

Sentiments Analysis of Twitter Data Using Machine Learning Methods.

Mr. Thok Jeevan Rajendra
UG Student,
Department of Computer
Engineering
SNDCOE & Rc Yeola

Mr. Salunke Gaurav Machindra.
UG Student,
Department of Computer
Engineering
SNDCOE & Rc Yeola

Mr. Pathan Yusuf Usman
UG Student,
Department of Computer
Engineering
SNDCOE & Rc Yeola

Prof. Pandav R.M.
Assistant Professor,
Department of Computer
Engineering
SNDCOE & Rc Yeola

ABSTRACT

Analysis of sentiments is the method of deciding whether the sentiments in the text is positive, negative or neutral. It is also known as material polarity or mining of opinions. The growth and advancement in social media platforms engaged a huge number of users. Social media platform like twitter where users can post their tweets in 280 characters. Because of the limited number of characters in tweets, it becomes easy for the sentiment analysis. On Twitter 550 millions of tweets are posted daily. Twitter also represents all age group people and also a fair representation of gender. Therefore, the sentiment analysis of twitter data becomes somewhat general sentiments of society. In this paper, we will compare various Machine Learning methods like the Naïve Bayes Classification method, Support Vector Machine Classification Method and Maximum Entropy Classification method. We will see how sentiments analysis is done by this classification algorithm and what is the accuracy and precision in these cases.

Keywords

Sentiments Analysis, Opinion Mining, Twitter Sentiments Analysis.

1. INTRODUCTION

What do you do when you want to express yourself or reach out to a large audience? We log on to one of our favorite social media services. Social Media has taken over in today's world, most of the methods we use to connect and communicate are using social networks, and Twitter is one of the major places where we express our sentiments about a specific topic or a concept. Twitter serves as a mean for individuals to express their thoughts or feelings about different subjects. [1] These emotions are used in various analytics for better understanding of humans. [2] In this paper, we have attempted to conduct sentiment analysis on "tweets" using different machine learning algorithms. We attempt to classify the polarity of the tweet where it is either positive or negative. If the tweet has both positive and negative elements, the more dominant sentiment should be picked as the final label

1.1 Objective

This project will be helpful to the companies, political parties as well as to the common people. It will be helpful to political party for reviewing about the program that they are going to do or the program that they have performed. Similarly companies also can get review about their new product on newly released hardware or softwares. Also the movie maker can take review³ on the currently running movie. By analyzing the tweets analyzer can get result on how positive or negative or neutral are peoples about it.

2. ANALYSIS OF THE SYSTEM

2.1 Existing System

The best results reached in sentiment classification use supervised learning techniques such as Naïve Bayes and Support Vector Machines, but the manual labelling required for the supervised approach is very expensive. Some work has been done on unsupervised approaches, and there is a lot of room of improvement.

2.2 Proposed System

In the proposed system, we will trained the module by providing twitter data which is download from Kaggle. The collected tweets will be subjected to preprocessing. We will then apply the supervised algorithm on the stored data. The supervised algorithm used in our system is naïve bayes. The results of the algorithms i.e. the sentiment will be represented in graphical manner (pie charts/bar charts). The proposed system is more effective than the existing one. This is because we will be able to know how the statistics determined from the representation of the result can have an impact in a particular field.

3. SENTIMENTS ANALYSIS

Sentiments Analysis is a task which mainly focuses on textual data and we expect there to be a huge amount of text data. This data is processed and analyses for sentiments that are

expressed. Our training set also required a lot of text data from twitter to analyses.

3.1 Data Gathering

Social Media Site like twitter uses a source of text. Twitter data set is used for preparing the test set we need to perform the following task: Register the twitter application for getting our credentials. First of all, make a twitter account and through it register to the twitter developer account. On twitter developer account citing the valid reason for our academic research creates an app. When we complete the creation of the app it will provide us with some keys like consumer key and access token key and their secret key. Using these keys in our program we can access data for our project work.

3.2 Pre-Processing

After collecting the textual data from twitter, the next step is pre-processing. That is implemented in python. There are several steps involved in pre-processing they are as follows:

Conversion of upper-case letter to lower case latter example: TOPPER to topper.

i) Tokenization: - Tokenization is done with the help of installing the NLP (Natural Language Processing) package. It means the removal of Hash Tags, and the conversion of text to token. And removing Numbers like(1, 2, 3, 4...) URL and Targets (@) etc.

ii) Removal of non- English words :Twitter supports many languages but since our project deals with English so, we remove non English words.

iii) Emotion Replacement : It is very important for evaluating the Sentiments of the users. Hence, we substitute the emotional terms with their polarity by contrasting them in the dictionary of emotions.

iv) Removal of stop words: For sentiments analysis we need to remove the stop words like (a, an, the... etc.) Which does not play any significant role in sentiments analysis.

4. FEATURE EXTRACTING

The selection of the words which are useful from the tweet of the user is called feature extraction. In feature extraction we extract the aspect from, pre-processed twitter data set

- There are three types of features mainly unigram, bigram, and n-gram feature.

- POS i.e. Parts of speech tags like adjective, adverb, verb, and noun is a good indicator of subjectivity and sentiments analysis.

- One of the difficult things to interpret is negation which usually changes the polarity of the content

5. CLASSIFYING METHODS

There are several Machine Learning Classifying method we will apply here are:

- Naïve Bayes Classifier
- Maximum Entropy Method

5.1 Naïve Bayes Classifier

Naïve Bayes is the supervised machine Learning algorithm which uses Bayes theorem for classification problems. It is mostly used in text classification which includes a dataset for high-dimensional training. It is one of the simplest and most powerful classification algorithms that helps to create fast machine learning models that can make predictions quickly. It

is mostly used for text analysis, sentiments analysis and classifying articles. It make use of following Bayes theorem.

$$\text{Prob}(A/B) = (\text{Prob}(B/A)*\text{Prob}(A))/\text{Prob}(B)$$

Where:

- Prob(A|B) is know as (posterior probability) of hypothesis A which occurs when some condition is already provided.

- Prob(B|A) it is also called likelihood probability it is the probability of evidence E when we presumes that given hypothesis is true.

- Prob(A) it is the prior known probability of A and does not include any condition.

- Prob(B) it is the prior known probability of A and does not include any condition.

5.2 Algorithm

Dictionary generation : Count occurrence of all word in our whole data set and make a dictionary of some most frequent words.

- **Feature set generation:** All document is represented as a feature vector over the space of dictionary words.

- For each document, keep track of dictionary words along with their number of occurrence in that document.

5.3.System Architecture

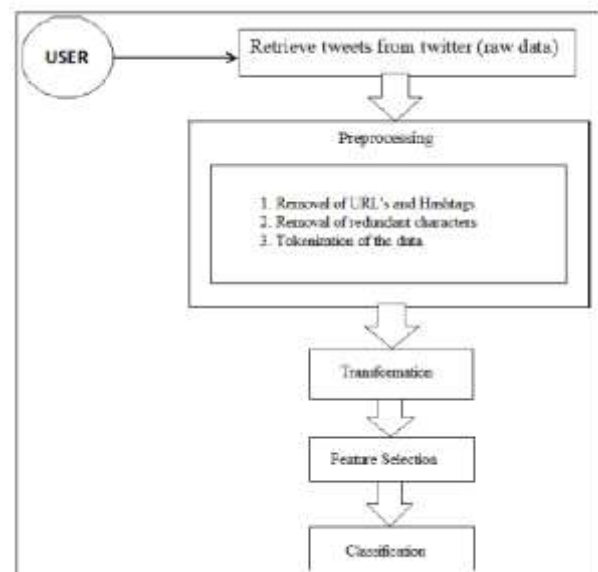


Fig No 1 System Architecture.

6. APPLICATIONS OF SENTIMENT ANALYSIS

6.1 Word Of Mouth(WOM)

Word of mouth (WOM) is deals with the process by which information is passed from one person to another. It helps people while making decisions. The major task of WOM is to classify documents based on WOM into the positive and negative. WOM has great effect on marketing strategies and customers behaviour. Consumers influence affects the informative quality of WOM. If the quality of WOM is high then it more influence the customer behavior the low quality

review product. Therefore most WOM focused model works on the quantity WOM . they do not consider the WOM quality. However, A document with good quality information is more reliable than a document with less quality information.

6.2 Voice of voters(VOV)

now a days, debate on politics is going on every corner of the country. It is going popular day by day. Every political parties spent a lot of money for the good advertising of their parties and they also spent money on voters before some days back from voting. Instead of doing a lot of 16 these thing if they took the people opinion and reviews ,that would be very effective. Therefore sentiment analysis not only help the political parties but also help in doing the analysis of news. People from other country has already took the sentiment analysis.

6.3 Online commerce

Online commerce has the huge collection of e-commerce website. These website takes the reviews of customer after buying the product. After obtaining such information from different areas such quality details of user and services, the user of the company experienced the suggestion, product review and features. These companies collect all the information such as reviews and opinion from the internet and convert these data into geographic form.

8.3 Government

sentiment analysis is doing help of organization to offer the different types of services to public. To analyze the negative and positive views of government, best result must be generated. Therefore sentiment analysis is pretty mush helpful in different field such as taxation, decision making policies, recruitments, and to evaluate the social strategies. There are some almost equal 17 techniques which provide the priorities and services as citizen. There is one major proble which could be apply on this in the country such as india where we speak different different languages. In india, sometimes we people speak the hybrid language i.e. the mixture of two or more language i.e. (English punjabi).and it is very common.

CONCLUSION

From the above seismic analysis of RCC building of G+ 10 In today's world, spacious amount of data is generated by various communication such as social media, organizations etc. these data may or may not be in structured form. Therefore to understand the polarity of data first we need to do the sentiment analysis of data. Opinion mining can be performed in various field such as marketing and customer feedback. large number of organizations are taking the valuable feedback of person and performing opinion mining on those data so that they could provide the better services to the customer and this data helps the organizations to enhance their future services. Furthermore, there are various scopes where we can perform the opinion mining such as sentence, paragraph, documents, sub sentences levels. In addition to this we took some sentiment classifiers such as support vector machine, naïve bayes, decision tree, K-nearest neighbor which performs best in terms of accuracy, precision, and recall. Out of these classifiers we conclude that DT performs best in finding accuracy of twitter dataset. It is best classifier on this dataset.

Acknowledgments

A very firstly we gladly thanks to my project guide Prof. Pandav R.M. for his valuable guidance for implementation of proposed system. We will forever remain a thankful for their excellent as well as polite guidance for preparation of this report. Also we would sincerely like to thank to HOD A.S. Chandgude and other staff for their helpful coordination and support in project work.

REFERENCES

- [1] Evolutionary Machine Learning Techniques: Algorithms and Applications (Algorithms for Intelligent Systems) by Seyedali Mirjalili, Hossam Faris, Ibrahim Alijarah.
- [2] Sentiment Analysis Using Support Vector Machine 1Aamera Z. H. Khan, 2Dr. Mohammad Atique, 3Dr. V. M. Thakare in International Journal of Advanced Research in Computer Science and Software
- [3] Engineering(http://ijarcsse.com/Before_August_2017/docs/papers/Special_Issue/ITSD2015/25).
- [4] J.Ren, S.D.Lee, X.Chen, B.Kao, R.Cheng, D.Cheung, "Naive Bayes Classification of Uncertain Data", Ninth IEEE International Conference on Data Mining, 2009. ICDM '09, pp. 944 – 949.
- [5] International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019, Sentiment Analysis Using Naïve Bayes Classifier Sentiment Analysis Using Naïve Bayes Classifier
- [6] Gurkhe D., Pal N. and Rishit B. "Effective Sentiment Analysis of Social Media Datasets using Naïve Bayesian Classification." (2014).
- [7] Bouazizi, M., Ohtsuki, T.: Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer. IEEE Access. 6, 64486-64502 (2018).
- [8] Gautam, G., Yadav, D.: Sentiment analysis of twitter data using machine learning approaches and semantic analysis. 2014 Seventh International Conference on Contemporary Computing(IC3) .(2014).
- [9] Amolik, Akshay, et al. "Twitter sentiment analysis of movie reviews using machine learning techniques." International Journal of Engineering and Technology 7.6 (2016): 1- 7.
- [10] International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-8 June, 2019, Sentiment Analysis Using Naïve Bayes Classifier Sentiment Analysis Using Naïve Bayes Classifier.
- [11] Ali Hasan, Sana Moin, Ahmad Karim and Shahaboddin Shamshirband" Machine LearningBased Sentiment Analysis forTwitter Accounts" 2018 by the authors. Licensee MDPI, Basel, Switzerland.
- [12] Sahar A. El_Rahman, Feddah Alhumaidi AlOtaibi ,Wejdan Abdullah AlShehri " Sentiment Analysis of Twitter Data".