

Sentimental Data Analysis on Social Media

Priyanka. B¹ J.T. Thirukrishna²

¹PG Scholar ²Assistant Professor

^{1,2}Department of Information Technology

^{1,2}SNS college of Technology, Coimbatore, India

Abstract— Social network is one of the admired communication medium and serves as the platform for tracking and analyzing public sentiment of various people on different domains or products. With the explosive growth of user generated messages, Twitter and Facebook has become a social site where millions of users can exchange their opinion. Sentiment analysis affords as an economical and effective way to expose public opinion timely, which is critical for decision making in various domains. Similarly, due to the large volume of opinions, rich web resources such as discussion forum, review sites, blogs and news etc are available in digital form; much of the current research is focusing on the area of sentiment analysis. The information is very useful for businesses for marketing, governments and individuals. While this content meant to be helpful, analyzing the bulk of user generated content is difficult and time consuming. In the existing system there is no way of analysis and ranking the user opinions and sometimes they consider the individual opinions without conducting any reviews. But in proposed there is a need to develop an intelligent system which automatically extract such huge content and classify them into positive, negative and neutral type using Artificial Neural Networks based on the specified criteria.

Key words: Twitter, Artificial Neural Network, public sentiment

I. INTRODUCTION

Social network emerges as a platform for connecting people to share information and it is an opportunity to study the propagation of ideas. Social networks pay the way for the big data that involves in self tracking of any kind of information by the individual or in groups. Twitter is worldwide popular website that offers a social network and micro blogging services, which enable the users to update their status in tweets and follow the people who are interested in retweeting posts and even communicate with them directly.

Sentiment analysis on Twitter has provided an economical and effective way to expose the public sentiment, which is used for decision making in various domains. Sentiment analysis is process that identifies positive and negative opinions, emotions and evaluations in text. Sentiment analysis is classified into three Levels i) at the document level ii) at sentence Level and iii) feature levels [1]. Twitter draws more and more attention from public and also researchers for different disciplines. There are several streams of research investigating the role of Twitter. Based on the stream it has attracted attention in both academia and industry. Previously there have been a many research studies and industrial applications in the area of public sentiment tracking and modeling. Previous research like focused on tracking public sentiment on Twitter and studying its correlation with consumer confidence and presidential job approval polls [2]. Similar

studies have been done for investigating the reflection of public sentiment on stock markets [4] and oil price indices [3] etc...The valuable analysis in twitter is to find possible reasons behind sentiment variation which provides important information's for making decision. Previous research was concentrated on modeling and tracking public sentiment.

The proposed work focuses one step further to interpret sentiment variations. A sentimental data analysis model is proposed using Neural Networks. The primary objective is to extract data from social media and to segregate the comments to analyze the sentimental data using sentimental data analysis. The project is to extract data from social media and to segregate the comments to analyze the sentimental data using sentimental data analysis. Sentimental data can be analyzed using negative and positive words given in the statement. Word analysis can be done through training the word match using artificial neural network. The repeated contents will be removed using clustering and classification. Finally both positive and negative feedbacks will be calculated here.

This paper next continues with literature review of the existing system. The second section will elaborate on the proposed system followed by result analysis in the third section. We will end this paper with conclusions and future work.

II. LITERATURE SURVEY

A process of analyzing the opinions that are extracted from social network on different sources like the comments on forums, products reviews, various policies and the topics mostly associated with social networking sites and tweets known as sentiment analysis. Pang [5] works on the supervised machine learning methods which are in existence for analyzing sentiments are analyzed.

Lilliam lee [6] provides an insight study on considering the ratio of positive words in total words to estimate the opinion on statistical inference from social network data using sentiment analysis models. M.Hu and B. Liu [7] Proposed a novel technique to summarize reviews of customers based on mine features of products on which customers have been commented, decision is made based on positive or negative opinion in each review and finally summarize the results of opinions.

T.Sakaki et al [8] proposed novel model which map the tweets in public segmentation like detecting real-time events in Twitter such as earthquakes, etc. The algorithm monitors tweets and detect event in which each Twitter user is considering as a sensor. Y.Hu et.al [9] developed a joint statistical model ET-LDA, which characterizes topical effects among the event and its related Twitter feeds. The model enables the topic modeling of the event dimensions such as sentiment and polarity.

Chakrabarti [10] proposed sophisticated techniques to summarize the relevant tweets which are used for some

highly structured and recurring events and also gives the hidden events.

Shulong Tan [11] proposed LDA model for interpreting public sentiment variations on Twitter. It contains two models 1. Foreground and Background LDA (FB-LDA) removes background topics and then extract foreground topics to show possible reasons. 2. Reason Candidate and Background LDA (RCB-LDA) ranks them with respect to their popularity within the variation period.

III. PROBLEM DEFINITION

The existing system

- Extracts the user opinions without accuracy and efficiency.
- The second disadvantage is topic mining.
- It will not handle the opinion strength and does not represent opinion with it verbs, adverbs, nouns.
- It does not rank document in certain order.
- It does not handle more general writings and crossing domains. It does not require detail features

IV. PROPOSED SYSTEM

The existing system's problems and requirements were overcome in the proposed architecture, through developing a dedicated application for this analysis model. This application will make a major impact among the market researchers like data analyst, data reporter, business analyst, business growth predictors and etc.

This application is easy, simple and user friendly to use all types of sentimental analysis model. Here raw data will be given as the input and out the output will a tremendous data reporting model.

Some marketers prefer leaving the analysis to dedicated methods, the methods behind sentiment analysis is nothing short from fascinating the various levels of analysis, the detail and the intricacy that make this analysis more accurate when performed by another human rather than a machine.

Nowadays, sentiment analysis is an integral part of social listening, although it can also be performed on its own. Sentiment analysis is more than just a feature in a social analytics method. This is a field that is still being studied. While this comment is general, it can be broken down into sentences. This comment has a number of opinions around Simply Measured, both positive and negative. Sentiment refers to the emotion behind a social media mention. It's a way to measure the tone of the conversation is the person happy, annoyed, and angry or neutrals. Sentiment adds important context to social conversations. If the requirement is to measuring mentions for a company's new product, user might assume a surge in mentions meant it was being well received. After all, more mentions more people talking about the product. Measuring sentiment will help you understand the overall feeling surrounding a particular subject, enabling you to create a broader and more complete picture of the social conversations that matter to you. Each sentence will be comes under word criteria for analyzing positive and negative words.

Criteria's are

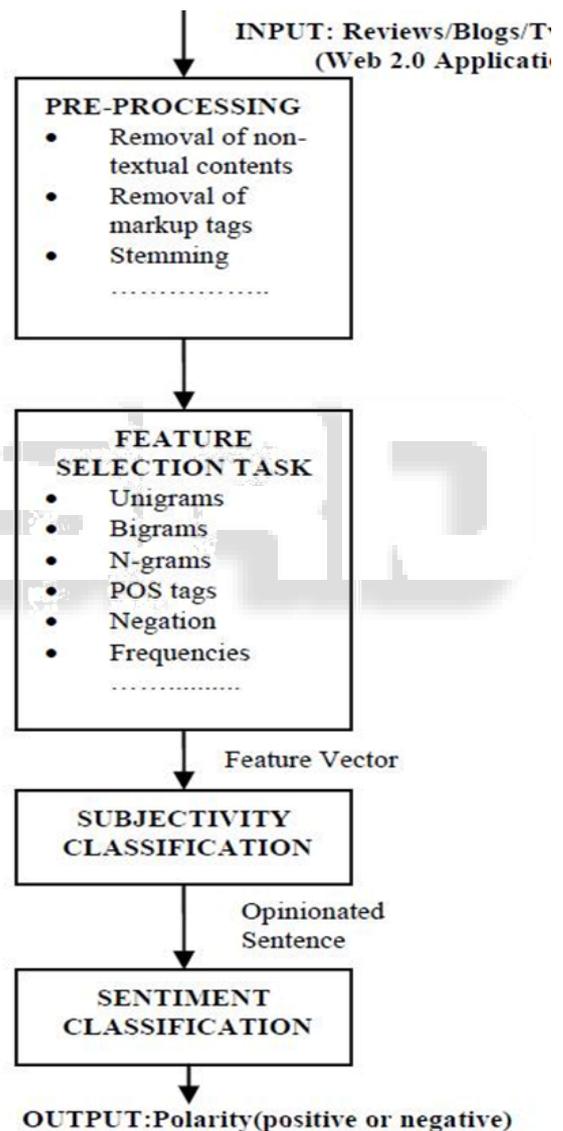
- Tested False Negative

- Tested True Positive
- Tested True Negative
- Tested False Positive

A. Advantages of Proposed System

- It has higher performance than a state-of-art opinion retrieval system
- Summarizing the reviews is useful to many product manufacturers.
- When training data is sufficient the result will be good.
- It can analyze possible reasons behind public sentiments.

V. SYSTEM ARCHITECTURE



VI. RESULT ANALYSIS

System analyzes the sentiment variation from the uploaded dataset.



Fig. 1: shows Index page for Sentimental data analysis.

Figure 1: shows index page for sentimental data analysis. The current date and time are shown in the index file. In the index file data can be uploaded, can also be viewed, the data analysis and ranking are done in the next phase.

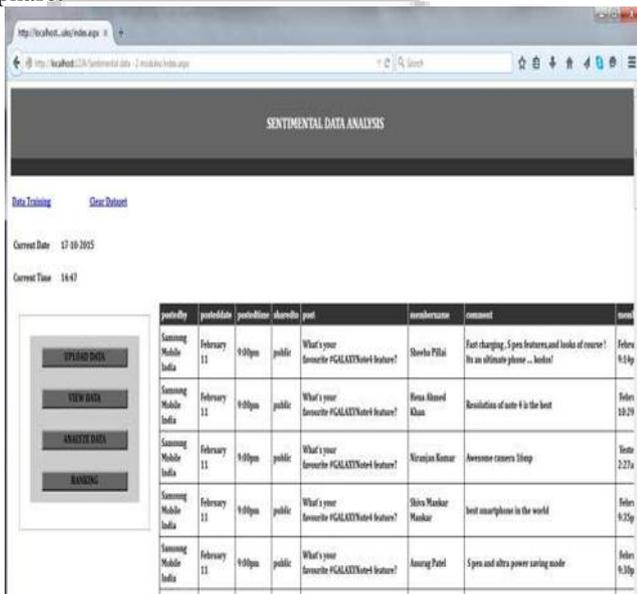


Fig. 2: shows uploaded files

Figure 2: shows uploaded files. The required file is uploaded by selecting upload data button. Select the file by browsing and click upload button, the dataset will be uploaded successfully. After uploading the dataset, we can view the corresponding dataset that has uploaded. The dataset is viewed and it is taken as input, it includes posted by, posted date, posted time, shared to, post, comments etc

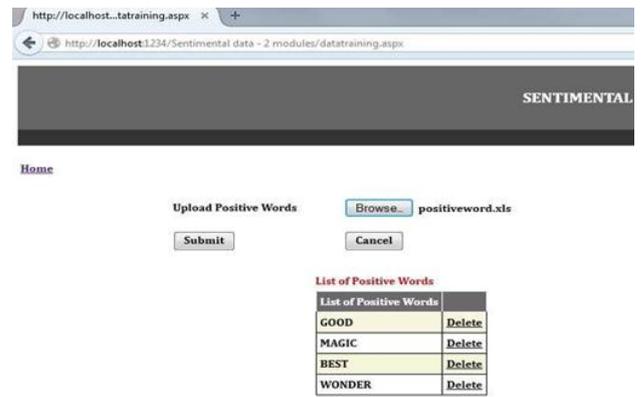


Fig. 3: shows data training- positive words

Figure 3: shows data training positive words. With artificial neural network, the data training is done. In this figure, the browse button is clicked and positive word file is selected. If positive word file is selected and given as submit, then the list of positive words are displayed.

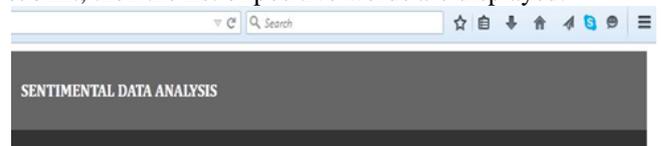


Fig. 4: Shows data training- negative words

Figure 4: shows data training- negative words. With artificial neural network, the data training is done. In the figure, the browse button is clicked and negative word file is selected. If negative word file is selected and given as submit, the list of negative words are displayed

A. Comparison Chart

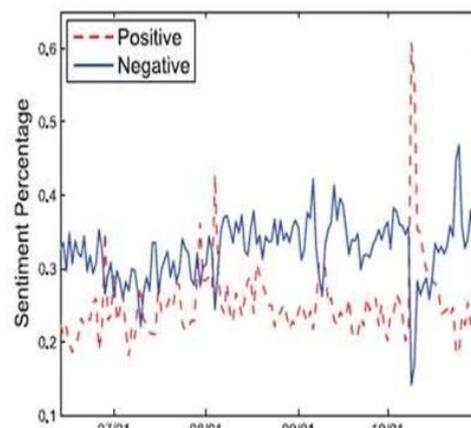


Fig. 4: Shows comparison chart

We analyze public sentiment variations on various on social network and mine possible reasons behind such variations. To track public sentiment; we combine two state-of-the-art sentiment analysis tools to obtain sentiment information towards interested targets (e.g., Canon camera) in each comment. Based on the sentiment label obtained for each comment, we can track the public sentiment regarding the corresponding target using some descriptive statistics (e.g., Sentiment percentage). Finally a chart is produced for the positive and negative comments.

VII. CONCLUSION

Social networks are demandable micro blogging service which has been built to discover what is happening at any moment of time and anywhere in the world. In the survey, we found that social media related features can be used to predict sentiment in various social networks. Our proposed system concludes the sentiments of comments which are extracted from social networks. The difficulty increases with the nuance and complexity of opinions expressed. Product reviews are relatively easy. Books, movies, art, music are more difficult.

This project reviews and summarizes the methodology for analyzing public sentiments. This survey presents various approaches to Opinion Mining and Sentiment analysis. It provides a detailed view of different applications and potential challenges of Sentiment Classification. The emerging topics are related to the actual or genuine reasons behind the variations are very important. So it is necessary to interpret sentiment variation and finding the reasons behind them to overcome above mentioned limitations on different models.

VIII. FUTURE WORK

The system can be further enhanced by adding new features and facilities. Now the system is platform dependent and it can be made as platform independent software. Once it is made as platform independent software, it is can be used by any type of social networks can be incorporated. The system can be also be added with other data mining techniques such as K-nearest neighbours search, neural network and genetic algorithm to find some interesting patterns in the data base.

We can also implement features like emoticons, neutralization, negation handling and capitalization/internationalization as they have recently become a huge part of the internet.

REFERENCES

- [1] B.Pang and L.Lee, "Opinion mining and sentiment analysis," *Found. Trends Inform. Retrieval*, vol.2,no.(1-2), pp.1-135,2008.
- [2] B. O'Connor, R. Balasubramanian, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proc. of the Fourth International AAAI Conference on Weblogs and Social Media*, Washington, DC, 2010.
- [3] Bollen, H. Mao, and A. Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proc. of the Fifth International AAAI Conference on Weblogs and Social Media*, Barcelona, Catalonia, Spain, 2011.
- [4] J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2011.
- [5] B. Pang and L.Lee, "Opinion mining and sentiment analysis," *Found. Trends Inform. Retrieval*, vol. 2, no. (12), pp. 1135, 2008.
- [6] Bo Pang, Lilliam Lee, "Seeing Stars: Exploiting class relationships for sentiment categorization with respect to rating scales", 2002.
- [7] M. Hu and B. Liu, —Mining and summarizing customer reviews, in *Proc. 10th ACM SIGKDD*, Washington, DC, USA, 2004
- [8] T. Sakaki, M. Okazaki, and Y. Matsuo, —Earthquake shakes twitter users: Real-time event detection by social sensors, in *Proc. 19th Int. Conf. WWW*, Raleigh, NC, USA, 2010.
- [9] Y. Hu, A. John, F. Wang, and D. D. Seligmann, —Et - lida: Joint topic modeling for aligning events and their twitter feedback, in *Proc. 26th AAAI Conf. Artif. Intell.*, Vancouver, BC, Canada, 2012.
- [10] D. Chakrabarti and K. Punera, —Event summarization using tweets, in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, Barcelona, Spain, 2011
- [11] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, —Interpreting the Public Sentiment Variations on Twitter, *IEEE Transactions on Knowledge and Data Engineering*, VOL. 26, NO.5, MAY 2014.