International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-5, March 2020

Enhanced Techniques for Opinion Mining

R. Cynthia Monica Priya, J.G.R. Sathiaseelan

Abstract*: *The exponential growth of online information sharing has led to the need to analyze them. Opinion Mining or Sentiment Analysis incorporates a systematic approach to favor analysis of the available online synthesized data .This paper includes a lexicon based approach towards analysis. The proposed SLN (Slang Negation) Algorithm follows a lexicon based approach to Opinion Mining. The slang lexicon provides the expansion of the acronyms and their scores. Three types of negations namely explicit, implicit and pseudo negation are handed. Slang and negation handling are necessary to support semantic orientation. The proposed SLN algorithm has shown significant improvement in the precision, recall, f-measure and accuracy values as compared to the use of sentiment lexicon approach.

Keywords: lexicon, negation, pseudo, SLN algorithm

I. INTRODUCTION

Opinion Mining or Sentiment Analysis analyses the opinion of people. The internet era has created various platforms for information exchange [1]. Blogs, Micro blogs and e-commerce websites and many online forums are data generators. The availability of user generated data invokes opinion mining to effectively extract knowledge. Online reviews analysis serves useful to customers, sellers and in turn the manufacturers. The customers are able to make decision on product purchase. The sellers and manufacturers are able to know the minds of buyers and their comments serve as a major factor for improvement [2].Opinion Mining involves the various phases of analysis namely pre-processing, subjectivity identification, feature extraction, polarity classification and evaluation phase [3]. The two approaches to opinion mining are supervised and unsupervised. The supervised approach involves with training the classifier with the data and with the known output. Various classifiers such as Support Vector Machine (SVM), Naïve Bayes can be involved. The drawback of this approach is that, it involves a lot of time for data training. The use of the second approach, namely unsupervised seems to overcome the above mentioned drawback. The lexicon based approach involves the dictionary-based and corpus- based approaches [4] that encompass developing lexicons and identifying the semantic orientation. The various levels of analysis include word level, sentence level and document level [5] .This paper deals with mining the opinion from the online product reviews at sentence level[6]. Semantic orientation deals with classifying the opinions as positive, negative or neutral. Handling negations [5,7,8] and slang words [9,10] are the identified problems in this work. The proposed algorithm SLN handles slang words and three types of negation namely implicit negation, explicit negation and pseudo negation. Slang words and abbreviated words do not provide meaning to the sentence and thus lead to inaccurate polarity classification [9,10]. It is indeed important to analyze the sentence after slang word replacement. Negation Handling [26,27] is one of the most important valence shifter

Sentence 1

I like the phone.

Sentence 2

I do not like the phone.

Sentence 3

Not only the phone has a camera of very high resolution but also has a perfect built in editor.

Sentence 4

The audio system needs to be improved.

Sentence 5

I am not disappointed with my purchase.

Sentence 1 is positive. The presence of not in sentence2, inverts the polarity. But, this is not the case always. Sentence 3 has not, still it is not negative. Sentence 4 has the word needs that invert the polarity. Sentence 5 has a not followed by another negative word. Hence, it is positive. From the above sentences it is clear that the mere presence of not alone does not shift the polarity. Hence, each of the above sentences has to be handled differently. Identifying the scope of negation is a very tedious task [28]. The proposed SLN algorithm includes a slang lexicon and handles three types of negations namely explicit, implicit and pseudo negations by identifying their scope. Explicit negations deals with the negative words like no, neither, not. Handling them is difficult, since their presence alone does not prove efficient for polarity inversion. Implicit negations deal with the inclusion of words that invert the polarity like needs, requires, could have [26].Pseudo negations are negative words that appear in pairs. The negative prefix along with a negative word appearing before it reverses the polarity.

II. LITERATURE REVIEW

Hunaida et.al [5] provided a preprocessing based approach to improve the lexicon based Sentiment Analysis for Arabic language. It was observed that the hybrid stemming approach involving root based and light stemmers provided better accuracy than when used individually. Neha et.al [11] used a logical methodology for opinion analysis.

Revised Manuscript Received on March 30, 2020.

***** Correspondence Author

R.Cynthia Monica Priya*, Department of Computer Science, Bishop Heber College , Tiruchirapalli, India.

Dr.J.G.R. Sathiaseelan, Department of Computer Science, Bishop Heber College , Tiruchirapalli, India.

The step-wise procedure included combinatory categorical grammar, lexicon building and semantic networks. Muhammad et.al [12] analyzed the word's semantic orientation by classifying them through modifiers, emoticons, general and domain related words from product reviews. Poria et.al [13] introduced a method of retrieving features from multimedia based data through deep convolutional network and observed a good performance. Severyn et.al [14] presented an unsupervised neutral network model for analyzing opinions at phrase and message levels and found high performance improvement. Cambria [15] identified that emotion recognition and semantic orientation identification as the two tasks of affective computing. Aminu et.al [16] reported that the major drawback of dictionary approach is the incorrect scoring done by the sentiment lexicon and introduced domain specificity based lexicon and observed. Wareesa et.al [17] pointed out that certain sentences that are assumed to be positive are actually negative. They calculated polarity by considering the dependencies of negative words and observed a good increase in precision, recall and accuracy.

Ayushi et.al [18] reported that lexicon based approach favors effective polarity classification. The features were extracted using emoticon lexicon, acronym lexicon, opinion lexicon, subjectivity lexicon and hash tag lexicon and the overall performance was satisfactory. Suresh et.al[19] introduced a framework for analyzing online reviews using hadoop and incorporated a lexicon based approach for polarity classification. Anna et.al [20] proposed an algorithm that uses lexicon based approach to analyse opinions. Sentiment normalization and evidence based combination function approach have been used. Nicolas et.al [21] analyzed the predictive performance using the dataset that was labeled manually and examined the advantages of the negative scope identification. The rule based algorithms were devised and observed to produce good results. Hussam et.al[22] examined the impact of clustering various feature groups for sentiment classification. A logistic regression classifier with weighing schema has been used for analysis and it was spotted out that the feature based sentiment lexicon had great impact in the overall performance. Rui et.al [23] pointed out that the valence shift problem affected the performance of the supervised sentiment analysis. The proposed three stages model to address the valence shift problem and outperformed the existing ones. Umar et.al [24] proposed a method to identify the scope of negation while finding the polarity. Thomas et.al [25] reported that the use of sentiment lexicons with polarity strengths perform better than dichotomous polarity based lexicons. They pointed that handling negations can improve accuracy. This was the motivation for this research. Fatima et.al [7] studied the impact of negation in Arabic and used a rule based approach to improve the accuracy.

III. METRICS USED FOR EVALUATION

$$\text{Precision (p)} = \frac{tp}{tp+fp} \quad (1)$$

$$\text{Recall (r)} = \frac{tp}{tp+fn} \quad (2)$$

$$\text{F-measure (f)} = \frac{2(p)(r)}{p+r} \quad (3)$$

$$\text{Accuracy (a)} = \frac{tp+tn}{tp+fp+tn+fn} \quad (4)$$

Fig. 1 Metrics for Evaluation

tp ,tn,fp,fn refer to true positive, true negative, false positive and false negative respectively[4].The equations 1,2,3 and 4 are the precision ,recall,f-measure and accuracy metrics.

IV. PROPOSED METHODOLOGY

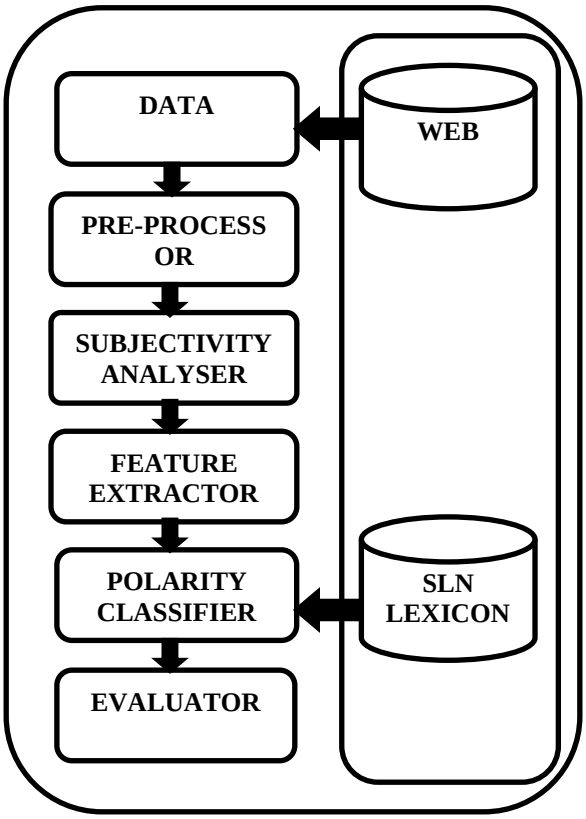

Fig. 2 Methodology Diagram

The online reviews of mobile phones and tablets collected from Amazon are used as the datasets. The first dataset involves 2563 mobile phone reviews and the second dataset involves 2647 tablet reviews. The pre-processed web data flows through the subjectivity analyzer and is categorized as subjective or objective [27].The objective sentences are factual in nature and do not bear any sentiment. The features from subjective sentences are extracted in the feature extractor phase. The semantic orientation of the extracted features that constitute the feature space are then calculated with the help of SLN lexicon. The SLN algorithm deals with handling slang words and negation. The busy world has made people to use different slang words, even in online forums. The SLN algorithm deals with the use of a lexicon based approach to address this issue. The second issue is negation and it is one of the major polarity shifter. The polarity shift involves in moving the polarity to the neighboring words. The SLN algorithm deals with handling negation that can be transferred either directly, indirectly or in a negative-negative form. Direct negations are called explicit negations, indirect negations are called implicit negations and negative-negative forms are called pseudo negations. The positive, negative and neutral scores are calculated and the sentences are classified based on the comparison between the obtained scores.

Pseudo Code for SLN Algorithm

Input: Review R={S1,S2,……..,Sn}
Output: Polarity Score Value S
Notations Used:
Slang Sl
Neg N= {n1,n 2,n3}
n1= {e1,e2,e3,…..en}
n2= {i1,i2,i3,……..in}
n3= {p1,p2,p3,…….., pn}
Total Score TS= {Tp,Tn,To}
Sp=0,Sn=0,So=0
for j=1,……,n
  for k=1,……|Sj|
    if wjk Є S1
    get the slang score and add it to Sp|Sn|So
    else if wjk Є n1
    get the implicit score value and add it to Sp|Sn|So
    else if wjk Є n2
    get the explicit score value and add it to Sp|Sn|So
    else if wjk Є n3
    get the pseudo score value and add it to Sp|Sn|So
    else
    get the sentiment score value and add it to Sp|Sn|So
    endif
  endfor
Calculate TS
endfor

V. RESULTS AND DISCUSSION

The online reviews are collected and after the preprocessing phase,the subjectivity analyzer helps to separate the subjective and objective sentences.The slang and negation features are then extracted to be within the feature space having greater than two occurences.The polarity classifier classifies into positive,negative and neutral

Table 1. Analysis of SLN

DATASET	METRICS	PRECISION	RECALL	F-MEASURE	ACCURACY
1	WITHOUT SLN	.8333	.8695	.8510	.8478
	WITH SLN	.8461	.9166	.8799	.8695
2	WITHOUT SLN	.8695	.8000	.8333	.8260
	WITH SLN	.8750	.8400	.8571	.8478
AVERAGE	WITHOUT SLN	.8514	.8348	.8422	.8369
	WITH SLN	.8606	.8783	.8685	.8587

The table 1 shows that both the datasets exhibit a significant improvement for precision, recall ,f-measure and accuracy.

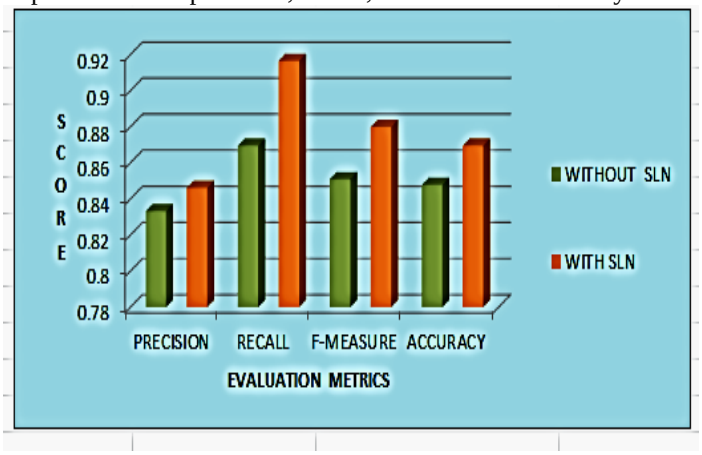

Fig.3 Dataset 1 Analysis

The Fig. 3 provides the analysis of mobile phone reviews. The precision values have shown improvement by 1.28%,the recall value has increased by 4.71%,the f-measure has increased by 2.89% and the accuracy by 2.17%.

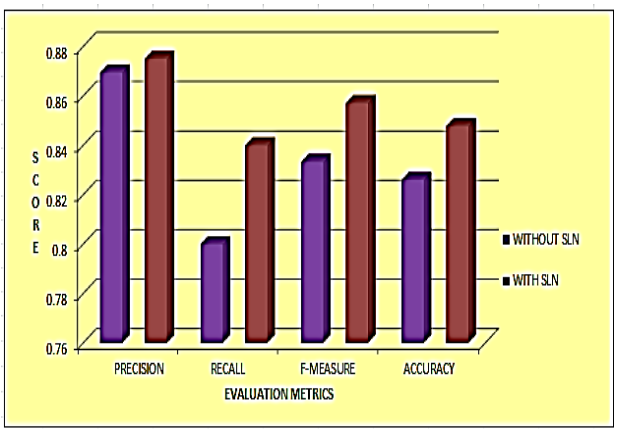

Fig.4 Dataset 2 Analysis

The Fig. 4 provides the analysis of tablet reviews. The precision values have shown improvement by .55%,the recall value has increased by 4%,the f-measure has increased by 2.38% and the accuracy by 2.18%.

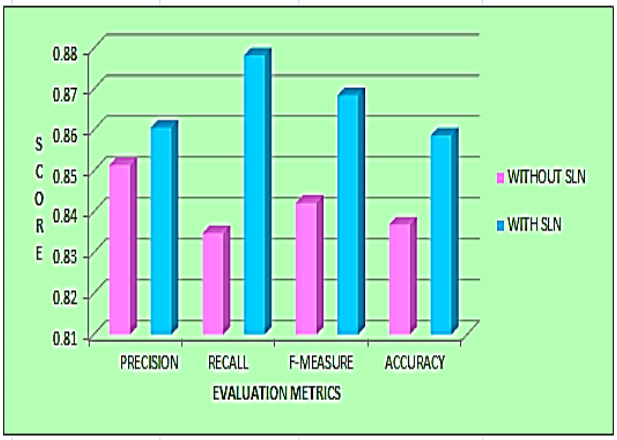

Fig.5 Analysis of the average of the datasets

The Fig. 5 provides the analysis of average of both mobile and tablet reviews. The precision values have shown improvement by .92, the recall value has increased by 4.35 %, the f-measure has increased by 2.63 % and the accuracy by 2.18%

VI. CONCLUSION AND FUTURE WORK

The use of slang words is highly increasing in all the online textual messages for saving time. The opinion mining can be effectively done only if these words are addressed. The negative words would inverse the polarity of the entire sentence. This paper has addressed the slang word and negation through a lexicon based approach. The results reveal that the use of SLN algorithm has improved the accuracy by 2.18% on the average. The future work would comprise using a lexicon based approach for analyzing the impact of amplifiers, diminishers, modal verbs and contrasts in reviews at sentence level .

REFERENCES

1. R. Cynthia Monica Priya and J.G.R.Sathiaseelan, "An Explorative Study on Sentiment Analysis," IEEE, 2017.
2. R. Cynthia Monica Priya, Dr. J.G.R. Sathiaseelan, "A Rule Based Stemmer", International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249–8958,Volume-9 Issue-1, 2019.
3. Maheswari S, K. Arthi, "Rule Based Morphological Variation Removable Stemming Algorithm", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, 2019
4. Vishal A.Kharde and S.S Sonawane, " Sentiment Analysis of Twitter Data :A Survey of Techniques", International Journal of Computer Applications,Vol. 139, No.11,2016.
5. Hunaida Awwad, Adil Alpkocak, "Using Hybrid-Stemming Approach to Enhance Lexicon-based Sentiment Analysis in Arabic", International Conference on New Trends in Computing Sciences",2017, pp. 229-235.
6. Vrushali K. Bongirwar , A Survey on Sentence Level Sentiment Analysis, International Journal of Computer Science Trends and Technology (IJCST) – Volume 3 Issue 3, May-June 2015.
7. Fatima T.AL-Khawaldeh, " A Study of the Effect of Resolving Negation and Sentiment Analysis in Recognizing Text Entailment for Arabic",World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741Vol. 5, No. 7, 124-128, 2015.
8. Jeremy Barnes, Erik Velldal, Lilja Ovrelid, " Improving Sentiment Analysis with Multi-task Learning of Negation", Natural Language Engineering , 2019
9. Da Li , Rafal Rzepka , Michal Ptaszynski, "A Novel Machine Learning-based Sentiment Analysis Method for Chinese Social Media Considering Chinese Slang Lexicon and Emoticons", Vol 2328/3,2019.
10. Liang Wu, Fred Morstatter, Huan Liu, "SlangSD: building, expanding and using a sentiment dictionary of slang words for short-text sentiment classification", Lang Resources & Evaluation 52,2018, pp. 839–852.
11. Neha R.Kasture,Poonam B.Bhilare, " An Approach for Sentiment Analysis on Social Networking Sites",IEEE,2015
12. Muhammad Zubair Asghar, Aurangazeb Khan, Shakeel Ahmad, Maria Qasim,Imran Ali Khan, "Lexicon-Enhanced Sentiment Analysis Framework using Rule- Based Classification Scheme", PLOS One,2017.
13. Soujanya Poria , Erik Cambria , Alexander Gelbukh "Deep Convolutional Neural NetworkTextualFeatures and MultipleKernel Learning for Utterance-Level Multimodal Sentiment Analysis", 2015, pp. 2539–2544
14. Aliaksei Severyn, Alexander Moschitti, " Twitter Sentiment Analysis with Deep Convolutional Neural Networks", ACM,2015,pp.959-962
15. E. Cambria, "Affective Computing and Sentiment Analysis," in IEEE Intelligent Systems, vol. 31, no. 2, 2016, pp. 102-107.
16. Aminu Muhammad Nirmalie Wiratunga Robert Lothian, "Contextual sentiment analysis for social media genres" ,Kowledge Based Systems, Volume 108, 2016, pp. 92-101
17. Wareesa Sharif, Noor Azah Samsudin, Mustafa Mat Deris, Rashid Naseem, Muhammad Faheem Mushtaq, "Effect of negation in sentiment analysis", IEEE,2017.
18. Ayushi Dalmia, Manish Gupta, Vasudeva Varma, " IIIT-H at SemEval 2015: Twitter Sentiment Analysis The good, the bad and the neutral!", SemEval ,2015, pp. 520–526.
19. R. Suresh Ramanujam, R. Nancyamala, J.Nivetha, J.Kokila, "Sentiment Analysis Using Big Data", IEEE,2015.
20. Anna Jurek, Maurice D.Mulvenna , Yaxin Bi, " Improved Lexicon-Based Sentiment Analysis for Social Media Analytics", Springer,2015
21. Nicolas Prollochs, Stefan Feuerriegel, Dirk Neumann, "Enhancing Sentiment Analysis of Financial News by Detecting Negation Scopes", IEEE 2015.
22. Hussam Hamdan ,Patrice Bellot,Frederic Bechet, "Lsislif: Feature Extraction and Label Weighting for Sentiment Analysis in Twitter", SemEval ,2015, pp. 568–573.
23. Rui Xia,FengXu ,JianfeiYua,YongQia, ErikCambria, " Polarity shift detection,elimination and ensemble:A three-stage model for document-level sentiment analysis", Information Processing and Management,2016,pp.36-45.
24. Umar Farooq, Hasan Mansoor, Antoine Nongaillard, Yacine Ouzrout, Muhammad Abdul Qadir,"Negation Handling in Sentiment Analysis at Sentence Level", Journal of Computers, Volume 12, Number 5, 2017,pp.470-478
25. Thomas Schmidt, Manuel Burghardt, "An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing", Humanities and Literature, 2018, pp. 139–149
26. Chetan Kaushik and Atul Mishra, "A Scalable,Lexicon Based Technique for Sentiment Analysis",International Journal in Foundations of Computer Science and Technology,Vol.4,No.5,2014.
27. Kumar Ravi and Vadlamani Ravi, " A Survey on Opinion Mining and Sentiment Analysis:Tasks, approaches and applications", Knowledge Based Systems,2015.

28. Vilares, David & Gomez-Rodriguez, Carlos & Alonso Pardo, Miguel, "Universal, Unsupervised (Rule-Based), Uncovered Sentiment Analysis",Knowledge-Based Systems,2017.

AUTHORS PROFILE

Mrs. R. Cynthia Monica Priya has completed MCA in 2008 and M.Phil in 2011.She has passed all the degree programs with distinction from Bharathidasan University. She has cleared State Eligibility Test in 2016 and National Eligibility Test in 2019. She is working as an Assistant Professor in the Computer Science Department, Bishop Heber College, Tiruchirappalli, since 2012. She has presented and published a number of papers in reputed journals. She is pursuing Ph.D in Computer Science from Bharathidasan University. Her area of specialization is Data Mining and in particular Web Mining. E-mail: cynmonpri@gmail.com

Dr. J.G.R. Sathiaseelan M.Sc., Ph.D is the Head of Computer Science Department, Bishop Heber College, Tiruchirappalli. He has three decades of teaching experience. He has presented papers in International Conferences and published more than 50 research papers in reputed journals. He has authored a book entitled, "Programming In C#.Net," published by PHI Learning, New Delhi in 2009.He is serving as a guide for M.Phil and Ph.D programs in Bharathidasan University. His research areas include Web Services Security, Data Mining, Image Processing and Internet of Things. E-mail: jgrsathiaseelan@gmail.com