

# A STUDY AND SURVEY ON SENTIMENT ANALYSIS

Mohd Shah Alam\*, Hari Om Sharan

Dept. of Computer Science and Engineering, FET, Rama University, Kanpur, India

**Abstract** - Opinion mining and sentiment analysis is rapidly growing area. Sentiment analysis or Opinion mining is one of the fastest growing fields with its demand and potential benefits increasing every day. With the onset of the internet and modern technology, there has been a vigorous growth in the amount of data. Everyone is able to express his/her own ideas freely on social media. All this data can be analyzed and used to draw benefits and quality information. One such idea is sentiment analysis, here, the sentiment of the subject is considered, and necessary information is drawn out whether it be a product review or his/her opinion on anything materialistic. One way to extract information is text mining and sentiment analysis, that include: data acquisition, data pre-processing and normalization, feature extraction and representation, labelling, and finally the application of various Natural Language Processing (NLP) and machine learning algorithms. This paper provides an overview of different methods used in text mining and sentiment analysis elaborating on all subtasks.

**Keywords** –Feature Extraction, Sentiment Analysis, Opinion Mining, Machine Learning

## I. INTRODUCTION

In this paper, a detailed analysis of various works related to sentiment analysis is taken. Sentiment analysis has its applications in numerous fields. Wherever there is a data attached to a subject there is always a sentiment. Sentiment analysis is usually considered to have either a positive, neutral or a negative result. Much research exists on sentiment analysis of user opinion data, which mainly judges the polarities of user reviews. In these studies, sentiment analysis is often conducted at one of the three levels: the document level, sentence level, or attribute level. In relation to sentiment analysis, the literature survey done indicates two types of techniques including machine learning and semantic orientation are important [1]. These techniques are shown in figure 1. There are several challenges in Sentiment analysis. The first is an opinion word that is positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in a same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very different from "the picture was not great". People can be contradictory in their statements [2]. Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time.

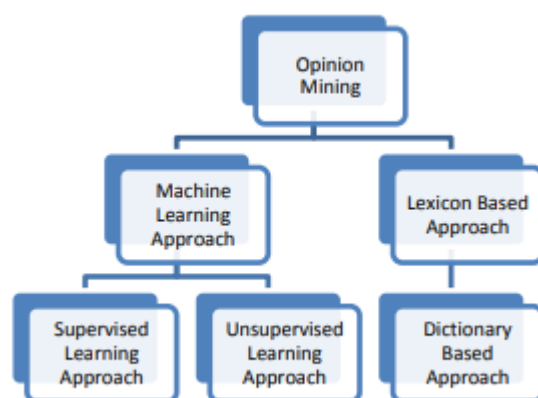


Fig. 1. Text Mining Techniques

## II. BACKGROUND AND LITERATURE SURVEY

The Machine Learning provides a solution to the classification problem that involves two steps:

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

B. 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 [3].

Lisette García-Moya, et al. [4] 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.

Richa Sharma, et al. [5] 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.

Dim En Nyaung and Thin Lai Lai Thein [6] 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.

Pak and Alexander et al. proposed [7] By using the corpus, Author builds a sentiment classifier, which is capable of determining positive, neutral and negative sentiments for the whole document. Experimental results show that the proposed techniques are more efficient and perform better as compared to previously proposed techniques.

Ortigosa and Alvaro et. al [8] proposed a novel method for sentiment analysis in social site giant Facebook that, starting from the messages written by its users, supports: (i) to extract useful information about the Facebook users' sentiment polarity (whether it is positive, neutral or negative), which reflected from the messages written by users; and (ii) to model the users' normal sentiment polarity and to analyze significant emotional changes in user.

Horakova and Marketa et al. [9] Present a model which collects tweets from social networking sites and thus provide a view of business intelligence. In our framework, there are two layers in the sentiment analysis tool, the data processing layer and sentiment analysis layer. Data processing layer deals with data collection and data mining.

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.

In [11] Kouloumpis et. al demonstrated the usefulness of linguistic features and existing lexical resources used in micro-blogging to detect the sentiments of twitter messages. From this paper the researchers concluded that microblogging features were more useful as compared to POS (Part-of-Speech) features and features from existing sentiment lexicon.

In [12], Tripathy et. al represented that the reviews and blog datasets obtained from the social networking sites were unsystematic and need classification for a meaningful information. They could be classified as positive, negative and neutral with the help of supervised machine learning methods.

In [13] Zheng et. al paid attention to the Chinese online reviews since they were directly affected by feature selections which includes n-char-grams and n-pos-grams as potential sentiment features. To select feature subset and to count feature weight enhanced document frequency method and Boolean weighting methods were used.

In [14], Basha et. al represented that because of the popularity of E-commerce product reviews for a product were also growing rapidly with an exponential factor. To decide among multiple option where time and money were precious, other people opinions would play an important role.

In [15], Alomari et. al represented that Arabic tweets pose a good opportunity for opinion mining research, but they were delayed due to shortage of sentiment analysis resources or challenges in Arabic language text analysis. It included Arabic Jordanian twitter corpus in which either the tweets were denoted as positive or as negative and these tweets were examined using different supervised machine learning approaches. For using different weight schemes, stemming and n-grams techniques experiments were conducted which showed that SVM classifier using TF-IDF through bigrams feature was better as compared to Naive Bayesian classifier.

### III. APPLICATIONS OF SENTIMENT ANALYSIS

- A. Business and e-commerce applications, such as product reviews and movie ratings [1]
- B. Opinions in the social and geopolitical context
- C. Predicting stock prices based on opinions that people have about the companies and resources [1]
- D. Determine areas of a product that need to be improved by summarizing product reviews to see what parts of the product are generally considered good or bad by users [2]
- E. Customer preference

### IV. CONCLUSION

Sentiment analysis or opinion mining is the study that is used to analyze people emotions, sentiments towards the product. In this paper, the latest developments in sentiment analysis are reviewed and the future possibilities for each of these developments are presented. In this survey we found that social media like twitter can be used to predict the sentiments of people. Sentiment analysis is a field which is catching up in the recent years and its applications are subject to increase to a broader range in near future.

### REFERENCES

- [1] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment Analysis Algorithms and Applications: A Survey", Ain Shams Engineering Journal, 2014.
- [2] Blessy Selvam1 , S.Abirami2, "A Survey On Opinion Mining Framework", International Journal of Advanced Research in Computer and Communication Engineering, 2013.
- [3] J.Ashok Kumar, S.Abirami, S.Murugappan, 2014, "Performance Analysis of the Recent Role of OMSA Approaches in Online Social networks", CS & IT-CSCP.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." LREC. Vol. 10. 2010.
- [8] Ortigosa, Alvaro, José M. Martín, and Rosa M. Carro. "Sentiment analysis in Facebook and its application to elearning." Computers in Human Behavior 31 (2014): 527-541.
- [9] Jebaseeli, A. Nisha, and E. Kirubakaran. "A Survey on Sentiment Analysis of (Product) Reviews." International Journal of Computer Applications 47.11 (2012).

- [10] Farhan Hassan Khan, Saba Bashir and Usman Qamar, 2013, "TOM: Twitter opinion mining framework using hybrid classification scheme", Elsevier.
- [11] Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! *Icwsm*, 11 (538-541), 164.
- [12] Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, 57, 117-126.
- [13] Zheng, L., Wang, H., & Gao, S. (2018). Sentimental feature selection for sentiment analysis of Chinese online reviews. *International journal of machine learning and cybernetics*, 9 (1), 75-84.
- [14] Basha, S. M., Zhenning, Y., Rajput, D. S., Iyengar, N., & Caytiles, D. R. (2017). Weighted Fuzzy Rule Based Sentiment Prediction Analysis on Tweets. *International Journal of Grid and Distributed Computing*, 10 (6), 41-54.
- [15] Alomari, K. M., ElSherif, H. M., & Shaalan, K. (2017, June). Arabic Tweets Sentimental Analysis Using Machine Learning. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 602- 610). Springer, Cham.