	12500	5000
49		Movie Review	35000	30000	5000
49		Book Review	90000	81000	9000

Table 3:- Comparative Analysis of Sentiment Analysis Technique

Classification Technique	Twitter API		Amazon		
	Twitter Sentiment Corpus Data set	Stanford Data Set	Smart Phone Review	Movie review	Book Review
SVM	53.77	48.45	55.67	61.46	65.68
Nave Bayes	49.71	45.23	53.78	60.85	64.66
Random Forest	54.51	51.37	52.67	58.68	63.22
Linear Regression	55.12	52.23	50.98	58.42	64.42
SVM + CW	67.23	69.56	63.44	63.24	67.82
Nave Bayes+ CW	67.89	68.78	62.56	62.68	66.60
Random Forest + CW	64.67	65.78	59.66	60.20	65.28
Linear Regression+ CW	65.12	66.54	56.67	60.48	64.76
SVM + PM	73.78	75.78	68.67	65.12	69.20
Nave Bayes+ PM	73.34	76.89	69.68	64.88	68.88
Random Forest + PM	69.87	73.56	62.84	61.42	66.46
Linear Regression+ PM	70.32	72.65	60.56	62.18	67.48
SVM + GDT	79.65	80.89	72.82	66.78	70.68
Nave Bayes+ GDT	78.98	81.76	70.45	65.66	70.22
Random Forest + GDT	75.45	78.67	68.20	62.24	69.44
Linear Regression+ GDT	74.34	77.89	66.86	63.44	68.26

First two campaigns has been carried out over twitter data set i.e. Stanford data set (TCD) [47] and Sanders Twitter Sentiment Corpus (TSD) [48] that scraped by twitter API. Stanford data set contain 160000 training tweets accompanied by 80000 both positive and negative tweets. Whereas Sanders Twitter Sentiment data set contain 570 positive and 654 negative tweets. However last three campaigns has been carried out over amazon online product reviews data set of smartphone (AS), movies (AM) and book (AB) [49]. Detail description of data set composition is summarized in table 2.

Performance evaluation of sentiment classification technique with and without negation control are described in Table 3. Accuracy of classifier has been increased after incorporating negation control over negative sentiment tweets or reviews.

The baseline classifier (SVM, Nave Bayes, Random Forest and Linear Regression) without negation scope control can yield approximate 45%-55 % and 50% - 65 % accuracy rate over twitter and Amazon data set respectively as shown in figure 3 . Where linear regression yield better performance and lead by approximate 1.4 % improvement over twitter data set. However over amazon data set SVM yield better performance and lead by approximate 2 % improvement.

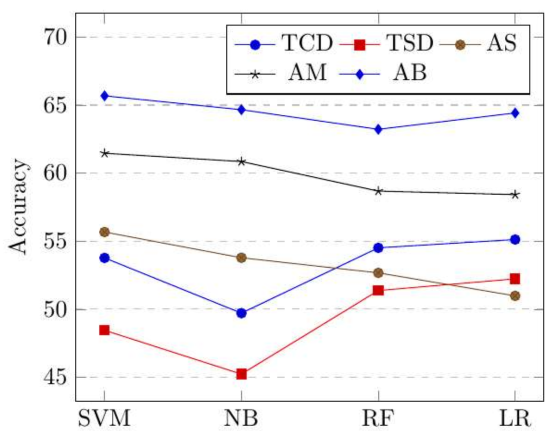

Figure 3: Sentiment Analysis with benchmark algorithm

The performance of baseline classifier is significantly increased after considering feature set of scope of negation for classification. With conjunction word scope detection technique classification rate of classifier increased by approximate 18%-37% and 27% -52% over two different variant twitter data set and 11%- 16 , 2.5%-3.5% and 0.5%-3.2% over all three different variant of amazon data set as shown in figure 4 & 5. Higher improvement in twitter data set is significantly because of higher number of negative twists i.e. approximate 50 % and 53 .43 % in Stanford and twitter corpus data set respectively. Similarly comparatively lower improvement in Amazon data set is because of lower number of negative twists i.e. approximate 28.57 %, 14.28% and 10 % in Smart phone, Movie and Book review data set respectively. With Conjunction word, SVM gives better performance over Twitter Sentiment Corpus, Smart phone, Movie and Book review data set whereas Nave Bayes lead the performance over Stanford Data Set. However highest improvement is gained by Nave Bayes in case of Conjunction word ie approximate 3%-36 % over different variant of data set as shown in figure 5.

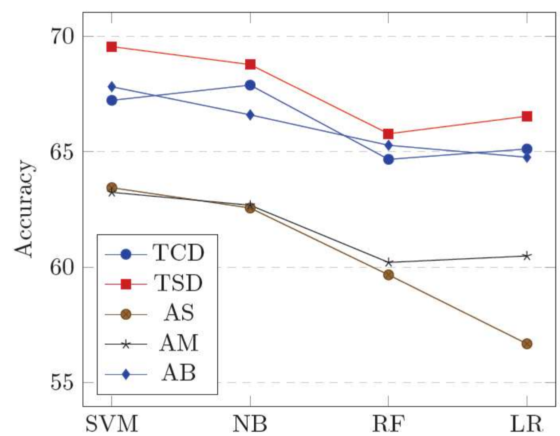

Figure 4:- Accuracy with Conjunction Mark Based Negation Scope detection Technique

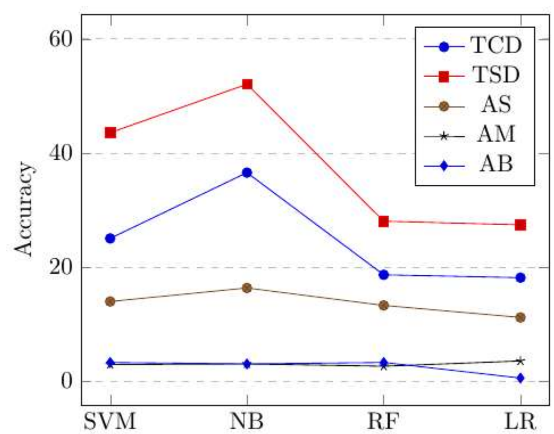

Figure 5:- Improvement with Conjunction Mark Based Negation Scope detection Technique

Punctuation mark enhance classification rate of classifier by approximate 27.6%-47.5% and 39% -70% over two different variant twitter data set and 18%- 30% , 4.5%-6.5% and 4.3%-6.25% over all three different variant of amazon data set as

Journal of Theoretical and Applied Information Technology
31st March 2019. Vol.97. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

shown in figure 6 & 7. SVM gives better performance over Twitter Sentiment Corpus, Movie and Book review data set whereas Nave Bayes lead the performance over Stanford and Smart phone review Data Set with punctuation mark. However highest improvement is gained by Nave Bayes in case of punctuation mark i.e. approximate 6.5%-70 % over different variant of data set as shown in figure 7.

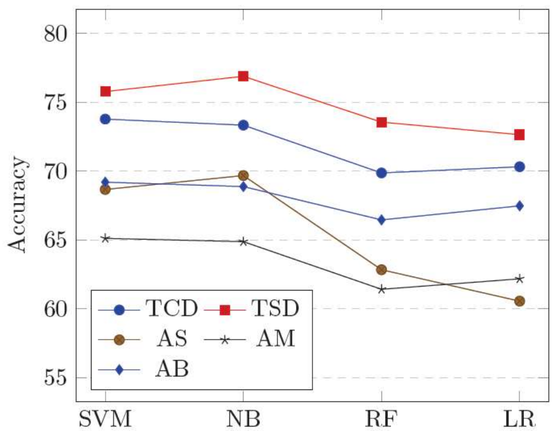

Figure 6:- Accuracy with Punctuation Mark Based Negation Scope detection Technique

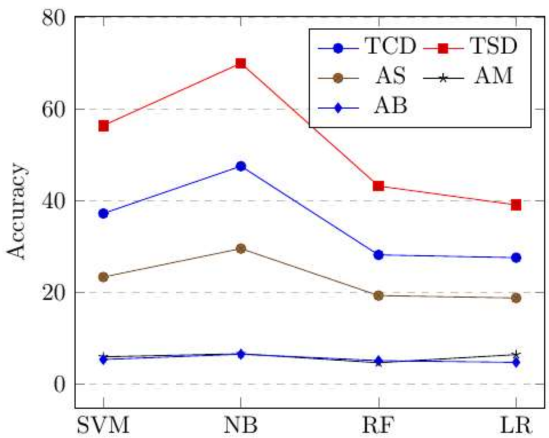

Figure 7:- Improvement with Punctuation Mark Based Negation Scope detection Technique

Grammatical dependency tree enhance classification rate of classifier by approximate 34.86%-58.80% and 49.12% -80.76% over two different variant twitter data set and 29.48%-31.14%, 6.06%-8.59% and 5.96%-9.85% over all three different variant of amazon data set as shown in figure 8 & 9.. SVM gives better performance over Twitter Sentiment Corpus, Smart phone, Movie and Book review data set whereas Nave Bayes lead the performance over Stanford Data Set with Grammatical dependency tree. However in case of GDT highest improvement is gained by Nave Bayes i.e. approximate 58.88% and 80.70% over different variant of twitter data set, linear regression i.e. approximate 31.14% and 8.59% over smart phone and movie review data set and random forest i.e. approximate 9.83% over Book review data set as shown in figure 9.

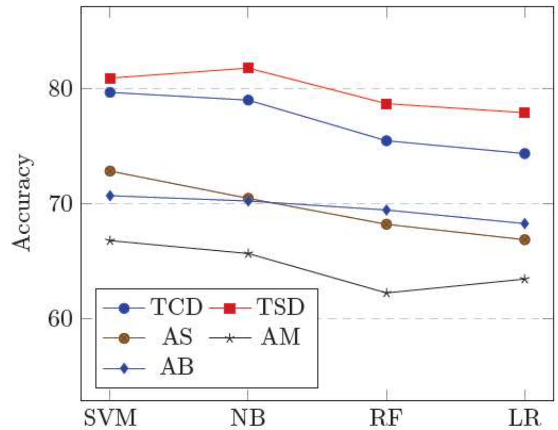

Figure 8:- Accuracy with Grammatical dependency tree Based Negation Scope detection Technique

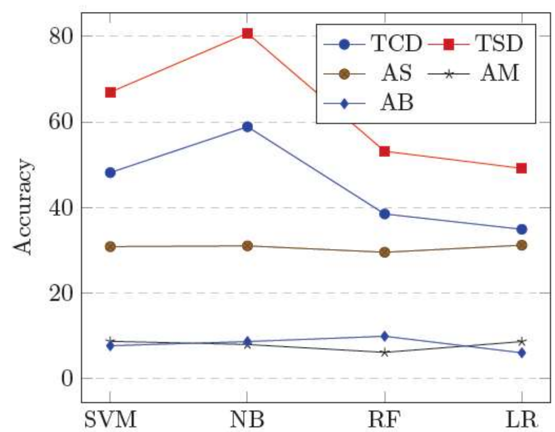

Figure 9:- Improvement with Grammatical dependency tree Based Negation Scope detection Technique

With different prospective of analyzing the performance of classifier for negative SA with scope detection techniques i.e. conjunction word, punctuation mark and Grammatical dependency tree. It is observed that classifiers gives better performance with GDT scope detection technique. SVM gain 63.24%-69.56%, 65.12%-75.78% and 66.78%-80.89% accuracy with CW, PM and GDT respectively over different variant of data set as shown in figure 10. SVM achieved highest improvement with GDT i.e. approximate 7.61% -

Journal of Theoretical and Applied Information Technology
31st March 2019. Vol.97. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

66.95% over different variant of data set as shown in figure 11.

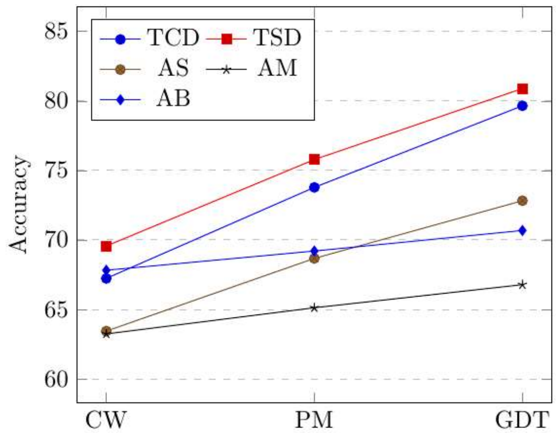

Figure 10:- Accuracy of SVM Sentiment Classifier after Scope detection Technique

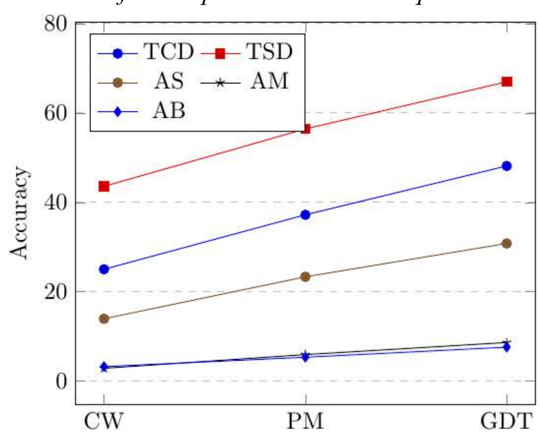

Figure 11:- Improvement over SVM Sentiment Classifier after Scope detection Technique

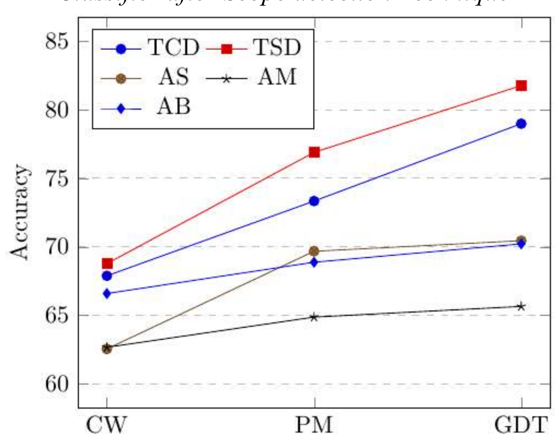

Figure 12:- Accuracy of NB Sentiment Classifier after Scope detection Technique

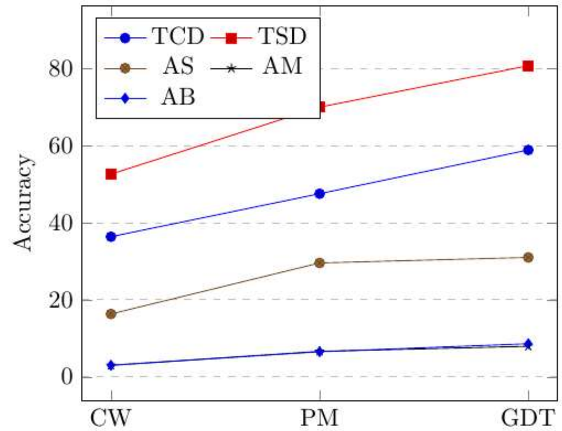

Figure 13:- Improvement over NB Sentiment Classifier after Scope detection Technique

Naïve Bayes gain 62.56%-68.78%, 64.88%-76.89% and 65.66%-81.76% accuracy with CW, PM and GDT respectively over different variant of data set as shown in figure 12. Naïve Bayes achieved highest improvement with GDT i.e. approximate 7.90% -80.76% over different variant of data set as shown in figure 13.

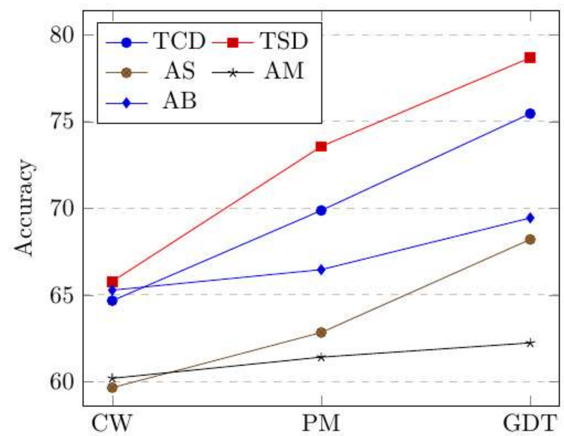

Figure 14:- Accuracy of Random Forest Sentiment Classifier after Scope detection Technique

Whereas Random Forest gain 59.66%-65.78%, 61.42%-73.56% and 62.24%-78.67% accuracy with CW, PM and GDT respectively over different variant of data set as shown in figure 14. Random Forest achieved highest improvement with GDT i.e. approximate 6.06% - 53.14% over different variant of data set as shown in figure 15.

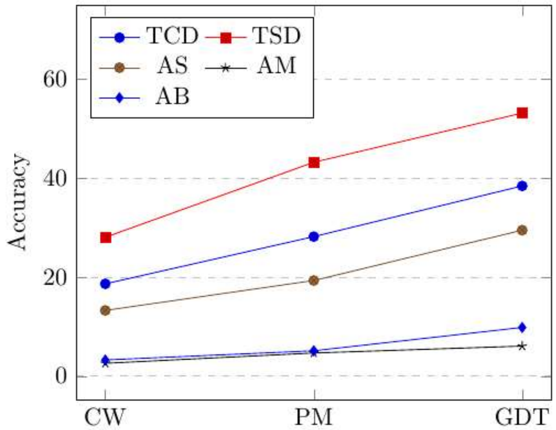

Figure 15:- Improvement over Random Forest Sentiment Classifier after Scope detection Technique

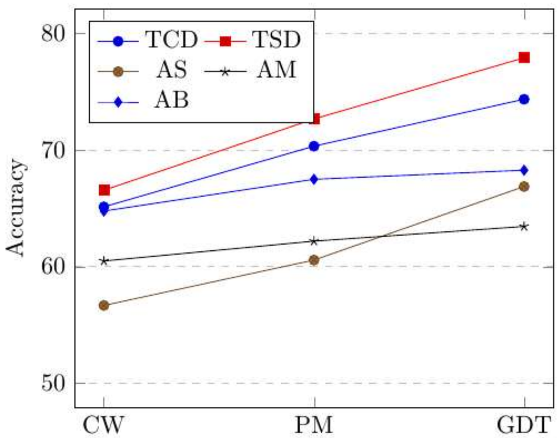

Figure 16:- Accuracy of Linear Regression Sentiment Classifier after Scope detection Technique

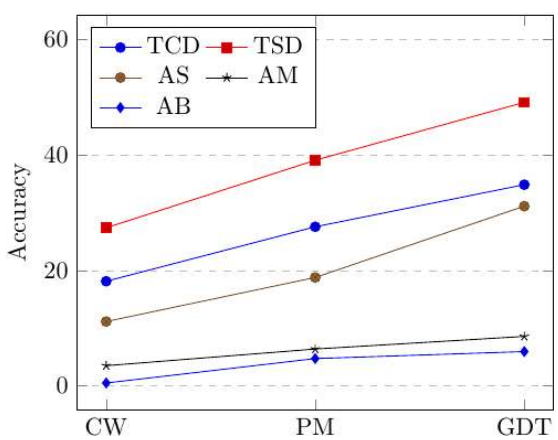

Figure 17:- Improvement over Linear Regression Sentiment Classifier after Scope detection Technique

However Linear Regression gain 56.67%-66.54%, 60.56%-72.65% and 63.44%-77.89% accuracy with CW, PM and GDT respectively over different variant of data set as shown in figure 16. Linear regression achieved highest improvement with GDT i.e. approximate 5.96% -49.12% over different variant of data set as shown in figure 17.

After evaluating the performance baseline sentiment classifier with feature set of scope detection technique following outcome has been acquired. GDT is best suited Scope detection technique to identify the range of influence marked by negation for negative sentiment Analysis. Whereas CW and PM gives biased result. SVM is best suited sentiment classification approach under negation whereas naïve Bayes achieved highest improvement after encapsulating scope detection with classification.

7. CONCLUSION

This paper incorporate a comparative analysis to evaluate the performance of recent negation cue and scope detection technique. Negation cue and scope detection technique generate negation cue feature vector that significantly improve the calculation of polarity score of review comment. The comparison of scope detection technique is carried out to identify the influence of negation. Perhaps it either invert or swift the polarity value of opinionated word within its scope. This paper present a framework to analysis the performance of recent negation cue and scope detection technique over social media data set. Social media data set contain the comment or review of end user regarding any specific topic globally, that may contain noise , misspelled and Slag sentences that need to be preprocess before proceeded to any operation. This framework initially preprocessed social media data set to handle noise, misspelled sentences and slag languages. And finally classify the review sentence according to their polarity value after incorporating scope of negation if negation cue are presented.

For the scope detection GDT is best suited as describe the grammatical dependency of negation cue with rest of term in sentences and build a tree for scope recognition. GDT improve the performance of SVM by approximate 7.61% to 66.95%, NB by approximate 7.90% -80.76%, RF by approximate 6.06% -53.14% and LR by approximate 5.96% -49.12% over different variant of data set. Whereas CW and PM having some exceptional condition i.e. `or' and `and' conjunction word and comma `,' punctuation mark allow to extend the scope of negation to next clause in some situation. However

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

this paper also evaluate the performance of supervised classification technique in presence of negation cues over social media data set. It is observed that SVM is best suited sentiment classification approach under negation whereas naïve Bayes achieved highest improvement after encapsulating negation cue feature vector for classification.

REFERENCE

[1] Gilad Katz, Nir Ofek, Bracha Shapira, "ConSent: Context-based sentiment analysis", *Knowledge-Based Systems*, Vol. 84, 2015, pp. 162-178.

[2] Farhan Hassan Khan, Usman Qamar, Saba Bashir, SWIMS: Semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis", *Knowledge-Based Systems*, Vol. 100, 2016, pp. 97-111.

[3] Nikhil Kumar Singh, Deepak Singh Tomar, Arun Kumar Sangaiah, "Sentiment analysis: a review and comparative analysis over social media" *Journal of Ambient Intelligence and Humanized Computing* , 2018, pp. 1-21

[4] E. A. Stepanov and G. Riccardi, "Detecting General Opinions from Customer Surveys," *IEEE 11th International Conference on Data Mining Workshops*, Vancouver, BC (Canada), December 11, 2011, pp. 115-122.

[5] Liang-Chih Yu, Jheng-Long Wu, Pei-Chann Chang, Hsuan-Shou Chu, "Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news" , *Knowledge-Based Systems*, Vol. 41, 2013, pp. 89-97.

[6] Ronen Feldman, "Techniques and applications for sentiment analysis", *Communications of the ACM*, Vol. 56, Issue 4, 2013, pp. 82-89

[7] Mishne, Gilad, Glance, Natalie, "Predicting Movie Sales from Blogger Sentiment", *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs, Standford* , California (USA),March 27-29, 2006, pp. 155-158.

[8] Cardie, Claire, Farina, Cynthia R., Bruce, Thomas R., "Using Natural Language Processing to Improve eRulemaking [Project Highlight]" *7th Annual International Conference On Digital Government Research*, Digital Government Society ,(North America), May 21-24, 2006,pp. 177-178

[9] Cambria, Amir Hussain, "Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis (1st ed.)" *Springer Publishing Company, Incorporated,*2015

[10] http://sentiment.christopherpotts.net/ , Potts, C. (2011), "Sentiment Symposium Tutorial", *Stanford Department of Linguistics, Stanford University*. Access date 10/10/2018.

[11] Murtadha Talib AL-Sharuee, Fei Liu, Mahardhika Pratama, "Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison", *Data & Knowledge Engineering*, Vol. 115, 2018, pp. 194-213.

[12] Sanjiv Das, Mike Chen, "Yahoo! for Amazon: extracting market sentiment from stock message boards", *Management Science,* Vol.539, 2001.

[13] A. Montejo-Ráez, M.C. Díaz-Galiano, F. Martínez-Santiago, L.A. Ureña-López, "Crowd explicit sentiment analysis", *Knowledge-Based Systems*, Vol. 69, 2014, pp.134-139.

[14] Raymond Y.K. Lau, Chunping Li, Stephen S.Y. Liao, "Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis", *Decision Support Systems*, Vol.65, 2014, pp.80-94.

[15] Xiaodong Li, Haoran Xie, Li Chen, Jianping Wang, Xiaotie Deng, "News impact on stock price return via sentiment analysis", *Knowledge-Based Systems*, Vol. 69, 2014, pp.14-23.

[16] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, David Chek Ling Ngo, "Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment", *Expert Systems with Applications*, Vol. 42, Issue 1, 2015, pp. 306-324

[17] Janez Kranjc, Jasmina Smailovic, Vid Podpecan, Miha Grcar, Martin Znidarsic, Nada Lavrac, "Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the ClowdFlows platform." *Information Processing and*

Management, Vol.52, issue 2, 2015, pp. 187-203.

[18] Alexandra Balahur, José M. Perea-Ortega, “Sentiment analysis system adaptation for multilingual processing: The case of tweets”, *Information Processing & Management*, Vol. 51, Issue 4, 2015, pp. 547-556.

[19] Luigi Di Caro, Matteo Grella, “Sentiment analysis via dependency parsing”, *Computer Standards & Interfaces*, Vol. 35, Issue 5, 2013, pp.442-453.

[20] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan, “Thumbs up?: Sentiment classification using machine learning techniques” . *In Proceedings of the acl-02 conference on empirical methods in natural language processing vol. 10*. Association for Computational Linguistics, Stroudsburg, PA (USA), 2002, pp.79-86.

[21] Orestes Appel, Francisco Chiclana, Jenny Carter and Hamido Fujita, “A hybrid approach to the sentiment analysis problem at the sentence level”, *Knowledge-Based Systems*, Vol.108, 2016, pp.110-124.

[22] Ioannis Korkontzelos, Azadeh Nikfarjam, Matthew Shardlow, Abeed Sarker, Sophia Ananiadou, and Graciela H. Gonzalez, “Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts”, *Journal of Biomedical Informatics*, Vol.62 Issue C, 2016, pp. 148-158.

[23] Arun Meena T. V. Prabhakar, “Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis” *European Conference on Information Retrieval (ECIR):Advances in Information Retrieval,* Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Vol. 4425, 2007, pp. 573-580.

[24] Ritesh Agarwal, T. V. Prabhakar and Sugato Chakrabarty, “I Know What You Feel: Analyzing the Role of Conjunctions in Automatic Sentiment Analysis” *International Conference on Natural Language Processing (GoTAL): Advances in Natural Language Processing*, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, Vol. 5221, 2008, pp. 28-39.

[25] Ramanathan Narayanan, Bing Liu, and Alok Choudhary, “Sentiment analysis of conditional sentences” In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 (EMNLP '09), Association for Computational Linguistics, Stroudsburg, PA (USA), Vol. 1. , 2009, pp.180-189.

[26] Umar Farooq, Antoine Nongaillard, Yacine Ouzrout, and Muhammad Abdul Qadir, Muhammad Abdul, “Product reputation evaluation: the impact of conjunction on sentiment analysis” *Int. Conference on Software, Knowledge, Information Management and Application*, 2013, pp. 242 - 250.

[27] Asad Abdi, Siti Mariyam Shamsuddin, Ramiz M. Aliguliyev, “QMOS: Query-based multi-documents opinion-oriented summarization”, *Information Processing & Management*, Vol. 54, Issue 2, 2018, pp. 318-338.

[28] Roser Morante “Descriptive Analysis of Negation Cues in Biomedical Texts.” *Conference: Proceedings of the International Conference on Language Resources and Evaluation*, Valletta (Malta), May 17-23, 2010, pp. 1429-1436.

[29] Nicolas Pröllochs, Stefan Feuerriegel and Dirk Neumann, “Negation scope detection in sentiment analysis: Decision support for news-driven trading”, *Decision Support Systems*, Vol. 88, 2016, pp.67-75.

[30] Tetsuji Nakagawa, Kentaro Inui, and Sadao Kurohashi, “Dependency tree-based sentiment classification using CRFs with hidden variables”, *In Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (HLT '10),* Association for Computational Linguistics, Stroudsburg, PA (USA), 2010, pp. 786-794.

[31] S. M. Jimenez-Zafra, M. T. Martin Valdivia, E. Martinez Camara and L. A. Urena-Lopez, "Studying the Scope of Negation for Spanish Sentiment Analysis on Twitter," *IEEE Transactions on Affective Computing*, Vol. 14, No. 8, 2015, pp. 1-14.

[32] Wei Li, Kun Guo, Yong Shi, Luyao Zhu and Yuanchun Zheng, “DWWP: Domain-specific new words detection and word propagation system forsentiment analysis in the tourism domain”, Knowledge-Based Systems, Vol.146, 2018, pp.203-214.

[33] Monireh Alsadat Mirtalaie, Omar Khadeer Hussain, Elizabeth Chang, and Farookh Khadeer Hussain, "Extracting sentiment knowledge from pros/cons product reviews: Discovering features along with the polarity strength of their associated opinions", *Expert Systems with Applications*, Vol.114, 2018, pp. 267-288.

[34] Fangzhao Wu, Yangqiu Song, and Yongfeng Huang, "Microblog sentiment classification with heterogeneous sentiment knowledge", *Information Sciences*, Vol. 373, 2016, pp.149-164.

[35] Muhammad Abdul-Mageed, "Modeling Arabic subjectivity and sentiment in lexical space", *Information Processing & Management*, Vol.56, Issue 2, 2019, pp.291-307.

[36] M. Ghiassi, S. Lee, "A domain transferable lexicon set for Twitter sentiment analysis using a supervised machine learning approach", *Expert Systems with Applications*, Vol.106, 2018, pp.197-216.

[37] Vicente Garca-Daz, Jordn Pascual Espada, Rubn Gonzlez Crespo, B. Cristina Pelayo G-Bustelo, and Juan Manuel Cueva Lovelle, "An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis" *Applied Soft Computing*, Vol. 67, 2018, pp.822-833.

[38] Claudia Diamantini, Alex Mircoli, Domenico Potena, and Emanuele Storti, "Social information discovery enhanced by sentiment analysis techniques", *Future Generation Computer Systems*, 2018, pp. 1-13.

[39] Tian Kang, Shaodian Zhang, Nanfang Xu, Dong Wen, Xingting Zhang, and Jianbo Lei, "Detecting negation and scope in Chinese clinical notes using character and word embedding", *Computer Methods and Programs in Biomedicine*, Vol.140, 2017, pp.53-59.

[40] Rui Xia, Feng Xu, Jianfei Yu, Yong Qi and Erik Cambria, "Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis", *Information Processing & Management*, Vol.52, Issue 1, 2016, pp.36-45

[41] Eric S. Tellez, Sabino Miranda-Jiménez, Mario Graff, Daniela Moctezuma, Oscar S. Siordia and Elio A. Villaseñor, "A case study of Spanish text transformations for twitter sentiment analysis", *Expert Systems with Applications*, Vol.81, 2017, pp.457-471.

[42] Salud María Jiménez-Zafra, M. Teresa Martín-Valdivia, M. Dolores Molina-González and L. Alfonso Ureña-López, "Relevance of the SFU Review SP-NEG corpus annotated with the scope of negation for supervised polarity classification in Spanish", *Information Processing & Management*, Vol.54, Issue 2, 2018, pp.240-251.

[43] Kennedy, Alistair and Inkpen, Diana, "Sentiment Classification of Movie Reviews Using Contextual Valence Shifters", *Computational Intelligence*, Vol 22, Issue 2, 2006, pp. 110-125.

[44] Ivan Habernal, Tomáš Ptáček, Josef Steinberger, "Supervised sentiment analysis in Czech social media", *Information Processing & Management*, Vol. 50, Issue 5, 2014, pp. 693-707.

[45] Alvaro Ortigosa, José M. Martín, Rosa M. Carro, "Sentiment analysis in Facebook and its application to e-learning", *Computers in Human Behavior*, Vol. 31, 2014, pp.527-541.

[46] Jin-Cheon Na, Christopher Khoo, Paul Horng Jyh Wu, "Use of negation phrases in automatic sentiment classification of product reviews", *Library Collections, Acquisitions, and Technical Services*, Vol. 29, Issue 2, 2005, pp. 180-191.

[47] Alec Go, Richa Bhayani, Lei Huang, "Twitter sentiment classification using distant supervision". *Processing*, Vol.150, 2009.

[48] D. Ziegelmayer and R. Schrader, "Sentiment Polarity Classification Using Statistical Data Compression Models," *IEEE 12th International Conference on Data Mining Workshops*, Brussels, 2012, pp. 731-738.

[49] Heeryon Cho, Songkuk Kim, Jongseo Lee, Jong-Seok Lee, "Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews", *Knowledge-Based Systems*, Vol. 71, 2014, pp. 61-71.