

Developing an Approach to Evaluate and Observe Sentiments of Tweets

PARINITA AGARWAL

Computer Science Department, Radha Govind Group of Institute, Meerut

Abstract -- Social media are computer-mediated tools that allow people or companies to create, share or exchange information, career interests, ideas and the form of text, audio, video, image in virtual communities and networks. Twitter is a trendy micro blogging service where many users procreate various status messages called tweets. Twitter is the fastest way to get real time information from the around the world. Tweets themselves are short and compact, like newspaper headlines. Analysis of sentiment is widely observed on numerous social networking websites. Nowadays, micro blogging sites deed as a base to perceive the actual social opinion. The task of sentimental analysis is often known by many other names as opinion extraction, opinion mining, sentiment mining, subjectivity analysis.

Indexed Terms: Twitter, sentiment analysis, tweepy, text blob, machine learning, sentiment classification

I. INTRODUCTION

Sentiment analysis is the measurement of positive, negative and neutral expressions of feeling which we can feel and understood. It is a way to evaluate the spoken text or written text whether the expression is neutral, favorable or unfavorable [4].

Sentiment analysis reveals the customer's opinion about different topic or product and services. Analysis of sentiment plays a crucial role as it determines what the customer like and dislikes about the brand. Consumers may use sentiment analysis to inquest products or services before making a purchase.

It is helpful for the trader or businessperson to know about the public opinion about their company or to examine client satisfaction [3].

Most of the organizations use this (sentiment analysis) to cluster evaluative feedback about the issues in newly released services.

Sentiment analysis is basically performed on the three main sections, first is on level of sentence or

phrase, it is the finest analysis of document. In this each word of the sentence is analyzed and its polarity is computed and then it determined whether each and every sentence as neutral, positive or negative. Second is Document level, at the document level complete document through-out is considered as a data set for the analysis for the product or services. Then the classification is done for the complete document as positive or negative. And another one is at the Feature level, at this level of sentimental analysis is done at the basis of certain features of the data, the system is trained about hose certain feature upon which this test has to done.

Sentimental analysis has a wider application area. It helps in differentiating various point of view of number of users [5]. Experience of customers with a company may be good one or a bad one, but the company doesn't know instantly. But with the sentimental analysis one can know instantly when there is a problem and take the necessary steps to keep their customers happy and satisfied.

With sentiment analysis, an organization need no to wait for the survey results. Sentiment analysis senses the tone of conversation using fuzzy logic between the customers and the service representatives. Sentimental analysis offers from instant damage control for any situation. Improve customer retention and customer satisfaction with the live sentimental analysis.

II. RELATED WORK

Micro blogging sites have emerged as a cause of diverse form of information. This is because people post real time messages on these micro blogs about their sentiment on various topics, examine current subject, grumble and signify unambiguous sentiment for products they use in day-to-day life. In fact, companies manufacturing such products have started

to poll these micro blogs to get a sense of general sentiment for their product [4]. Many times, these companies study user reactions and reply to users on micro blogs. One challenge is to build technology to detect and summarize an overall sentiment.

Alec Go, Richa Bhayani, Lei Huang in their research work named “Twitter Sentiment Classification using Distant Supervision” introduces a novel approach for analyzing the sentiment of tweets. These twitter messages are assorted as each of two as positive or negative with respect to a query term. This is helpful for consumers who want to examine the effect of products before purchase, or companies who want to audit the public sentiment of their brands. There is no investigation done on classifying sentiment of messages on micro blogging services like Twitter. They present the outcomes of machine learning methods for classifying the sentiment of tweets using distant supervision. The training data consists of tweets with emoticons, which are used as noisy labels. This form of training data is amply feasible and can be obtained through automated means. They show that machine learning algorithms (Naive Bayes, Maximum Entropy, and SVM) have accuracy above 80%. They even describe the preprocessing steps essential in order to accomplish high precision. The main contribution was of the idea of using Twitter messages with emoticons for distant supervised learning [9].

The key functions are: Defining the Sentiment, they defined sentiment to be “a personal positive or negative feeling. Most of the times it is not clear that a tweet contains a sentiment or not. For such cases, they perform litmus test i.e. if the tweet could ever emerge as a FrontPage newspaper headline or as a sentence in Wikipedia, then it belongs to the neutral sentiment. For instance, the following tweet is considered neutral because it could have appeared as a newspaper headline, even though it projects an overall negative feeling about General Motors: RT @Finance Info Bankruptcy filing could put GM on road to profits (AP) <http://cli.gs/9ua6Sb> #Finance. In their work, they didn’t consider neutral Twitter messages in training or testing data. Apoorv Agarwal, BoyiXie, Hia Vovsha, Owen Rambow, Rebecca Passonneau contributed their work for sentiment analysis in their paper “Sentiment Analysis on Twitter Data”, they focus on two task a) Binary task

for labelling sentiment into positive versus negative category. b) 3-way task for classifying the opinion into positive versus negative versus neutral category. They study three types of models: unigram model, a tree kernel-based model and a feature-based model. For the feature-based model, they use the features proposed in past research and propose new features for further study, they perform the feature analysis which acknowledge the important feature for the analysis [6].

III. MACHINE LEARNING

As we are living in the world of humans and machines. In today’s world these machines or the Roberts are like they need to be programmed before but what if machine started learn on their own, this is where the machine learning comes into existence. There are many examples or implementation of machine learning around such as Tesla, Sophia AI Robot, Self-Driving Car, and many more [11].

Machine Learning is subfield of artificial intelligence, it focuses mainly on the designing of systems, thereby allowing them to learn and make predictions based on some experience which is data in case of machines. Machine learning enables the computer to act and make data driven decisions rather than being explicitly programmed to carry out a certain task. These programs are design to learn and improve over time and expose to new data.

Machine learning is categorized into two different types:

1. Supervised Machine:

Learning: In mathematical term, supervised learning is when the input variable ‘x’ and an output variable ‘y’ and an algorithm to learn the mapping function from input to the output i.e. ‘ $y = f(x)$ ’. The goal is to approximate the mapping function such that whenever a new input data ‘x’ is inserted the prediction of output variable i.e. ‘y’ can be made.

In simple terms, supervised learning can be defined as where each instances of training dataset are composed of different input attributes and an expected output. The input attributes of expected

dataset can be of any kind of data it can be of pixel of image, value of database or an audio frequency histogram. For each input instance and expected output value is associated, a value can be discrete representing a category or can be a real or continuous value, in either case the algorithm learns the input patterns and generate the expected output. Once the algorithm is trained it can be used to predict the correct output of the unknown input.

This type of learning is known as supervised learning because the process of an algorithm learning from training dataset can be thought as teacher supervising the learning process, the correct answers are known, the algorithm iteratively makes while predicting on the training data and is corrected by the teacher, the learning stops when the algorithm achieve the acceptable level of performance.

The popular supervised learning algorithm is Linear Regression, Random Forest and Support Vector Machines. Famous use cases and applications of supervised learning are, Cortana or any speech automation tool trains using the voice and once train it start working based on the training action is performed by the machine. Many more examples are like weather app, biometric attendance, in banking sector for predicting the credit worthiness of credit card holders and so on.

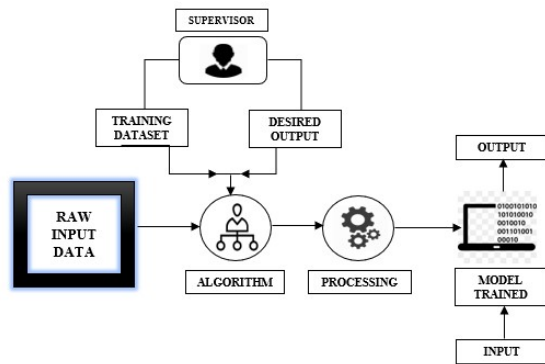


Fig. 1: Supervised learning

2. Unsupervised Machine Learning:

Mathematically, unsupervised learning is when there is input data 'x' and no corresponding output variable to it. The goal for unsupervised learning is to model the underlying structure or distribution in the data in

order to learn more about the data. In simple terms, unsupervised learning is also defined as the data instances of training dataset do not have an expected output associated to them instead unsupervised learning algorithm detects pattern based on inherent characteristics of input data. An example of machine learning task that applies unsupervised learning is clustering; in this task similar data instances are group together in order to identify clusters of data.

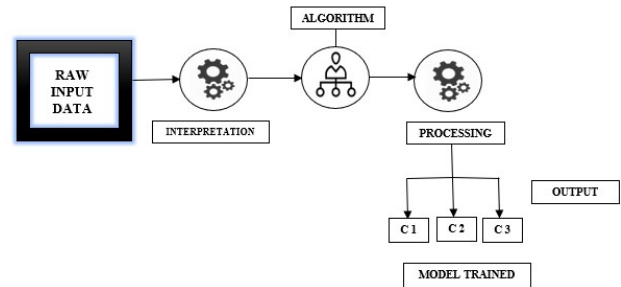


Fig. 2: Unsupervised learning

IV. FRAMEWORK OF TWITTER

Twitter is an engrossing product for studying its structural design. The Twitter web interface works on the Ruby on Rails framework. Twitter message handling is done by software written in the programming language Scala. The framework allows for additional web services and applications to interact and combine with Twitter [10].

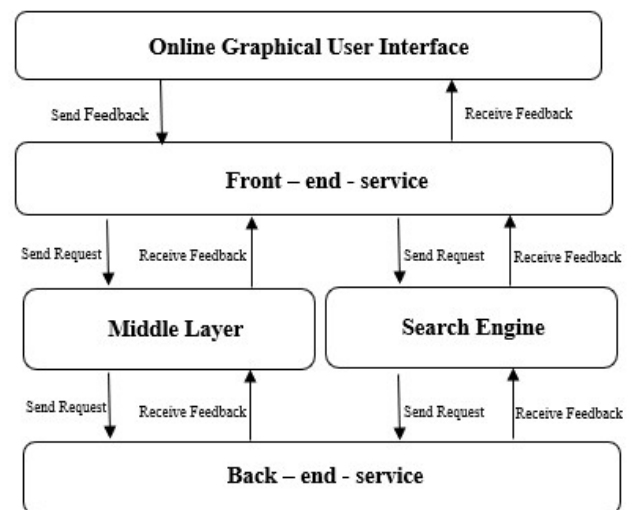


Fig. 3: Operation diagram of twitter

V. TECHNICAL TERMS USED IN TWEETER COMMUNICATION

1. #Hash Tag:

Twitter uses #hash tags to state a particular topic or a subject. With the help of hash tags the twitter service of searching about a particular topic is quite faster. For example, if one wants to search tweets about Narendra Modi, then we can use #narendramodi hash tag. As there is a limitation of characters to be used, we can make use of hash tags in order to shorten the length of the post like if one wants to include the link of any website, then we can make use of hash tags to shorten the URLs

2. Retweeting:

We can even share the tweets of other people by clicking RETWEET button. If anyone wants to add some information with some post then type “RT @USERNAME and the message. When followers retweet, us they will tweet what we tweeted along with an @ourusername tag. This gets our name out and helps more people to potentially follow. As more and more one’s tweets are retweeted, more people started following that person and it results in socially growth on micro blogging sites [2].

3. @ Name Tag:

If one wants to address a tweet to someone, then type @ and username of that person. Name tag is also use to find someone on the twitter with their name tags. For example, if we want to send a message to someone or tag someone through our post, then type @username and the message like ‘@john, join twitter its great’.

4. Like:

When a tweet receives a ‘heart’ from a specific person or user, it is formally known as favorite or star.

5. Lists:

A list is an organized group of twitter accounts. One can create its own lists or even can subscribe to the lists that are created by other users on twitter. When one view his or her list timeline then a bunch of tweets from the accounts that are present in the list are visible. It’s easy to create a list by just clicking on create a new list option which is under the profile

drop down menu, then select the name for the list and a short description and finally we need to choose it either we want to make it public or private.

VI. TWITTER APPLICATION PROGRAMMING INTERFACE (API)

Twitter provides two types of APIs namely Streaming API and Representational State Transfer API. In Streaming API to get the search results and to create post request to the twitter API it uses Create Streaming Connection() function. The Streaming API finds for the hash tag, keywords simultaneously. Streaming API supports longer connections and provides real time data and grant access the data since it is being twitted [1]. While REST API support short-lived connections and are rate-finite that is one can download limited to data posted not more than a week. REST API enables access to assemble data that is twitter data like status updates and user info. REST API consists of two APIs, one just calls the REST API and another Search API.

VII. TEXT BASED SENTIMENTAL ANALYSIS

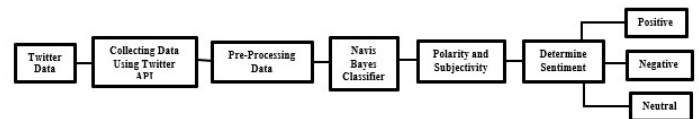


Fig. 4: Block Diagram of Text Based Sentimental Analysis

Collecting Data, for accessing the twitter data via twitter API first twitter application is to be created and get the access token keys. Pre-Processing Data, the twitter data is pre-processed, it consists of three steps first is tokenization, second is normalization, third is parts-of-speech tagging. Navis-Bayes Classifier [7]

We started our work by creating a new python file and importing certain dependencies such as TextBlob, Tweepy, Matplotlib these modules are already installed in the system.

Next, we create a function which will calculate the percentage, once the function is created then we

move forward to establish the connection with our twitter API with the help four keys i.e. consumer key, consumer secret key and access token, access token secret. The connection is built with the help of tweepy functions.

```
def percentage(part, whole):
    return 100 * float(part)/float(whole)

consumer_key = 'BT3F8gFDg75zptKzjZMhx1DZ'
consumer_secret = 'qszu15vUbmearrVOYnkiV6bQ8DFNC3j0kgkFzrKNDI38ZC13V5'
access_token = '2896401224-Jm15iE3MqXrvcWcNvfhvx5t14yrIRSm0nLaqGL1'
access_token_secret = 'cQcBBXMIASbi.dhW2TVKIqdvqPIHCrjHrQNTGqUw4C0N'
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)
```

Fig. 5: Accessing tweeter API

Average numbers of tweets that are positive, negative or neutral are calculated by the three variables define below i.e. *positive*, *negative*, *neutral*. To print the tweets *tweet.text* is used. This text of the tweet is passed into TextBlob as a parameter for analysis and polarity is computed. Based on the polarity that is computed the variable positive, negative, neutral is incremented by one.

```
positive = 0
negative = 0
neutral = 0
polarity = 0

for tweet in public_tweets:
    print(tweet.text)
    analysis = TextBlob(tweet.text)
    polarity += analysis.sentiment.polarity

    if(analysis.sentiment.polarity == 0):
        neutral += 1
    elif(analysis.sentiment.polarity < 0.00):
        negative += 1
    elif(analysis.sentiment.polarity > 0.00):
        positive += 1
```

Fig. 6: State Of Polarity

Percentage of the positive, negative and neutral variables are calculated and formatting it up to two decimal places. To print the results i.e. after analyzing particular number of tweets about the keyword which we want to analyze by checking the polarity of the tweet. If polarity is equal to zero then the tweet is neutral and if polarity is greater than zero then post is referred to as negative and positive when the polarity is less than zero. Polarity contains the overall and average results.

```
print("HOW PEOPLE ARE REACTING ON " + searchTerm + " BY ANALYZING " + str(noOfSearchTerms) + " TWEETS. ")

if(polarity == 0):
    print("neutral")
elif(polarity < 0.00):
    print("negative")
elif(polarity > 0.00):
    print("positive")
```

Fig. 7: Result analysis of polarity

The results are represented in the form of pie charts with the labels positive, negative and neutral along with their percentages. All the three labels are shown with three different colors with equal axis.

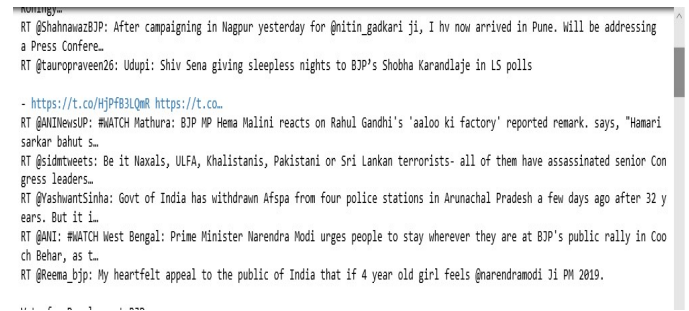


Fig. 8: Fetched tweets

HOW PEOPLE ARE REACTING ON bjp BY ANALYZING 50 TWEETS.

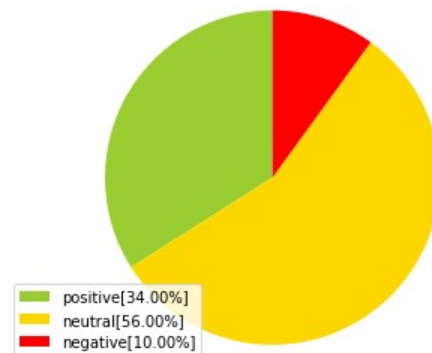


Fig. 9: Pie chart for sentiment analysis

VIII. CONCLUSION

Analysis based on sentiments is widely used to analyze the public opinion. There can be number of reactions which are obtained on any particular event. These reactions include anger, sadness, happiness, aggressiveness but can be broadly categorized into

three sentiments that is positive, negative and neutral. The major challenge we come across is when both adjective and adverbs are used simultaneously. For performing sentiment analysis, initially the tweets are retrieved using twitter API then by applying various classification techniques of machine learning, results can be obtained. In our analysis we have obtained that tweets polarity after analysis of set of tweets related various topics. And result is helpful to classify the text into positive, negative and neutral.

REFERENCES

- [1] M. Trupti, Suresh Pabboju and G. Narasimha: "Sentiment Analysis On Twitter Using Streaming API". Proceedings of the 2017 IEEE 7th International Advance Computing Conference.
- [2] Po-Wei Liang, Bi-Ru Dai, "Opinion Mining on Social Media Data", IEEE 14th International Conference on Mobile Data Management, Milan, Italy, June 3- 6, 2013, pp 91-96, ISBN: 978-1494673-6068-5, <http://doi.ieeecomputersociety.org/10.1109/MDM.2013>.
- [3] Hussein, D.-M.E.D.M. A survey on sentiment analysis challenges. Journal of King Saud University – Engineering Sciences(2016), <http://dx.doi.org/10.1016/j.jksues.2016.04.002>
- [4] A Kowcika and Aditi Guptha "Sentiment Analysis for Social Media", International Journal of Advanced Research in Computer Science and Software Engineering, 216-221, Volume 3, Issue 7, July 2013.
- [5] G. Vinodini and R M. Chandrashekar, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, 283-294, Volume 2, Issue 6, June 2012.
- [6] Apoorv Agarwal BoyiXie Ilia Vovsha Owen Rambow Rebecca Passonneau, "Sentiment Analysis of Twitter Data", Columbia University, New York.
- [7] Pablo Gamallo and Marcos Garcia "A Naive-Bayes Strategy for Sentiment Analysis on English Tweets" Proceedings of the 8th International Workshop on Semantic Evaluation (Sem Eval 2014), pages 171–175, Dublin, Ireland, Aug 23-24 2014.
- [8] Ms. K. Mouthami, Ms. K. Nirmala Devi, Dr. V. Murali Bhaskaran, "Sentiment Analysis and Classification Based on Textual Review"
- [9] Alec Go, Richa Bhayani and Lei Huang, "Twitter Sentiment Classification using Distant Supervision", Stanford University, California.
- [10] Rohan D.W Perera, S. Anand, P. Subbalakshmi and R. Chandramouli, "Twitter Analytics: Architecture, Tools and Analysis". Proceedings of the 2010 Military Communications Conference - Unclassified Program - Cyber Security and Network Management.
- [11] Jasper Snoek, Hugo Larochelle: "Practical Bayesian Optimization of Machine Learning Algorithms".
- [12] Spencer Chang, Timothy Cohen and Bryan Ostdiek: "What is machine learning", Oregon, USA.