ACCURACY IN BINARY, TERNARY AND MULTI-CLASS CLASSIFICATION SENTIMENTAL ANALYSIS-A SURVEY

Anjume Shakir
M. Tech(CSE) scholar
Desh Bhagat University, Punjab, India

Jyoti Arora
Assistant professor
Desh Bhagat University, Punjab, India

Abstract: Sentiment analysis is nowadays quite a hot topic for research. Since most of the research is been done on the data acquired from the social networking sites mostly twitter and is subsequently classified into binary classification ("positive" and "negative") or the Ternary classification ("positive", "negative", and "neutral"). The Binary and Ternary classification is not going to serve the sole purpose of sentimental analysis. Multi-Class classification can help in getting the essence and core message from the data. Whether it is Binary, Ternary or Multi-Class classification, the main objective always remains the accuracy of finding the actual sentiments. Since ample work has been done on Binary and Ternary classification and the better accuracy has been achieved but in case of Multi-Class classification accuracy is still a challenge. In this paper, we will analyze different machine learning algorithms and techniques that have been used in the sentimental analysis and the accuracy achieved using those algorithms and techniques.

Keywords: Sentiment analysis, Twitter, Multi-Class classification, Machine Learning, Accuracy.

1. INTRODUCTION

Data has become an important part of economy of any country or organization. The raw data introduces the challenges of storage, noise and errors. Therefore, there is need of cleaning the data before any useful information is mined from it. With every passing minute, the amount of data grows exponentially. Data mining helps the organizations in taking the right decisions at right time and taking an advantage over the other competent organizations. The Organization can serve its customers better and fulfill their needs by analyzing what they actually want.

Twitter, Facebook and other Online Social Networks (OSN) have become the biggest communication platforms for the people to express their views and thoughts about the products [1], movies [2], , politics [3] etc. Some of the other social networking sites have feature of video and multimedia also but twitter is having some important property that makes it really an interesting subject of data mining [4]. Twitter allows users to post short texts messages not more that 140 characters long. The constraint turned out to be attractive for posting quick real-time updates regarding ones activities as well as replying to them quickly [5].

Sentimental Analysis also known as opinion mining is the process of identifying the sentiment or opinion that a person holds towards an object. The factual data is processed, searched or analyzed with the help of textual information retrieval methods. The subjective properties of the components can be presented on the basis of various textual contents within the actualities [6]. The base of sentiment analysis (SA) includes opinions, attitudes, emotions, appraisals and so on. In order to develop new applications, various challenges have been faced while applying these techniques. The major reason due to which issues arise is the regular generation of huge types of data on various online platforms. The different types of positive or negative opinions are given by the users related to various objects which can help organizations in providing feedbacks that can be used in enhancing the quality of those objects. With the utilization of Natural Language Processing (NLP), various tweets, speech or text available on sources can be processed in sentiment analysis.

2. RELATED WORK

Mondher Bouazizi, et.al[7] developed a novel technique to classify the texts into multi-class classification using SENTA tool. On the basis of experimental results achieved it is seen that the multi-class classification is achieved with 60.2% of accuracy. **Ankit Kumar Soni [8]** Naïve Bayes and Maximum Entropy classifiers are combined to generate one algorithm. Amongst various algorithms, the results are compared which can help in analyzing the performance of various algorithms amongst each other and show which has provide to be better. **Wiraj Udara Wickramaarachchi, et.al[9]** presented in this paper that one significant method of expressing the opinion of the users of social network is expressing genuine feelings or emotions through chats and comments for images, status or recordings that has been uploaded to social The research accomplished more efficiency than previous works because of light weight of the methodology, additionally presented a prototype of GUI. Likewise the topic is open for future enhancement also. **N.Moratanch,et.al (2017)** presented in this paper [10] provides survey on extractive summarization approach by categorized them in: Supervised learning approach and unsupervised learning approach. Then different methodologies, the advantages are presented in the paper. The author also includes various evaluation methods, challenges and future research direction in the paper. **Pierre Ficamos, et.al (2017)** studied in paper [11] proposed a feature extraction method that relays on Part Of Speech (POS) tags. That helps in selection of the unigram and bigram features. The paper focuses on the sentiment analysis of the Chinese social media. The grammatical relations

between the different words are used in construction of the bigram and unigram features. The experiment shows that the proposed method provides the better results with the Naïve Bayes. **Aldo Hernandez, et.al (2016)** [12], proposed a sentiment analysis technique that can help in predicting any kinds of future attacks that can possible arise within the web applications.

Venkata Sasank Pagolu, et.al [13] proposed the technique of utilization of sentiment analysis and supervised machine learning principles in combined manner to analyze the sentiments about the stock market movements of a company It is seen that the proposed technique provides better evaluation results in comparison to existing techniques. **Manisha Gupta, et.al**[14] developed a novel approach for text summarization of Hindi text document based on some linguistic principles. Dead wood words and phrases are likewise removed from the original document to generate the lesser number of words from the original text. **Wen Hua, et.al**[15] proposed a prototype framework his paper in order to understand the short messages. This will help in providing semantic knowledge which can further be utilized in order to provide automatic harvesting of the web content generated. On the basis of results achieved, it can be seen that the proposed technique provides better results and helps in analyzing the short messages in better way. **Ankur Goel, et.al (2016)** [16] found that the SentiWordNet along with

Naïve Bayes can improve the accuracy of the tweets classification The implementation is done in Python with NLTK and the python twitter APIs are used. The final experiment shows the classification accuracy improved to a considerable extent.

3. APPROACH

The proposed work is based on data comparison of the accuracy achieved in case of different class sentimental analysis which has been achieved till now using different machine learning techniques In Sentimental Analysis, algorithms are used for feature extraction and correlation factor is used for classification. This leads to reduction in accuracy of classification and increase execution time. The process starts with feature extraction of the dataset taken from the social networks, the machine learning technique is applied to classify the sentiments. In case of classifier used for the sentimental analysis similarity is calculated using many techniques like Euclidian distance, POS-tagging, clustering etc. are used and classifies the features which are approximately equal. The approach of Sentimental Analysis is presented in **Figure 1**. This approach shows the step by step procedure used for the Sentimental Analysis and calculates the execution time and reduce fault de
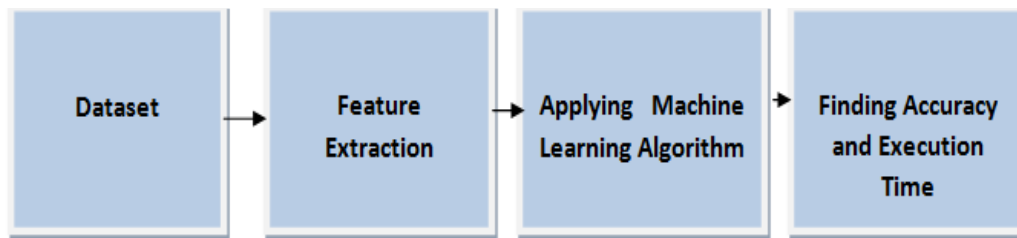

Figure 1

4. ANALYSIS

The Binary, Ternary and Multi-Class classification were analyzed in different papers and the results varied by applying various machine learning techniques. It was found that the average accuracy achieved for Binary-Class classification and ternary class classification is pretty good

but the accuracy is still a challenge. The accuracy for the Binary class sentimental analysis goes up to average of 80% and the accuracy of the Ternary class sentimental analysis is average of 70%.But for the Multi-Class sentimental analysis, the average accuracy is around 60% only. The average accuracy for the respective classification is given in graph below in **Figure 2.**
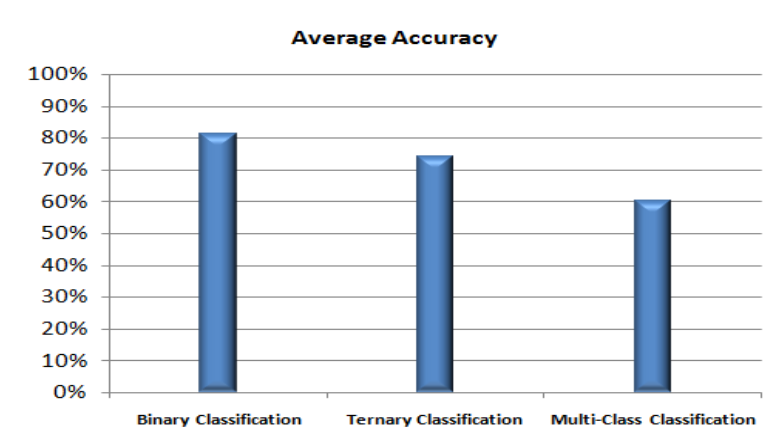

Figure 2

5. CONCLUSION

The study of numerous methods and techniques used for the sentimental analysis shows that the acceptable accuracy is achieved in case of Binary class and Ternary class classification but the accuracy is still a challenge in Multi-Class classification for sentimental analysis.

6. FUTURE SCOPE

New techniques can be used to increase the accuracy of Multi-Class sentimental analysis of twitter data which has been challenge till data. The use of hybrid algorithms can help in achieving the accuracy of the sentimental analysis.

7. REFERENCES

[1] B. O'Connor, R. Balasubramanyan, B. Routledge, and N. Smith, ``From tweets to polls: Linking text sentiment to public opinion time series,'' in Proc. Int. AAAI Conf. Weblogs Social Media, May 2010, pp. 26_33.

[2] U. R. Hodeghatta, ``Sentiment analysis of Hollywood movies on Twitter,'' in Proc. IEEE/ACM ASONAM, Aug. 2013, pp. 1401_1404.

[3] J. M. Soler, F. Cuartero, and M. Roblizo, ``Twitter as a tool for predicting elections results,'' in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2012, pp. 1194_1200.

[4] M. Edkrantz, S. Truve, and A. Said Alan, "Predicting Vulnerability Exploits in the Wild," in 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing, pp. 513-514, 2015

[5] A. Java, X. Song, T. Finin, and B. Tseng, ``Why we Twitter: Understanding microblogging usage and communities,'' in Proc. 9th WebKDD 1st SNA- KDD Workshop Web Mining Social Netw. Anal., Aug. 2007, pp. 56_65.

[6] Feng S., Wang D., Yu G., Yang C. and Yang N., "Sentiment Clustering: A Novel Method to Explore in the Blogosphere", in Advances in Data and Web Management, vol. 5446, pp. 332-344 , 2009

[7] Mondher Bouazizi and Tomoaki Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", in IEEE, vol. 5, pp. 20617-20639, 2017

[8] Ankit Kumar Soni, "Multi-Lingual Sentiment Analysis of twitter data by using classification algorithms", in Second International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 2017

[9] Wiraj Udara Wickramaarachchi and R. K. A. R. Kariapper, "An Approach to Get Overall Emotion from Comment Text towards a Certain Image Uploaded to Social Network Using Latent Semantic Analysis", in 2nd International Conference on Image, Vision and Computing, pp.788-792, 2017

[10] N. Moratanch, S. Chitrakala, "A Survey on Extractive Text Summarization", in International Conference on Computer, Communication and Signal Processing (ICCCSP), 2017

[11] Pierre Ficamos, Yan Liu, Weiyi ChenA, "Naive Bayes and Maximum Entropy approach to sentiment analysis: Capturing domain-specific data in Weibo",in IEEE International Conference on Big Data and Smart Computing (BigComp),pp.336-339,2017

[12] Aldo Hernández, Victor Sanchez, Gabriel Sánchez, Héctor Pérez, "Security Attack Prediction Based on User Sentiment Analysis of Twitter Data", in IEEE International Conference on Industrial Technology (ICIT),pp.610-617, 2016

[13] Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda and Babita Majhi, "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", in International conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016

[14] Manisha Gupta, Dr.Naresh Kumar Garg, "Text Summarization of Hindi Documents using Rule Based Approach", in International Conference on Micro-Electronics and Telecommunication Engineering(ICMETE, pp.366-370, 2016.

[15] Wen Hua, Zhongyuan Wang, Haixun Wang, Kai Zheng, and Xiaofang Zhou. "Understand Short Texts by Harvesting and Analyzing Semantic Knowledge", in IEEE Transactions on Knowledge and Data Engineering , Vol. 29, Issue:3, pp.499-512, 2016.

[16] Ankur Goel, Jyoti Gautam, Sitesh Kumar, "Real time sentiment analysis of tweets using Naive Bayes", in 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 257-261, 2016.