International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 6; Issue 7; July 2017; Page No. 4941-4945
DOI: http://dx.doi.org/10.24327/ijcar.2017.4945.0618

Research Article

SENTIMENTAL ANALYSIS OF DEMONETIZATION IN INDIA USING MACHINE LEARNING ALGORITHMS

Ravinder Kumar[2] and Anil Kumar[1*]

[1]USICT, GGSIPU, Dwarka, Delhi
[2]HMRITM Affiliated to GGSIP University, Delhi

A R T I C L E I N F O

Article History:

Received 4th April, 2017
Received in revised form 11th May, 2017
Accepted 12th June, 2017
Published online 28th July, 2017

Key words:

Hastag, Sentiment Analysis, Demonetization, Twitter, Machine Learning

A B S T R A C T

This paper analyzes demonetization that took place in India on 9 Nov 2016-30 Dec 2016 by honorable prime minister Narendra Modi So now we are presenting in public domain the Results From India's Demonetization Campaign using tweets posted on Twitter from 9 Nov 2016 – 30 Dec 2016. Twitter is a social network where users post their feelings, opinions and sentiments for any event. This paper transforms the unstructured tweets into structured information using open source libraries. Further objective is to build a model using Bayesian Network classification on unseen tweets on the same context. This paper collects tweets on this event under seven hashtags. This study explores three freely available resources / Application Programming Interfaces Python (APIs) for labeling of tweets for academic research. This paper proposes three sentiment prediction models using the sentiment predictions provided by three APIs. BN classifier is used to models. The performances of these models are evaluated through standard evaluation metrics. The experimental results reveal that the TextBlob API and proposed Preference Model outperformed than the other four sentiment prediction models.

INTRODUCTION

On 8 November 2016, the Government of India announced the demonetisation of all ▪500 (US$7.40) and ▪1,000 (US$15) banknotes of the Mahatma Gandhi Series. The government claimed that the action would curtail the shadow economy and crack down on the use of illicit and counterfeit cash to fund illegal activity and terrorism. The sudden nature of the announcement-and the prolonged cash shortages in the weeks that followed-created significant disruption throughout the economy, threatening economic output. The move was heavily criticised as poorly planned and unfair, and was met with protests, litigation, and strikes. Prime Minister of India Narendra Modi announced the demonetisation in an unscheduled live televised address at 20:00 Indian Standard Time (IST) on 8 November. In the announcement, Modi declared that use of all ▪500 and ▪1000 banknotes of the Mahatma Gandhi Series would be invalid past midnight, and announced the issuance of new ▪500 and ▪2000 banknotes of the Mahatma Gandhi New Series in exchange for the old banknotes.

The plan to demonetise the ▪500 and ▪1000 bank notes began six to ten months prior, and was kept highly confidential with only about ten people aware of it completely. The logistical processes and preparations for printing the new ▪500 and ▪2000 bank notes began in early-May. The cabinet was informed about the demonetisation on 8 November 2016 in a meeting called by the Prime Minister of India Narendra Modi which was followed by Modi's public announcement about the demonetisation in a televised address.Microblogging websites have evolved to become a source of varied kind of information. This is due to nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. Many times these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment. In this paper, we look at one such popular microblog called Twitter and build models for classifying “tweets” into positive, negative and neutral sentiment. We build models for two classification tasks: a binary task of classifying sentiment into positive and negative classes and a 3-way task of classifying sentiment into positive, negative and neutral classes. We experiment with three types of models: unigram model, a feature based model and a tree kernel based model. For the feature based model we use some of the features proposed in past literature and

Corresponding author:* **Anil Kumar
USICT, GGSIPU, Dwarka, Delhi

propose new features. For the tree kernel based model we design a new tree representation for tweets. We use a unigram model, previously shown to work well for sentiment analysis for Twitter data, as our baseline. Our experiments show that a unigram model is indeed a hard baseline achieving over 20% over the chance baseline for both classification tasks. Our feature based model that uses only 100 features achieves similar accuracy as the unigram model that uses over 10,000 features. Our tree kernel based model outperforms both these models by a significant margin. We also experiment with a combination of models: combining unigrams with our features and combining our features with the tree kernel. Both these combinations outperform the unigram baseline by over 4% for both classification tasks. In this paper, we present extensive feature analysis of the 100 features we propose. Our experiments show that features that have to do with Twitter-specific features (emoticons, hashtags etc.) add value to the classifier but only marginally. Features that combine prior polarity of words with their parts-of-speech tags are most important for both the classification tasks. Thus, we see that standard natural language processing tools are useful even in a genre which is quite different from the genre on which they were trained (newswire). Furthermore, we also show that the tree kernel model performs roughly as well as the best feature based models, even though it does not require detailed feature engineering. As mentioned earlier, the aim of sentiment analysis is to "define automatic tools able to extract subjective information from texts in natural language." The first choice when one is applying sentiment analysis is to define what text (ie, the analyzed object) means in the case of study considered.

Message level: The aim is to classify the polarity of a whole opinionated message. For example, given a product review, the system determines whether the text message expresses an overall positive, negative, or neutral opinion about the product. The assumption is that the entire message expresses only one opinion on a single entity (eg, a single product).

Sentence level: The aim is to determine the polarity of each sentence contained in a text message. The assumption is that each sentence, in a given message, denotes a single opinion on a single entity.

Entity and aspect level: Performs a finer-grained analysis than message and sentence level. It is based on the idea that an opinion consists of a sentiment and a target (of opinion). For example, the sentence "The iPhone is very good, but they still need to work on battery life and security issues" evaluates three aspects: iPhone (positive), battery life (negative), and security (negative).

The Role of Semantics

The semantics of the language used in social networks is fundamental to accurately analyze user expressions. The context of a textual expression is therefore a crucial element that should be taken into account to properly deal with the underlying sentiment. A sentence "taken as it is" can appear as negative or positive, but if it is properly analyzed from a semantic point of view it can be completely different. For instance, the sentences

"Demonetization should be considered just as a beginning of fight against corruption in India.... @narendramodi @PMOIndia"can be interpreted as positive and some lexical cues that are typical of the social network language, we should derive a (real) positive judgment. Lexica, corpora, and ontologies need to be properly constructed and used for us to have a deep understanding of the semantics of the natural language in online social networks.

Sentiment Analysis and Process

Sentiment analysis process can be depict with the help of following figure:

It includes the process data collection, pre-processing, feature extraction, sentiment classification and projection of output. Various steps that are used in pre-processing are: Tokenization, Data filtering, stop word removal and stemming;

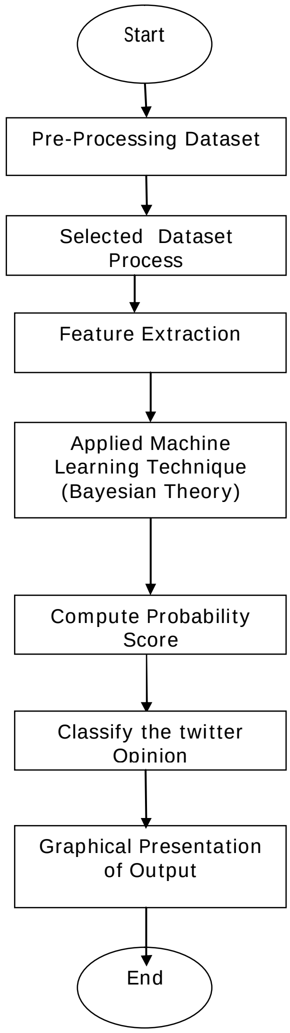


Fig 1 Flow chart of Sentimental Analysis Process of Twitter Demonetization

Data Collection

Demonatiztion related data was collected from user generated content from twitter. Manual analysis is almost impossible. Therefore, text analytics and natural language processing are used to extract and classify.

Pre-Processing

Pre-processing starts the text preparation into a more structured representation. This includes the following steps.
Tokenization: Tokenization is used to identify all words in a given text.

Data Filtering: People use a lot of casual language on twitter. For example, 'happy' is used in the form of 'haaaaaaappy'. Though this implies the same word 'happy', the classifiers consider these as two different words. To improve this and make words more similar to generic words, such sets of repeated letters are replaced by two occurrences. Thus demonatize losee would be replaced by demonetization effext etc..

Stop Word Removal: It used to eliminate that words that occurs frequently such as article, prepositions, conjunction and adverbs. These stop words depends on language of the text in questions. For example, words like the, and, before, while, and so on do not contribute to the sentiment. Remove all URLs (e.g. www.xyz.com), hash tags (e.g. #topic), targets (@username). Remove all punctuations, symbols, and numbers. Expand Acronyms (we can use a acronym dictionary).

Stemming: In information retrieval, stemming is the process of reducing a word to its root form. For example, walking, walker, walked all these words are derived from the root word walk. Hence, the stemmed form of all the above words is walk` The raw data is pre-processed to improve quality.

Feature Extraction

Features in reviews are extracted so that it helps to know which feature has positive comment and which one has negative.

Sentiment Classification

In this step, subjective sentences are classified in positive, negative, good, bad; like, dislike, but classification can be made by using multiple points.

Presentation of Output

The main objective of sentiment analysis is to convert unstructured text into meaningful information. When the analysis is finished, the text results are displayed on different types graphs. Also time can be analyzed and can be graphically displayed constructing a sentiment time line with the chosen value (frequency, percentages, and averages) over time.

Problematic figures that we faced in sentimental analysis

A figure of speech is any artful deviation from the ordinary mode of speaking or writing [4]. In the tradition of Aristotle, figures of speech can be divided into two groups: schemes and tropes. The function of schemes and tropes is to carry out a transference of some kind; schemes are characterized by a transference in order, while tropes are characterized by a transference in meaning.

For example, the most problematic figures of speech in natural language processing are irony and sarcasm, which are collocated under the tropes group. While irony is often used to emphasize occurrences that deviate from the expected, such as twists of fate, sarcasm is commonly used to convey implicit criticism with a particular victim as its target [5]. Examples of sarcastic and ironic sentences are:

1. RT @Timcast: Youtube flags my videos for demonetization then after most people have already watched them they will reinstate ads,
2. RT @yksingh26: Wasn't #DeMonetisation supposed to curb stone pelting ?Didn't we hear voices propagating theories about Demonetization
3. @narendramodi Dada,Im suffering with financial crisis & I use RupayCard everytime after u announce demonetization but still waiting fr prize

Experimental Setup

The experimental setup of this approach presents the research methodology employed the tools used to analyse the opinion We used a Mac OSX i7 2.9 GHz with 8GB DDR3 RAM in this study open source libraries,packages.API are extensively used such as:

Data Extraction

Tokenisation

We start our analysis by breaking the text down into words. Tokenisation is one of the most basic, yet most important, steps in text analysis. The purpose of tokenisation is to split a stream of text into smaller units called tokens, usually words or phrases. While this is a well understood problem with several out-of-the-box solutions from popular libraries, Twitter data pose some challenges because of the nature of the language.

Fig 2 Screenshot of twitter dataset on demonetization

	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSII	id	replyToUI	statusSou	screenNa	retweetC	isRetweet	retweete
1	RT @rssurjewala: Critical question: Was PayTM info	FALSE	0	NA	23-11-2016 18:40	FALSE	NA	8.01E+17	NA	<a href="h	HASHTAGI	331	TRUE	FALSE
2	RT @Hemant_80: Did you vote on #Demonetization	FALSE	0	NA	23-11-2016 18:40	FALSE	NA	8.01E+17	NA	<a href="h	PRAMODK	66	TRUE	FALSE
3	RT @roshankar: Former FinSec, RBI Dy Governor,	FALSE	0	NA	23-11-2016 18:40	FALSE	NA	8.01E+17	NA	<a href="h	rahulja13(	12	TRUE	FALSE
4	RT @ANI_news: Gurugram (Haryana): Post office en	FALSE	0	NA	23-11-2016 18:39	FALSE	NA	8.01E+17	NA	<a href="h	deeptiyvd	338	TRUE	FALSE
5	RT @satishacharya: Reddy Wedding! @mail_today c	FALSE	0	NA	23-11-2016 18:39	FALSE	NA	8.01E+17	NA	<a href="h	CPIMBadli	120	TRUE	FALSE
6	@DerekScissors1: India's #demonetization: #Blackm	FALSE	0	DerekScissors	23-11-2016 18:39	FALSE	NA	8.01E+17	2.59E+09	<a href="h	ambazaari	0	FALSE	FALSE
7	RT @gauravcsawant: Rs 40 lakh looted from a bank i	FALSE	0	NA	23-11-2016 18:38	FALSE	NA	8.01E+17	NA	<a href="h	bhodia1	637	TRUE	FALSE
8	RT @Joydeep_911: Calling all Nationalists to join...	FALSE	0	NA	23-11-2016 18:38	FALSE	NA	8.01E+17	NA	<a href="h	KARUNAS	112	TRUE	FALSE
9	RT @sumitbhati2002: Many opposition leaders are	FALSE	0	NA	23-11-2016 18:38	FALSE	NA	8.01E+17	NA	<a href="h	sumitbhat	1	TRUE	FALSE
10	National reform now destroyed even the essence o	FALSE	0	NA	23-11-2016 18:38	TRUE	NA	8.01E+17	NA	<a href="h	HelpIndia	0	FALSE	FALSE
11	Many opposition leaders are with @narendramodi	FALSE	1	NA	23-11-2016 18:37	FALSE	NA	8.01E+17	NA	<a href="h	sumitbhat	1	FALSE	FALSE
12	RT @Joydas: Question in Narendra Modi App where	FALSE	0	NA	23-11-2016 18:37	FALSE	NA	8.01E+17	NA	<a href="h	MonishGa	120	TRUE	FALSE
13	@Jaggesh2 Bharat band on 28??<ed><U+00A0><U+0	FALSE	0	Jaggesh2	23-11-2016 18:37	FALSE	8.01E+17	8.01E+17	1.23E+09	<a href="h	yuvaraj_k	0	FALSE	FALSE
14	RT @Atheist_Krishna: The effect of	FALSE	0	NA	23-11-2016 18:36	FALSE	NA	8.01E+17	NA	<a href="h	PMKejri	45	TRUE	FALSE
15	RT @sona2905: When I explained #Demonetization	FALSE	0	NA	23-11-2016 18:36	FALSE	NA	8.01E+17	NA	<a href="h	hkgupta16	50	TRUE	FALSE
16	RT @Dipankar_cpiml: The Modi app on #DeMonetiz	FALSE	0	NA	23-11-2016 18:35	FALSE	NA	8.01E+17	NA	<a href="h	aazaadpar	45	TRUE	FALSE
17	RT @roshankar: Former FinSec, RBI Dy Governor,	FALSE	0	NA	23-11-2016 18:35	FALSE	NA	8.01E+17	NA	<a href="h	darkdestir	12	TRUE	FALSE

The tokenization is probably far from perfect, but it gives you the general idea. The tokenisation is based on regular expressions (regexp), which is a common choice for this type of problem. Some particular types of tokens (e.g. phone numbers or chemical names) will not be captured, and will be probably broken into several tokens. To overcome this problem, as well as to improve the richness of your pre-processing pipeline, you can improve the regular expressions, or even employ more sophisticated techniques like Named Entity Recognition.

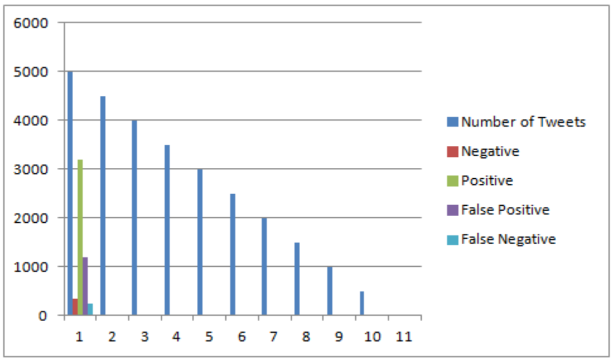
Sentiment Analysis Techniques

Sentiment analysis has two main techniques: machine learning based and lexicon based techniques. The machine learning approach (ML) applies the famous ML algorithms and uses linguistic features. The lexicon-based approach depends on a sentiment lexicon, a collection of known and precompiled sentiment terms. It is divided into dictionary based approach and corpus-based approach. This corpus-based approach used statistical or semantic methods to find sentiment polarity. [8].The text classification methods using ML approach can be divided into supervised and unsupervised learning methods.

Implementation on Twitted Dataset

Here, Python language used for implementation. Python language offers maximum support for sentiment analysis using Python twitter API. The reason for using Python language is, when the dataset is big, it is fast and efficient in terms of performance. This language has a wide scope of performing the analysis using different machine learning alogorithms such as Bayesian theory pf probability, SVM, by its rich set of libraries. The packages in the python languages tools are updated regularly and have greater number of probabilistic and statistical functions. The reason we are using Python language is when the dataset is big, it is fast and efficient in terms of performing classification and feature extraction espically when dataset is updated regularly as in twitters messages. In this paper the recent Demonetization process that takes place in India is centred and the twitter dataset for the demonetization was collected for doing sentiment analysis. Demonetization dataset consists of more than 10000 thousands of tweets. After collecting the dataset, pre-processing steps is done on demonetization dataset. We remove the retweeted data sets and the dataset is reduced to 4000 in number. Then we have to create two separate dataset called positive and negative. , Later the dataset is compared with the text data. If the word is available on positive entries of the text means it considered as a positive text. If it is available in negative entries of the text means it is considered as a negative tweet. Otherwise it is considered as a neutral polarity. This implementation is done by using Python scripts with necessary libraries installed for analysing tweets blogs. The result is computed and is projected in the figure shown below.

As shown in above graph almost the positive tweets span almost 64% and 7% of the tweets falls on the negative scale. Though with this we can derive a conclusion about the impact created by the demonetization, the graph also shows details about false positive scale is 24% and false negative 4% tweets. Both false positive and false negative impact have a huge gap on the scale. The false positive and false negative impact can be studied better using specialized algorithms Bayesian theory approach and lexicon or dictionary based approaches.



CONCLUSION

Applying sentiment analysis to data mining on large number of data set especially when they are unstructured in nature is the primary focus of find the best system for sentiment analysis. Demonetization is one of the current debate topics in Indian context where economy growth are analysed it GDP and impact of Modi government bold decision on demonetization. Government of India have interested to find the answer to the question, by analyzing the micro-blogs to understand the impact and feeling of the people of India on their demonetization move?. Demonetization dataset used to find out the view of different people on the demonetization by analyzing their tweets from twitter.

References

1. B.Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool, SanRafael, CA, 2012.
2. J.M. Wiebe, R.F. Bruce, T.P. O'Hara, Development and use of a gold-standard data set for subjectivity classifications, in: Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, ACL '99, Association for Computational Linguistics, 1999, pp. 246-253.
3. N.Jindal, B.Liu, Identifying comparative sentences in text documents, in: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '06, ACM, 2006, pp. 244-251.
4. E.P.J. Corbett, Classical Rhetoric for the Modern Student, second ed., Oxford University Press, Oxford, UK, 1971.
5. S. McDonald, Exploring the process of inference generation in sarcasm: a review of normal and clinical studies, Brain Lang. 68 (3) (1999) 486-506.
6. F.Lazarsfeld, R.K.Merton, Friendship as a social process: a substantive and methodological analysis, in: M. Berger, T. Abel, C.H. Page (Eds.), Freedom and Control in Modern Society, Van Nostrand, New York, 1954, pp. 8-66.
7. B.Pang, L.Lee, Opinion Mining and Sentiment Analysis, Found. *Trends Inf.Retr.*2(1-2)(2008)1-135.
8. A.B. Goldberg, X. Zhu, S.J. Wright, Dissimilarity in graph-based semi-supervised classification, in: AISTATS, vol. 2, 2007, pp. 155–162.

9. D.Hopkins, G.King, Extracting systematic social science meaning from text, Manuscript available at http:// gking. harvard.edu/ files/words.pdf, (2007).
10. M.Bansal, C.Cardie, L.Lee, The power of negative thinking: exploiting label disagreement in the min-cut classification framework, in: COLING (Posters), 2008, pp. 15-18.
11. S.C. Greene, Spin: Lexical Semantics, Transitivity, and the Identification of Implicit Sentiment, ProQuest, Ann Arbor, MI, 2007.

How to cite this article:

Ravinder Kumar and Anil Kumar (2017) 'Sentimental Analysis of Demonetization in India Using Machine Learning Algorithms ', *International Journal of Current Advanced Research*, 06(07), pp. 4941-4945.
DOI: http://dx.doi.org/10.24327/ijcar.2017.4945.0618

****