An overall aspect of sentiment analysis-A critical review

Niti Dey[1], Sucharita Samanta[1], Souvik Brahma Hota[2,*]
[1]*Department of Computer Science Engineering, Techno India University, Kolkata-700091*
[2]*Department of Mechanical Engineering, Techno India University, Kolkata-700091*

(E-mail: souvik.hota@gmail.com)*

Abstract—Opinion Mining which is popularly known as sentiment analysis is a field of data mining process that helps to understand peoples emotional condition or their attitude towards a particular subject, service or product. Emotions are nothing but the feelings that are created from conscious and unconscious processing and can be expressed in positive, negative or neutral ways. With the immense advancement of wireless technology, the internet gains so much popularity as a place for reviewing a product, exchanging ideas or online learning. For a product or service at least thousands of reviews could be available which make it tough to recognize customer opinion. By applying NLP, text analytics and computational linguistic which mainly classify the opinions polarity subject information extracted from the source materials i.e. opinions of reviewer. Sentiment Analysis is an emerging area of research with a wide variety of use applications like e-commerce, review sizes, online food ordering, online cab service etc. This paper provides and overall aspect about sentiment analysis and its recent trends. Brief idea about machine learning algorithms like ANN, Naïve Bayes, SVM and decision tree are discussed in this review for better understanding of the context.

Keywords— *Sentiment Analysis; Opinion Mining; NLP; ANN; Naïve Bayes; SVM*

I. INTRODUCTION

Modern day's people search for others reviews or comments from online before choosing any service or purchasing any product, when they are not fully familiar about that service or product. With the help of others opinion, reviews, rating etc. available in internet customers decide to go for that product or service. Sentiment analysis or opinion mining may be analyzing a person's perception, interpretation or attitude which is nothing but gathering of human understanding and electronic knowledge for mining the content of the reviews and arranging reviewer's conceptions for believe, likes, dislikes and demands [1].Sentiment analysis plays a tremendous role in social media too get reviews or opinions expressed by the users. It often use the mode of online feedback forms, online popup for feedback or review, customer retention form, emails ,messages and different social network such as twitter, facebook , LinkedIn various blogs.

Basically opinion mining is done to deduce the logical orientation of the given opinion [2].Sentiment analysis categorized reviews or specifically the words into three classes i.e. positive, negative and implied class of neutral. Opinion expressed in a review statement could be either a direct opinion or a comparative opinion. In expressions of a direct opinion there should be some target objects like products, services, topics. For an example "The test of the chocolate is great". For the second case the expression of opinion is nothing but stating differences or similarities of more than one product. E.g."Veg burger is cheaper than chicken burger". Comparative expressions are of different types i.e. equative (same) or superlative (greater than) and or non-equal great(less than).Sentiment analysis can be performed in three levels: document level, sentence level and aspects level [3].

- **Document level:** In document level sentiment analysis a review is classified on the overall sentiment expressed by the reviewer.
- **Sentence level:** Sentence level sentiment analysis is delicate than document level sentiment analysis in which sentence's polarity is classified as positive, neutral and negative.
- **Aspect level:** Aspect level sentiment analysis or feature level sentiment analysis is analysis of product attributes to identify opinion of the review. It is the most sophisticated analysis model among the other model.

Machine learning system is an architectural set up of advance computer science which learns from training, experiences etc. that are gathered during the operation. Any machine learning algorithm like Naïve Bayes, SVM, ANN, and Decision Tree can be used. There are different tools available to trace the sentiment of the reviews' which are Red Opal, Web Fountain, Review Seer Tool and opinion observer.

II. LITERATURE SURVEY

Movie review mining by using machine learning and semantic orientation has been investigated by Lina Zhou et al. (2005)[4].They used supervised classification and text classification techniques in their proposed machine learning approach for classification of movie review. All the classifiers are trained using a corpus in their model which makes the proposed technique more efficient. Results showed that the supervised classification approach got 84.49% of accuracy whereas an accuracy of 66.27% on hold-out samples is

achieved. Semantic approach achieved 77% accuracy in movie reviews. Therefore, the study concluded with more accuracy of supervised machine learning than semantic orientation approach which makes supervised machine learning more practicable to mine opinions from the unstructured data automatically.

Liang-Chih et al. (2013) [5] reported that extraction of opinions depending on features of content is the most vital task of sentiment analysis. They proposed a new method to deal with feature-level sentimental analysis problems. Extraction of features for the whole review is a significant task to be performed. The new technique uses the relating opinion words for extracting the features and in accordance to the mutual support it filters the noise. However this method has some pit falls.

Wilson T (2005) [6] has figured out the nature of opinion expressions may not be really subjective for all the time. He has find out sentence level or document level classes of sentiment analysis did not provide the needed information. So, one must choose aspect level for better accuracy in opinion mining.

Zhu et al. (2011) [7] observed feature based sentiment analysis from customer reviews. They have proposed a feature based or aspect based segmentation model that segments multi feature sentence into single feature unit. This single feature unit was used for sentiment analysis. They have tested a real Chinese restaurant reviews by applying sentiment analysis algorithm resulted 77.5% accuracy in feature level sentiment analysis. They have suggested the same method can be applicable for movie reviews also.

Agarwal et al. (2005)[8] reported a machine learning method that used linguistic knowledge gathered through synonymy graphs. This method has much efficacy and can be an effective sentiment classification too that shows degree of influence among relationships of documents have on their opinion mining. Their experiment showed an accuracy of 90% and more.

Sharma et al. (2014) [9] stated an aspect level sentiment analysis system for classification of the reviews according to its polarity i.e. positive, negative or neutral. They have used unsupervised technique as their methodology. With the help of WordNet the determination of the opinion words and their synonym and antonyms has been done. Majority of the review words aids to find out the sentence's polarity.

Buche et al. (2013) [10] stated some fact about machine learning. They also reported about the data source which has been used by sentimental analysis and tasks of opinion mining for classification of sentiment. They have reviewed the tools available for opinion classification and their performances. Some of the tools they have highlighted are Red Opal Tool, Review Seer Tool, and Naïve Bayes.

Bhatt et al. (2015) [11] reported a system that performs as a classifier of opinions given by reviewer that is followed by fetching sentiment for the review statement given in online. Classification of opinions and opinion mining process improved the accuracy of the process which helped to supply correct review to the user. They also stated about the available

tools for making such applications with its advantage and limitations.

Pang (2008) [12] has analyzed the concept behind sentiment analysis. According to him sentiment analysis on online reviews gained much popularity and is a multidisciplinary research in nature which comprises multifarious areas like natural language processing (NLP), information recovery, computational language rules and artificial intelligence (AI).

Khan et al. (2013) [13] reported a new twitter sentiment analysis frame work for predicting word's polarity in a review. Their model has improved accuracy level of the sentiment classification. They have constructed their analysis model using various stages.

III.   MACHINE LEARNING ALGORITHMS

Machine Learning is a key tool to perform sentiment analysis. Extraction of sentiments from a review is not an easy task to be performed always. Here machine learning helps to ease the task. [14] Various machine learning algorithms are available with their distinct function to perform their desired task for sentiment analysis. Some of those algorithms are described below:

A. Artificial Neural Network (ANN):

An artificial neural network acts the same way how the neurons of a human system work. It generally has three layers [15]:

- **Input layer:** It text input.

- **Hidden layer:** It processes the input.

- **Output layer:** it sends the derived output.

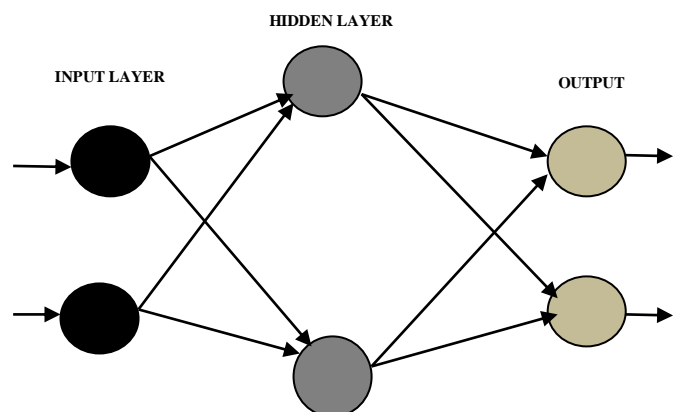A schematic representation of ANN is presented in fig.1.

Fig.1.Schematic diagram of Artificial Neural Network

Artificial neural networks are of three types [16]:

- **Supervised Neural Network:** Here output of the system is already known.

- **Unsupervised neural network:** It has no prior information about the output.

- **Reinforced neural network:** This network acts like a human communicated with the surrounding. This network also has no previous information about the output. A schematic diagram of reinforced neural network is presented in fig.2.
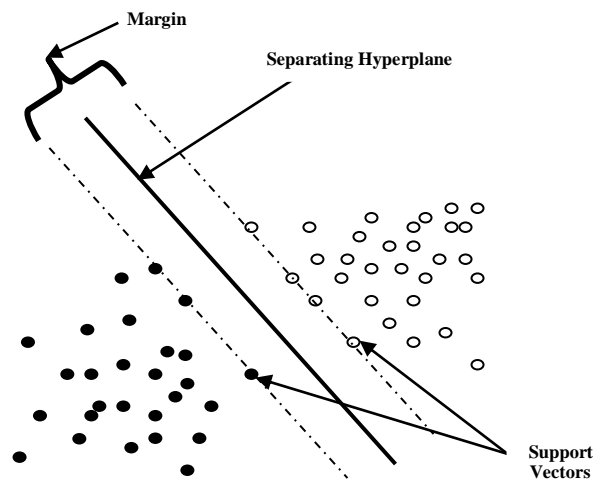

Fig.2.Schematic diagram of reinforced neural network [16]
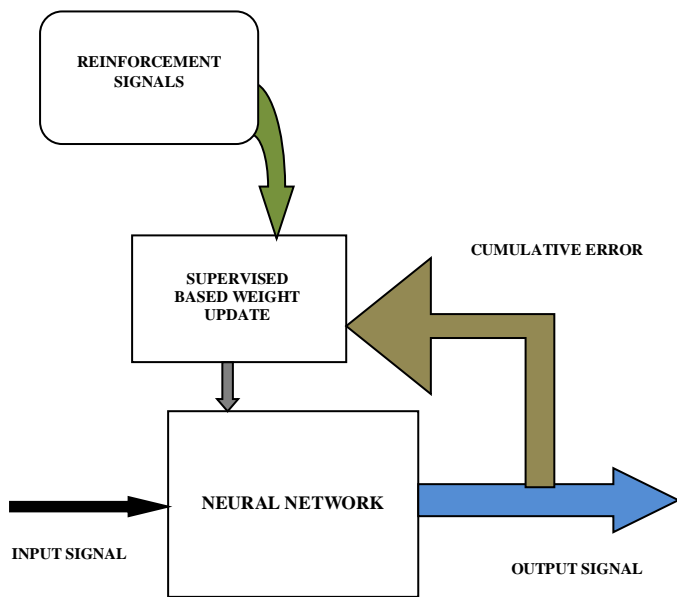
B. Naïve Bayes:

This algorithm is a supervised machine learning algorithm which targets the text classification job. Its main purpose is to do clustering and classification [17]. The purpose of Naïve Bayes architecture mainly is to create trees depending on their provability of happening. These trees are also called as Bayesian network.

C. Support Vector Machine (SVM):

Another popular supervised machine learning which is likely to be called as state of art machine learning technique is support vector machine. Support vector machine is generally used to classify which is done by principle of margin calculation. SVM use to separate the classes by drawing margins. A schematic diagram of support vector machine's working is represented in fig.3.

Fig.3.Schematic diagram of support vector machine [18]

D. Decision Tree:

The purpose of using decision tree is to make a training model that can predict value or class of target variables with the help of decision rules deduced from prior data. Basically it learns some kind of patterns like the values from the training data set and implements them or feed them to the test data set. Each tree consists of nodes and branches in which nodes represents feature in group that is to be classified and each and every branches is nothing but representation of a value that a node can take [19].A representational example of decision tree is shown in fig.4.
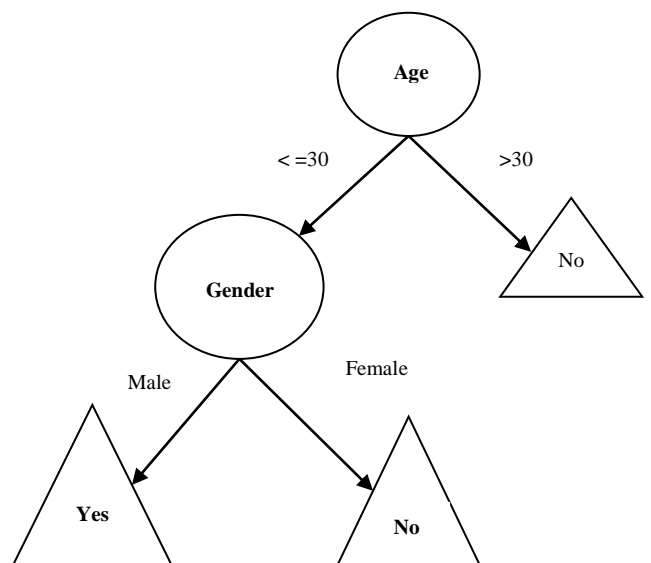

Fig.4. Schematic representation of decision tree [20]

IV.   OPINION MINING PROCESS

For opinion mining classification of sentiment is a vital step to be performed which is a task that classify and reviewed document, sentence or feature according to their polarity i.e. positive, neutral or negative. It is a handy process to analyze useful information available in a large number of tests. Machine learning is the never failing fiend for execution of the sentiment classification. According to the data available from various literatures, research work in the field of sentiment analysis, we can generalize the whole process as a chain of some sequential   steps as follows:

- **Selection of data set:** selection of data can be done perform from multiple sources like online reviews (e.g. Amazon, Flipkart, Zomato, Food Panda etc.), blogs, social media (e.g. facebook, twitter, YouTube etc.).

- **Collection of data:** The opinions are collected from some specific data set.

- **Preprocessing:** After data collection has been done, the collected data is proposed for analysis. In this process the collected is actually cleaned to gate fair review. Cleaning is nothing but removing special character such as &, @, #, $ etc. For getting the best result stop word are also been removed in this step.

- **Classification:** In this step the reviews or opinions are classified according to their polarity i.e. positive, negative or neutral .Machine learning algorithm are used in this step.

A brief description of various machine learning algorithms are already been discussed in this paper. Implementation of any of these algorithms is needed for a good classification. This classification may be document level, sentence level or aspect level. A general diagram containing opinion mining steps is shown in the fig.5.
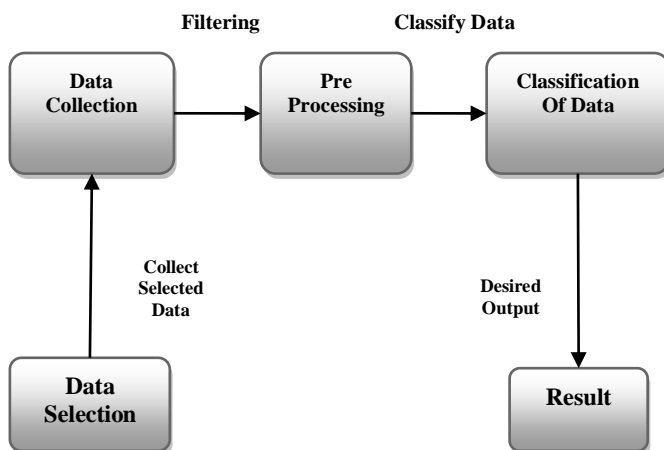

Fig.5. General steps in sentiment analysis

V.   CONCLUTION

Opinion mining or sentiment analysis not only focuses on text mining but also it retrieves information. This review gives an overall idea of sentiment analysis and the steps involved in it. In sentiment analysis determination of sentiment is much easier while determining object with their corresponding features is much harder. Thus, combining these two considerations the job of sentiment analysis seems to be very long process which needs desired level of accuracy. Major challenge in this process is selection of correct machine learning algorithm which plays a vital role to get excellent accuracy level. Sentiment analysis is an amazing field of research which tracks the mood of a customer towards a particular product or service. It assists to find a best product that is predicted from the user review. That is why use of opinion mining in the field of E-commerce is increasing day by day. Though there are numerous problems like accuracy, quality, scalability and standard of data associated with opinion mining. Other challenges in sentiment analysis are limited accessibility of standard data set, context dependency of natural language processing etc. which makes it difficult. Opinion mining from unstructured data is also a great challenge. As a concluding remark it can be stated that sentiment analysis is a challenging field of study and its future trend depends on the upcoming E-commerce based uses. Further research can be initiated for areas like poor spelling, grammatical mistake, incomplete sentences etc.

ACKNOWLEDGEMENT

REFERENCES

1. Kaur, A., & Gupta, V. (Nov. 2013). A survey on sentiment analysis and opinion mining techniques. Proceedings of Journal of Emerging Technologies in Web Intelligence , Vol.5, No. 4. 62.

2. 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

3. 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.

4. Lina Zhou, Pimwadee Chaovalit, (2005), "Movie Review Mining: a Comparison between Supervised and Unsupervised

Classification Approaches", Proceedings of the 38th Hawaii International Conference on system sciences.

5. Yu Liang-Chih, W. J.-L.-C. (2013). Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock. KnowlBased Syst , (pp. 41:89–97).

6. Wilson T, W. J. (2005). Recognizing contextual polarity. In: Proceedings of HLT/EMNLP.

7.Zhu, Jingbo Wang, Huizhen Zhu, Muhua Tsou, Benjamin K. Ma, Matthew, (Jan-June 2011),"Aspect-Based Opinion Polling from Customer Reviews", IEEE Transactions on Affective Computing, Volume: 2,Issue:1 On page(s): 37.

8. Alekh Agarwal and Pushpak Bhattacharyya,(2005)"Sentiment analysis: A new approach for effective use of linguistic knowledge and exploiting similarities in a set of documents to be classified", In Proceedings of the International Conference on Natural Language Processing (ICON).

9. 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

10. 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.

11. 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.

12. Pang B, L. L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval , 1-2 (2), 1-135.

13. Farhan Hassan Khan, Saba Bashir and Usman Qamar, (2013),"TOM: Twitter opinion mining framework using hybrid classification scheme", Elsevier.

14. W. Richert, L. P. Coelho, "Building Machine Learning Systems with , Packt Publishing Ltd., ISBN 978-1-78216-140-0

15. V. Sharma, S. Rai, A. Dev,(October 2012) "A Comprehensive Study of Artificial   Neural Networks", International Journal of Advanced Research in computer Science and Software Engineering, ISSN 2277128X, Volume 2, Issue 10.

16. S. B. Hiregoudar, K. Manjunath, K. S. Patil,(May, 2014) "A Survey: Research summary on Neural Networks", International Journal of Research in Engineering and Technology, ISSN: 2319 1163, Volume 03, Special Issue 03, pages 385-389.

17. D. Lowd, P. Domingos, "Naïve Bayes Models for Probability Estimation.

18. D. Meyer, (August 2015)"Support Vector Machines – The Interface to libsvm in package e1071"

19. S.B. Kotsiantis, (2007) , "*Supervised Machine Learning: A Review of Classification Techniques*", Informatica 31 249-268

20. L. Rokach, O. Maimon, "*Top – Down Induction of Decision Trees Classifiers – A Survey*", IEEE Transactions on Systems.

Niti Dey has completed her B.tech in CSE from Maulana Abul Kalam Azad Technical University in 2015.Currently she is just about to finish her M.Tech from Techno India Univesity.Her research area includes Python,NLP,Machin Learning, Big Data,DatabaseManagement System, Cryptography etc.She has a keen interest to continue her research in opinion mining area.

Sucharita Samanta has completed her B.tech in CSE from Maulana Abul Kalam Azad Technical University in 2016.She is currently pursuing M.Tech in CSE from Techno India University. Her areas of research are JAVA, Cryptocurrency, NLP, Machine Learning, Python etc. She is involved in her M.Tech Research on Text Summarization Techniques.

Souvik Brahma Hota has completed his Aircraft Maintenance Engineering on Radio Navigation specialization in 2011.He has pursued his B.E. in Aeronautical Engineering from Anna University with university rank in 2016.He is on the verge of completing his M.Tech in Mechanical Engineering from Techno India University, West Bengal. His area of interests are Artificial Neural Network, Deep Learning Model, Internet of things, Robotics, Modern machining, Computational Fluid Dynamics etc. He is currently involved in surface coating techniques oriented research work.