REVIEW ON SENTIMENT ANALYSIS TECHNIQUE (OPINION MINING)

Sashi Jangra

shashijangra101@gmail.com

M. Tech Scholar, Department of Computer Science & Engineering, BRCM CET, Bahal, (Haryana)

Mrs. Neha

neha@brcm.edu.in

Assistant Professor, Department of Computer Science & Engineering, BRCM CET, Bahal, (Haryana)

ABSTRACT

In this research paper, Review are described over previous research on sentiment analysis technique after reading many authors research over sentiment analysis technique all authors have done research on their perspective. One thing common in all that is all have taken a dataset ,preparing modal for applying on dataset, taken an algorithm like Naïve Bayes, Linear Regression, Support Vector Machines, Deep Learning etc. All have described that Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. All have given modals and their result over sentiment analysis how we can analyse text over social network by the help of machine learning. Analysis of sentiment is widely used in mining of subjective information from internet content using various techniques including NLP, Statistical techniques and Machine Learning methods. The application of Sentiment Analysis is very powerful and broad field which mainly focuses on huge text analysing process rather analysing manually. The Opinion Mining technology is used for extracting social data and makes the business process profitably. This survey describes about the research accomplished in Sentiment Analysis in last decade to show how the problems have how the problem is solved until.

*Keywords: Sentiment Analysis Techniques, NLP, Machine Learning, Accuracy.*

INTRODUCTION

The emotions in textual data is analysed and processed in Sentiment Analysis. In other words the Sentiment Analysis specifies that the given information is positive, negative or neutral about a specific topic or product. Due to this scenario it is widely termed as Opinion Mining. For processing the textual information Sentiment Analysis adapts the approaches of NLP (Natural language processing), AI (Artificial Intelligence) and ML (Machine Learning). Conclusively analysis of sentiment deals with the people opinion about a particular topic or product. Sentiment Analysis is competent in understanding the people's opinion and to provide a basis of evidence and reasoning on particular product. Sentiment Analysis analyses abundant textual information presented in internet to provide future insight for the organization and aid the public to take decision on their purchase. For instance, in an e-commerce website, people purchase product and give their reviews with or without using the product. These reviews must help the customers who desire to purchase the product or item. But the issue is, as the number of reviews is more the customers may find it difficult to read all reviews. So, there is a need of automated process to give an appropriate conclusion for

particular product or topic and this task is known as Sentiment Analysis. □

Case Normalisation: - The entire documents are changed to lower case or upper case.
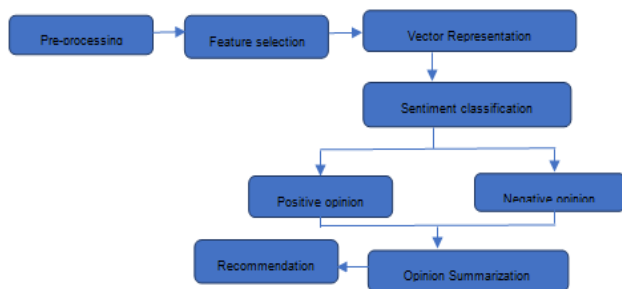
Figure 1: Framework for Sentiment Classification

a) **Preprocessing**: In Preprocessing the unanalyzed data is handled for feature extraction. It is further divided into below step:

- **Tokenisation:** White spaces, symbols and special characters are removed and a sentence is divided into words.
- **Stop Word Removal:** Articles are removed.
- **Stemming:** Token or words are reduced for root forms.

b) **Feature extraction:** Feature extraction handles the following task:

- **Feature Type:** In this step features are identified like the term frequencies, term co- occurrences, Opinion word, OS information, Negation Syntactic Dependencies.
- **Selection of Feature:** Good features are selected for classification using the following ways like Information gain, Document frequency, Odd ratio and Mutual Information.
- **Feature Weighting Mechanism:** The features are ranked by computing the weight using term presence, term frequency and Inverse document frequencies.

- **Reduction of Feature:** To optimize the classifiers performance the vector size is reduced.

c) **Sentiment Analysis:** Polarity of text is classified by Sentiment Analysis. This process is done in 3 different levels.

d) **Sentiment Classification:** Sentiment classification uses two approaches to classify the nature of documents/sentence.
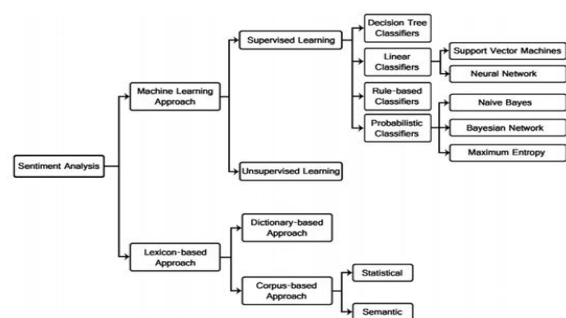
Figure 2: Sentiment classification techniques.

PARAMETERS AND MEASUREMENTS IN SENTIMENT ANALYSIS

The pervious review and various sentiment analysis technique are discussed in this section. The sentiment of the reviews is identified by approaches (1),(2),(3),(4),(5),(6),(7),(8). The accuracy of the classifier is increased by reducing the noise in textual data by pre-processing it. Bo Pang and Lillian Lee proposed that Naïve Bayes and SVM is the efficient method for providing highest accuracy. Xiangjie Kong, Huuizhen Jiang, Zhuo Yang, Zhhenzhen Xu, feng Xia, Amr Tobla [10] embraced Random Walk model for giving highest accuracy in academic domain. Huakang Li, Yixiong Bian, Xiuying Xu, Guozi Sun suggested Monte Carlo decision tree algorithm for mining interest similarity with highest performance when compared with other techniques. G. Vinodhini proposed NB is the best technique for estimating the quality of the document. Jayashri Khairnar and Mayura Kinikar put forward that SVM excelled when comparing with other techniques in sentiment classification. Faruk and Arnab initiated a model for trust

management with highest accuracy. Rudy Prabowo explored a new approach for improving the performance of the classifier. Dongjoo Lee et al says the when dealing with huge volume of data, PMI give better accuracy. Soudamini Hota, Sudhir Pathak recommended SVM and KNN as a best method for handling noisy datain textual information.

Algorithms used in research work:

There are various Algorithms used in research these are following:

- **Rule Based: -** It extract knowledges in the form of rules from the classification model, which are easy to comprehend and very expressive. It Decrease in Recall rate, difficult to list the rules.

- **Naive Bayes: -** It is carried out to identify the polarity of the textual information. It assumes only uncongenial attribute.

- **Support Vector Machine (SVM):-** It is used to split the data points of the classes from one another. In SVM, Transparency of the result is lacking.

- **Decision Tree: -** It has the potential of describing the decision-making knowledge from the given data. It is not efficient for regression test and for predicting continuous values.

- **Convolutional Neural Network (CNN): -** It classifies unorganised data in textual information. It also Loses Phrase level labels.
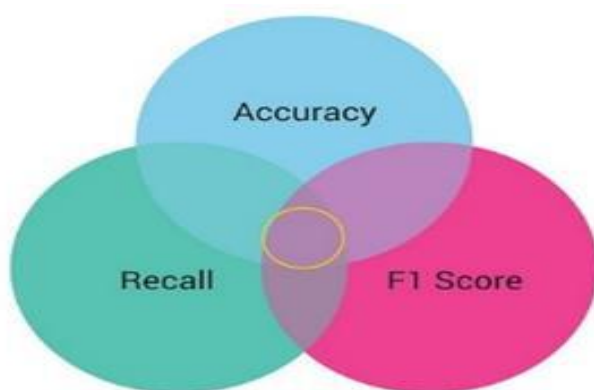
Figure.2: Factors in Sentiment Analysis.

CONCLUSION

Sentiment Analysis Technique (Opinion mining) is a field where a large volume of data is being generated through person-to-person communication. Opinion Mining in market analysis and teaches what changes is necessary for fore coming product generation with the help of historical market data. The product designing and development may further construct in an efficient way using Opinion mining. Usually customers use Opinion Mining platform for purchasing a product which they never used before. The customization and review process reveal the product in an efficient way and makes the customer satisfaction level high through Opinion Mining. This paper presented the various parameters and measurements for determining the efficiency of the classifier and concludes classifier alone cannot give complete efficiency in accuracy since the result is based on number of factors.

FUTURE OUTLOOK

There is huge scope of sentiment analysis technique in future because it is no doubt that there are need of many machine learning algorithms which can analysis the sentiment of written text, deployed picture over the social media and internet it may be about of movie, social sites etc. Some algorithms have been designed which is based on NLP and machine learning that will be used to analysis the sentiments of text, sentences, picture which all work is done by machine learning. Volume of data over the internet is being increased day by day that is why we need efficient algorithm and modal to analysis that large amount of data.

REFERENCES

[1] E. Kouloumpis, T. Wilson, J. Moore, Twitter sentiment analysis: The good the bad and the omg!,Proc. 5th Int. AAAI Conf. Weblogs Social Media, pp. 538-541, 2011.

[2] D. Terrana, A. Augello, G. Pilato, Automatic unsupervised polarity detection on a Twitter data stream,Proc. IEEE Int. Conf. Semantic Comput., pp. 128-134, Sep. 2014.

[3] H. Saif, Y. He, M. Fernandez, H. Alani, Semantic patterns for sentiment analysis of Twitter,Proc. 13th Int. Semantic Web Conf., pp. 324-340, Apr. 2014.

[4] H. Saif, Y. He, H. Alani, Alleviating data sparsity for Twitter sentiment analysis,Proc. CEUR Workshop, pp. 2-9, Sep. 2012.

[5] H. G. Yoon, H. Kim, C. O. Kim, M. Song, Opinion polarity detection in Twitter data combining shrinkage regression and topic modeling,J. Informetrics, vol. 10, pp. 634-644, 2016.

[6] F. H. Khan, U. Qamar, S. Bashir, SentiMI: Introducing point-wise mutual information with SentiWordNet to improve sentiment polarity detection,Appl. Soft Comput., vol. 39, pp. 140-153, Apr. 2016.

[7] A. Agarwal, B. Xie, I. Vovsha, Sentiment analysis of Twitter data,Proc. Workshop Lang. Social Media Assoc. Comput. Linguistics, pp. 30-38, 2011.

[8] Bo Pang, Lillian Lee, A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts, ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, Article No.271, 2004.