Sentiment Analysis of Product Reviews: A Review

Shivaprasad T K
Dept. of CSE
NMAM Institute of Technology Nitte
Nitte,Karkala,India
shivathenkila92@gmail.com

Jyothi Shetty
Dept. of CSE
NMAM Institute of Technology Nitte
Nitte,Karkala,India
jyothi_shetty@nitte.edu.in

Abstract: **Now a day's internet is the most valuable source of learning, getting ideas, reviews for a product or a service. Everyday millions of reviews are generated in the internet about a product, person or a place. Because of their huge number and size it is very difficult to handle and understand such reviews. Sentiment analysis is such a research area which understands and extracts the opinion from the given review and the analysis process includes natural language processing (NLP), computational linguistics, text analytics and classifying the polarity of the opinion. In the field of sentiment analysis there are many algorithms exist to tackle NLP problems. Each algorithm is used by several applications. In this paper we have shown the taxonomy of various sentiment analysis methods. This paper also shows that Support vector machine (SVM) gives high accuracy compared to Naïve bayes and maximum entropy methods.**

Keywords-Sentiment analysis, Opinion mining, Product reviews, Natural language processing.

I. INTRODUCTION

Sentiment analysis is a process of extracting and understanding the sentiments defined in the text document. The explosion of data in the various social media channels like twitter, facebook, and linkdin has given consumer new way of expressing their opinion on a particular product, person and places. The user opinion is always in the form of textual information. Per day millions of textual message data is sent over social media or online shopping website. Investigating and analyzing the sentiment of the opinion is a very critical task to perform. The NLP with artificial intelligence capability and text analytics are used to determine whether the sentiment of the opinion is positive, negative and neutral. The opinion mining and sentiment analysis is doesn't depend on any particular domain or platform. It spreads to all the social media networks, healthcare, management, economy and many more and also it is very useful for the growth of many companies and organizations. Sentiment analysis is also providing a business intelligence which can be used to make good impactful decision. Sentiment analysis and sentiment classification are the two methodologies used in opinion mining. Whereas both have its own independent features, but sometimes it may be used interchangeably. Sentiment classification indicates the sentiment orientation by assigning the class labels to the document or segment. Sentiment orientation is a kind of text classification that classifies text data based on the sentiment orientation of opinion. Sentiment orientation indicates the polarity of the opinion either true or false based on subjectivity [1]. Subjective analysis is a process of identifying whether the given text or reviews data is subjective or objective in nature.

In this paper several sentiment analysis methods have been discussed. Even though we find several papers proposed by different researchers in this area, there is a need to make sentiment analysis more accurate and easy to understand. Sentiment analysis is extremely useful in various situations. But it is very difficult process because of the complexity involved in the human language. It has got several variants like grammatical, cultural etc. Humans can easily interpret statements like "My order been delayed. Very good". But it may be difficult for the machine to understand. Similarly word "thin" may be taken as positive with respect to laptop but it may be negative when it comes to apartment wall. So to give the correct decision sentiment analysis must be sometime more business specific.

II. LEVEL BASED SENTIMENT ANALYSIS

In general the process of sentiment analysis is divided into 3 levels.

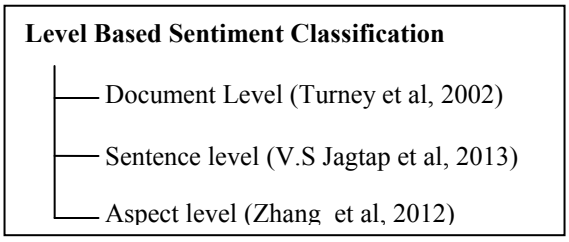

Fig 1: Different levels of Sentiment Classification

A. Document level analysis: Document level sentiment analysis determines the overall opinion of the document. Apart from the sentiment orientation of the individual sentences it classifies the sentiment expressed by the whole document. The classification is expressed by either positive or negative sentiment [2]. This level analysis is helpful only if the document related to a single entity. Because it expresses opinion on a single entity (E.g. Product, Person). Thus it is not applicable to the documents which contain the comparison of the multiple entities.

B. Sentence level analysis: The sentence level sentiment analysis is also called as subjectivity classification [3]. It distinguishes subjective information from the objective information. It considers each sentence as a separate unit and it predicts that the sentence must contain only one opinion. It iterates each sentence and determines whether the sentiment orientation of the sentence is positive, negative or neutral.

C. Entity and Aspect level analysis: Entity and aspect level sentiment analysis captures the mixture of emotion from the review sentence. It performs fine-grained sentiment analysis and it differentiates what actually user wants and do not-want. One main feature of aspect level sentiment analysis is that, it directly looks into the opinion instead of paragraph, sentences, phrases and document. The goal is to find sentiments on entities and their aspects. Meant to say that aspect level sentiment analysis concentrate on sentiment (positive or negative) and a target (opinion).

A sentence or a document may contain a combination of positive and negative opinions.

- Sentiment level opinion mining is performed by two tasks i.e. subjective and objective.

  Objective: I purchased a new Android Smartphone few months ago.

  Subjective: It is such a nice Gadget.

- The polarity of the opinion is determined by using the subjective sentence.

  Positive: It is such a nice Gadget.

  Negative: It has poor radio signal reception quality.

In a document level sentiment analysis it determines the overall document and a sentence. The classes of the opinion are determined by polarity. i.e. either positive or negative. The binary classification method is used for this purpose. Most of the existing sentiment analysis algorithm uses the binary classification methods. Mean to say that they assign review or opinions to bipolar classes such as positive or negative. Binary sentiment classification classifies reviews or opinions by using multi-point rating scale. Multi-point rating scale uses the rating inference. Using this rating inference class labels are assigned as scalar rating such as 1 to 5 "stars". The detailed description of binary approach is given in section 4.

III. SENTIMENT ANALYSIS PROCESS

The basic sentiment analysis framework consists of following steps. Review collection process, Data preparation, Review analysis and Sentiment classification.

A. Data Preparation: Data preparation is a process of collecting the reviews of a particular product from internet sources or any online shopping websites. The collected data may be either unstructured. There are many publically available datasets which provide the huge collections of reviews which can be are used in sentiment analysis process. Sometimes the set of reviews contains unwanted information such as HTML tags, URL information etc. Removing such useless information is done at the stage of preprocessing of reviews.

B. Review Analysis: Review Analysis step analyze the special features of the reviews and then identifies the interesting information in the review including opinion. Review analysis step first applies various computation tasks to the review and then extracts the opinion and features of the product. Two mainly used methods in review analysis are POS tagging and Negation tagging.

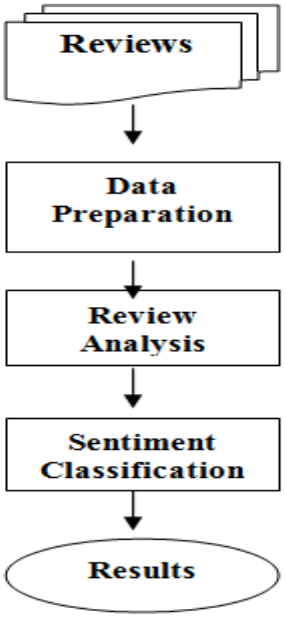

Fig 2: Sentiment Analysis process Model

C. Sentiment Classification: The two major approaches used for classifying review are Sentiment orientation approach and machine learning approach. Sentiment orientation consists of two subtasks. First task is to extract the opinions from the review. While the second subtask is to determine the overall sentiment orientation of a sentence. The Sentiment Orientation contains two seed adjective, "excellent" and "poor". The machine learning approach depends on the classification classes such as positive and negative.

IV. POLARITY BASED SENTIMENT CLASSIFICATION

A typical approach to sentiment analysis is to begin with a lexicon of negative and positive words and phrases. Identification of polarity from this lexicon decides the performance of sentiment analysis. Sentiment classification can also be used to analyze the perspective of the different sources of data or information. Especially in the field of reviews the numerical rating system makes the categorization process into fine grained scales other than positive and negative categories. This richer information can be used for rank and compare the reviews quantitatively.

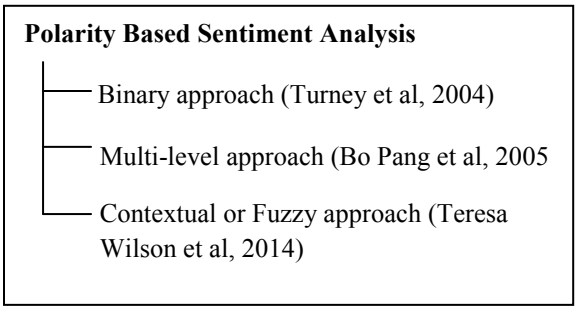

Fig 3: Polarity Based Sentiment Analysis

A. Binary Approach: Binary approach is also called as two class categorization process. Where the sentiment classification is categorized into two classes namely positive and negative. Turney [4] proposed counting of positive and negative terms in a product review to determine the polarity. Binary approach is very useful for brief analysis for any type of reviews. This approach is implemented by including the other metrics called negation words, intensifier and diminishes

B. Multi-class approach: Instead of using the binary classification system multiclass approach is used to convey the sentiment expressed by the piece of text document. Multi class method is also called as star based inference approach where the classification includes categorization process as high (5 stars) and low (1 star) inference [5]. We can easily categorize the reviews based on these star rating. But the sentiment based multi-class approach naturally formulated as regression problem. This is happened because each class corresponding to a different point of scale. More specifically each class may have its own vocabulary and meaning of the sentence will be different depending on domain and situations.

C. Contextual or Fuzzy approach: Contextual or fuzzy approach is used to find whether the review text document is in neutral form or polar form. Contextual approach is very useful when user unable to find the actual intension of people about the particular product. In simple this approach explains about whether the review text documents belongs to the polarity measures (positive, negative and neutral) or it belongs to the neutral polarity (only neutral). And this creates dis-ambigutiy in polarity of polar expression. Sometime, it may be a combination of the both approaches. In order to tackle the problem of ambiguity the 4 way approach is used like positive, negative, both and neutral[6].

V. SENTIMENT CLASSIFICATION TECHNIQUE

Sentiment classification methods usually divided into machine learning approach and lexicon based approach. Machine Learning Approach uses linguistic methods and applies common ML algorithms. The lexicon based approach depend on sentiment lexicon and a set of commonly known and a precompiled sentiment words. The lexicon based approach is sub divided into dictionary based approach and corpus based approach. The corpus based approach contains two sub methods called statistical method and semantic method.

The sentiment classification method using machine Learning approach is divided into supervised learning and unsupervised learning approaches. The supervised method is a sophisticated approach where it uses only labeled training documents. There are many supervised classifiers in literature. Most common used classifiers are probabilistic classifiers and Naive Bayes Classifiers. If supervised methods unable to find the labeled training document then unsupervised method is used to solve the problem.

The lexicon based approach is used to discover the opinion lexicon which is used to analyze the text document. It includes two sub approaches. The dictionary based approach depends on the dictionary and it searches the antonyms and synonyms for the seed words in the dictionary. The corpus based approach started with list of opinion words and find out some other opinion words in the large corpus to retrieve semantic orientation.

VI. SENTIMENT ANALYSIS METHODS

The main branches of Sentiment Analysis include Machine learning approach and Lexicon based

approach. Each of these approaches includes sub-approaches. The Taxonomy of Sentiment analysis methodologies are listed below in graphical model. We listed some techniques and procedures used by these methodologies. The two important sentiment classification methods are

1) Machine learning approaches

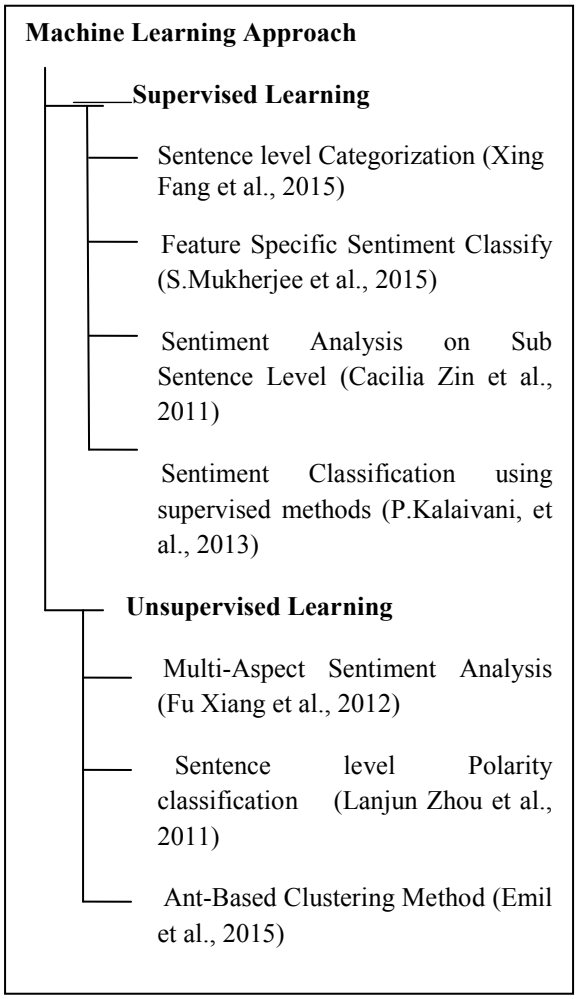

Fig 4: Machine learning approaches

In Fig 4, we listed some of the methods which are used by the machine learning approaches. The main branches of the machine learning methods are supervised learning and unsupervised learning. The goal of the supervised learning method is to find the desired output for the desired input. The mapping between the input and output is identified in supervised learning method. In unsupervised method there is no suitable input for learning process. Each input and desired output identification is done by the algorithm itself. The table contains the list of research done on the machine learning approach. In Fig 5, it contains the methods of lexicon based approaches. Lexicon based approaches uses text classification approach and it mainly aims to find the semantic orientation of a document or a text. Main branches of the lexicon based approaches are dictionary based approach and corpus based approach. The corpus based approach contains two sub methods called statistical and semantic based approaches.

2) Lexicon based approaches

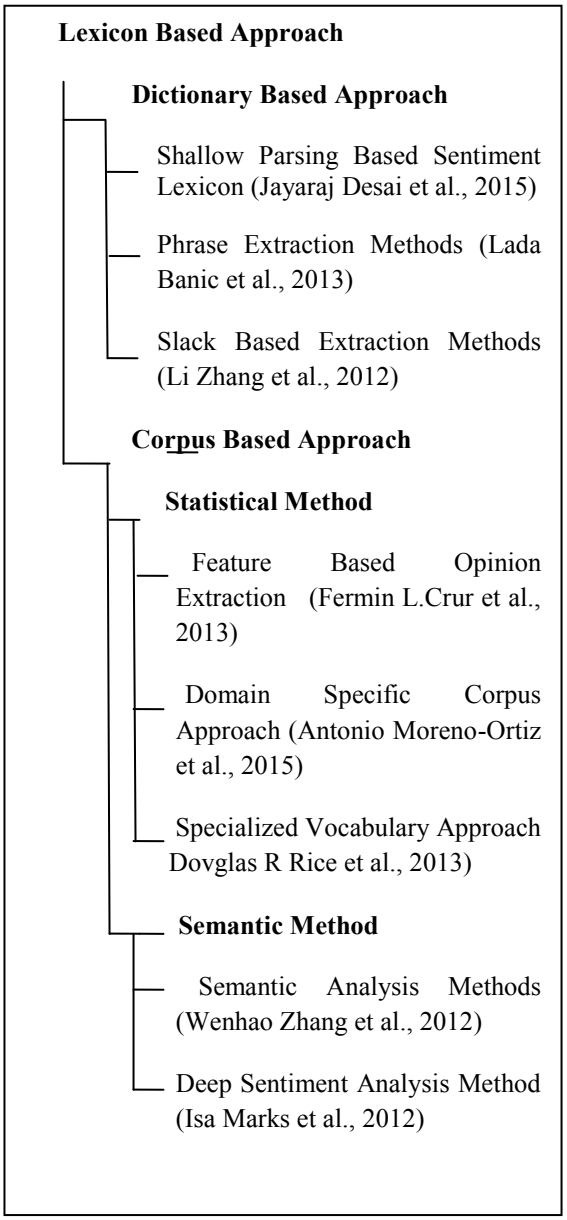

Fig 5: Lexicon based approaches

VII. FEATURE SELECTION METHODS

The most frequent feature selection methods are used in the sentiment analysis is a lexicon based methods which needs manual annotations. While Statistical methods are fully automatic. The feature selection method treat text document as a Bag of Words (BOW's). Because of their simplicity BOW's are most commonly used in the feature selection methods. The common use of BOW's is stop words removal and stemming in sentiment analysis process.

VIII. LITERATURE SURVEY

Xing Fang et al., [7] represent a sentiment analysis system for product reviews which uses the sentiment polarity categorization process. The overall process is split into 3 phases. The classification methods selected for categorization are naive Bayesian, support vector machine and random forest. During phase 1 the subjective content is extracted from the review data and objective content is removed. After the extraction perform the POS tagging to extracted content. During phase 2 identify the sentiment phase either negative of adjective (NOA) and negation of verb (NOV). And perform the sentiment score computation for sentiment tokens. Using the sentiment score formula feature vector for sentiment are generated. Finally in phase 3 they performed the sentiment polarity categorization.

Subhabrata mukherjee and pushpak Bhattacharyya et al., [8] investigate the product review using feature specific sentiment analysis. A dependency parsing method is used to capture the relation between the features and their associated opinions. They developed a system that extracts the potential features from review and clusters opinion expression describing each of the features. To achieve good experimental result they used two datasets Lakkaraju et. al., [9] and Hu and Liu et. al., [10]. The proposed system performs average accuracy of 80.98% for dataset 1 and 70% of accuracy for dataset 2.

Jian jian and ping ji et. al., [11] presents a framework for the identification of product feature aspects and consumer detailed reasons from online product reviews. The proposed co-clustering algorithm provides a short summary about consumers concerns on different aspects of product features. And reason of consumers to provide a concise description regarding customer requirement for product designers. The approach of conditional random fields is employed to detect aspects of product features and detailed reasons jointly.

Lada Banic and Ana mihanovic et al., [12] investigate hotel review mining using machine learning and natural language processing. They develop a product evaluation system configuration which works on a KNIME sentiment analysis framework. The sentiment analysis process includes retrieving data from the database, dictionary development and review scoring. They represents two evaluation system it evaluated each term or phrase positive or negative. Cumulative information about the number of positive and negative terms or phrases was generated on the level of final evaluation. In second evaluation they maintain a grade from 1 to 5. In which 1 is referred as bad and 5 is referred as excellent.

Haruna isah and Paul Trundle et al., [13] represents a lexicon based and machine learning based sentiment classification approach which uses naïve bayes, maximum entropy and support vector machine machine learning classifiers. They have developed a framework which processes the experiences of customers of popular brands of drugs and cosmetic products. Naïve bayes classifier are used to obtain baseline result for accessing other classifiers.

IX. ANALYSIS OF SOME IMPORTANT METHODS

Researchers use sophisticated methods to predict the sentiments of product features. The datasets of review may crawl from online or it may be use offline data. The classifiers play very important role in sentiment analysis. Classifiers work independent to all domains (product, movie or place). The widely used classifier like naïve bayes (NB), support vector machine (SVM) and maximum entropy (Max-Ent) ease the process of sentiment analysis. Naïve bayes classifiers are highly scalable and they are very popular method used in text categorization. With its appropriate results it gives competition to the very advanced method called support vector machine. Many researchers use naïve bayes for their first preference and for better results they move on to other advanced classifiers. Support vector machine builds a set of hyper planes in a infinite dimensional space which can be used for classification and other processes. SVM algorithms are widely used in advanced classifier, and experimental results show that SVM achieves high accuracy in classification process as compared to any other methods. Maximum entropy is suitable for finding the best probability distribution. Max-Ent is also widely used method in sentiment analysis. The graph shown in fig (6) depicts the accuracy of support vector machine, naïve bayes and maximum entropy for the different product datasets [14]. In the graph naïve bayes and maximum entropy shows moderate accuracy with small variation and support vector machine shows the highest accuracy for all the dataset.

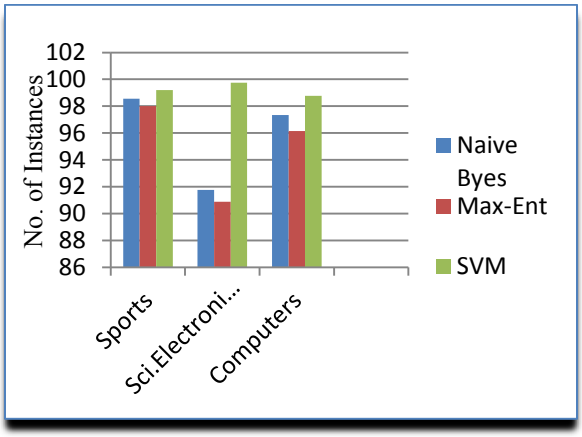

Fig 6: Accuracy of sentiment classifiers over different dataset

X. CONCLUSION

In this paper it is observed that sentiment analysis or opinion mining plays important role while making a decision towards a particular product or a service. But it is very important to consider certain quality measures like helpfulness, usefulness and utility while analyzing each review. In the literature survey there are many sophisticated methods explained which defines the sentiment analysis with respect to different aspects.

In future, more research work is needed to improving the performance measures further. Sentiment analysis or opinion mining can be applied for any new applications which follow data mining rules. Although the techniques and algorithms used for sentiment analysis are advancing fast and giving high quality results, lot of problems in this field of study remain unresolved and also it is hard to find the fake review by reading. Sometimes fake reviews also seen as good quality review and it was modified like no one can identify their actual intension. So fake review detection is another important field which requires deep data mining techniques.

REFERENCES

[1] Bing Liu,"Exploring User Opinions in Recommender Systems",Proceeding of the secondKDD workshop on Large ScaleRecommender System and the Netflix Prize Competition",April 2012,LasVegas,USA.

[2] Antonio Moreno-Ortiz, Javier Fernández-Cruz," Identifying polarity in financial texts for sentiment analysis: a corpus-based approach", 7th International Conference on Corpus Linguistics: Current Work in Corpus Linguistics: Working with Traditionally-conceived Corpora and Beyond (CILC 2015)

[3] Zhang Wenhao, Hua Xu, Wan Wei. Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Syst Appl 2012.

[4] Peter D. Turney." Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews", Proceedings of the 40th Annual Meeting of the Association for ComputationalLinguistics (ACL), Philadelphia, July 2002, pp. 417-424.

[5] Bo Pang and Lillian Lee. "Seeing stars Exploiting class relationships for sentiment categorization with respect to rating scales ". Proceedings of the ACL, 2005

[6] TheresaWilson, JanyceWiebe, Paul Hoffmann," Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", Advanced Research and Development Activity (ARDA)

[7] Xing Fang ,Justin Zhan ,"Sentiment Analysis using product review data",Springer: Journal of Big data",2015,North Carolina A& T State university,Greensboro,NC,USA.

[8] Subhabrata Mukherjee,Pushpak Bhattacharyya,"Feature Specific Sentiment Analysis for product Reviews",IET,2015,IIT Bombay.

[9] Himabindu Lakkaraju, Chiranjib Bhattacharyya,Indrajit Bhattacharyya and Srujana Merugu,"Exploiting Coherence for the simultaneous discovery of latent facts and associated sentiments",SIAM International Conference on Data Mining (SDM),April2011.

[10] Minqing Hu and Bing Liu,"Miming and Summarizing customer reviews", KDD 04: proceedings of the tenth ACM SIGKDD international Conference on knowledge discovery and data mining.

[11] Jian Jin and Ping Ji,"Mining online productreviews to identify consumers FineGrinedConcerns",IET,2015,Hong Kong SAR,Chaina.

[12] Lada Banic,Ana mihanovic,Marko Brakus,"Using Big Data and Sentiment Analysis in Product Evaluation "MIPRO,2013,Croatia.

[13] Haruna isah,Paul Trundle,Daneiel Neagu,"Social Media Analysis for Product Safety and using Text Mining and SA",IET,2015,University of Bradford,UK

[14] R. RajKumar,V.P. Kallimani, Lam H. Lee, Dino Isa," Text document processing with naïve bayes and support vector machine", published by IEEE computer society.

[15] Zirn C, Niepert M, Stuckenschmidt H, Strube M. Fine-grained sentiment analysis with structural features. In: Presented at the 5th International Joint Conference on Natural Language Processing (IJCNLP'11); 2011.

[16] Maks Isa, Vossen Piek. A lexicon model for deep sentiment analysis and opinion mining applications.

[17] Cruz Fermı´n L, Troyano Jose´ A,Enrı´quez Fernando, Javier Ortega F, Vallejo Carlo G.Long autonomy or long delay?' The importance of domain in opinion mining. Expert Syst Appl 2013.

[18] Zhou L, Li B, Gao W, Wei Z, Wong K. Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities. In: Presented at the 2001 conference on Empirical Methods in Natural Language Processing (EMNLP'11); 2011.

[19] Emil Şt. Chifu , Tiberiu Şt. Leţia Viorica R. Chifu," Unsupervised Aspect Level Sentiment Analysis Using Ant Clustering and Self-organizing Maps",IEEE,2015

[20] Jayraj M. Desai, Swapnil R. Andhariya," Sentiment analysis Approach to adapt a shallow parsing based sentiment lexicon", IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems, ICIIECS'15

[21] Maks Isa, Vossen Piek. A lexicon model for deep sentiment analysis and opinion mining applications. Decis Support Syst 2012.