Impact of Demonetization in India using Opinion Mining over Twitter Data

Mrugendrasinh Rahevar[1], Martin Parmar[2], Rekha Karangiya[3]

[1,2]Assistant Professor, U & P U. Patel Department of Computer Engineering, CSPIT, CHARUSAT, Changa, Anand, Gujarat, India

[3]Assistant Professor, Information Technology, CSPIT, CHARUSAT, Changa, Anand, Gujarat, India

ABSTRACT

In recent years, the utilization of Internet has turned out to be one of the everyday activities in our life. Social networks constitute a noteworthy segment of the Web and made an upheaval. It incorporates social media, forum conversations, blogs and micro-blogs like twitter. Due to this, large numbers of comments are produced on daily basis. So, nowadays most of the researchers or analyzers are concentrating on extracting significant data from social networks in order to understand the public viewpoint. This research has been reached out outside the computer science to cover other areas like business, political and social science. Hence, Sentiment analysis and Opinion mining are popular field of research in Data mining. This paper delineates various aspects of sentiment analysis in detail inclusive of important concepts, classification, process, importance, challenges and applications. The following paper presents experiment on sentiment analysis of public opinion on demonetization in India. Sentiment analysis is performed on tweets related to demonetization in India extracted from twitter. Polarity of the opinion is observed through the experimental analysis. Through the outcome of this analysis, the sentiments of the citizens that are determined help the government in improving their decisions and work for the welfare of the citizens.

Keywords : Sentiment analysis, Demonetization, Twitter data, opinion mining, polarity, WordCloud.

I. INTRODUCTION

Sentiment analysis is an automated process of analyzing opinion that largely influences businesses and government systems. Nowadays large repositories can be prepared by extracting data through Twitter, Facebook, LinkedIn and many other social platforms and e-commerce websites to analyze public opinion and demands. These results are used in the business to scrutinize and maintain quality of their product. Feedback requirement in recent years is increasing to know about people's point of view. These can be observed through social media platform such as twitter, as through twitter public opinion about government policies, international policies, products and various other trending issues can be known [1]. Sentiment analysis through twitter data provides positive and negative trait observed through people's opinion on different topics. Sentiment analysis provides frequencies of positive, negative or neutral polarity on a particular issue. Demonetization was a trend in 2016-2017 among citizens of India and it had mixed opinion both positive and negative, large number of citizens disliked this move while many of them supported it. Twitter had massive response on demonetization and to perform sentiment analysis demonetization was one of the best issue from the availability of data and experiment point of view. The following paper discusses detail information on sentiment analysis. Also, the paper includes

sentiment analysis on demonetization of twitter data using python as a language and Jupyter notebook as platform.

II. METHODS AND MATERIAL

Important concept

(A) Text Classification: The first step to perform sentiment analysis is to classify the text in two types, Subjectivity and Objectivity.

1) Subjective Text: This type of text comprises of sentiments.

Example: Taj Mahal is a beautiful monument. The example contains sentiment (beautiful), hence it is subjective. This type of text is significant for performing sentiment analysis.

2) Objective Text: This type of text comprises of factual information and does not contain any sentiments.

Example: Taj Mahal was built by Shah Jahan. The example contains only information and has no sentiments in it, hence it is objective.

(B) Polarity: Subjective text can be further classified on the basis of sentiments delivered from the text. The sentiments can be positive, negative or neutral as per the feelings, emotions and expressions.

1) Positive: It can be identified by sentiments bearing good connotations.

Example: Sachin Tendulkar is one of the best cricketer.

2) Negative: It can be identified by sentiments bearing evil/bad connotations.

Example: I do not like Classical Dance.

3) Neutral: The sentiment that does not bear good as well as bad connotations can be termed as a sentence with Neutral polarity.

Example: I usually feel sleepy by noon. The example conveys the feelings of a person but does not contain any positive or negative connotations.

(C) Levels of Sentiment Analysis: Sentiment analysis can be performed on three different levels:

1) Document level: This is the least difficult type of classification. The entire document comprising of text with opinions is considered as a basic unit of data [2]. A single polarity, either positive or negative, is given to the entire document. For Example, the document contains opinions about a single object like a film or a book. This classification is not feasible when it contains reviews of more than one product.

2) Sentence level: As each sentence contains different opinion, it is necessary to find polarity of each and every sentence of the document. So sentence level is considered as fine-grained analysis of the document [2].

3) Feature/Aspect level: The above two classification levels do not discover exact information about likes and dislikes of people. So, there is need of aspect level classification because different feature of a single product provides different opinion. This level considers the product, feature of that product, opinion of feature, reviewer and time. Because of these parameters this level can find what actually consumer like means which feature of product is liked by most of the customer and on which time [3]. For Example, the cell phone has long battery life but has poor camera resolution. The above examples states two different features of single product having opposite polarity.

Sentiment Analysis

Sentiment analysis is the field of study that examines individual's feelings, assumptions, assessments, evaluations, mentalities. Furthermore, their feelings towards substances, for example, items, administrations, associations, people, issues, occasions, themes, and their properties. It speaks to a vast issue space. It is a progression of strategies, systems, and devices about recognizing and extracting subjective data, for example, feeling and states of mind, from dialect [4].
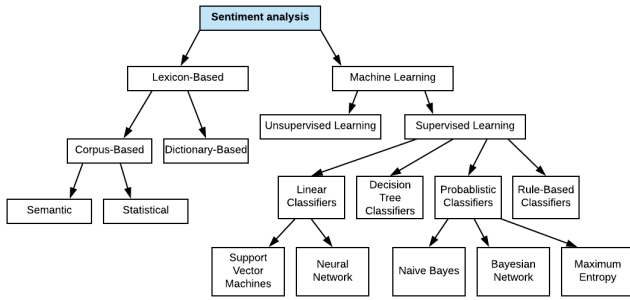
Figure 1. Sentiment Analysis Classification

Sentiment analysis is classified based on the two approaches.

(A) Lexicon-Based Approach:

In Lexicon-based approach, a message is denoted as a sack of words where each word is associated with a score. Following this portrayal of the message, sentiment values from the lexicon are allocated to each and every positive and negative word inside the message. A merge function, for example, summation or average, is used so as to make the last prediction with respect to the general estimation for the message [5].

1) Dictionary-based method:

The procedure of this method is as follows. A little arrangement of supposition words (seeds) with known positive or negative introductions is gathered first, which is simple. The calculation at that juncture develops this set via rummaging in the lexical database for their equivalent words and antonyms. The recently discovered words are added to the seed list. The next emphasis starts. The iterative procedure closes when not any more new words can be found. After the process finishes, a manual assessment step is utilized to clean up the list [6]. This method is not feasible when we are trying to search for domain oriented words.

2) Corpus-based method:

Corpus is gathering of works, frequently on a particular subject. In this approach, seed list is made and is extended with the assistance of corpus text [7].

Thus it overcomes the limitation that was faced by Dictionary-based method. Further this method can be classified in two ways.

(i). Statistical method:

This method mainly focuses on the co-existence of the words in the corpus. If the word exists generally in positive content then the polarity of that word is positive. If it mostly occurs in negative content then the polarity is negative.

(ii). Semantic method:

This approach computes sentiment values by utilizing the principle of similitude between words. Lexical database can be utilized for this reason. Equivalent words and antonyms of given word can be discovered utilizing this and sentiment values can be computed [8].

(B) Machine Learning: This approach utilizes various algorithms that effectively help in classification of text. It is further divided as:

1). Supervised Learning: In this method of learning, the algorithm is trained by using labelled data and then the testing data is given as an input to the algorithm. It can be done through various classifiers such as Linear (SVM and Neural Network), Probabilistic (Bayesian Network, Naive Bayes and Maximum Entropy), Rule-based and Decision tree classifiers.

2). Unsupervised Learning: In this method of learning, the algorithm draws inferences from the input data without any label.

Process of Sentiment Analysis

The process of Sentiment analysis requires the text to pass through different number of phases which can be depicted from Figure 2. The first step is to prepare a dataset by collecting the reviews of user. Then the next step is to identify the type of text as subjective or objective. The useful text in this process is subjective text. This subjective text is then passed to the next phase of this process which is text pre-processing.
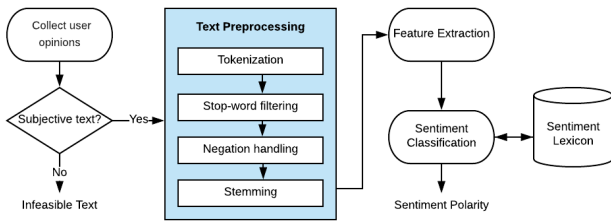
Figure 2. Sentiment Analysis Process

Text Pre-processing converts the unstructured text into a structured format. It includes the following steps:

(A) Tokenization: It is the first step in this phase. It is the process of dividing a sentence on each space that forms tokens [9]. These tokens formed can be words, phrases and keywords as well. In this step, some symbols like punctuation marks are removed.

(B) Stop-Word filtering: Stop-words do not provide any type of information about the sentiments. Pronouns, articles and prepositions are some examples of stop-words. Eliminating this words permits to lessen the number of features extracted [9].

(C) Negation handling: Negation handling if not handled will alter the entire meaning to its opposite. Some of the negation words are not, no, never, cannot and many more. For example, The marks obtained are not good. In this example, the word "good" indicates positive polarity but "not" changes the polarity to negative.

(D) Stemming: It is the process of eliminating prefixes and suffixes. For example, 'running', 'runs' can be transformed to 'run'. Stemming helps in characterization yet once in a while prompts diminish in accuracy of classification [10].

After going through this phase, features need to be extracted from a large complex dataset that are informative and unique leading to a better human understanding. After extracting the features, the sentiment of the content needs to be classified which can be done with the help of sentiment lexicon. Finally, the polarity of the text is obtained.

III. Experiment and Results

This experiment is performed using Python language and Jupyter Notebook. In this work, the data is collected from twitter related to demonetization that is used for performing sentiment analysis. For accessing the tweets, the first step is to create a twitter application and obtain the required credentials. Next step for extracting the tweets is to connect to the Twitter API through a library provided in Python-tweepy that also supports authentication. The extracted data is stored in a .csv file. The data generated is often dirty and inconsistent. So as to make it suitable for analysis, the data needs to be cleaned (removal of hashtags, unnecessary white spaces, retweets, links and punctuations) [11]. This cleaned data is passed to TextBlob that provides common text processing functionalities.

The experiment can be easily visualized through charts and plots. This can be done by using Matplotlib-a python library known for constructing visualization graphs. Figure 3 is a histogram showing the frequency of tweets of different trending hashtags related to Demonetization.
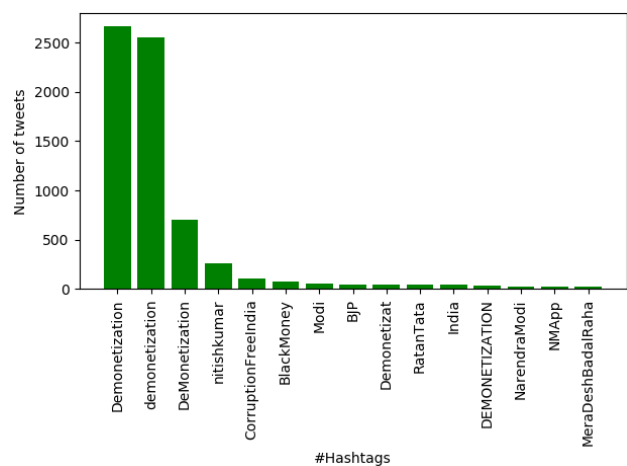

Figure 3. Most Trending Hashtags

The dataset contains total 7915 tweets on Demonetization. There are 7915 rows (one for each

tweet) and 15 columns. Now sentiment analysis is performed on this pre-processed dataset in Python. Table 1 shows the number of tweets classified on the basis of polarity.

Table 1. Frequency of Tweets based on Polarity

Total tweets	7915
No. of Positive tweets	1914
No. of Neutral tweets	3196
No. of Negative tweets	2805

Figure 4 is a bar graph showing the number of tweets of different polarities from the total number of tweets used in the experiment.

Figure 4. **Bar Chart Visualization**

Figure 5 is a pie chart showing the percentage of the positive, negative and neutral polarities obtained from the analysis of the total number of tweets.

Figure 5. Pie Chart Visualization

It can be depicted from Figure 4 and Figure 5 that the proportion of tweets with Neutral polarity is more than Positive and Negative polarities. Tweets containing Positive polarity are the least.

WordClouds are one of the important outcome of text analysis. In a WordCloud, the bold and big words are those with the maximum occurrence. Figure 6 and Figure 7 shows the WordCloud of Positive words and Negative words respectively. It can be depicted from the figures that words like 'well', 'better', 'best', 'top' are some of the frequently occurring positive words and words like 'anti', 'ill', 'annoyance', 'concession' are some of the negative words frequently used in the tweets.

Figure 6. Positive WordCloud

Figure 7. Negative WordCloud

Importance of Sentiment analysis

With the help of Sentiment analysis, one can grasp out important information from the available loads of information that is present in an unstructured manner. It also decreases the manual effort required

in processing of data. It increases scalability and works very efficiently in a cost effective manner. It is used for real-time analysis. It also reduces the rate of errors and at the same time improves the consistency of data.

Major Challenges:

The major concern in the field of sentiment analysis is to deal with language and its context which is dynamic and ambiguous in nature. Hence, there arise several difficulties and challenges in sentiment analysis. Few of them are:

A) Word Interpretation Ambiguity: It is often referred to as Word Sense Disambiguation (WSD). Actual meaning of the word is depended on the context associated with the word as single word may have many different meanings with respect to different context. For example, the party was extravagant. Here the word extravagant is polarity ambiguous word. Extravagant can be interpreted positivetly as luxurious party and negatively as spendthrift party.

B) Comparative Analysis: Determining the polarity of comparative sentences is challenging. For example, camera of company x is better than camera of company y. Here 'better' is a word with positive context but the entity selected creates difficulty in determining the polarity and entity is most important fraction of detail in comparative analysis.

C) Negations: Negations if not took care of appropriately can give totally wrong outcomes. For instance there is a decent possibility that this telephone won't break effectively. This survey appears positive extremity however existence of negation changes the impact totally [2].

D) Intensity: Deciding polarity based on intensity is quite difficult whether it results highly positive or highly negative. This is also called degree of polarity.

E) Sarcasm and Irony: Irony is conveyed as difference between actual and intended meaning where as sarcasm is more intense version of irony. It is

challenging to state polarity of these type of statements where the intended meaning is unclear. So, the expressions need to be analyzed at more fine-grained level.

Applications

(A) In Decision making: Opinions extracted from Movie, Hotel, Product, and Restaurant reviews.
(B) In technology: Work as a sub part in mailing system for spam detection.
(C) User's voice: Helps in identifying the nuances of client that occurs over time and in configuration of better informative questions for futuristic surveys.
(D) Various organizations: In analysis of social media and brand, research and analysis in marketing, and Product evaluation.
(E) Other areas: Political Science, Sociology, Psychology and in building norms.

IV. CONCLUSION

The experimental analysis performed in this paper calculates sentiment values of public tweets on demonetization. Using the techniques available in python, the sentiment values are plotted. From the outcome of this analysis, it can be concluded that majority of the people are neutral towards demonetization. Gathering opinion on individual basis consumes a lot of time and is expensive. Hence, such type of analysis is helpful for assessment of government schemes and monitoring the development of schemes from individual's viewpoint.

V. REFERENCES

[1]. K.Arun, A.Srinagesh and M.Ramesh, "Twitter Sentiment Analysis on Demonetization tweets in India Using R language", International Journal of Computer Engineering In Research Trends, Volume 4, Issue 6, June-2017, pp. 252-258.

[2]. Seema Kolkur, Gayatri Dantal and Reena Mahe, "Study of Different Levels for Sentiment Analysis", International Journal of Current Engineering and Technology, Vol.5, No.2 (April 2015).

[3]. Bharat R.Naiknaware, Seema Kawathekar and Sachin N.Deshmukh, "Sentiment Analysis of Indian Government Schemes Using Twitter Datasets", IOSR Journal of Computer Engineering (IOSR-JCE), Vol. 2, 2017.

[4]. Mika V. Mäntylä, Daniel Graziotin and Miikka Kuutila, "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers", Computer Science Review, Volume 27, February 2018.

[5]. Anna Jurek , Maurice D. Mulvenna and Yaxin Bi, "Improved lexicon-based sentiment analysis for social media analytics", Security Informatics, 9 December 2015.

[6]. Reshma Bhonde, Binita Bhagwat, Sayali Ingulkar and Apeksha Pande, "Sentiment Analysis Based on Dictionary Approach", International Journal of Emerging Engineering Research and Technology Volume 3, Issue 1, January 2015.

[7]. Keshtkar Fazel and Inkpen Diana, "A bootstraping method for extracting paraphrases of emotion expressions from texts", Comput Intell, vol. 0, 2012.

[8]. Medhat, Walaa, Ahmed Hassan and Hoda Korashy, "Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal 5.4:1093-1113, 2014.

[9]. Pierre FICAMOS and Yan LIU, "A Topic based Approach for Sentiment Analysis on Twitter Data", International Journal of Advanced Computer Science and Applications, Vol. 7, No. 12, 2016

[10]. Harpreet Kaur, Veenu Mangat and Nidhi, "A Survey of Sentiment Analysis techniques", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2017

[11]. CH.R.Vinodkumar, Prof.D.Lalitha Bhaskari, "Sentiment Analysis of #MakeInIndia & #Demonetization using R", IOSR Journal of Computer Engineering, Volume 19, Issue 6, Ver. II (Nov.- Dec. 2017)