# Web Objects Opinion through Sentiment Engineering

**Shrinivas Biradar, G. T. Raju**

***Abstract:*** ***Sentiment Analysis is the analysis of thoughts, feelings and qualities of people towards an object. Automatically recognizing user-generated content views is of great help for commercial and political use. Sentiment Analysis / Opinion Mining lets us gather information about the positive and negative characteristics of any given object / product, and we recommend the favorable and highly scoring views on the object / product to the user. Although researchers have contributed a lot towards objects review through sentiment analysis, still there are open issues needs to be addressed such as Negation Handling, Domain Generalization and Detection and Removal of Fake Reviews. This paper presents a review on the various algorithms used for Negation Handling, Domain Generalization and Detection and Removal of Fake Reviews along with a comparative study against performance metrics along with their limitations.***

***Keywords: Domain Generalization, Fake Reviews, Negation Handling and Sentiment Analysis.***

## I. INTRODUCTION

If a person has to make a decision in ancient days, he usually asks for advice from friends and families. The organization has conducted surveys and focus groups to find public views on its products and services. Perhaps the awful job is to scan for opinion sites and track them on the internet, as there are huge numbers of pages. All related sites are difficult to search, read and arrange in open formats. It is therefore important to have automated opinion discovery and description systems [2].

Analysis of sentiments is characterized as the computational analysis of the opinions, assessments and emotions of people towards objects and their characteristics or attributes [1] On the other hand, it is a question of measuring the subjectivity, polarity (positive or negative) and polarity of a piece of text (slightly positive, mildly positive, strongly positive, etc.). Via forums, comment forms etc. [3] opinions and comments on products and services are expressed on the Internet. Sentiment analysis will help us gather information about the positive and negative qualities of any particular object, so the consumer will be told the positive and highly scoring feelings about a particular product. For marketing the products many companies and businessmen are making uses of opinion mining. Fig 1 shows the proposed architecture for Web Objects Review through Sentiment Analysis.

There are three key stages in the study of feelings that are used by any sentiment analysis architecture.

**Revised Manuscript Received on December 22, 2019.**

**Shrinivas Biradar,** Assistant Professor, Department of CSE RNS Institute of Technology, Bengaluru, Karnataka, India.

**Dr G. T. Raju,** Vice-Principal and Head, Department of CSE RNS Institute of Technology, Bengaluru, Karnataka, India.

Natural Language Processing concepts are widely used to handle the negation as it mainly involves separation of stop wards, special symbols etc. from the document or sentence or aspect, other phases are also shown in the Fig 1 that are equally important for making the appropriate decision.

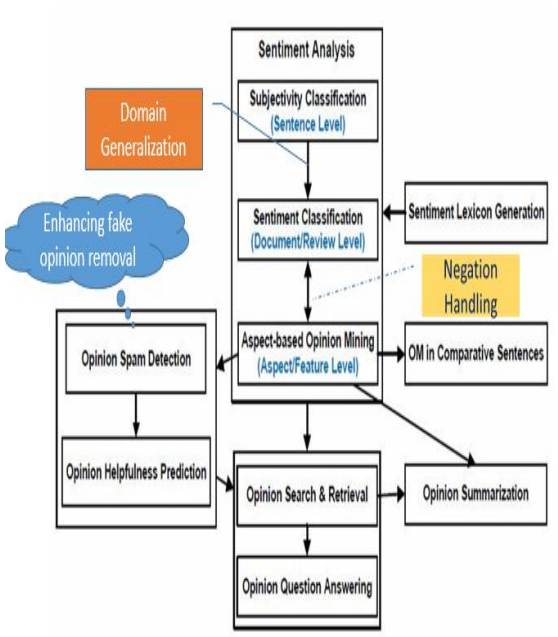


**Fig 1. Proposed System Architecture**

Sentiment Analysis/ Opinion mining has three granularity grades:

1. *Document Level:* The entire document is treated as a single entity at this level and the analysis is extended to the entire document. Sometimes the findings are not acceptable at the level of the report.
2. *Sentence Level:* The sentence is treated as an object at this level of Opinion mining and evaluation is applied to individual sentences, then their result is summarized to provide the document's overall result.
3. Aspect Level: This is also known as the Opinion mining stage or feature level. The aim of this level is to discover feelings about item aspects[1 ].There are two approaches to handle the sentiment analysis:

i. Lexicon-based approach (such as SentiWordNet)

ii. Machine Learning approach (such as SVM, NN etc.)

*Lexicon*-based techniques are usually unable to handle negation. Essentially, these methods are based on the Bag of Words technique. Perhaps the Bag-of-Words design would lead to key features of loss, such as "I don't like that movie" using the Bag-of-Words method we can't reverse the polarity of the word "like."

Perhaps using *machine learning approach* (Support Vector Machine, Naïve Bayes, Logistic Regression etc.), it is possible overcome the limitations of Lexicon-based techniques and would handle the negation in a more appropriate way.

Although researchers have contributed towards objects' review through sentiment analysis, but still the following are the open issues needs to be addressed:

i. Handling Negation

ii. Domain generalizations
Generalizing the sentiment words in particular domain

iii. Dealing with fake reviews
Identifying and removal

The methods used to monitor the user-generated content polarity of opinion are: Review Seer App, Web Fountain, Red Opal, Opinion Observer, etc. [9].

## II. NEGATION HANDLING

Negation plays an important role in determining the corresponding adjective's polarity and therefore the text's polarity. Negation words in general, includes not, neither, nor. For example: The restaurant is good (should be classified "positive").The restaurant is not good (should be classified "negative").However this solution fails to certain cases like "No wonder the restaurant is good". The use of pure language processing technique or pure use of mathematical models fails to tackle negation completely [5]. Negation is the opposite of something which is actually positive or true and vive-versa. Negation can be represented by using negation words like No, Not, Never etc. for example "Movie was not good" is a negative statement and negation word used here is "Not".

Stanford Parser is a 92% accuracy external source library that can parse more than 1000 phrases per second. There are several algorithms to handle negation and some of them are:

1. *Rest of Sentence (ROS):* This algorithm will change the sentiment polarity of the entire word after the negation word.
2. *First Sentiment Word (FSW):* This algorithm will change the sentiment polarity of the first after negation word.
3. *Fixed Window Length (FWL):* This algorithm will change the polarity of sentiment words that exist in the scope of its window size [19].

The performance of different algorithms for negation handling with respect to Precision, Recall and Accuracy is summarized in Table I.

Word Sense Disambiguation (WSD) algorithm can identify the meaning of ambiguous terms automatically by analyzing their phrases. The combination of Word Sense Disambiguation and Negation Handling will provide greater accuracy compared to the standard lexicon-based approach, but this distinction is also seen in conjunction with other machine learning methods in Table I. Stacking algorithm requires two groups of learners, namely first-level students and second-level students. First-level learners are called single learners and combiner is called two-level learners or meta-learners [13].

**Table I. The Performance of different Methods/ Algorithms for Negation Handling.**

| Authors/ Ref. No. | Approach/ Methodology | Dataset | Performance | | | Limitations |
|---|---|---|---|---|---|---|
| | | | Precision | Recall | Accuracy | |
| Claudia amantini, Alex Mircoli, omenico Potena [11], 2016 | WSD+NH | Corpus of 497 tweets of different topics | - | - | 62.57% | Less accurate |
| | NH | | | | 61.12% | |
| | WSD | | | | 58.02% | |
| Swastika Pandey, Santwana Sagnika and Bhabani Shankar Prasad Mishra [12], 2018 | DPT+PA | 100 movie reviews | - | - | 88.37%(P) 94.73%(N) | Less accurate to handle ironic statements |
| | Senti-Lexical | | | | 72.64%(P) 74.47%(N) | |
| | Feature-based | | | | 90.9%(P) 51.11%(N) | |
| Sayali Zirpe and Bela Joglekar [13], 2017 | Logistic Regression | Airline Reviews | 71% | 76% | 84% | More complex polarity shift structures are not considered |
| | Naïve Bayes | | 46% | 54% | 88% | |
| | Linear SVM | | 52% | 40% | 90% | |
| | Stacking with Negation Intensity | | 71% | 98% | 97% | |

## III. DOMAIN GENERALIZATION

The biggest challenge facing the study of opinion mining and sentiment is the generalization of the field terms of sentiment [10]. A text or a paragraph may be connected with multiple entities. Knowing the individual to which the opinion is directed is extremely important [6]. Sentiment Analysis is usually done for a particular field, and this has shown very good accuracy results.

Consider two fields, digital camera and car, for example. The way consumers express their thoughts, opinions, and prospective digital camera may vary from vehicles.

Real world machine learning algorithms can cover a number of different domains in natural language processing and usually require a lot of work to annotate domain-specific training data. For this reason, in recent years, the strategies of domain adoption have gained a lot of popularity [14].

The problem of training and testing models on different distributions is known as domain adoption [16].

The performance of different algorithms for domain generalization with respect to accuracy is summarized in Table II.

The baseline is a reference model based on the bag-of-words space of the source data, which is informed and tested on the target data. The gold standard on the bag-of-words functionality of the same domain is a professional and tested linear vector support system [20].

**Table II. The summary of different domain generalization algorithms/methods**

| Authors/ Ref.No | Approach/Methodology | Dataset | Accuracy | Limitations |
|---|---|---|---|---|
| Songbo Tan, Xueqi Cheng, Yuefen Wang and Hongbo Xu [15], 2011 | Naïve Bayes | Reviews of education, stock reviews and review of computers | 60.60% | Entropy is often not the best way to select generalization features |
| | EM-based Naïve Bayes | | 59.30% | |
| | Naïve Bayes Transfer Classifier | | 64.88% | |
| | Frequently Co-occurring Entropy | | 82.62% | |
| Debora Nozza, Elisabetta Fersini and Enza Messina [20], 2016 | Baseline | 340000 ratings of 25 different Amazon products | 78.40% | There will be a correlation with Convolution Neural Networks |
| | Gold Standard(SVM) | | 75.30% | |

## IV. RESULTS

The site contains content that is both genuine and spam. Before delivery, this spam content should be extracted for better opinion mining identification. This is also called Fake review and hence refers to bogus or fake reviews which definitely mislead the reader by giving them negative opinions related to any product or the object [8]. These fake opinions make sentiment analysis useless in different application areas. Perhaps, this is a social challenge faced by opinion mining.

One of the main challenges of fake review identification is the lack of ground-truth, as it is difficult to find fake reviews even for human readers [17]. The performance of different algorithms for handling fake review with respect to Precision, and Recall is summarized in Table III.

**Table iii. The summary of different algorithms/methods for detection of fake reviews.**

| Authors/Ref.No | Approach/ Methodology | Dataset | Performance | | Limitations |
|---|---|---|---|---|---|
| | | | Precision | Recall | |
| Yuming Lin,Hao Wu, Jingwei Zhang, Xiaoling and Aoying Zhaou 17], 2014 | SVM | 155080 reviews from Liu's Dataset | 93.90% | 90.80% | Does not work effectively for threshold-based values. |
| Arjun Mukherjee, Bing Liu and Natalie Glance, Vivek Venkataraman [18], 2013 | Unigrams | 400 Reviews from Hotel and Restaurant domains | 86.70% | 89.40% | Does not work effectively for multi-behavioral features. |
| | Bigrams | | 88.90% | 89.90% | |

Support Vector Machine (SVM) model will generate a better result as compared with Unigram and Bigram models, however for the threshold-based values this model does not work effectively and this will be addressed in the future models. The other two models i.e. unigram and Bigram are relatively giving better results where as these models work effectively with single feature and does not work so effectively for multi-behavioral features.

## V. CONCLUSION

Sentiment Analysis has gained lot of interest in reviewing web objects. In this paper, we presented a comprehensive review in various methodologies for handling Negation, Domain generalization and Fake opinion detection.

Although these contributions made significant impact on reviewing of web objects, but still they lack with Negation handling, Domain generalization and detection of Fake opinion. Future work should address these issues and better evaluate negative sentences by analyzing typed dependencies to overcome the shortcomings of modern window-based algorithms.

## REFERENCES


1. Vidisha M.Pradhan, Jay Vala, Prem Balani,"A Survey on Sentiment Analysis Algorithms for Opinion Mining", International Journal of Computer Applications, Volume 133-No.9, Jan 2016.
2. Dr. S. P. Khandait, 2 Dr. P. D. Khandait, 3 Pravin D. Jambhulkar," An Approach of Cross-Domain Sentiment Analysis for Opinion Mining", International Journal of Computer Science and Network,Volume 5,Issue 2, April 2016.
3. Jeevanandam Jotheeswaran, Dr.S.Koteeswaran, "Sentiment Analysis: A Survey of Current Research and Techniques", International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 5, May 2015.
4. M K Prakash and G Dinesh," A Perspective Survey on Opinion Mining Tools and Techniques", International Journal of Engineering Science, Volume 8, Issue 3, 2018.
5. Amrita Kaur, Neelam Duhan, "A Survey on Sentiment Analysis and Opinion Mining", International Journal of Innovations and Advancement in Computer Science, Volume 4, Special Issue, May 2015.
6. S.Chandrakala, C.Sindhu,"Opinion Mining and Sentiment Classification: A Survey", 2012.
7. David Osimo, Francesco Mureddu,"Research Challenge on opinion Mining and Sentiment Analysis", 2012.
8. Chandni, Nav Chandra, Sarishty Gupta, Renuka Pahade,"Sentiment Analysis and its Challenges", International Journal of Engineering Research and Technology, Volume 4, Issue 03, March 2015.
9. P.Kalarani, Dr.S.Selva Brunda,"An Overview on Research Challenges in Opinion Mining and Sentiment Analysis", International Journal of Innovative Research in Computer and Communication Engineering, Volume 3, Issue 10, October 2015.
10. Haseena Rahmath P, "opinion Mining and Sentiment Analysis-Challenges and Applications", International Journal of Application or Innovation in Engineering and Management, Volume 3, Issue 5,May 2014.
11. Claudia Diamantini, Alex Mircoli, Domenico Potena, "A Negation Handling Technique for Sentiment Analysis", International Conference on Collaboration Technologies and Systems, 2016.
12. Swastika Pandey, Santwana Sagnika and Bhabani Shankar Prasad Mishra, "A Technique to Handle Negation in Sentiment Analysis on Movie Reviews", International Conference on Communication and Signal Processing, April 2018.
13. Sayali Zirpe and Bela Joglekar, " Negation Handling using Stacking Ensemble Method", Third International Conference on Computing, Communication, Control and Automation, 2017.
14. Debora Nozza, Elisabetta Fersini and Enza Messina, "Deep Learning and Ensemble Methods for Domain Adaprtion", International Conference on Tools with Artificial Intelligence, 2016.
15. Songbo Tan, Xueqi Cheng, Yuefen Wang and Hongbo Xu, "Adapting Naïve Bayes to Domain Adaption for Sentiment Analysis", Springer-Verlag Berlin Heidelberg, 2009.
16. Xavier Glorot, Antoine Bordes and Yoshua Bengio, "Domain Adaption for Large-Scale Sentiment classification: A Deep Learning Approach", International conference on Machine Learning, Bellevue, WA, USA, 2011.
17. Yuming Lin,Hao Wu, Jingwei Zhang, Xiaoling and Aoying Zhaou, International conference on Advances in Social Networks Analysis and Mining, 2014.
18. Arjun Mukherjee, Vivek Venkataraman, Bing Liu and Natalie Glance, Technical Report, Department of Computer Science, University of Illinois at Chicago,2013.
19. Annisa Maulida Ningtyas and Guntur Budi Herwanto, IEEE, 2018.
20. Debora Nozza, Elisabetta Fersini and Enza Messina,"Deep Learning and Ensemble Methods for Doman Adaption", IEEE 28th International Conference on Tools with Artificial Intelligence, 2016.


## AUTHORS PROFILE

**Shrinivas Biradar** is currently working as Assistant Professor in CSE department at RNSIT, Bengaluru, affiliated to VTU. He is pursuing PhD at VTU and his research interests include Machine Learning, Web Mining and data mining

**Dr G T Raju** currently working as Vice-Principal and Head of CSE Dept. RNSIT. He has 25+ years of teaching and 12+ years of research experience. He has published 55 papers in international Journals, conferences. He has organized 28 workshops/FDP/conferences & delivered 30 invited talks. He has received 21 lakhs research grants from AICTE and VTU . He is life member of ISTE and CSI. His research interests are pattern recognitions, web mining, image processing & information retrieval, machine learning and artificial intelligence. He has held various positions at university.