Words Reflect Man - A Review on Opinion Mining

Nidhi N. Solanki[1], Dr. Dipti B. Shah[2]

[1]Department of Computer Science, M.K. Institute of Computer Studies, Bharuch, Gujarat, India
[2]G.H.Patel Post Graduate Department of Computer Science and Technology, Sardar Patel University,
V.V.Nagar, Gujarat, India

ABSTRACT

Opinion mining plays a great role to understand the customers more whether he is happy or not. Today's formula of success is the satisfactory customer. Users express their opinion on various social sites. This paper describes a brief overview of techniques, challenges, and the basic flow of the opinion mining process. Less work is done on code mix language. Unstructured data and lack of the right algorithms and packages result in accuracy compromise. The development of an optimal model will help in providing better services to viewers and empowering relationships.

Keywords : Opinion Mining, Machine Learning, Deep Learning, Lexicon, Code-mix language.

I. INTRODUCTION

Opinion mining which is also synonymous with sentiment analysis is the study of estimating the views of netizens from their written comment's sentences, which can be given in a single or hybrid language. Happy customer results in happy developer, mining helps in recognizing whether a user is pleased, unpleased, or in middle. Opinion mining is one of the promising choices of many researchers because of its applications like choice selection, future planning, and business intelligence.

A. levels of opinion mining
Fundamentally Opinion mining is done on the following three levels. [12]

1) Document Level:
It is done entirely on a single subjective paragraph. [2; 10]

2) Sentence Level:
It categorizes the opinions into positive negative or neutral from the user's expressions, and emotions. [4]

3) Aspect level:
It provides more integrated analysis over document and sentence level. [13]

II. LITERATURE SURVEY

Muhammad Zubair Asghar[5] states that detection of opinion polarity of YouTube comments is a challenging task for researchers. He presented the problems of opinion dictionaries, informal language

style, community-created words, labeling, and classification performance.

The work by Sonali Rajesh Shah [1] found that when a machine learning or lexicon-based approach is used with deep learning techniques, the accuracy of the opinion mining model is improved. She founds BERT and ELMo classifier removes the complexity and cost problem of deep learning. Her work used 10 supervised learning algorithms and cross-validation, parameter tuning, and statistical testing for result confirmation. For the Marglish dataset, MLP worked best with all vectorizers with an accuracy of approx 64%. For the Devanagari dataset, BNB performed best with the count Vectorizer. Work concluded MLP works well on a large dataset. Data has been labeled manually into 5 different categories. Count Vectorizer is the best one according to the study.

Shilpi chawla[7] has listed 4 features of opinion mining Part of speech information, Term presence and their frequency, Negations, and Opinion words, phrases. The author has performed opinion mining of smartphone reviews obtained from Amazon. In his study SVM classifier proved far better than Naïve Bayes with an accuracy of 90%.

Gagandeep Kaur[6] has proposed a model for opinion mining of Hinglish YouTube comments using a semi-supervised approach, where two top Indian cookery channels were considered. The work proved that the Logistic regression classifier worked well with the term frequency vectorizer in both datasets. They have used nine machine learning algorithms including both parametric and non-parametric types.

Siddhi Patni [3] proposed various data sources as review sites, blogs, forums, social networks, and Google play store.[8] The author listed opinion mining techniques as Parser, spell checking, and feature extraction with Bag of words, wordnet synsets, lexicons and stop words.

III. BASIC FLOW OF OPINION MINING

Opinion mining can be performing in the following three basic steps.

A. Dataset Creation

Sometimes datasets are readily available online and mostly in the case of code mix language, it is to be created. Raw data can be obtained from diverse sources like blogs, channels, review sites [12].it can be fetched from different APIs presented by social media sites or online websites, and then after categorizing it into predefined labels, it is stored in files like CSV. In most of the studies, all the labeling has been done manually. [1]

B. Pre-Processing

The natural language tool kit provides methods for data preprocessing such as stemming, lemmatization, and others [11].Sklearn packages were used for vectorization, training, testing, and finding the accuracy of the model.
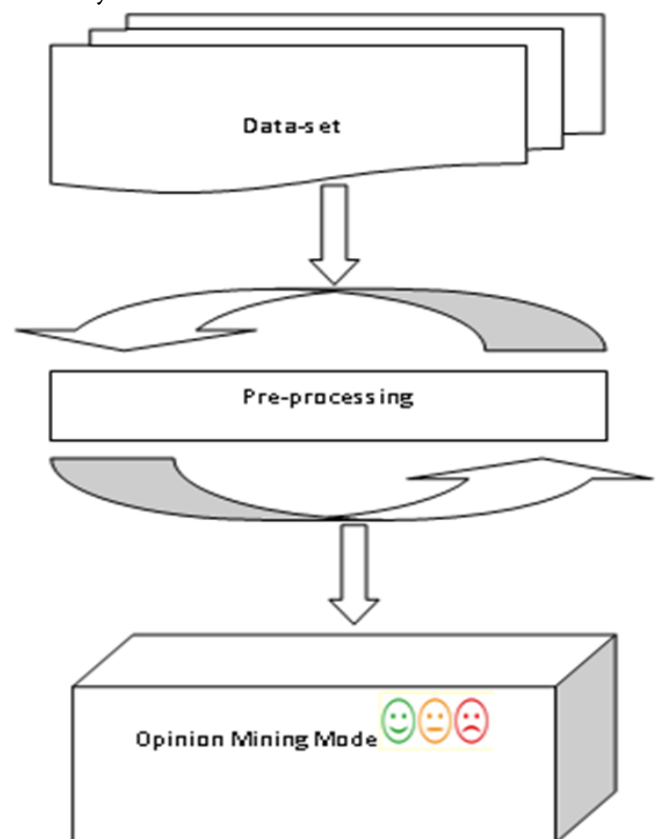

Figure 1. Opinion Mining Flow

C. Model Development

The model was developed on the input of cleaned data, with the help of a combination of different machine learning algorithms. Scikit-learn of python 3 was used for model creation. [15]

IV. OPINION MINING TECHNIQUES

Opinion mining can be done with three techniques.

A. Machine learning

It is categorized into supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning includes algorithms of classification and regression and unsupervised learning includes algorithms of clustering and association. Reinforcement learning performs predictions based on past output and response. Based on parameters, machine learning algorithms can be parametric (logistic regression, naïve Bayes, support vector machine) or non-parametric (Decision tree, random forest).

B. Lexicon-based

It identifies the polarity and subjectivity of input data with the help of the user-created dictionary. It consists of various words and their polarity value. [9]

C. Rule-based

In this conventional approach, opinions are analyzed with help of a predefined set of rules. Lexicon-based and machine learning are the most used approaches in many types of research [14].

V. LIMITATIONS

After studying a variety of papers following limitations are obtained:

- Incomplete or no list of stop words and community words for code-mix languages.[5; 13; 15]
- In some cases, input data have to be cleaned manually.[1]
- Opinions were classified on only 2 polarities negative and positive. The absence of neutral affect the accuracy.[13]
- Small dataset or absence of dataset for analysis.[6]
- Insufficient POS tagger for tagging the part of speech.[15]
- Limited research on aspect-level opinion mining and deep learning techniques.[13]
- Absence of optimal language-specific lemmatizer and stemming algorithm.[6; 15]

Hence to reduce the above-mentioned tribulations there is ample scope of research for code-mix languages.

VI. CONCLUSION

This paper presents a brief overview of various techniques and levels of opinion mining. The study has been done on the opinion analysis process as opinions of viewer helps in decision making and service monitoring. We also focused on the probable enrichment task for the opinion mining of regional language. For future work, we would like to analyze some new code-mix language and will explore more of its application, challenges, and solutions.

VII. REFERENCES

[1]. Shah, S.R., Kaushik, A., Sharma, S. and Shah, J., 2020. Opinion-mining on marglish and devanagari comments of youtube cookery channels using parametric and non-parametric learning models. Big Data and Cognitive Computing, 4(1), p.3.

[2]. Alessia, D., Ferri, F., Grifoni, P. and Guzzo, T., 2015. Approaches, tools and applications for opinion analysis implementation. International Journal of Computer Applications, 125(3).

[3]. Patni, S. and Wadhe, A., 2014. Review Paper on Opinion Analysis is–Big Challenge. International Journal of Advance Research in Computer Science and Management Studies, 2(2).

[4]. Patel, N.S., Shah, D.B. and Bhadka, H.B., 2017. Design of Opinion Mining Model for Opinions Review on World Wide Web by Mobile Computing.

[5]. Asghar, M.Z., Ahmad, S., Marwat, A. and Kundi, F.M., 2015. Opinion analysis on youtube: A brief survey. arXiv preprint arXiv:1511.09142.

[6]. Kaur, G., Kaushik, A. and Sharma, S., 2019. Cooking is creating emotion: A study on hinglish opinions of youtube cookery channels using semi-supervised approach. Big Data and Cognitive Computing, 3(3), p.37.

[7]. Chawla, S., Dubey, G. and Rana, A., 2017, September. Product opinion mining using opinion analysis on smartphone reviews. In 2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 377-383). IEEE.

[8]. Parashar, P. and Sharma, S., 2016. A Literature Review on Architecture Classification Technique and Challenges of Opinion Analysis. International Journal of Engineering Research, 5(5).

[9]. Goyal, M.J. and Dadhich, M.A., 2015. Procedure of Opinion Mining and opinion analysis techniques: A survey. Expansion, Impact and Challenges of IT & CS, p.57.

[10]. Kaushik, A. and Naithani, S., 2015. A study on opinion analysis: methods and tools. International Journal of Science and Research (IJSR), 4(12).

[11]. Chaudhari, M. and Govilkar, S., 2015. A survey of machine learning techniques for opinion classification. International Journal on Computational Science & Applications (IJCSA), 5(3), pp.13-23.

[12]. Janaki, M. and Akilandeswari, K., 2019. A Study on Fundamental and Techniques of Opinion Analysis.

[13]. Shah, S.R. and Kaushik, A., 2019. Opinion analysis on indian indigenous languages: a review on multilingual opinion mining. arXiv preprint arXiv:1911.12848.

[14]. Thakur, V., Sahu, R. and Omer, S., 2020. Current State of Hinglish Text Opinion Analysis. Available at SSRN 3614442.

[15]. Singh, G., 2021. Opinion Analysis of Code-Mixed Social Media Text (Hinglish). arXiv preprint arXiv:2102.12149.

Cite this Article