

Social Media Opinion Analysis: Based on Public Interest

Mr.Rafeekh AP

Assistant Professor, Department of computer Science and Engineering, College of Engineering Trikaripur, Kasaragod, Kerala, India

Abstract - It is important for many different applications such as government and business intelligence to analyze and explore the diffusion of public opinions on social media as it is one of the key players in this modern world. It is changing the existing information behavior by giving users access to real-time online information channels without the constraints of time and space. This generates a huge unstructured data for data mining. This gives scientists an enormous scope for data analysis challenges. In this project, the system is going to study the public response in various social scenarios such as politics, government policies, entertainment, sports etc. Based on the user tweets the system analyses the public perspective on certain issues. It also helps to rate various personalities and their responses on these issues. Ultimately, I design an opinion flow visualization that combines sentiment analysis to extract the public opinion on various issues, for a certain duration of time. Through this project I aim to develop a system that perform sentiment analysis through user responses on social media on diverse topics. By using visualization and analytical tool the system aims to summarize public opinions on diverse topic the system can be mapped.

Key Words: Social Media Opinion Analysis, Tweet Analysis, Sentiment Analysis, Trend Analysis, Volume Analysis

1.INTRODUCTION

The effective tracing and analysis of opinion diffusion on social media is valuable in many different scenarios as one of the key players in this modern world is social media. Social media is changing existing information behaviour by giving users access to real-time online information channels without the constraints of time and space. This generates a huge unstructured data for data mining giving scientists an enormous scope for data analysis challenge. In this project system is going to study the general user tweets from the different point of view. Here the System will study the user view on any preferred topic of the user.

With all the real time tweets collected over a period of time, the system analyzes them and draws meaningful inferences from these data. Proposed system will analyze tweets data from different user perspective to draw meaningful conclusion to represent them graphically. Trend analysis, individual tweet analysis, sentiment analysis and volume analysis are major parts of proposed system.

Two main approaches are used to accomplish this task: classifying words into Positive, negative and neutral based on twittered output. K-means is more suitable algorithm for clustering of tweet data and to find trends. Volume analysis of tweets will give idea of popularity of particular topic or person over a period of time. Sentiment analysis of tweets will help to draw conclusion for political orientation of overall users respective to political parties, topics.

2.Related Work

2.1 Use Twitter Sentiment to Forecast The 2013 Pakistani Election and 2014 Indian Election

This technique is implemented by Vadim Kagan and Andrew Stevens, Sentimetrix, V.S. Subrahmanian, University of Maryland, Using a sophisticated mix of social network analysis and methods to automatically learn diffusion models from the Twitter data, They could accurately project the winner of the election and, because it was real time, they were able to update the projections frequently. They were even able to predict the most influential individuals on specific topics on social media, allowing potential candidates in the future to use this knowledge to better shape their campaign strategy. The goal of their, they learned statistical models from the Twitter data. Such models are called sentiment diffusion models, and they used these models to forecast the eventual winner of these elections.

2.2 Analysis of Indian Election Using Twitter

This technology has been developed by Gayatri P. Wani Department of Computer Engineering, Phule Pune University, India and Nilesh V. Alone Department of Computer Engineering, GES s R. H. Sapat College of Engineering, Savitribai Phule Pune University. Modern civilizations one of the key players is social media. Social media is changing existing information behaviour by giving users access to real-time online Information channels without the constraints of time and space. This generates a huge unstructured data for data mining. This gives scientist an enormous scope for data analysis challenge. In this project system is going to study the general user tweets from the election point of view. Here the System will study the user view of Indian election. Based on the users tweets system analyses if there exist a pattern between the tweets and to analyze and draw meaningful inferences from the collection of these tweets collected over a certain period; the proposed system identify the feasibility of development of a

classification model to identify the political orientation of the twitter users based on the tweet content and other user based features.

2.3 Demonetization: A Visual Exploration and Pattern Identification of People Opinion on Tweets

This technique was developed by M BalaAnand , Dr.N Karthikeyan², Dr.S Karthick³, C B Sivaparthipan. Improvement of the nation is regularly dictated by different sort of exercises, for example, black cash, counterfeit notes and debasements. The plan is named as Demonetization which is actualized by the decision party BJP on November 8, 2016. This paper proposes envisioning online networking investigation utilizing huge information: web-based social networking information on Twitter. Framework gives numerous planned perspectives to outwardly, instinctively, and at the same time investigate changes in individuals and situational clarifications from genuine voices of people groups, for example, dissensions, on demonetization from online networking i.e., twitter. It also sees to both sides from the minute Demonetization was instituted. The general population dialog was recovered from the online networking "Twitter" with the page Demonetization. Millions began to talk about the integrity and troubles which individuals going to look in the upcoming years. As a piece of it, this paper gathered the information of various conclusions from people groups with qualities like tweet id, name, answer, retweet, reply, likes, timestamp and etc.

In proposed framework datasets of online networking information on twitter about demonetization were chosen. The information was in different frame, for example, content, pictures, and recordings. The general population conclusion about the demonetization framework changed step by step. Utilizing two sorts of calculations are coordinated effort channel and affiliation run mining to dissect the general population sentiment. Community oriented separating is a strategy for making programmed expectations (sifting) about the interests of a client by gathering inclinations or taste data from numerous clients. K means algorithm is a machine learning strategy for finding similarity relations between the factors in the data.

3. PROPOSED SYSTEM

Social media is changing existing information behavior by giving users access to real-time online Information channels without the constraints of time and space generating a huge unstructured data for data mining. With all the real time data collected over a period of time, the system will analyze and draw meaningful inferences from the collection of tweets. Proposed system will analyze tweets data from many perspectives to make meaningful inferences. Trend analysis, sentiment analysis volume analysis are major parts of the proposed system. In trend analysis, the system will try to find

discussions based on popular, mixed, recent emotions expressed in user tweets over a period of time. From literature, K-means is a more suitable algorithm for clustering of tweet data and to find trends. Sentiment analysis of tweets will help draw conclusions for particular topics of overall users with respect to various topics.

The major problems identified while reviewing the areas related. Major challenges identified:

- Big data problem
- Data warehousing is difficult
- Need a long time
- Heterogeneity of data
- Data quality
- Precision and accuracy
- Representation of variety of tweets

Hence, we need an efficient mechanism to find the social media opinion and sentiment analysis and to plot them.

4. METHODOLOGY

System development is the process of developing the system according to the design specification. During the system development phase, the system is constructed from the specification prepared in the design phase as shown in Fig-1. Coding structure translates detailed representation of software into a programming language realization. HTML has been selected as the front tool for software development and Python has been selected as the back end.

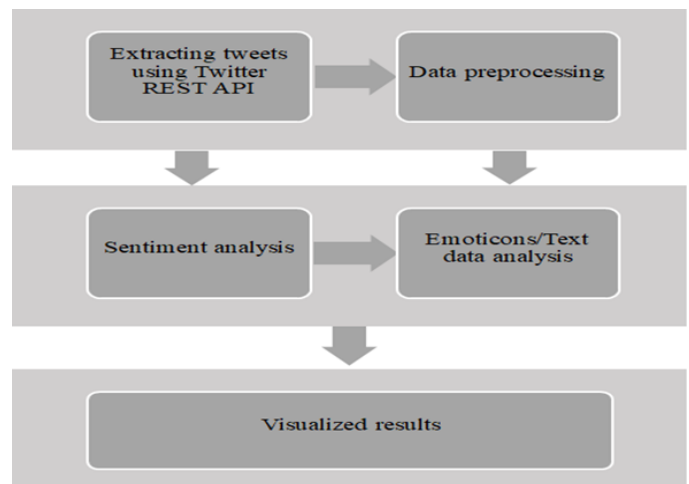


Fig -1: System Overview

4.1 Frameworks

4.1.1 Natural Language Processing

Natural language processing is a field concerned with the ability of a computer to understand, analyze, manipulate, and potentially generate human language. Natural Language Processing (NLP) linguistic pre-processing methods and tools are adapted for social media texts. I survey the methods used we for adaptation to this kind of text. I briefly defined the evaluation measures used for each type of tool in order to be

able to mention the state-of-the-art results. In general, evaluation in NLP can be done in several ways:

- Manually, by having humans judge the output of each tool;
- Automatically, on test data that human has annotated with the expected solution ahead of time;
- ask-based, by using the tools in a task and evaluating how much they contribute to the success in the task.

I primarily focused on the second approach here. It is the most convenient since it allows the automatic evaluation of the tools repeatedly after changing/improving their methods, and it allows comparing different tools on the same test data. Care should be taken when human judges annotate data. There should be at least two annotators that are given proper instructions on what and how to annotate. There needs to be a reasonable agreement rate between the two or more annotators, to ensure the quality of the obtained data. When there are disagreements, the expected solution will be obtained by resolving the disagreements by taking a vote (if there are three annotators or more, an odd number), or by having the annotators discuss until they reach an agreement (if there are only two annotators, or an even number). A supervised text classification model predicts the label c of an input x , where x is a vector of feature values extracted from document d . The class c can take two or more possible values from a specified set. The training data contain document vectors for which the classes are provided. The classifier uses the training data to learn associations between features or combinations of features that are strongly associated with one of the classes but not with the other classes. In this way, the trained model can make predictions for unseen test data in the future.

There are many classification algorithms. I name three classifiers most popular in NLP tasks. Decision trees take one feature at a time, compute its power of discriminating between the classes and build a tree with the most discriminative features in the upper part of the tree; decision trees are useful because the models can be easily understood by humans. Naive Bayes is a classifier that learns the probabilities of association between features and classes; these models are used because they are known to work well with text data. Support Vector Machines (SVM) compute a hyperplane that separates two classes and they can efficiently perform nonlinear classification using what is called a kernel to map the data into a high-dimensional feature space where it become linearly separable ; SVMs are probably the most often used classifiers due to their high performance on many tasks.

NLP tools are important because they need to be used before, we can build any applications that aim to understand texts or extract useful information from texts. Many NLP tools are now available, with acceptable levels of accuracy on texts that are similar to the types of texts used for training the models embedded in these tools. Most of the tools are trained on carefully edited texts, usually newspaper texts, due to the wide availability of these kinds of texts. For example, the Penn TreeBank corpus, consisting of 4.5 million words of

American English, was manually annotated with part-of-speech tags and parse trees, and it is often the main resource used to train part-of-speech taggers and parsers. Current NLP tools tend to work poorly on social media texts, because these texts are informal, not carefully edited, and they contain grammatical errors, misspellings, new types of abbreviations, emoticons, etc. They are very different from the types of texts used for training the NLP tools. Therefore, the tools need to be adapted in order to achieve reasonable levels of performance on social media texts.

4.1.2 Sentiment Intensity

Analyzer One of the challenges of Sentiment Analysis is defining the objects of the study opinions and subjectivity. Originally, subjectivity was defined by linguists, most prominently, Randolph Quirk. Quirk defines a private state as something that is not open to objective observation or verification. These private states include emotions, opinions, and speculations, among others. The very definition of a private state foreshadows difficulties in analyzing sentiment. Subjectivity is often implied in conversation, it is highly context-sensitive, and its expression is often peculiar to each person. Sentiment Analysis has many names. It's often referred to as subjectivity analysis, opinion mining, and appraisal extraction, with some connections to affective computing (computer recognition and expression of emotion). The field usually studies subjective elements, defined by Wiebe et. al. as linguistic expressions of private states in context. These are usually single words, phrases, or sentences. Sometimes whole documents are studied as a sentiment unit, but it's generally agreed that sentiment resides in smaller linguistic units. Since sentiment and opinion often refer to the same idea, this paper will use the terms interchangeably. Sentiment that appears in text comes in two flavors: explicit where the subjective sentence directly expresses an opinion (It's a beautiful day), and implicit where the text implies an opinion (The earphone broke in two days). Most of the work done so far focuses on the first kind of sentiment, since it is the easier one to analyze. Sentiment polarity is a particular feature of text. It is usually dichotomized into two positive and negative but polarity can also be thought of as a range. A document containing several opinionated statements would have a mixed polarity overall, which is different from not having a polarity at all.

5. RESULT AND DISCUSSION

The system analyzed tweets data from different user perspectives and draw meaningful conclusions to represent them graphically. Trend analysis, individual tweet analysis, sentiment analysis and volume analysis are part of our system. Topic wise and duration-based analysis also included in the system. User can choose the topic of preference and enter the type of the user preference, whether the user needs a mixed emotion or recent emotions or popular emotions via selecting the type of the tweet. The

user will also be able to select the duration for which the tweet analysis needs to be done. This enables to give priority to the use preferences to analyze the tweets from the area of interest of the user adding on to the user perspective, making it closer to the user.

The graphical representation (Fig -2) enables the user to get a clear idea about the end review of the analysis. This also enables the user to view the top positive negative tweets as in fig-3.

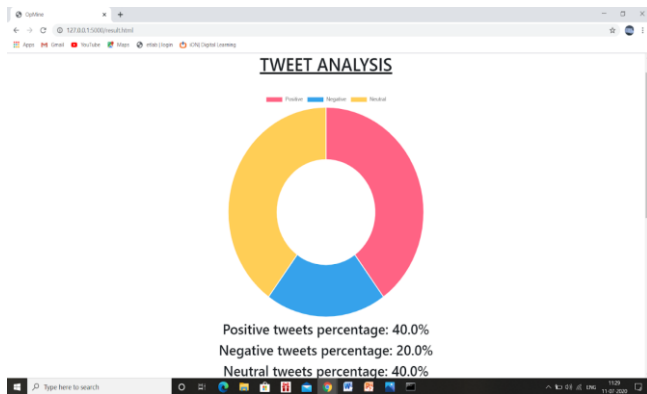


Fig -2 Graphical representation of tweets

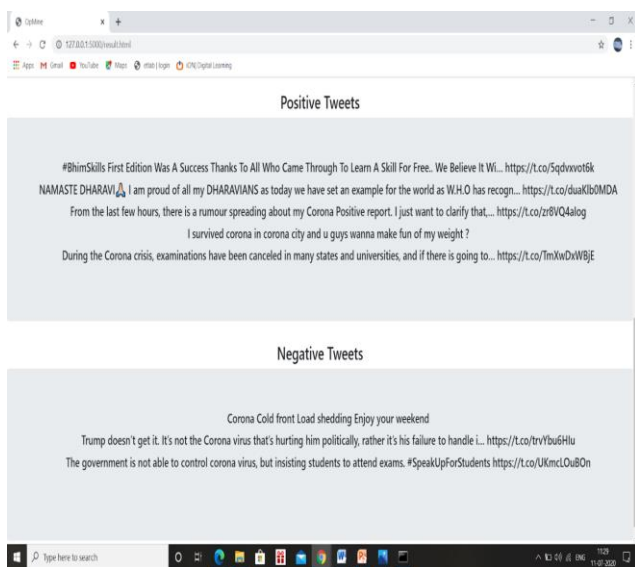


Fig -3 Analysed tweets

6.CONCLUSION

With the increased use of social media, the current paper focused mainly on use of social media as a tool for analyzing public interest. The advent of social media and the rapid development of mobile communication technologies have dramatically changed the way to express the feeling, attitude, mood, passion etc. People often express their reactions, fancies and predilections through social media by means of short texts of epigrammatic nature rather than writing long text. Many micro blogging services like Twitter enable

people to share and discuss their thoughts and views in the form of short texts without being constrained by space and time. Millions of tweets are generated each day on multifarious issues. In the new era, social media has become very important and interesting for all types of services, companies and organizations, so it is of great interest to understand in what way social media data can harvest important tasks and information needed for making decisions at organization level. Till date, the majority of social media analysis is focused on individual companies, products or services.

REFERENCES

- [1] Yingcai Wu(2014) Opinion Flow:Visual analysis of opinion diffusion on social media. IEEE Transactions on Visualization and Computer Graphics R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
- [2] Mr.M.BalaAnand, Dr.N.Karthikeyan, Dr.S.Karthick, Mr.C.B.Sivaparthipan (2008) Demonetization: A Visual Exploration and Pattern Identification of People Opinion on Tweets, 2018 International Conference on Soft-computing and Network Security (ICSNS)
- [3] Gayatri P. Wan Department of Computer Engineering, Phule Pune University, India and Nilesh V. Alone Department of Computer Engineering, GES R. H. Sapat College of Engineering, Savitribai Phule Pune University, Analysis of Indian politics using twitter. IEEE paper
- [4] B.Pan,Y.Zheng,D.Wilkie, and C.Shahabi,"Crowd sensing of traffic anomalies based on human mobility and social media,"in proc.21st ACM SIGPATIAL Int. Conf. Adv. Geographic Inf. Syst.,2013.
- [5] R.Kruger,D.Thom,and T.Ertl,"Visual analysis of movement behavior using web data for context enrichment," in Proc.IEEE pacific Vis.Symp., 2014.
- [6] Vadim Kagan and Andrew Stevens, Sentimetrix, V.S. Subrahmanian , Use twitter sentiment to forecast the 2013 pakistani election and 2014 indian election. Journal paper
- [7] Nick Anstead Department of Media and Communications, London School of Economics and Political Science,and Ben O'Loughlin Department of Politics and International Relations, Royal Holloway, University of London, Egham, Surrey. , Social media analysis and public opinion:The 2010 UK general election. IEEE paper

BIOGRAPHIES



Rafeekh AP
Assistant Professor
Department of Computer Science &
Engineering
College of Engineering Trikaripur
Kasaragod, Kerala