



## Opinion Mining and Analysis of the Techniques for User Generated Content (UGC)

<sup>1</sup>Reema Verma, <sup>2</sup>Dr. Kiranjyoti

<sup>1</sup>M.tech. Student, <sup>2</sup>Assistant Professor

<sup>1,2</sup>Department of Information Technology

Guru Nanak Dev Engineering College, Ludhiana, Punjab, India

---

**Abstract-** *Opinion mining is the field of study that analyzes the opinions of users, sentiments, evaluations, attitudes, and emotions from social media such as reviews, forum discussions, blog, micro-blogs, Twitter, and social networks. This has led to the emerging field of opinion mining and sentiment analysis, which deal with information retrieval and knowledge discovery from text using data mining. These opinions are subjective information which represents sentiments of user's, feelings or appraisal related to the same. The main purpose of this study is to focus on the various classifications of opinion mining techniques that convey user's opinion i.e. positive or negative at different level.*

**Keywords-** *Opinion, UGC (User Generated Content), OM (Opinion Mining), Approach.*

---

### I. INTRODUCTION

Opinion is person viewpoint about an object whereas mining is the extraction of knowledge from facts or raw data. Thus, in another word it is a technique which detects intelligent information from data accessible on web. The people who express their opinion on web has dramatically day by day. They can express their opinion almost based on User Generated Content eg review sites, forums, discussions groups, blogs, products etc.

#### *Diverse Stage of Opinion Mining*

Document Level OM: the task at this level is to classify the overall document express by the opinion holder as positive, negative or neutral about certain object. There are two main approaches to document level opinion mining :

1. Supervised learning: this type of document level opinion mining is done by testing and training the document expressed by the authors opinion. This approach assume that document should be classify into finite set of classes that data is available for each class eg the product review with two classes which are two star(negative) and 5 star ranking.(positive).
2. Unsupervised learning: this approach is also known as automatic document level classification based on determining the semantic orientation (SO).

Sentence Level OM: This level analyzes the document at sentence level. This level associated with two task, first task deal with identifying whether the given sentence is subjective (that contain opinion) or objective. Example: I brought an Audi. (Objective), Example: It is such a nice car. (Subjective):Second task is to find positive, negative or neutral based on sentence. This level has two popular approaches first corpus based approach and dictionary based approach.

Feature Level OM: This level is also called Aspect level. The analysing going much deeper and deals with identify the feature in sentence for a given document and analyze feature and classify them according to optimistic, pessimistic and impartial.

### II. STRUCTURAL DESIGN OF OPINION MINING

It is a process of finding user's opinion towards a topic/a product. Opinion mining concludes whether user's view is optimistic, pessimistic or impartial about a product, topic or event etc. It process involves three main steps:

#### *Opinion Retrieval*

This is a process to collect review text from different review website which contain reconsiders for product, hotel, news, movies, blogs and store all these review text data in the database review

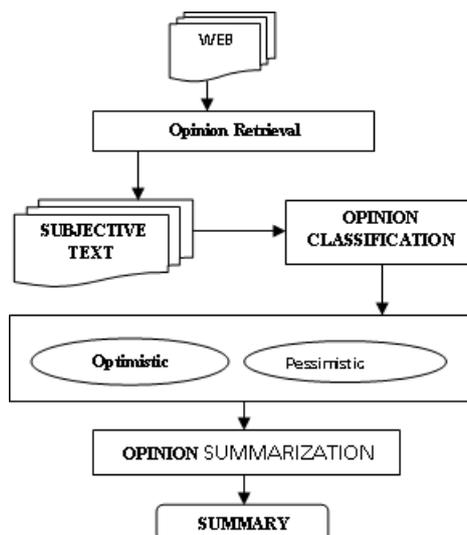


Fig.1 Structural Design of Opinion Mining

#### Opinion Classification

The key step in opinion analysis is classification of review text. The classification of review text is supervised by learning method. Using sample reviews the classifiers are trained. The trained classifier model is used to predict category of new text reviews. Support Vector Machines (SVM) and Naive Bayes (NB) are popular machine learning classifiers for text categorization. Lexicon based approach: in this approach predicts sentiment of review text using database which contain word polarity values. Review text is classified by calculating and averaging polarity score of individual score in sentences.

#### Opinion Summarization

This process mainly involves the following two approach:

First, feature based summarization involves collecting the feature term that are appearing in various reviews. The summary is done by selecting sentences that contain only feature information. The features in review can be identified using LSA method. Second, Term frequency is count of term occurrences in a document. If a term has higher frequency it means that term is more import for summary presentation. In this method sentences are scored by term frequency. The summary is presented by selecting sentences that are relevant and which contain highest frequency terms.

### III. SURVEY OF THE OPINION MINING BASED ON UGC (USER GENERATED CONTENT)

[1] Ganapathivhotla,M. and Liu,B.(2008) studied sentiments expressed in comparative sentences that is to determine which entities in a comparison are preferred. This paper proposed a new measure called One-side Association (OSA) to solve the problem, which deals with context based sentiments by gathering external data available on the net.

[2] Pak,A. and Paroubek,P.(2011) presented a method for an automatic collection of corpus that can be used to prepare a classifier of sentiment. They used tree tagger for POS tagging and identify the difference along with neural, negative and positive sets. The classifier is based on multinomial naïve based classifier that uses N-gram and POS-tags. In this paper the authors used syntactic structures to describe state facts or emotions.

[3] Liu,B. et al.(2005) discussed about two contributions. First they proposed a framework for analyzing and comparing consumer opinions of competing products. They implemented opinion observer as a prototype system which clearly gives the strength and weakness of each product. Second they propose a new technique of pattern mining to extract pros and cons in particular type of review.

[4] Mishra,N et al.(2012) presented a theoretical model of opinion mining, which discussed the problem and supply a common outline in different research directions. In section second they uses opinion mining task at various levels which determine whether a document or sentence carries positive or negative opinion. In section third they discussed comparison of various tasks of opinion mining. They also detect fake opinions.

[5] Virmani,D. et al.(2014) proposes a sentiment analysis a collaboration with opinion extraction summarization and tracking the student's record. They modify the existing algorithm in order to obtain the collaboration opinion about the student. The algorithm compares each word with sentiment and negation in database. The algorithm is implemented on the basic of score assigned to each sentiment word in the database. The collaboration opinion is evaluated by analysis teacher's aside by text. The resultant opinion is represented like very high, high, moderate, low and very low.

[6] Hu,M. et al. (2004) used number of techniques for opinion mining from product review based on NLP and data mining methods. The experimental result technique is not more effective in performing their task. In future they are planning to improve the task.

[7] Yu .H, and Hatzivassiloglou .V (2003) discussed about detecting opinions at sentence level as positive, negative or neutral. The necessary components for an opinion question answering system is distinguish the opinion from the fact at both document and sentence level. They used bayesian classifier for discriminating between documents with a

preponderance of opinions such as editorials from regular news stories and describe three unsupervised, statical techniques for detecting opinions at sentence level. This paper mostly focused on characterizing opinions and facts in a generic manner.

[8] Hu and Bing.L (2006) discussed several problems and different types of assessment on web; he has given few initial approaches and result on the basic of format based on web. In this survey different reviews format used different techniques to perform the extraction of opinion. Format 1 Pros and Cons (the reviewer is asked to describe experts and defrauds separately), Format 2 Pros and Cons and detailed review (the reviewer is asked to describe pros and cons separately and also write a facet review), Format 3 free format(the reviewer can freely write, no separation of experts and defrauds). For evaluation they had used two corpora from web for two categories of reviews. In corpora first they had used five electronic products and corpora second they had used fifteen electronic products. The result is discussed on feature extraction and opinion identification.

[9] Dave .K et al.(2003) implemented in their research paper opinion mining tool which process a set of search result for a given item and generating a list of product attributes (poor,mixed,good). In their research they develop a method for automatically distinguish between positive and negative reviews. Our team draws an information retrieval method for feature extraction and scoring the result depends upon the various matrixes on testing situation. This method results better in complete web based tool rather than individual sentence collected from web

[10] Thawakar,M. et al(2014) proposed an approach that mines the opinion of customer according to product feature. They proposed the method for rating products according to feature. Their approach compares product according to the rating that helps customer to take decision regarding product purchasing.

[11] Sandhu et al. (2011) focused on designing a system to organize web opinions at the time when users are posting, before actually being extracted by expertise. New system (opinion organization system) provides four stages. In first stage, it provides a list of several product categories and user select as a minimum one. In second stage, a list of selected product relevant features is demonstrated to the user. In third stage, user firstly selects features for which wants to convey opinions, then uses polarity based P set and N set containing adjective words list and uses thumb selection table to add opinions, in fourth stage.

[12] Chandarkala et al. (2011) have discussed opinion mining or sentiment analysis is a natural language processing and information extraction task identifies the user's views or opinions explained in the form of neutral, negative or positive comments and quotes underlying the text. Text categorization generally classifies the documents by topic.

[13] Bafna. K and Toshniwal .D (2013) proposed a dynamic system for feature based summarization system of customer's opinions for products, which work according to domain of the product. Each time after extraction of product's review, they carry out following steps: Firstly, identification of feature of product from customer view is done. Next, for each feature, its corresponding opinion's are extracted and their orientation of schism is detected. The final polarity of feature-opinions pairs is calculated. At last, feature based summarization of reviews generated by extracting the relevant excerpts with respect to each feature-opinions pair and placing it into their respective feature based cluster. Results indicate that the proposed methods are highly effective and efficient in performing their task.

[14] Angulakshmi,G. and Chezian,R.M. (2014) discussed tools and techniques of opinion mining in detail. They mine opinions in form of document, sentence and feature level. They discussed supervised and unsupervised machine learning and case base reasoning technique for opinion mining. They discussed various tools to track the opinion and polarity from the user generated contents, Review seer tool on naive bayes approach, Web Fountain on Base Noun Phrase heuristic approach, Red Opal, Opinion Observer.

[15] Yi,J. and Ninlack,W.(2005) described algorithm sentiment mining system and web fountain text analytic platform on which the sentiment miner is running. The proposed prototype system support both research and a set of customers who are involved in use of applications hosted in the environment of production. NLP based sentiment mining system consistently achieved high quality results on various data sets. The performance of the algorithms was verified on online product review article, web pages and news articles.

#### IV. CONCLUSION

In this paper, we have presented a survey of all the opinion mining techniques that has been used to extract the opinions from the social networking sites to identify the opinion of the online users i.e. positive or negative. The papers used in this study describe the importance of opinion expressed in social media in different domains. Various phases of opinion mining such as opinion Extraction, opinion Analysis, and opinion Classification have been presented with their methods. Still the optimized framework to evaluate the text to predict the opinions on the basis of exact feelings/appraisal of the user has to be set up.

#### REFERENCES

- [1] Ganapathibhotla. M and Liu,B., "Mining Opinions in Comparative Sentences", Proceedings of the 22<sup>nd</sup> International Conference on Computational Linguistics, pp.241-248, 2008.
- [2] Pak.A and Paroubek,P., "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", University de Paris-Sub, Laboratoire LIMSIS-CNRS, Batiment 508, France,pp.1320-1326, 2011.
- [3] Liu,B., Hu,M. and Cheng,J., "Opinion Observer:Analyzing and Comparing Opinions on the Web",05 Proceedings of 14<sup>th</sup> International Conference on World Web, pp.342-351, 2005.
- [4] Mishra,N. and Jha, C., "Classification of Opinion Mining Techniques", International Journal of Computer Application,Vol-56,No.13,pp.1-6, 2012.

- [5] Virmani,D., Malhotra,V. and Tyagi,R., “Sentiment Analysis Using Collaborated Opinion Mining”, International Journal of Soft Computing and Engineering,Vol-4,pp.34-37, 2014.
- [6] Hu,M. and Liu,B., “Mining Opinion Features in Customer Reviews”, American Association for Artificial Intelligence 04 Proceedings of the 19<sup>th</sup> National Conference on Artificial intelligence,pp.775-760,2004.
- [7] Yu,H. and Hatzivassiloglou,V., “Towards Answering Opinion Question: Separating Facts from Opinions and Identifying the polarity of Opinion Sentences”, EMNLP’03 Proceedings of the 2003 Conference on Empirical Methods in Natural Processing, pp.129-136,2003.
- [8] Hu,M. and Liu,B., “Opinion Extraction and Summarization on the Web”, American Association for Artificial Intelligence’06 Proceedings of the 21<sup>st</sup> National Conference on Artificial Intelligence, pp.1621-1624, 2006.
- [9] Dave,K., Lawrence, S., Bowman, D. and Pennock, M., “Mining a Peanut Gallery: Opinion Extraction and Sentiment Classification of Product Review”, World Wide Web’03 Proceedings of 12<sup>th</sup> International Conference ACM,NY,USA, pp.519-528, 2003.
- [10] Thawakar,M., Patil,S. and Bamnote,G.,“A Survey on Feature Mining in Customer Review Using Soft Computing”,pp.10-18, 2014.
- [11] Sandhu,R.D. and Mehta,R ,“Applying Opinion Mining To Organize Web Opinions”, International Journal of Computer Science, Engineering and Appliation,Vol.1 ,No.4,,pp.82-89, 2011.
- [12] Chandar Kala,S. and Sindhu,C., “Opinion Mining and Sentiment Classification: A Survey”, ICTACT Journal of Soft Computing,Vol.3,pp.420-427, 2012.
- [13] Bafna,K. and Toshniwal,D., “Feature Based Summarization of Customer Reviews of Online Products”,17<sup>th</sup> International Conference in Knowledge Baed and Intelligent Information and Engineering System,pp.142-151, 2013.
- [14] Angulakshmi,G.,and Chezian,R.M , “An analysis on Opinion Mining: Techniques and Tools”, International Journal of Advanced Research in Computer and Communication Engineering , Vol.3 Issue 7,pp.7483-7487, 2014.
- [15] Yi,J.and Niblack,W.,“Sentiment Mining in Web Fountain”, Proceedings of 21<sup>st</sup> International conference on Data Engineering,pp.1073-1083, 2005.