Sentiment Analysis by Python using Twitter's Data

Dr. Isha Singh[1], Deepa[2], Vasu Garg[3], Shubh Singla[4]
[1,2,3,4]*Department of Information Technology, HMR Institute of Technology and Management*

Abstract **- The rise of web technology has created a huge amount of data that is collected and used by internet users. The web has become a place for people to discuss and exchange ideas. Social networking sites like Facebook and Twitter have also gained popularity due to their ability to allow people to connect and share their opinions. This survey mainly focuses on the sentiment analysis of Twitter data. It is helpful to analyze the various types of opinions expressed in the tweets and identify their biases.**

I.INTRODUCTION

The age of Internet has changed the way people communicate with each other. Instead of just posting a blog post, people are now using social media to share their opinions and emotions. Through these platforms, they can connect and influence others.

Social media is also generating a huge amount of data that can be used to advertise. Most people rely on the content generated by users to make decisions. For instance, if a person is planning on buying a product or service, they would first check their reviews and then decide whether they should buy it or not.

A Sentiment Analysis is a technique that tells users whether the information about a product or service is satisfactory. This type of analysis data is used by marketers and firms to understand their customers' requirements in a way that's consistent with their marketing strategies. Due to the huge amount of information that's available on the Internet, developing new applications related to sentiment analysis is challenging.

II. RELATED METHODOLOGY

The goal of this dissertation is to first extract live twitter data, then analyze these and extract root words in each tweet. The final sentiment analysis consists of calculating the numerical attributes like the sum of positive/negative/neutral terms. In order to achieve the above, this paper designed an experiment following the below series of steps:

A. Data Collection

1. In steps it uses twitter oAuth API to extract live twitter data.
2. This data will be in JSON format. So json data will be retrieve and store in database and then converted into CSV format.

B. Data Pre-processing

1. After storing data in database. Before done sentiment analysis pre-process the data it includes.
2. Covert uppercase letter to lowercase letter.
3. Removing stop words like (a, an, the, where…etc)
4. Removing extra character like (, # $ % ^etc)

C. Feature Determination/Sentiment Labelling

1. The data set that is achieve after pre-processing needed to be labelled with its respective category of a Positive/Negative/Neutral sentiment. So, it used the counts of the positive and negative words in a tweet to categorize it too Positive/Negative/Neutral
2. If the count of positive words is greater than the negative words, then the tweet is classifying as 1, else its value is -1, and it has value 0 for a tie.
3. This data is then split into 70% and treated as training data and 30% which is treated as test data.

D. Rating Calculation

For each food chain, a count of these classifiers is taken to get a total rating. It takes a sum to ensure both positives and negatives are considered while calculation of the ratings.

III. CONCLUSION

Our console application has provided a way of selecting the place they want to go by seeing the rating of a particular restaurant, etc.

High customer base of Twitter gives an improved representation of public sentiments and the most

recent updates . This model can be further extended to perform an analysis in areas like elections, movies, music, sports and so on. Further, the emoticons that are repeatedly used by twitter users can also be explored to determine the sentiments of the responses.

REFERENCES

[1] D. Kim, Y. Jo, I-C. Moon, and A. Oh, Analysis of Twitter Lists as a Potential Source for Discovering Latent Characteristics of Users, Workshop on Microblogging at the ACM Conference on Human Factors in Computer Systems (CHI 2010).

[2] B. Pang, L. Lee, and S. Vaithyanathan, Thumbs up? Sentiment Classification using Machine Learning Techniques, Proc. Of the Conf. on Empirical Methods in Natural Language Processing (EMNLP), July 2002, pp. 79- 86.

[3] A. Go, R. Bhayani, and L. Huang. Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 2009.

[4] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis"

[5] A. Ortigosa, J. M. Martin, and R. M. Carro, "Sentiment analysis in Facebook and its application to e -learning," Computers in Human Behavior, vol. 31, pp. 527 – 541, 2014.

[6] Balakrishnan Gokulakrishnan, Pavalanathan Priyanthan, AShehan Perera, Nadarajah Prasath, Thiruchittampalam Ragavan, "Opinion Mining and Sentiment Analysis on a Twitter Data Stream" International Conference on Advances in ICT for Emerging Regions Dec 2012.

[7] Yun Xu, XINHUI Wu, QINXIA WANG "Sentiment Analysis of Yelp's Ratings Based on Text Reviews, Stanford University.

[8] I. Hwmalatha, Dr. G. P saradhi varma, Dr. A. Govardhan "preprocessing the informal text for efficient sentiment Analysis" International Journal of Emerging Trends of Technology in Computer Science (IJETTCS), July-August 2012.