

SENTIMENT ANALYSIS OF PRODUCT REVIEWS ON E-COMMERCE SITE: A SURVEY

*Seenia Joseph**, *S. Hemalatha*

Abstract

The Internet has become the most valuable source for getting ideas and reviews for a particular product or service. Millions of reviews are generated every day about the product that we are searching for and it finds difficult in handing out such a huge review. Sentiment Analysis plays a vital role in analyzing huge reviews and finding polarity of opinion to help the customers in their decision-making. This is the process of collecting, extracting, analyzing, and classifying text based on customer opinions about a product or service from an online platform. It extracts subjective information from reviews collected using natural language processing and text mining processes. This article aims to analyze different methods used for sentiment analysis and also discussed the importance of aspect based sentiment analysis.

Keywords: Sentiment Analysis, Aspect Level, Lexicon Approach, SVM, Machine Learning

I INTRODUCTION

In recent years' Sentiment Analysis has gained widespread acceptance among businesses, government and organizations as its helps them in their growth. Sentiment Analysis is the process of extracting, analyzing and classifying the polarity of sentiments defined in the textual data into positive, negative or neutral. Every day millions of review data is generated over social media and online shopping sites. Analyzing this data is a critical problem. Artificial intelligence and natural language processing techniques are needed to classify the polarity of opinions. Sentiment analysis and classification are the two main

methodologies used in opinion mining. This study focuses on the survey of different methodologies used for Sentiment Analysis. This also concentrates on the importance of aspect based sentiment analysis as it helps both consumers and businesses to know more about the product and to take decisions.

II METHODS USED FOR SENTIMENT ANALYSIS

Many researchers have been done on different methods and technologies used in Sentiment Analysis and each one has its own importance in different domain. In the same year 2016, a researcher C. Thellaamudhan [1] presented a survey on aspect level sentiment analysis and discussed several techniques, approaches, methods used for sentiment analysis. The paper explained different phases of Aspect Level Sentiment Analysis. Different evaluation parameters like Recall, Precision and Accuracy are also discussed in the paper. Researcher D.M.E.D.M. Hussein [2] published on the challenges of sentiment analysis. The Researchers discussed various works that has been done in this field. They point out some challenges in areas of NLP Overheads, Bipolar words, Spam and Fake, Negation etc. The work presented in this paper has shown that area of bipolar words, NLP overhead, and Negation has least accuracy and there is still scope of improvement in this. Same year, a researcher N. F. A. Rahim [3] presented a comparative work on mainly two approaches of sentiment analysis i.e. Machine Learning approach and NLP based approach. They also presented utilization of ontology in sentiment analysis which can affect results in high manner. The work they presented explained the performance of machine learning and NLP based approach shows effective results and that depends on training and testing data on the basis of different domains. The extended

¹Department of Computer Science

Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

*Corresponding Author

research work by E. Kim [4] discussed about emotions and Sentiment Analysis study. This research made a study on Natural Language Processing and then emotion's model along with other approaches.

2.1 Lexicon Based Approach

Lexicon based approach is one of the important techniques used in sentiment analysis. This finds sentiment orientation of the text from semantic orientation of texts. Many researchers have been done on different lexicon based approaches. An Author, Q. Rajput [5] with his co researchers presented a survey of lexicon based approaches and methods of Sentiment Analysis and discussed several approaches, methods and techniques in the field and also the latest updates in the area of lexicon based approach. The table they have presented includes different methods used, datasets used and the scope of the work that has been done in the past in the area of lexicon based approach of Sentiment Analysis. An Author Khin Zezawar Aung [6] with his co researcher analyzed students feed back towards teacher using lexicon based approach. They have analyzed both qualitative and quantitative data set and also made a survey on related work in education field. In 2019 a researcher Venkateswarlu Bonta [7] and his co researchers made a study on lexicon based approaches of NLTK, Text blob and VADER on movie reviews and they have concluded VADER is more accurate than other two.

2.2 Machine Learning Approach

To handle the data effectively different machine learning techniques are available and can be used effectively. Mostly for the task of text classification support vector machine algorithm is used. Text features can be extracted using unigrams or bigrams and for better result POS tagging can be used[8]. Researchers performed machine learning approach on Dataset of Twitter. They have used feature set based on Information Gain, Bigram for social Networking sites. Their proposed model utilized classifiers like Naïve Bayes and

Support Vector Machines and they have presented very effective result. Another work on Twitter data by Researchers A. P. Jain [9] presented machine learning approach i.e. micro blogging site. They utilized apache spark for their work. The tool they have used is much fast and very flexible. Decision trees and Naïve Bayes Classifiers are used for machine learning and text classification. J. Singh [10] presented a work on how to improve the performance of Sentiment Analysis task by using machine learning classifiers. They utilized four classifiers of machine learning i.e. Naïve Bayes, J48, BF Tree and OneR for performance optimization of Sentiment Analysis. They have applied these classifiers on different datasets and evaluated the performance for all classifiers. With the result, they concluded Naïve Bayes as very good learner and OneR as highly accurate classifier. Researcher Munir Ahmad [11] applied different machine learning techniques. They have taken datasets from various research works and compared the result with different techniques of machine learning.

III ASPECT LEVEL SENTIMENT ANALYSIS

Finding different aspects of the product and extracting its polarity helps both the customers and business to know more about the product. Aspect Level Sentiment Analysis focused mainly on the important aspects of the product to find out the sentiment polarity and then it analyzes and extracts polarities on them. Nowadays, the research work on Aspect Level Sentiment Analysis is very much popular in organizations and research centers. SemVal task, the computational workshops and tasks in sentiment analysis area, has given great contributions towards Aspect Level Sentiment analysis. In 2015, an author M. Pontiki [12] showed SemEval-2015 in which he presented all the important aspects of Aspect Level Sentiment Analysis was a great change in this area. Researcher Sebastian Ruder [13] suggested a hierarchical model for aspect based sentiment analysis and they worked on multi domain datasets with the conclusion that hierarchical model outperforms non-

hierarchical model. Another work by Qingnan Jiang[14] suggested two effective models CapsNet and CapsNet-BERT for ABSA and have taken challenging dataset MAMS, Multi-aspect multi-sentiment data set with five multiple aspects on restaurant reviews. They have shown the result in a table explains that CapsNet and CapsNet-BERT shows better performance than the base line aspect based sentiment analysis models. Classification of aspect sentiments and its extraction are two main processes in aspect based sentiment analysis.

3.1 Aspect Extraction

Aspect extraction also needs to take into account commonly used nouns and noun phrases that are a major part of aspect extraction. Many methods have been suggested by researchers and in 2010 C. Long [15] presented an improved approach (23.2%) in which nouns are filtered based on its frequency and Naïve Bayes Classifier and SVM are utilized for its classification.

In today's scenario sequential methods of learning are much utilized for extraction of aspect and, labeled data is essential for training phase in supervised learning methods .In 2016, Lu Chen [16] presented a clustering approach that simultaneously identifies implicit and explicit product features and groups them into aspect categories and this approach combines feature and aspect discovery instead of chaining them. In the same year R. Panchendrarajan [17] concentrated on implicit aspect identification and double propagation technique was used to extract opinion target to check whether the identified aspect is correct or not. Relationships between different entities with different aspects are modeled as a hierarchy, which improves the accuracy of implicit aspect identification in the presence of a large number of interrelated aspects.

3.1.1 Categorization of Aspect Sentiment

The key role at this level is to detect the given opinion is

highly neutral for or against the different aspects. The techniques of Aspect Sentiment can be differentiated as Lexicon based or supervised learning based. Supervised learning approaches has been observed to be outperformed Lexicon based methods in some application groups, but it does not increase and also the groups scale up. The Lexicon-oriented approaches are not fair but supervised approaches are having an upper hand when it is about open-domains [18]. Any learning method carries a big challenge of proving its ability for evaluating the extent of any sentiment expression. It happens by detecting whether the sentence having an aspect which one is interested about. With the help of dependency parser creation, scientists were able to set up that a features group can be dependent on aspects, which leads to the need of their classification. The steps to identify the path aspect include the following

- Keyword and phrase identification
- Use the sentiment shifter to find the polarity of the aspect.
- Use of shifters to find aspect polarity.
- Consider how the "but" clause affects the aspect polarity.
- Embed an aggregate function to create an opinion.

3.1.2 Aspect Term Extraction

Determination of the exact terms to extract in compilation makes aspect sentiment analysis a success. So, it becomes a challenging task to perform aspect term extraction in the area of Sentiment analysis.

Dividing the task of extracting data into smaller tasks is one of the popular options that show the possibility of using sequence labeling techniques. Conditional random field methods are also widely used, as shown in SemEval2014 by M. Chernyshevich[19].Aspect analysis performed in this way has also been shown to give better results compared to other information extraction methods. This addition to machine learning allows researchers to look up each word in

the context of that aspect [20]. Of all the approaches, the combination of four emotional lexicons has proven to be the best [21].

IV SUMMARY

The Research papers studied has proven of great effect in formulating the background for the work. The summary of the papers studied has been provided in table.

Citation	Domain	Algorithm Used	Remarks
C. Long et. al. (2010)	Hotel Reviews (3cities: Boston, Sydney and Las Vegas)	Novel Review Selection Approach , SVM Classifier, Naïve Bayes Classifier	Overall Feature Rating improved by 23.2% by using their Novel Review Selection Approach.
A. Agarwal et. al. (2011)	Twitter data	State-of-art Unigram Model, Tree Kernel, Feature Based Approach	Their Previously proposed State-of-art unigram model report an improvement of 4% for their classification tasks.
D. Grabner et. al. (2012)	Customer Reviews of Hotel	Corpus of Twitter data	Accurate high classification
A. Baladur (2013)	Tweets	SVM, n-grams	Using Unigram and bigram with generalized affective words provided best results.
P. Palanisamy et. al. (2013)	Tweets	Lexicon Based Approach, Serendio Taxonomy	Precision for Positive words: .9361 and for Negative: .8864 Recall for Positive words: 7132 and for Negative: .7912
C. Musto et. al. (2014)	Twitter data, Micro Blog Posts	Lexicon Based Approaches, Lexicon Resources: Sentic Net, Word Net, MPQA, Senti Word Net	MPQA and Senti Word Net Produced best results, WordNet-affect showed not so good results.
A. Severyn et. al. (2014)	Youtube data	Tree Kernel Technology, FVEC sentiment classifier, STRUCT model	30% improved classification rate.
L. Augustyniak et. al. (2014)	Movie Reviews	Lexicon and Machine learning Approach	Classification approach easily won over Lexicon approach.
M. D. Miranda et. al. (2014)	Customer Reviews on Brazilian Online Job Search Company	Repustate API	Positive Polarity towards services : 42%
M. Potinki et. al. (2014)	Restaurants and Laptops Reviews	Aspect term extraction, Polarity Extraction	F-measure and Precision evaluated for 4 sub-tasks Best F Score for Laptops achieve with 74.55%
A. Jurek (2015)	Tweets, IMDB movie reviews dataset	Novel Lexicon based approach	Novel proposed approach performed better than traditional approach.
A. Tripathy et. al. (2015)	Movie Reviews	Naïve Bayes, SVM	SVM classifier outperforms Naïve Bayes Classifier.

B. Le et. al. (2015)	Twitter, 200000 Tweets	Naïve Bayes and SVM, Feature set based on Information gain, bigram and object oriented extraction	Analysis said that emoticons were noisy and highly accurate results came with the proposed model.
H. Zhou et. al (2015)	Movie Review and Trip Advisor comdata	Part-of-speech LDA method	For Sentiment analysis: Improved accuracy upto 0.881 For Aspect Level Sentiment analysis: Accuracy at document level is 0.881
C. Yang et. al. (2015)	36000 comments from Jing Dong Website	POS with ICTCLAS (Segmentation technique)	Polarity Accuracy: 80.91%, Recall: 78.96%, FA Value:79.92%
D. Jhaveri et. al. (2015)	Tweets	Lexicon Based Approaches	Bag-of-words: 57.65% BOW + New Words: 61.11% BOW+Affinn: 44.44% SentiWordNet: 62.56%
J. Mei et. al. (2015)	Extremist webpages, news sources discussing extremism, anti-extremist webpages, and webpages that have nothing to do with extremism	Sentiment Analysis, TENE web Crawler, Sentiment Classification	92% Accuracy - Extremist Webpages, 80% Rate of Success - class differentiation
B. Alvares (2016)	e-commerce websites	Web crawler technology, POS tagging, Naïve bayes Classifier, Maximum Entropy Text classifier.	Naïve bayes classifier is efficient than Maximum Entropy because it need less training data. Maxent is more accurate than Naïve bayes but long time to compute.
D. Effrosynidis (2016)	Twitter Data	Preprocessing techniques, Linear Support Vector Classifier, NaïveBayes, Linear Regression	Results for 15 pre processing techniques presented
O. Bharti et. al. (2016)	Mobile reviews data (500)	Naïve Bayes, K-Nearest Neighbour and Modified K-means Clustering	Naive Bayes 79.66 % KNN : 83.59 % Modified K-Means +NB :89 % Modified K-Means + NB + KNN : 91 %
R. N. Behera et. al. (2016)	20 Videos from social media website YouTube	Naïve Bayes Classifier	Improved Performance of the sentiment analysis system because of incorporated sentic features for computation.
G. Kaur et. al. (2016)	Flipkart mobile reviews	Naïve Bayes, Decision Tree Algorithm	Results evaluated in terms of F - measure, Recall and Precision
Sebastian Ruder, Parsa Ghaffari and John G. Breslin (2016)	Multi domain datasets (restaurants, hotels, laptops, phones, cameras)	Hierarchical LSTM model	Hierarchical model outperforms non - hierarchical model
Lu Chen , Justin Martineau , Doreen Chen , Amit Sheth (2016)	Cell phone, GPS,TV	Classification	Implicit and explicit aspect features are analysed.

SENTIMENT ANALYSIS OF PRODUCT REVIEWS ON E-COMMERCE SITE: A SURVEY

U. R. Babu et. al. (2017)	Reviews on e-shopping websites: Amazon, Snapdeal, HomeShop18, Flipkart and Jabong	POS tagging, Sentiwordnet Approach	Positive, Negative and Neutral results for every website is represented in table.
Khin Zezawar Aung, Nyein Nyein Myo (2017)	Students feedback towards teacher	Lexicon based approach	Compared the result with Our Lexicon and Afinn Lexicon. Our Lexicon produced more accurate result
J.C. Collados (2018)	BBC News Dataset, IMDB, Stanford Sentiment Dataset	Lowercasing, Lemmatization, Multiword Grouping,	Less effective performance of Lemmatization and Lowercasing in cross - preprocessing
Hasan et. al. (2018)	Tweets	Sentiment Analyzers (TextBlob, SentiWordnet, W- WSD), Machine Classifiers (SVM and Naïve Bayes)	W-WSD + Naïve Bayes Classifier: 79% Accuracy TextBlob: 76% Sentiwordnet: 54.75%
G. S. Brar et. al. (2018)	Online Movie Dataset	Feature based opinion mining and Supervised Learning	Accuracy: 81.22%
S. SG et. al. (2018)	Youtube data	Google API, Linear Regression Technique	Accurate results using correlation.
N L Jahnvi et. al. (2018)	Laptop Product Reviews from Flipkart.com	Flipkart API, ROCK and CART algorithm	Result: Positive and Negative Percentage of reviews
S. A. Otaibi et. al. (2018)	Tweets	Support Vector Machines	Accuracy: 87%
T. U. Haque et. al. (2018)	Amazon dataset	Supervised Learning approach	Higher accuracy in results.
K. Kulkarni et. al. (2019)	Viral Ads online	ANOVA test	Active Sharers (17.2 %) Brand-fabatic Sharers (30.2%) Content-hungry Sharers (29.9%) Dormant Sharers (22.7%)
H. H. Do et. al. (2019)	Product Reviews (Tweets)	Deep Learning Approaches	Deep Learning is still infancy
T. Feilonga et. al. (2019)	Reviews on Electronics and Restaurants	JABST and Max-Ent JABST	JABST and Max Ent- JABST performed better.
S. Wua et. al. (2019)	Text from different sources. 5 already available datasets online	ABSA	Method proved to be effective and efficient in results.
S. M. Basha et. al. (2019)	User Reviews	Gibbs Sampling Technique of LDA i.e. Latent Dirichlet Allocation LARR i.e. Latent Aspect Rating Regression	Evaluation of F-measure, RMAE and results came effective.
Venkateswarlu Bonta, Nandhini Kumaresh and N. Janardhan (2019)	Movie Reviews	NLTK, Text blob and VADER Sentiment analysis tool	Accuracy of VADER is 77%, Text blob has 74% and NLTK has 62%

R. Panchendrarajan, M. N. Nazick Ahamed, B. Murugaiah, S. Prakash, S. Ranathunga, A. Pemasiri (2016)	Restaurant Reviews	Machine learning techniques	F1-measure of 0.842
Munir Ahmad, Shabib Aftab, Syed Shah Muhammad and Sarfraz Ahmad (2017)	Products and movie reviews	Machine learning techniques	Compared the results of different research works.
H.M. Zin (2017)	Internet Movie Database (IMDb) https://www.cs.cornell.edu/people/pabo/movie-review-data/ .	SVM(Linear and Non Linear), Vector Space Model (TF and TFIDF)	Pre Processing Strategies improved classification results significantly, (SVM with RBF achieved best Performance)
J. Singh et. al. (2017)	Amazon Reviews, IMDB movie reviews	Naïve Bayes, J48, BFTree, OneR	Naïve Bayes: Fast Learner One R: 91.3% Precision, 92.34% accuracy in instance classification, 97% F-measure.
M. K. Hassan et. al. (2017)	Amazon Reviews, Mongo DB	Machine Learning Approach, Bayesian Networks	Accuracy for Probabilistic Model: 73%
O. R. Llombart et. al. (2017)	Movie Reviews, Stanford Twitter Sentiment	Unigram, Bigram, Naïve Bayes, SVM, Random Forest	Decision trees, and Naïve Bayes showed better performance than SVM.
R. Safrin et. al. (2017)	Reviews on self-created online shopping website	POS tagging, Feature extraction, K-means Clustering	Recall: 90% Precision: 87% Accuracy: 90.47%
R. Kaur et. al. (2017)	Movie Reviews dataset	Random Forest Classifier, Support Vector machines	RF showed more accurate results, better RMS error, recall, F-measure and precision than SVM.
S. Shah (2017)	Twitter	SVM, K- nearest neighbor	SVM(Accuracy): 86.67% SVM+KNN (Accuracy): 93.33%
S. Maghilnan et. al. (2017)	21 Audio files Twitter Dataset Movie Reviews	Sphinx4, Bing Speech API, Google Speech API (speech recognition systems) Naïve Bayes, Linear SVM, VADAR.	Accuracy for Naïve Bayes: 84 Linear SVM: 88 VADER: 95.2

4.1 Table 1: Summary of Literature Review

V CONCLUSION

Review Sentiment Analysis research helps public and private companies and developers gain insight into customer sentiment towards the positive or negative aspects of a product. This helps to understand the need to be involved in product PR, marketing, and promotional activities to enhance positive emotions and manage negative emotional

restrictions. Focusing on the main aspects of the product helps consumers make effective purchasing decisions and allows organizations to improve product standards and quality, which also contributes to product reputation and marketing. Product aspect-based rankings help keep customers long-term. Focuses on difficult areas of Sentiment Analysis such as bipolar words, NLP overhead, and negation are given more concern.

REFERENCES

- [1] Thellaamudhan, C.R.Suresh and P.Raghavi. "A comprehensive Survey on aspect based sentiment analysis." *International Journal of Advanced Research in Computer Science and Software Engineering* 6.4 (2016): 442-447
- [2] Doaa Mohey El-DinMohammed Hussein."Challenges on sentiment Analysis",*Journal of King Saud-University Engineering Sciences*(2018),Volume 30,Issue 4,pp 330-338
- [3] Nur Farahin Abdul Rahim, Mohd Ridzwan Yaakub," A Review of Techniques used in Sentiment Analysis"(2016)
- [4] Evgeny Kim, Roman Klinger,"A Survey on sentiments and Emotion Analysis for Computational Literary Studies". *Zeitschrift Für Digitaleb Geisteswissenschaften* Wolfenbüttel.(2018)
- [5] Quratulain Rajput,Sajjad Haider,and Sayeed Ghani,"Lexicon Based Sentiment Analysis of Teachers Evaluation". *Applied Computational Intelligence and Soft Computing*,Volume 2016
- [6] Khin Zezawar Aung and Nyein Nyein Myo," Sentiment analysis of students' comment using lexicon based approach". *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*(May,2017) 978-1-5090-5507-4
- [7] Venkateswarlu Bonta , Nandhini Kumaresh and N. Janardhan," A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis". *Asian Journal of Computer Science and Technology* ISSN: 2249-0701 Vol.8 No.S2,2019,page. 1-6
- [8] Bac Le and Huy Nguyen," Twitter Sentiment Analysis Using Machine Learning Techniques". *Advanced Computational Methods for Knowledge Engineering*(2015), volume 358, 978-3-319-17996-4,page 279–289
- [9] Anuja P Jain; Padma Dandannavar," Application of machine learning techniques to sentiment analysis".*IEEE*(2016) , 2nd International Conference on Applied and Theoretical Computing and Communication Technology, 978-1-5090-2399-8
- [10] Jaspreet Singh, Gurvinder Singh and Rajinder Singh," Optimization of sentiment analysis using machine learning classifiers". *Human-centric Computing and Information Sciences*(2017)Volume 7, Issue 3
- [11] Munir Ahmad , Shabib Aftab , Syed Shah Muhammad and Sarfraz Ahmad," Machine Learning Techniques for Sentiment Analysis: A Review". *INTERNATIONAL JOURNAL OF MULTI DISCIPLINARY SCIENCES AND ENGINEERING*, VOL. 8, NO. 3, APRIL 2017
- [12] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Suresh Manandhar and Ion Androutsopoulos, "SemEval-2015 Task 12: Aspect Based Sentiment Analysis". *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*,

pages 486–495.

(SemEval 2014), pages 309–313

- [13] Sebastian Ruder, Parsa Ghaffari and John G. Breslin, "A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis". EMNLP 2016, 1609.02745
- [14] Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao and Min Yang, "A Challenge Dataset and Effective Models for Aspect-Based Sentiment Analysis". Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Pages 6280–6285
- [15] Chong Long, Jie Zhang and Xiaoyan Zhu, "A Review Selection Approach for Accurate Feature Rating Estimation. Proceedings of International Conference on Computational Linguistics(Coling)2010, Poster Volume, pages 766–774.
- [16] Lu Chen , Justin Martineau , Doreen Cheng and Amit Sheth, " Clustering for Simultaneous Extraction of Aspects and Features from Reviews". Proceedings of NAACL-HLT 2016, pp. 789–799
- [17] R. Panchendrarajan, M. N. Nazick Ahamed , B. Murugaiah, S. Ranathunga, A. Pemasiri, " Implicit Aspect Detection in Restaurant Reviews". Proceedings of NAACL-HLT 2016, pp 128–136
- [18] Bing Liu and Lei Zhang, " A Survey of Opinion Mining and Sentiment Analysis". Mining Text Data, pp 415–463, DOI: 10.1007/978-1-4614-3223-4_13
- [19] Maryna Chernyshevich, " IHS R&D Belarus: Cross-domain Extraction of Product Features using Conditional Random Fields". Proceedings of the 8th International Workshop on Semantic Evaluation
- [20] Deepak Kumar Gupta and Asif Ekbal, " IITP: Supervised Machine Learning for Aspect based Sentiment Analysis". Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 319–323
- [21] Joachim Wagner, Piyush Arora, Santiago Cortes and Utsab Barman, " DCU: Aspect-based Polarity Classification for SemEval Task 4". Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 223–229.