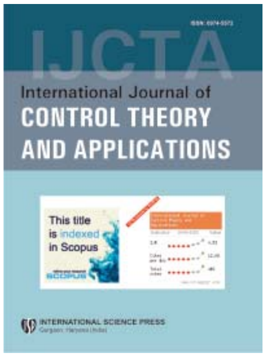

International Journal of Control Theory and Applications

ISSN : 0974–5572

Volume 9 • Number 40 • 2016

Measuring the Response of Citizens to New Initiatives by Analysis of Twitter Data

Subburaj Ramasamy[a] Akash Govind[a] Abhishek Khari[a] and Srishti Mishra[a]

[a]SRM University, Chennai, Tamil Nadu – 603203

E-mail: subburaj.r@ktr.srmuniv.ac.in, Corresponding Author

Abstract: Life changing events happen all around the world which can affect human beings socially, economically, environmentally or politically. During these events, a large number of people tend to express their views and opinions on micro blogging websites. One such micro-blogging tool is Twitter. It is an online News and social networking service where registered users canpost and interact with messages called "tweets," that are restricted to 140 characters. We aim to perform a sentimental analysis on the data received from Twitter, concentrated on one specific topic, to realize people's sentiment towards an event. Also, we studied how the opinion of the masses change with time. This mechanism can help policy makers and decision makers to take informed and beneficial decisions.

Keywords: *Micro-blogging, Sentiment Analysis, Twitter*

1. INTRODUCTION

Sentiment analysis has emerged as one step solution to multiple problems that concern data scientist and researchers. The inconsistent nature of the micro-blog content - such as Twitter makes difficult for sentimental analysis to be practically implemented. Some of these challenges brew from the exponentially increasingvolume of data generated by users put together with the lack of content resulting from restriction on number of characters of the text and a tendency to use abbreviated language slangs to express sentiments.Multiple companies observe user actions and reactions and replythem on Twitter. The task is to develop technology to observe and quantify an overall sentiment [1].

Demonetization review on Twitter : The demonetization of 500 and 1000 banknotes was a step taken by the Government of India on 8 November 2016, declaring a ban on the usage of all 500 and 1000 INR notes of the Mahatma Gandhi Series as legal currency in India from 9 November 2016. The announcement was made by Prime Minister Narendra Modi, in a live Television program at 20:15 IST on 8 November 2016.

1.1. Machine Learning

Machine learning is a sub-domain of computer science that provides computers with the ability to learn without being explicitly programmed. It was evolved from a study of pattern recognition and computational learning theory. Machine learning focuses on the development of computer programs that can act accordingly when exposed to new data. The process of machine learning is similar to that of data mining. Both systems aim at finding patterns for recognition. However, machine learning doesn't extract data for human understanding, unlike data mining applications. Data extracted by machine learning is used to detect patterns in data and process as per the need.

These algorithms can be classified into two broad categories which are supervised and unsupervised. Supervised algorithms apply previously learned knowledge from datasets. Unsupervised algorithms can process data sets to draw inferences.

1.2. Natural Language Toolkit

NLTK is a leading platform for building Python programs to work with human language data. It provides an easy interface to more than 50 corpora and lexical resources and guides, along with a collection of data processing libraries for classification, tagging, parsing, and semantic reasoning and wrappers for industrial-strength NLP libraries.

The pattern is a web mining module for Python, and the Pattern.en module is a natural language processing (NLP) toolkit [6].

2. LITERATURE SURVEY

Microblog data like Twitter, allows registered users to post opinions about "everything," gives rise to newer challenges.

Earlier approaches used distant learning as a methodcollect data. They used tweets that ended withemoticons like ':)' ':-)' as positive and emoticons like' :(' ':-('as negative. Models were developed using Naive Bayes, Support Vector Machines (SVM), etc. Unigram and bigram model with POS was used as features. It was observed that unigram goes beyond all other models.

Another approach collected data following a similar distant learning paradigm. Classification task is done by: subjective vs. objective. For subjective data, tweets ending with emoticons were collected. For objective data, Twitter accounts of newspapers like "New York Times", "Washington Posts" etc. were scraped. Their report stated that POS as well as bigrams were helpful.

Another significant effort was made on Twitter data was made in 2010. For training purpose, data from three websites was collected labels and used almost 1000 manually labeled tweets for tuning, and another 1000 tweets were manually labeled for testing. They propose the use of syntax features of tweets like retweet, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words. Researchers have attempted to develop systems to automatically label words that indicate opinions as being either positive or negative [2].

A study classified couples into happy and unhappy category depending upon their communication on instant messaging platforms [7].

An earlier approach, The SentiWordNet lexicon was very noisy. SentiNet is and affective resource for opinion at concept-level and sentiment analysis. It is also available for public use[3][4].

2.1. Vader Algorithm - Valence Aware Dictionary for sentiment Reasoning

VADER is a lexicon and rule-based sentiment analysis tool that is specifically constructed to suit sentiments expressed in social media and works well on texts from other sub-domains [12].

Following are some parameters considered by Vader Algorithm:

1. Negations (e.g., "*not* good")
2. Use of contractions as negations (e.g., "*wasn't* very good")

3. Use of **punctuation marks** to realize increased sentiment intensity (e.g., “Good!!!”)
4. The conventional use of **word-shape** to notify emphasis (e.g., using ALL CAPS for words/phrases)
5. Using **degree modifiers or intensity boosters** to alter sentiment intensity, e.g., “very” and intensity *dampeners* such as “kind of”)
6. Use of many sentiment-laden **slang words as modifiers** such as ‘uber’ or ‘friggin’ or ‘kinda’
7. Use of many sentiment-laden **emoticons** such as :) and :D
8. Understanding**acronyms** (for example: ‘lol’)

2.1.1. Vader_lexicon.txt

It has been empirically validated by multiple independent human judges, VADER incorporates an incredible sentiment lexicon that is especially aligned to micro-blog-like contexts, e.g., Twitter.

The VADER sentiments take care of both, Polarity as well as Intensity in the context of data collected from micro blogging websites and also works correctly on other domains of text.

It becomes a necessity to rely on existing lexicons as creating new ones can be really labor intensive as well result in errors. Existing word-banks like LIWC, ANEW etc. helped in creating a new list.LIWC does not include acronyms, emoticons, or slangs, which are considered necessary for sentiment analysis of twitter data [5][8].

9,000 lexical feature candidates were constructed. Wisdom-of-the-crowd13 (WotC) approach to get an estimated value for the intensity of each context-free candidate feature was used. Lexical features were rated on a scale from “–4to +4”, with a “0” for Neutral.

Every feature that had a non-zero mean rating, and whose standard deviation was less than 2.5 as determined by the aggregate of ten independent raters. This reduced approximately 1500 features and left only 7,500 lexical features.

2.1.2. VaderSentiment.py

The Python code for the rule-based sentiment analysis engine implements all the rules including grammatical and syntactical rules, making an impact of each rule on the intensity of sentiment in tweets. Importantly, these computations go beyond the limits of a bag-of-words model.

For example, degree modifiers like intensifiers and booster words, impact sentiment intensity by altering the intensity. Consider these examples:

1. “This place is extremely good.”
2. “This place is good.”
3. “This place is marginally good.”

3. IMPLEMENTATION

Our proposed method aims to realize the trend how the people’s opinion changed with time. To achieve this, firstly data from Twitter focusing on a single topic is extracted and saved in a CSV format. This is done using TwitterAPI which always a registered user to scrape Twitter data.

```
FromTwythonimportTwython
TWI_KEY = ‘xxxxx’
TWI_SECRET = ‘xxxxx’
```

```
TWI_TOKEN = 'xxxxxx'
TWI_TOKEN_SECRET = 'xxxxx'
t = Twython(app_key=TWI_KEY,
app=TWI_SECRET,
oauth=TWI_TOKEN,
oauth_token_secret=TWI_TOKEN_SECRET)
search = t.search(q='#omg',
count=100)
tweets = search['statuses']
```

Secondly, sentiment polarity of each tweet is computed. This score is calculated by summing the valence scores of each word present in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and +1 (most extreme positive). This is probably the effective metric if you want a single uni-dimensional measure of sentiment for some data.

Thirdly, a count of all positive, negative and neutral tweets is done.

```
tweets.sentiment_type.value_counts().plot(kind='bar',title="sentiment analysis")
```

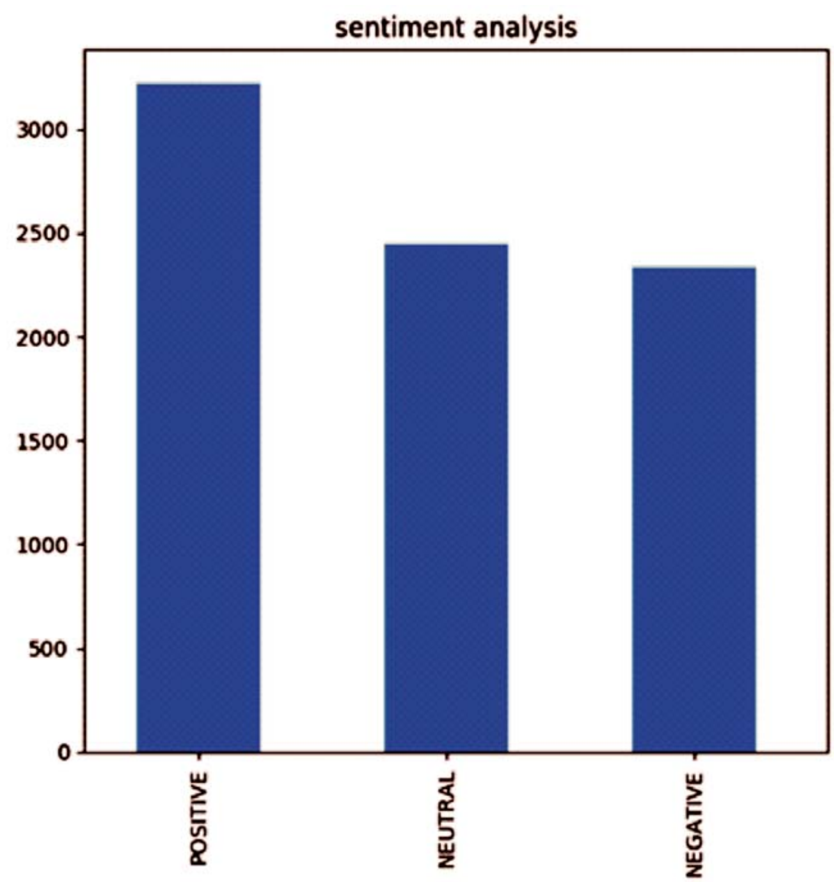

Figure 1

Lastly, a graph is plotted, mean of (sentiment polarity) versus hour is plotted to predict the trend of change of opinion.

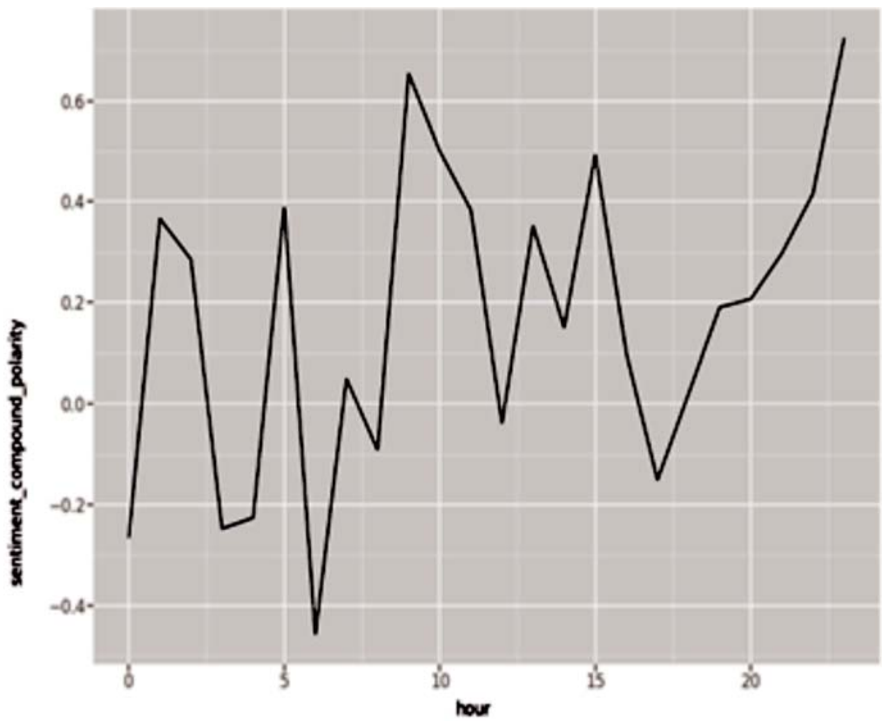

Figure 2

The above graph portrays the trend how human opinion changed with time.

4. SUMMARY & CONCLUSIONS

Sentiment Analysis is a powerful tool for decision makers from the lowest to the highest level. We developed a tool for sentiment analysis which can work on any Twitter data and provide sentiments regarding positive, negative and neutral. The tool also provides long-term variations in the sentiments of human beings on a chose topic. In this study, we look at a popular micro-blog called Twitter and build models for classifying "tweets" into three categories that are positive, negative and neutral. The major focus of this project was to carry out an experiment to realize the trend of variation in human emotion with time. We used Vader Algorithm for the need of a lexicon and rule-based sentiment analysis tool that is specifically aligned to sentiments expressed on Twitter. This task was successfully executed by our proposed approach,

REFERENCES

[1] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proc. WLSM-11s.

[2] Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0. In Proc. of LREC-10.

[3] Cambria, E., Havasi, C., & Hussain, A. (2012). SenticNet 2. In Proc. AAAI IFAI RSC-12.

[4] Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010). SenticNet. In Proc. of AAAI SCK-10.

[5] Davidov, D., Tsur, O., & Rappoport, A. (2010). Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. ICCL-10.

[6] De Smedt, T., & Daelemans, W. (2012). The pattern for Python. Journal of Machine Learning Research, 13, 2063–2067.

[7] Hancock, J. T., Landrigan, C., & Silver, C. (2007). Expressing emotion in text-based communication. In Proc. CHI-07.

[8] Hutto, C. J., Yardi, S., & Gilbert, E. (2013). A Longitudinal Study of Follow Predictors on Twitter. In Proc. CHI-13.

[9] Liu, B. (2010). Sentiment Analysis and Subjectivity. In N. Indurkhya & F. Damerau (Eds.), Handbook of Natural Language Processing (2nd ed.). Boca Raton, FL: Chapman & Hall.

[10] Liu, B. (2012). Sentiment Analysis and Opinion Mining. San Rafael, CA: Morgan & Claypool.

[11] Lu, Y., Castellanos, M., Dayal, U., & Zhai, C. (2011). Automatic construction of a context-aware sentiment lexicon: an optimization approach. In Proc. WWW-11.

[12] https://github.com/cjhutto/vaderSentiment