A Review on Opinion Mining and Sentiment Analysis

Tahura Shaikh
Department of Computer Science,
JNEC College of Engineering,
Aurangabad - 431001, India

Deepa Deshpande, PhD
Assistant Professor,
Department of Computer Science,
JNEC College of Engineering,
Aurangabad - 431001, India

ABSTRACT
Opinion Mining or Sentiment Analysis is a field of data mining. Opinion Mining is a form of Natural Language Processing which is used to record the attitude of people towards a particular subject or product. Mainly Opinion Mining classifies the given review as positive, neutral or negative. Recently Opinion Mining has accomplished much focus due to availability of vast amount of opinion rich web resources in digital form such as discussion forums, review sites, blogs etc. As the use of e-commerce websites is increasing profusely, users not only buy a product on websites but also give their feedback and suggestions that will be beneficial to other users. The collected user reviews are examined, analyzed and organized to make better decision.

The paper reviews the recent research work carried out in the area of opinion mining. It also outlines framework and the steps which are carried out in opinion mining. There are distinct kind of Opinion Mining such as sentence level, document level, and aspect or feature level. It aids consumers in better decision making. For a business it helps to predict brand perception, reputation management, and new product perception. An Organization gets to know their manufacture from perspective of end user. An Opinion can be direct opinion or comparative opinion. Different Machine Learning algorithms like Naïve Bayes, SVM, ANN, Maximum Likelihood, and Decision Tree are used for various tasks which are carried out in sentiment analysis.

General Terms
Data Mining, Opinion Mining, Sentiment Analysis

Keywords
Opinion, Sentiment, Machine Learning Algorithm, Reviews, E-commerce

1. INTRODUCTION
Nowadays, people search for other people's opinions from the Internet before purchasing a product, when they are not familiar about a specific product. With the help of reviews, ratings etc. online data presents useful information to customers for buying a product and for manufacturers to improve the quality of product. Opinion mining is used to identify and select useful information from text. Sentiment may be a person's perception, attitude or interpretation. Sentiment classification is a sub discipline of text classification concerned with the opinion a topic expresses. [11] Sentiment analysis also has different names, among which are opinion mining, sentiment analysis, sentiment extraction, or affective rating. Sentiment analysis is performed to find the semantic orientation of the given review or comment. [3]

Opinion Mining plays a crucial role in social media to obtain opinions expressed by the users. It uses the mode of online feedback forms, emails and different social networking sites such as Facebook, Twitter, LinkedIn, YouTube, Blogs etc. The sentiments are usually studied at the sentence level, document level, and aspect level. [12]. Selection of feature is considered very important in opinion mining. An aspect or a feature is an explicit reference of an entity for which an opinion is given. An entity is the hierarchical representation of components and sub-components. Each component is associated with set of attributes. Opinion can be positive, neutral, or negative. Opinion can be evaluated in two ways viz. direct and indirect opinion. In direct, a review is given directly for a product e.g. camera X is good. In indirect review a comparison is made between two or more products e.g. camera X is better than camera Y.

Various ML algorithms can be used for sentiment analysis tasks. For example "The resolution of this phone is good". In this task, opinion holder is the user who has given the review. Opinion object is the "resolution" of the phone and the opinion word is "good" that is positive in its orientated. The main objective of sentiment analysis is polarity classification. Semantic orientation determines whether a sentence has positive, neutral or negative orientation. Machine Learning system is a system which learns from observations, training, experiences etc. Supervised learning generates a function that maps inputs to desired outputs also known as labels as they are training examples labelled by human experts.[4] Any supervised learning method can be used e.g., Naïve Bayes classification, and support vector machines.

The sentiment analysis can be performed at one of the three levels: the document level, sentence, feature level [14].

1. Document Level Sentiment Classification: In document level sentiment analysis main challenge is to extract informative text for inferring sentiment of the whole document.
2. Sentence Level Sentiment Classification: The sentiment classification is a fine-grained level than document level sentiment classification in which polarity of the sentence can be given by three categories as positive, negative and neutral.
3. Feature Level Sentiment Classification: Product features are defined as product attributes or components. Analysis of such features for identifying sentiment of the document is called as feature based sentiment analysis. In this approach positive or negative opinion is identified from the already extracted features. It is a fine grained analysis model among all other models.

People post their views feedback experience about product then collect the corpus containing views given by people and then the corpus is processed. Pre-processing is done. Feature Extraction is performed to extract the relevant features. The

review given by person is classified as positive negative or neutral by applying machine learning algorithm and the output is given. Because of the huge number of reviews in the form of unstructured text, it is impossible to summarize the information manually. Accordingly, efficient computational methods are needed for mining and summarizing the reviews from corpuses and Web documents. Different tools are available to track the sentiment of users which are Review Seer Tool, Red Opal, Web Fountain, Opinion Observer etc.

2. LITERATURE SURVEY

The aim of Machine Learning is to develop an algorithm to increase the performance of the system using the given data or past experience. The Machine Learning provides a solution to the classification problem that involves two steps:

1. Learning the model from a corpus of training data.

2. Classifying the unseen data based on the trained model. Sentiment analysis of natural language texts is an emerging field. Converting a part of text to a feature vector is an essential step in any data driven approach to Sentiment analysis [13].

Lisette García-Moya, et al. [3] Addresses the aspect-based summarization task by introducing a novel methodology for retrieving product features from a collection of free-text customer reviews about a product or service. Their proposal relies on a language modeling framework that combines a probabilistic model of opinion words and a stochastic mapping model between words to approximate a language model of products. Their work extends a preliminary approach introduced which addresses the modeling of a language of product features from customer reviews. They provide a formalized methodology for the retrieval of product features from the estimated language model of features.

Bing Liu [12] used a distance based approach to extract opinion words and phrases after extracting aspects. To calculate the polarity of each extracted opinion word WordNet was used, proposed a set of techniques for mining and summarizing product reviews based on data mining and natural language processing methods. Objective was to provide a feature-based summary of a large number of customer reviews of a product sold online.

Su Su Htay and Khin Thidar Lynn [3] proposed a unique way which helps to locate opinion words or phrases from customer reviews for each feature in a useful way. By using adjective, adverb, verb, and noun they get the patterns of opinion words/phrases of product feature from the given review. The extracted features and opinions help to generate a useful summary which provides significant informative resource to help the user. It is also useful for merchants to track the best choice of product. A part-of-speech tagger is used which identifies phrases in the given text that contains adjective or adverb or verb or nouns as opinion phrases.

Richa Sharma, et al. [4] proposed an aspect based opinion mining system to classify the reviews as positive, neutral or negative for each feature. In their system negation is also handled. They used unsupervised technique to carry out their work. Dictionary used to determine the opinion words and their synonyms and antonyms is WordNet. All product features on which reviews are given by customers are first identified. Orientation of sentence for every feature is then given. Majority of opinion words helps to find the polarity of the given sentence

Dim En Nyaung and Thin Lai Lai Thein [5] worked mainly on opinion summarization. For Opinion summarization the task of product feature and opinion extraction is very important. To determine polarity and numeric score of all product features Senti-Word Net Lexicon is used. It helps to find intensity of opinions for positive and negative features. Their main objective is to provide a feature-based summary of customer reviews of product which are sold via e-commerce sites.

Madhavi Kulkarni and Mayuri Lingayat [6] proposed a technique which ranks efficiently the products by mining the genuine reviews of the product. System provides a method which allows only those users to write a review about a product who have purchased from the website. Other users are not allowed to give review. This reduced the wrong reviewing of product and customer get reliable product. They have proposed product ranking system which can take product information in query form and system provides the product matching with customer's requirements along with product ranking.

Arti Buche, et al. [7] paper provides information about the data sources that opinion mining uses, machine learning and tasks of sentiment analysis for sentiment classification, text classification and the tools available for sentiment classification, and the performance evaluation. Red Opal tool helps users to recognize products which rely on their features. Opinions of products are given in graph format feature by feature. Review Seer tool is used for automation and aggregation of sites. Naïve Bayes classifier is used to assign score to the extracted features which are done on classified reviews.

Set of inputs like clustering are used for unsupervised learning during training, labels are unknown. To perform classification fixed syntactic patterns which are used to express the opinions are used. To compose syntactic patterns part of speech tags are used. Semi-supervised learning generates an appropriate function or classifier in which both labeled and unlabeled examples are combined.

Aashutosh Bhatt, et al. [8] proposed a system that performs the classification of reviews given by customer which is then followed by finding sentiment for the reviews given. A rule based extraction of the feature for the given product is done. Their system visualizes the review's sentiment which is then presented in the form of a chart. Classification of reviews along with sentimental analysis increased the accuracy of the system which in turn provided accurate reviews to the user. The main aim of the system is to ensure fair results of sentiments and saving time, that user used to spend on reading long textual descriptions of reviews, by providing summarized results in charts. Asmita Dhokrat, et al. [9] gave review about the techniques and tools used in opinion mining. Paper shows the basic requirements of opinion mining to explore the present techniques used to developed a full fledge system. It highlights the opportunities or deployment and research of such systems. The available tools for building such applications are presented with merits and limitations.

Farhan Hassan Khan, et al. [10] proposed a new Twitter Opinion Mining framework to predict the polarity of words into positive or negative feelings in tweets. It improves the accuracy level of the classification. It is constructed using different stages.

As there is tremendous use of Internet nowadays the scope in the area of information retrieval is also expanding. As Opinion Mining is one of the subareas related to information retrieval and knowledge discovery, much research work can be done in this area. Opinion Mining is considered as an interesting area of research due to its various applications. Over the past few years special attention has been given to mining of customer reviews for a given product.

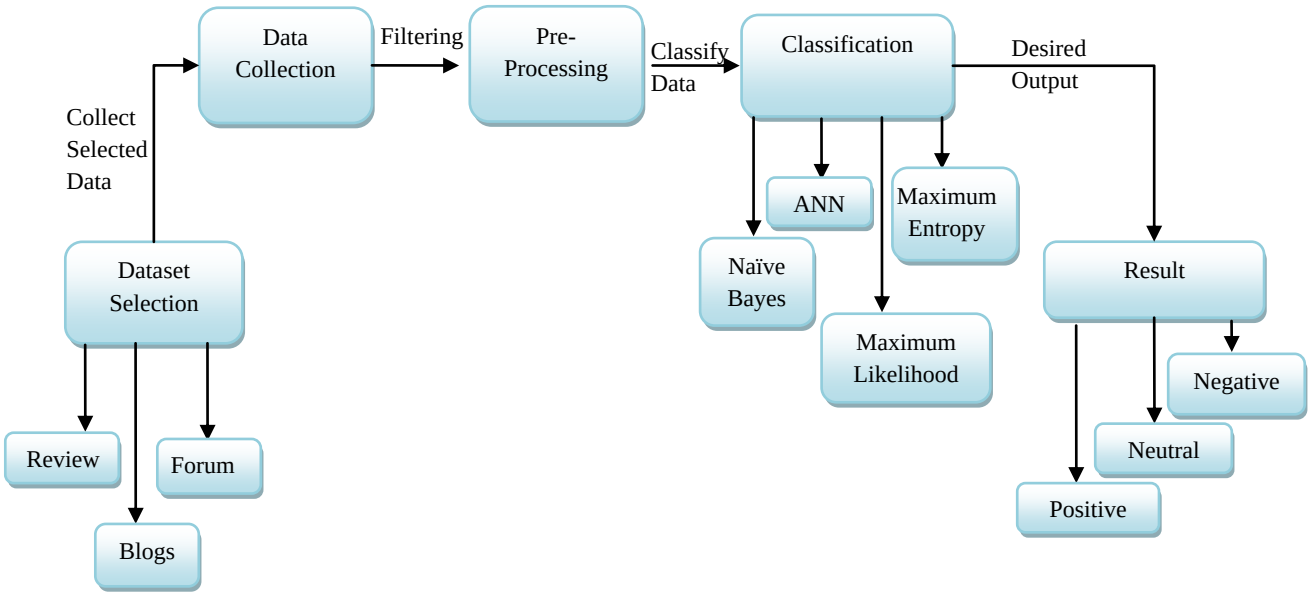

Figure 1: General Opinion Mining Steps

3. STEPS IN OPINION MINING

Polarity or Sentiment classification is a classification task which labels an opinionated document, sentence or feature as positive, negative or neutral. It is a technique for analyzing useful information in a large number of texts. A typical approach for sentiment classification is to use machine learning algorithms. Based on the review of literature, typical steps involved in opinion mining are shown in Figure 1. The different steps are data selection, data collection, pre-processing, classification and result or output.

1. Data-set Selection: Data selection can be done from various data sources such as blogs, online reviews e.g. amazon, yelp, etc., forums and micro-blogging sites e.g. twitter.
2. Data Collection: The reviews are collected from the chosen data-set.
3. Pre-Processing: The collected data is pre-processed or cleaned for analysis to get fair text review. Cleaning of data is done by removal of special characters (such as :“:/.,’#$*^&-) to retrieve best results. Stop words are also removed.
4. Classification: The given Reviews are classified as positive, neutral or negative. ML Algorithms are used for classification. Different ML algorithms which can be used for this step are SVM, Naïve Bayes, Maximum Likelihood, ANN and Decission Tree. By applying any of these algorithms the document, sentence or aspect level can be oriented positively, or negatively.
5. Result: After classification is done then the given review is classified as positive, negative or neutral. Ratings can also be given. The given result can be given in bar or pie chart.

4. CONCLUSION

This paper give brief information on Opinion Mining and the recent literature review putting light on the steps involved in Opinion Mining. Opinion Mining is the field of study that tracks the mood or attitude of people towards a particular product or topic. It helps to find the product with best quality, which is predicted based on the user reviews. It is also used for recommendation system and business intelligence. There are many challenges and future developments possible in Opinion Mining approach like short length and irregular structure of the content such as named entity recognition, parsing, sparsity, abbreviations. Opinion mining is to find the opinion of a person from sentences and then classify them on the basis of polarity. Opinion mining helps people to know the semantic orientation for a product by classifying it as positive, negative or neutral.

Retrospective study of various methods led to observation that current studies on opinion mining still have limitations and scope for improvement. It has numerous problems such as accuracy, scalability, quality, and standard of data. Context dependency, semantic relatedness and ambiguity are the major challenges of Natural Language Processing which makes opinion mining difficult. Another challenge in opinion mining research is non availability or limited accessibility of standard data-set. Mining of opinions from unstructured data and lack of standardized measures for evaluation of results pose further challenge. Thus, Opinion Mining is emerging as a challenging field of research and the future research work can be done in various approaches like irregular structure, short length, poor spelling, punctuation, grammar, incomplete sentences, sarcasm etc.

5. REFERENCES

[1] J. Ashok Kumar1, S. Abirami, 2015,"An Experimental Study Of Feature Extraction Techniques In Opinion Mining", International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.4, No.1.

[2] Lisette García-Moya, Henry Anaya-Sánchez, and Rafael Berlanga-Llavori, 2013,"Retrieving Product Features and Opinions from Customer Reviews", Intelligent Systems, IEEE ,Volume:28 , Issue: 3

[3] Su Su Htay and Khin Thidar Lynn, 2013,"Extracting Product Features and Opinion Words Using Pattern Knowledge in Customer Reviews", Hindawi Publishing Corporation, The Scientific World Journal, Volume 2013, Article ID 394758

[4] Richa Sharma,Shweta Nigam and Rekha Jain, 2014,"Mining Of Product Reviews At Aspect Level", International Journal in Foundations of Computer Science & Technology , Vol.4, No.3.

[5] Dim En Nyaung, Thin Lai Lai Thein, 2014,"Feature Based Summarizing From Customer Reviews", International Journal Of Scientific Engineering and Technology Research, Vol.03, Issue.46.

[6] Madhavi Kulkarni,Mayuri Lingayat, 2015,"Effective Product Ranking Method based on Opinion Mining", International Journal of Computer Applications (0975 – 8887),Volume 120 – No.18.

[7] Arti Buche, Dr. M. B. Chandak, Akshay Zadgaonkar, 2013,"Opinion Mining And Analysis: A Survey", International Journal on Natural Language Computing (IJNLC), Vol. 2, No.3.

[8] Aashutosh Bhatt, Ankit Patel, Harsh Chheda, Kiran Gawande, 2015, "Amazon Review Classification and Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 6.

[9] Asmita Dhokrat, Sunil Khillare, C. Namrata Mahender, 2015,"Review on Techniques and Tools used for Opinion Mining", International Journal of Computer Applications Technology and Research, Volume 4– Issue 6, 419 – 424.

[10] Farhan Hassan Khan, Saba Bashir and Usman Qamar, 2013,"TOM: Twitter opinion mining framework using hybrid classification scheme", Elsevier.

[11] Jayashri Khairnar and Mayura Kinikar , 2013,"Latent Semantic Analysis Method used for mobile rating and review summarization", International Journal of Computer Science and Telecommunications, Volume 4– Issue 6.

[12] Bing Liu, 2012, Sentiment analysis and opinion mining, Morgan & Claypool Publishers

[13] J.Ashok Kumar, S.Abirami, S.Murugappan, 2014, "Performance Analysis of the Recent Role of OMSA Approaches in Online Social networks", CS & IT-CSCP.