Demonetization in India: Looking Through the Mirror of Twitter Sentiments

Kamalpreet Kaur, Dr. Neeraj Sharma

Abstract— **Twitter, the most popular and free micro blogging site nowadays which had a great hand in voicing public opinion by posting tweets on Twitter on a wide range of topics or domain like drugs, elections, etc. which is used for sentiment analysis to identify its polarity. Recently implemented Demonetization in India had a drastic change on the Indian economy. People post their opinions on the Twitter that whether they are in favor of it or against it. The task of sentiment analysis is used to find the sentiments of tweets about Demonetization in India. There have been four approaches used i.e. Lexicon approach, Naïve Bayes approach, Naïve bayes with binary features approach and Extension of Naïve bayes using lexicons approach to classify the tweets on Demonetization in terms of positive, negative and neutral polarity. After this, the standard measure of evaluation (Precision, Recall and F-measure) is calculated for all approaches and the comparison between all the approaches help us to find the best approach of sentiment analysis.**

Index Terms— **Lexicon approach, Naïve Bayes approach Sentiment analysis, Twitter.**

I. INTRODUCTION

Sentiment analysis, task of Natural Language Processing (NLP) is the most important for text analysis. It identifies positive and negative opinions and measure how positively or negatively an entity is required. Here, an entity can be any product, service, event, topic and their attributes. It can be say that it is a field of study that broadly examines sentiments & opinions of people on the given entities. In short, a sentiment can be termed as a feeling of any person which can be positive or negative towards anything and an opinion can be defined as a belief or emotion of somebody about something. Opinion mining is the process of tracking of opinions of people about certain product. Opinion is the thinking of any person about anything. It is not necessary that everybody has same opinion on the same entity as different people have their own views to interpret the results.

Twitter

Nowadays, there is an increase in number of people who are willing to post their opinions on Twitter (popular and free micro blogging service where users create status messages), which is now considered a valuable online source for opinions [1]. Here, status messages are known as tweets. A tweet is a text-based post which is considered short in length because it has only 140 characters. In conclusion, sentiment analysis using twitter data is an effective way of determining public opinion for any social study.

In general, original tweets are very noisy i.e., it may contain hash tags (#), emoticons, target, numeric data, abbreviations and many other noisy words. These original tweets need to be normalized first and after that use for the purpose of sentiment analysis in NLP systems. After normalization, tweets are known as normalized tweets. The process of normalization of nonstandard data is often a necessary preprocessing step to enable NLP tools which require clean and standardized data as input to perform on their expected quality levels.

Moreover, tweets have very unique features like the maximum length of a tweet is 140 characters, its availability as it is very easy to extract a large corpus of tweets with the help of Twitter API's using keywords like demonetization, note ban, etc. Furthermore, tweets can be extracted about any domain.

In this work, we have collected corpus of tweets with the help of twitter API about a query term which is demonetization here. The data set had been used for training the models of learning approaches after annotation for finding their polarities (positive, negative or neutral). Here, *positive* means that a tweet is in favor of demonetization, *negative* means that a given tweet which is analyzed is against demonetization and *neutral* means that neither a tweet is positive nor it is negative. After this, we have implemented four approaches of sentiment analysis on the dataset to get the polarity of opinion given by people on the Twitter. The implemented approaches are Lexicon approach, Naïve Bayes approach, Naïve Bayes with binary features approach and Extension of Naïve Bayes using lexicons approach. Their outputs are used for calculating the standard measures of evaluation i.e. Precision, Recall and F-measure and compare all the techniques to get the better result.

Sentiment Classification

In sentiment analysis, the polarity of the text or sentiment can be classified at three different levels [2]:

1. Document level: At the document level, opinion holder will consider each document as a single opinion target or object and gives opinion about it. The main goal of document

Manuscript received August, 2017.
Kamalpreet Kaur*, Department of Computer Science, Punjabi University, Patiala, Patiala, India, +91-842-749-4432*
Dr. Neeraj Sharma, *Department of Computer Science, Punjabi University, Patiala, Patiala, India,*
+91-981-591-8075

level sentiment classification is to determine the overall sentiment orientation of the document.

2. Sentence level: The next level of sentiment classification is sentence level which will consider each sentence as a separate unit and expressed it as positive, negative or neutral.

3. Entity level: Entity level is also known as feature level which aims to produce a feature-based opinion summary of multiple reviews on an opinion target. This level has mainly three tasks. The first task is to identify object features and extract it commenting by opinion holder. The next is to find the polarity of opinions which is followed by grouping feature synonyms.

II. TRAINING DATA SET

The data set includes the twitter messages about a query term which is Demonetization here. For the training, it had been collected with the help of Twitter API from Twitter using multiple keywords of Demonetization process such as, *demonetization, demonetisation, notebandi, note ban* etc. Further tweets had been extracted after the announcement of Demonetization by the Prime Minister on 8th November 2016. There are 1500 tweets which are selected out of total tweets and used for training dataset.

Initially, when tweets are collected with the help of twitter API, they may involve emoticons, target, hash tags, website addresses, numeric data and other noisy words. These are known as original tweets which need to be clean by filtering out all noisy words. This filtering process is called normalization and after normalization, we get normalized tweets which are used for the purpose of training data set.

Example of original tweet:
"impreetsbakshi: RT @AshiQuotes: Positive Effects of Demonetization which any corrupt channel won't discuss.\n#NamoBestPMOfIndia

After normalization:
Positive Effects of Demonetization which any corrupt channel won't discuss

The next thing is to annotate the corpus sentences with two or three annotators so that higher inter annotator agreement should be maintained. The annotation must be keeping in mind about three classes: Positive, Negative and Neutral. With respect to demonetization, positive means for demonetization, negative means against demonetization and neutral means neither positive nor negative. For this annotation process, a team of two annotators work on manual annotations for 1500 sentences. While annotating a corpus, large efforts are required to maintain its consistency.

A team of annotators must follow the guidelines for the annotation process to get the higher inter annotator agreement. Inter annotator agreement is a measure of how well two or more annotators can annotate the same annotation decision for demonetization [3]. The special instructions to annotators were given to be as consistent as they can be.
There were many sentences which actually contain some kind of information or reporting of fact, state, custom, action or suggestion or expectation or state of hope, etc. which can be considered different than an opinion expressed. Some of these have positive undertone and while some have negative undertone. We treated these as ***neutral*** because they are not actually reflecting the sentiments of the opinion holder.

Example:
- *"Live Updates on Currency Demonetization News, Reports and Developments."* (Information, but not a direct sentiment of opinion holder).
- *"Yes, more power and loot pretty much explains everything rulers do, whatever the propaganda may say."* (Positive undertone, but not a direct positive sentiment of opinion holder).
- *"A list of those who were informed about demonetization by modi forehand, deposited more than 25lakh before 8th Nov."* (Action, but not a direct sentiment of opinion holder).
- "*Indians must compromise and support demonetization so that we can make our country cashless and digital*" (Suggestion, but not a direct sentiment of opinion holder).
- *Hope this demonetization ends up in some success. I am keeping my fingers crossed. Hope situation will situation change.* (Fact, but not a direct sentiment of opinion holder).

Sentiment analysis models

After annotation process, we have implemented various approaches of sentiment analysis on the training data set.

(a) Lexicon Baseline
One of the simplest approaches of machine learning is lexicon approach [4] which uses the lists of positive and negative words to determine the polarity of tweets and output the most selected class which is associated with the maximum tokens. If positive and negative tokens are same in number then neutral is considered as a most selected class.

(b) Naive Bayes
The another simplest technique of machine learning is Naïve Bayes which assign a class to a particular tweet, say d, by the following given formula:

$$c^* = arg\ max_c\ P_{NB}\ (c/d)$$

where $P_{NB}\ (c/d)$ is given as:

$$P_{NB}\ (c/d) = (P(c)\ \Sigma_i\ P(f_i / c)^{n_i(d)})/P(d)$$

In the above formula, f is an attribute, $n_i(d)$ is the count of an attribute i in tweet d. Naïve Bayes approach is a probability based classifier. It is also called Multinomial Naïve Bayes Classifier. Suppose there is a document 'a' and out of all classes c϶C, the classifier returns the class ĉ, this class ĉ has maximum probability in the document.

(c) Naive Bayes with Binary Features
The next step is to apply Multinomial Naive Bayes with binary features approach. In this model, the frequency of a given word in a given document is either 0 (if the word does not occur in the document) or 1 (if the word occurs 1 or more times in the document).

(d) Extension of Naive Bayes with Usage of Lexicons

In this approach, two types of Lexicons using sentiment analysis, i.e. Negative Lexicon list and other one is Positive Lexicon list are created which lead to more generalized results. Negative Lexicon list contains list of words with negative meaning and Positive Lexicon list contains list of words with positive meaning. The way of using lexicons in classifier is to use the occurrences of positive and negative words in lexicon.

III. STANDARD MEASURE OF EVALUATION

The next phase is to evaluate the classification based on standard measures of evaluation (Precision, Recall and F-measure).

Precision: Precision measures the correctness of a classifier. It is denoted by P and a fragment of appropriate documents from the retrieved one. Precision can be calculated as:

Precision = tp/(tp+fp)

where *tp* is the true positives and *fp* is the false positives More precision means less number of false positives while less precision means more false positives.

Recall: Recall measures the completeness, or sensitivity, of a classifier. Higher recall means less number of false negatives, while lower recall means more false negatives. It is denoted by R and calculated as:

Recall = tp/(tp+fn)

where *tp* is the true positives and *fn* is the false negatives. Recall is named as true positive rate (TPR) or sensitivity for positive classifier and specificity or true negative rate (TNR) for negative classifier.

F-Measure: Precision and recall can be combined to produce a single metric known as F-measure, which is the weighted harmonic mean of precision and recall. It can be calculated as:

F-measure = 2*(precision*recall)/(precision+recall)

IV. EXPERIMENTATION AND RESULTS

In this research work, we calculated the measure of agreement between two annotators and precision, recall and f-measure for all the classes i.e. positive, negative and neutral using different approaches and give the results of all approaches and compared the results.

Table I shows the results for the agreement of sentiment classification annotations. Inter-annotator agreement of annotations between two annotators A and B is 87.73%. And the agreement goes up to more than 92% in both annotation sets after discussions and many iterations of cleaning.

Table I. The agreement between two annotators

Comparison	**Total number of sentences**	**Total agreement between**		
		A and B	**Final and A**	**Final and B**
Sentiment Classification	1500	1316 (87.73%)	1383 (92.20%)	1388 (92.53%)

Lexicon approach

By looking at the results in the table II, it is observed that accuracy of lexicon baseline approach is 98.80% and also the classification based on this model gives a Precision of 100%, Recall of 96.93% and F-score of 98.44% for “Positive” class. Moreover, this model had given a Precision of 99.38%, Recall of 100% and F-score of 99.68% for “Negative” class. For “Neutral” class, lexicon baseline approach had given a Precision of 96.62%, Recall of 100% and F-score of 98.28%.

Naïve Bayes approach

The Naïve Bayes approach had given an output for three classes as shown in the table III. This is observed that this model had given an accuracy of 97.26%. Furthermore, the standard measure of evaluation for “Positive” class is 93.48% of Precision, 100% of Recall and 96.63% of F-score. For “Negative” class, Precision is 100%, Recall is 97.71% and F-score is 98.84% and for “Neutral” class, Precision is 100%, Recall is 93.02% and F-score is 96.83%.

Naïve Bayes with binary features approach

The next step is to apply Multinomial Naive Bayes with binary features approach.

The result of Naïve Bayes with binary features approach is similar with the naïve bayes approach shown in table IV. It also gives an accuracy of 97.26%. Precision of 93.48%, Recall of 100% and F-score of 96.63% had been evaluated for “Positive” class. Likewise, Precision of 100%, Recall of 97.51% and F-score of 98.73% had been evaluated for “Negative” class and for “Neutral” class, Precision is 100%, Recall is 93.25% and F-score is 96.51%.

Extension of Naïve Bayes using lexicons

The given table V shows the output of extension of naïve bayes using lexicons approach. It is clear from the table that the most accurate approach is naïve bayes extension using lexicons because it had given an accuracy of 99.40%. For “Positive” class, the Precision is 98.49%, Recall is 100% and F-score is 99.24%. For “Negative” class Precision is 100%, Recall is 98.34% and F-score is 99.16%. And for “Neutral” class, Precision is 100%, Recall is 99.76% and F-score is 99.88%.

Table II. Result of lexicon approach

Model		**Lexicon approach**		
Class	**Accuracy**	**Precision**	**Recall**	**F-score**
Positive	0.9880	1.0	0.9693	0.9844
Negative		0.9938	1.0	0.9968
Neutral		0.9962	1.0	0.9828

Table III. Result of naïve bayes approach

Model		**Naïve Bayes approach**		
Class	**Accuracy**	**Precision**	**Recall**	**F-score**
Positive	0.9726	0.9348	1.0	0.9663
Negative		1.0	0.9771	0.9884
Neutral		1.0	0.9302	0.9683

Table IV. Result of naïve bayes with binary features approach

Model		**Naïve Bayes with binary features approach**		
Class	**Accuracy**	**Precision**	**Recall**	**F-score**
Positive	0.9726	0.9348	1.0	0.9663
Negative		1.0	0.9751	0.9873
Neutral		1.0	0.9325	0.9651

Table V. Result of naïve bayes extension using lexicons approach

Model		**Extension of Naïve Bayes approach**		
Class	**Accuracy**	**Precision**	**Recall**	**F-score**
Positive	0.9940	0.9849	1.0	0.9924
Negative		1.0	0.9834	0.9916
Neutral		1.0	0.9976	0.9988

Comparison between approaches

The next thing is to compare the different approaches based on the standard measures of evaluation (Precision, Recall and F-measure). The table VI shows the comparison between the four implemented approaches of sentiment analysis i.e., Lexicon approach, Naïve Bayes approach, Naïve bayes with binary features approach and extension of naïve bayes using lexicon approach.

From the table VI of comparison, it is clear that the approach of extension of naïve bayes using lexicon performs best in all cases and have a highest accuracy from all other approaches as it has 99.40% of accuracy.

Table VI. Comparison between approaches

Model	**Accuracy**	**Positive**			**Negative**			**Neutral**		
		Precision	**Recall**	**F-score**	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-score**
Lexicon	0.988	1.0	0.9693	0.9844	0.9938	1.0	0.9968	0.9662	1.0	0.9828
Naïve Bayes	0.9726	0.9348	1.0	0.9663	1.0	0.9771	0.9884	1.0	0.9302	0.9683
Naïve bayes with binary	0.9726	0.9348	1.0	0.9663	1.0	0.9751	0.9873	1.0	0.9325	0.9651
Naïve bayes extension	0.994	0.9849	1.0	0.9924	1.0	0.9834	0.9916	1.0	0.9976	0.9988

V. Conclusion

In this paper, we had built sentiment lexicons for Positive and Negative polarities and a corpus of tweets sentences are created by crawling tweets from Twitter about *Demonetization* in India. After that, this corpus of tweets is annotated by three annotators manually and is lead to high degree of inter-annotator agreement. Model named Sentiment lexicon, Naïve Bayes, Naïve Bayes with binary features and extension of Naïve bayes are used to determine accuracy and
standard measures of evaluation. Aftermath, most of the accurate and best approach in all cases is the Naïve bayes extension using lexicon approach.

References

[1]. Alec Go., Richa Bhayani, Lei Huang. "Twitter Sentiment Classification using Distant Supervision"

[2]. V.S. Jagtap Karishma Pawar (2013). "Analysis of different approaches to Sentence-level Sentiment Classification" *International Journal of science Engineering and Technology (ISSN 2277-1581), 1 April 2013*

[3]. "Positive and Negative Sentiment Lexicons in Punjabi", Available from SemTADA group of Gurinder Gosal, Department of Computer Science, Punjabi University, Patiala, 15 September 2016.

[4]. Maite Taboada, Julian Brooke*,* Milan Tofiloski*,* Kimberly Voll*,* Manfred Stede (2010). "Lexicon - Based Methods for Sentiment Analysis" *Association for Computational Linguistics (Volume 37 Number 2).*

Kamalpreet Kaur is a student of M.Tech (CSE) in Department of Computer Science, Punjabi University, Patiala carrying out her research work under the guidance of Dr. Neeraj Sharma. Her main research interest is analyzing sentiments of tweets on the topic of demonetization in India.

Dr. Neeraj Sharma is the Head of Department (HOD) at Department of Computer Science, Punjabi University, Patiala. He is holding a PhD degree in computer science. His current research interest is sentiment analysis.

