General Purpose Opinion Mining System for Social Media Platform using Machine Learning Based Api

Urmita Mehta, Dhanraj Verma

Abstract: With the growth of technology there is lot of data available on the internet. Social media platform like Twitter, FaceBook,Google+,whats app,instagram etc are the platform that allow people to share and express their views, ideas, thoughts and experiences about any topics, post messages across the world. There are mainly two types of textual information available on social media platforms. One is fact and another next one is sentiments or more formally it can also called opinion. The social media is a platform where people gives their opinion regularly. These opinions may contain some factual information. For the analysis of sentiments we required some tools. Mostly text based mining is used for opinion mining. Text mining required lots of different tools and research work. This paper, provides a machine learning techniques for opinion calculation in Twitter..

Keywords: matplotlib, nltk, Opinion Mining, tkinter,

I. INTRODUCTION

Social Media is a platform where people express and share their feelings, thoughts, experiences, ideas and views. These expressions could be for any events, places, Political parties, leaders, celebrities, community, countries, or of any items.OM is the study of opinion, sentiments and emotions presented in the form of text, images, video or symbols. Opinion mining is the process of determining the opinion or filling of a piece of text. Opinion Mining is the process of determining the opinion of an individual. For this purpose people use social media like Twitter, face book, Blogs, Instagram, different social media web sites and so on. The Opinion mining extracts the data and analyse the sentiments or opinions of people. During last few decades' researchers, academics and various organizations are now focuses on sentiment Analysis. This Opinion Mining analyses mood and views of people by extracting their opinions posted on social media platform and classifies them on the basis of their polarity.

This extracted information can be useful for manufacturers to analyse their products. Companies across the word implementing machine learning to do it automatically. It is a super use of opinion inside the text. Once company understand how the customer feels after analysing their comments and reviews. They can identify what kind of things they like or dislike and built things like recommendation system. Or more target marketing systems for them. Opinion mining is the task to classify a short paragraph of text is being positive or negative. So the OM is a system that gathers and analyse the data and gives the desired results. This paper proposes an opinion mining tool for twitter data. Many work already been done in this field. Opinion mining calculates opinions and subjectivity of data. OM also classifies the polarity of a text at different levels. Contents could be +ve, -ve or neutral. That means polarity is classified into positive (means good or opinion that favours), negative (means bad or opinion that neglect or disfavour) or neutral (zero preferences). So the OM tool calculates the opinions by calculating text through algorithms and gives opinion. Opinion mining system has many applications like in stock market, political issues, news article or marketing etc.

Sentiment analysis analyse the public opinion and determine the Following 3things;

1. Subjectivity: The text is analysed to determine whether it is subjective in nature, means the

Analyse text contain any opinionated contents.

2. Polarity: Polarity is classified into 3types i.e. positive opinion or negative opinion and neutral.

3.PolarityStrength: The polarity can be divided into different levels e.g. very high polarity, high polarity, low polarity, very low polarity and so on.

Polarity strength can be represented as strongly positive, mildly positive, weekly positive, weekly negative, mildly negative, strongly negative.

II. RELATED WORK

Bo Pang and et al (2008) [1] presented a survey about different techniques and approaches used in opinion mining. Authors also focused of challenges raised during opinion mining. Four main challenges were discussed in this paper, first is opinion mining for subjective material. Second problem of document relevancy, third was calculating overall sentiments of documents and final one was representation of opinion information in summarized form.

   Urmita Mehta*, Assistant Professor in Department of Computer Science at S.J.H.S Gujarati Innovative College of Commerce and Science

   Dhanraj Verma, Currently working as Professor in Department of Computer Science & Engineering, Dr APJ Abdul Kalam University,

General Purpose Opinion Mining System for Social Media Platform Using Machine Learning Based Api

Various areas of application were also included in this article like sub-component technology (recommendation systems, detection of flames, ads on websites etc.), business intelligence, government intelligence, politics etc. Challenges included in this article were contracts with standard fact based textual analysis, some factors that makes. Opinion mining difficult, like use of double mining words. Ketan Sarvakar et al (2018) [2] proposed a method of classification with add some pre-processing tasks to filter the text. So that more accurate results can be generated. Proposed method was feature-based sentiment classification and it was used for movie review classification. The enriched dictionary and NGram tokenizer were used to improve accuracy of classification. Author used a Weka tool for experimental purpose. Shriji Madhu (2018) [3] proposed an opinion mining method which used the concept, bag of words, part of speech and NLP. This proposed method is used to identify probability of suicide attempt by analysing persons tweets and messages. Author used Textblob module of python programming language to develop the proposed system. Soujanya Poria et al. (2016) [4] presented the deep learning approach in extracting opinions from text. Opinion holder parses the text to calculate opinion about products or service. Authors also developed some linguistic patters and integrated them with neural network. In this research work author proposed a deep CNN architecture with seven layers. One input layer, two convolution layers, two max-pooling layers, one fully connected layer and one output layer. Each convolution layer is followed by one max-pooling layer. This article also covers some background details of Dee CNN. They used back propagation algorithm in to train the neural network. Model is trained using two datasets, Google news and Amazon product review. Two main features are selected for the network one is word embedding and second is part of speech tags. These two features vectors were concatenated and fed to CNN. Through their experimental results they achieved 95% of confidence level. Shiliang Sun et al. (2017) [5] presented a study of natural language process (NLP) techniques for sentiment analysis. They mentioned general information science technique that is needed for text pre-processing like tokenization, word segmentation, a part of speech (PoS), tagging, parsing etc. in step with authors sentiment analysis is divided into 3 levels, first document level, second sentence level and third is fine-grained level. Authors conjointly enclosed completely different OM techniques in their analysis paper. Information fusion in terms of opinion mining, that integrates information from different sources to give some analytical result. In this research information fusion and its parameters for OM is discussed like, data sources, formatting of information, integration of data, methods of fusion. Most data available is unstructured so author also defined text pre-processing method to convert unstructured data into structured data. Different NLP tools were also explored in this research paper like NLTK, Open NLP, Core NLP, Gensim, Fuan NLP, LTP and Niu Parser.M. Rathan et al. (2018) [6] presented a feature level sentiment analysis model for micro-blogging site (Twitter). Emoji detection, spelling correction and emotion detection features were included for data mining. By the experimental results author shows that automated training data provides good accuracy then manual training data set. Multiple model of SVM classifier is used to get accuracy. Nazan Ozturk et al (2018) [7] investigated the public views and filling about Syrian refugee crisis. Author selected 2381,297 tweets in two different languages (Turkish and English). Author analysed both Turkish tweets and English tweets separately. Turkish tweets are more positive than English tweets. 35% Turkish tweets were positive and 12% English tweets were positive. Thien HaiNgayen et al. (2015) [8] modeled a stock prices movement prediction system using sentiments on social media. Topic Sentiment Latent Dirichlet Allocation (TSLDA) feature was introduced in model. Author also compared TSLDA base sentiment analysis with two other methods, Latent Dirichlet Allocation (LDA) based and, Joint Sentiment/Topic (JST) based. TSLDA is extended model of LDA. Two different datasets, Historical price dataset (from yahoo finance) and Message Board Dataset (from yahoo finance message board) were used for development of stock prediction model. Forhan Hasson Khan et al. (2014) [9] proposed a hybrid classification scheme for opinion mining on twitter feeds. Their research was focused on problems in classifications like, classification accuracy, data sparsity and sarcasm. Their proposed method introduced various pre-processing steps before tweets are feed into actual classifier. Amit Gupte et al. (2014), [10] presented the comparative study of classification algorithms used in sentiment analysis. Authors included all the major steps of sentiment analysis. Steps are text extraction, text refinement, text classification and score aggregation. This literature also included four different classifiers used in sentient analysis.

III. OPINION MINING APPROACHES

There are generally two approaches to do opinion mining.

A. Lexicon Based Opinion mining

The first one is the lexicon based approach. In this approach given text are divided into smaller words, phrases and sentences called token and this process is known as tokenization. Then number of words and their frequency is counted. This resulting tally is called Bag of Word model. Next processes look up the subjectivity of each word from an existing lexicon, which is a database of emotional values of words. These words are pre-recorded by researchers by this values analyser can compute the overall subjectivity on text. There are mainly 2 different approaches identified for lexicon based opinion mining.

i)Dictionary based approach

ii)Corpus-based approach.

B. Machine Learning Based Opinion mining

The other approach uses machine learning. If there are some text labelled with positive or negative.

9728

We can train a classifier on it and give a new text to classifier to take decision weather text are positive or negative. This approach works on semantic of word. Machine learning based sentiment analyser is difficult to implement with compare to lexicon based sentient analysis system. Machine learning based opinion mining can be further classified in supervised and unsupervised learning. Supervised learning based on existing labelled documents. These documents give supervision platform to upcoming text. There are many classifiers which may be used in supervised learning like Probabilistic classifiers (Naive Bayes Classifier [18], Bayesian Network [19], and Maximum Entropy Classifier [20]), linear classifiers (Support Vector Machines Classifiers [21] and Neural Network [22]), Decision tree classifiers and Rule-based classifiers [16]. The text classification is mainly used to classify text documents in certain predefined categories. In a machine learning techniques, two sets of data are needed:
1.Training Set
2.Test Set.

IV.TWITTER SENTIMENT ANALYSIS WITH PYTHON

Python is a Language, which is created by Guido van Rossum. This is very popular for its code and readability. Python provides a large standard library, which can be used for various applications for example NLP, machine learning, and data analysis etc

TextBlob is a Python library that processes data in a text form. It provides an Application programming Interface for diving NLP like speech tags, noun and phrase extraction, S.A, classification etc. **Sentdex** is an algorithm for analysis of Sentiments, which is a combination of "sentiment" and "index." Sentdex recognizes the emotions, which is used by people in their messages posts on social media and converts it to the computer language.

A. Python Modules Used

i)tweepy- module provides all supporting classes and method that can be used to access twitter data.

ii)tkinter- module is used to design the interface of application.

iii)nltk- is natural language toolkit. It provides dataset for training the algorithm.

iv)Textblob- module provides the various text operations on text. Sentiment calculation is one of them.

v)matplotlib- module is used to represents the results in graphical form like pi chart, bar chart etc.

V METHODOLOGY

A Dataset

Proposed sentiment analysis system works on real time twitter posts (tweets). So that in this implementation there is no pre stored datasets. This proposed solution is taking real time tweets directly from the twitter.

To access twitter post and other relevant information like source of tweets, date of post, time of post etc. We need a twitter account, which work as a gateway to for application to access twitter data

In order to calculate sentiments, we need to collect data some source, here we are taking from Twitter. This data undergoes various steps of pre-processing which makes it more machine sensible than its previous form.

Proposed work proceeds in following four steps:
1. Get the data from Twitter
2. Access the data
3. Calculate the Opinions
4. Represent results

To run social media campaigns on Twitter , API key and access token must requires for every profile that is employed. API stands for Application Programming Interface. API keys square measure series of code created by Twitter to permit Twitter profile to poll Twitter feeds. For that a Twitter application has to be created by progressing to the http://apps.twitter.com Associate in natural language to make an application and wish to 5 info as requested by the dialog boxes. Once giving all the fundamental knowledge Twitter application is formed. That enable user to maneuver social profile to perform. After obtaining this Twitter certificate, Twitter knowledge has to be retrieve. As during the implementation the Python language is employed. Python provides Application Programming Interface known as "Tweepy" to access the Twitter account. Thus first of all Tweepy API should be putting in on system. Once installed 3 main categories of Tweepy has to be import in program. These are Stream, O.AuthHandler and StreamListener. During this planned implementation "Sentdex" is employed for opinion mining. Rather than Sentdex API one will use the other API like TextBlob or anybody will produce their own custom perform to calculate opinions of Tweets.
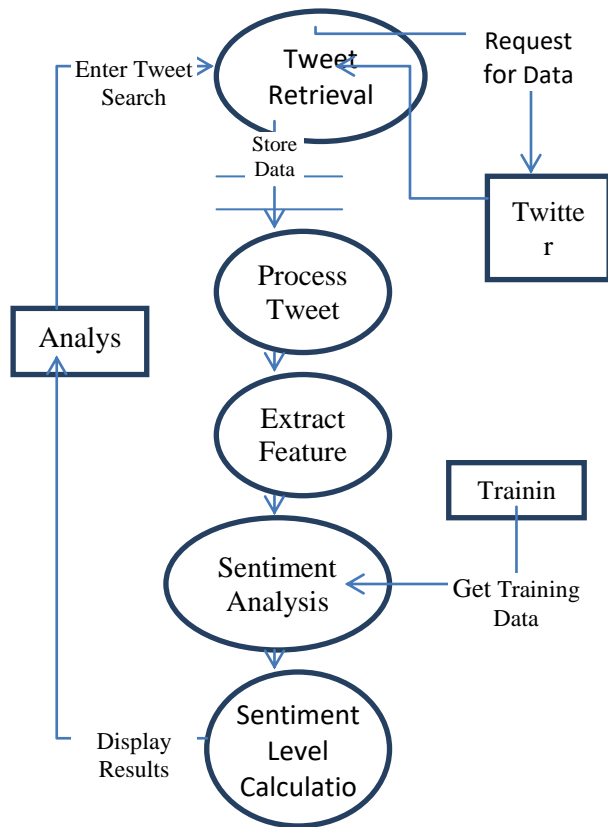
Figure 1

The implementation of tool can work for any type of subject, which is discussed on Twitter. Proposed tool work for real time tweets, so that execution time depends on two factors. First one is time taken in extracting the tweets and second, time taken in calculating opinion

.**Figure 2**

B Tools Features

- Real Time*:* implemented tool takes real time tweets from twitter handler so that it will give real time opinion about any search.
- Generalized: This tool is not specific for any domain. This works for any issue, person product or topics. Anyone can randomly picked up a topic and start OM.
- Text based Opinion Mining: Proposed tool uses twitter text material for OM. This current version of tool does not consider the images, audio, video and other multimedia content for Opinion Mining.

- Uses Supervised machine learning approach: Proposed tool uses NLTK data set to train the algorithm. Then algorithm classified the tweets in different categories (positive, negative and neutral). Classification algorithm is based on natural language processing approach.
- *Opinion level*: Algorithm not only classifies the tweets into positive, negative and neutral. It also calculates the opinion level in numeric form. Positive number for positive polarity, negative number for negative polarity, and zero represents the neutral.
- Overall and individual tweets analysis: Proposed tool give both types of analysis. Individual tweets analysis (polarities and subjectivity) and overall opinion analysis of all sample tweets used for mining

Figure 2

There are mainly three information is represented, Opinion value of proposed algorithm (Sentdex based), opinion value of existing algorithm (Textblob based) and accuracy of proposed algorithm. Figure 3 shows the overall analysis section of tool.
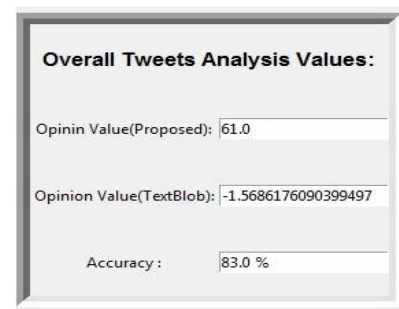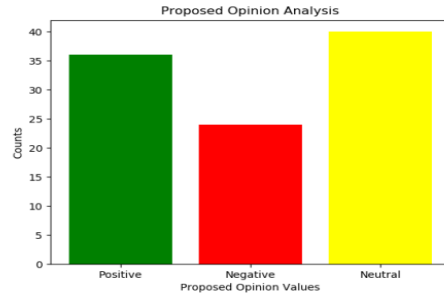

Figure 3

VI ANALYSIS

The implemented tool provides generalized OM. Through this tool analysis of any trendy topic can be done easily. Proposed tool gives real time OM. It uses supervised machine learning algorithm to calculate opinion. Two different algorithms comparative results are calculated and presented in this tool. NLTK dataset is used as a training dataset. Textblob and Sentdex API is used for calculating opinion. 88% accuracy is archived through Sentdex API. For calculation of accuracy we have considered Textblob OM as ideal analysis. Proposed tool works on real time data. So internet connectivity is required to run tool.

Table 1: Experimental Dataset Information

S.No.	Date	Start Time	End Time	Analysis Text	Number of Tweets
1	23-10-18	1:38:00 PM	1:41:00 PM	Bitcoin	100
2	23-10-18	1:43:00 PM	1:46:00 PM	Bitcoin	50
3	23-10-18	1:48:00 PM	1:53:00 PM	Bitcoin	100
4	15-12-18	1:35:11 PM	1:46:01 PM	Bitcoin	300
5	15-12-18	2:03:13 PM	2:03:13 PM	cricket	90
6	15-12-18	2:05:04 PM	2:09:22 PM	cricket	100
7	15-12-18	2:18:21 PM	2:20:57 PM	cricket	100
8	15-12-18	2:34:24 PM	2:36:02 PM	AQUAMAN	100
9	15-12-18	2:36:12 PM	2:47:47 PM	iPhone x	93
10	15-12-18	2:41:46 PM	3:11:46 PM	Mortal Engines	39

Table 2 Experimental Results of Textblob API based OM

S.No.	Analysis Text	Positive	Negative	Neutral
1	Bitcoin	37	17	46
2	Bitcoin	19	5	26
3	Bitcoin	38	26	36
4	Bitcoin	77	85	138
5	cricket	45	19	26
6	cricket	45	19	26
7	cricket	44	27	29
8	cricket	36	37	27
9	AQUAMAN	24	12	64
10	iphone x	17	13	63
	Mortal Engines	9	11	19
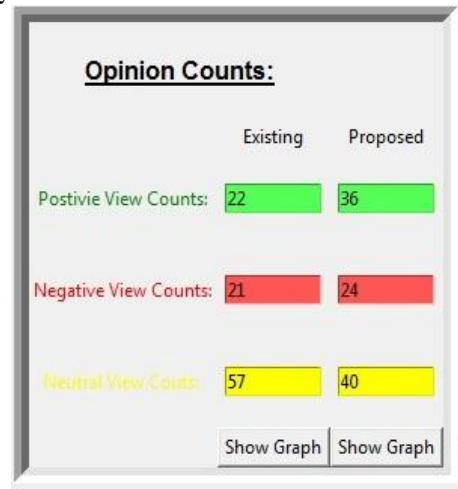
Table 3 Experimental Results of Sentdex API based OM

S.No.	Analysis Text	Positive	Negative	Neutral
1	Bitcoin	50	19	31
2	Bitcoin	23	7	20
3	Bitcoin	57	16	27
4	Bitcoin	121	69	110
5	cricket	36	14	40
6	cricket	45	20	35
7	cricket	35	27	38
8	AQUAMAN	19	11	70
9	iphone x	20	12	61
10	Mortal Engines	7	9	23

Proposed Opinion Analysis

This consists two different approach based opinion mining. One is "Textblob" based which is well known python module which used in many text manipulation applications. Another next one is "Sentdex" based opinion mining which calculates opinion value through NLTK data set Input data is taken form twitter. Data on twitter are vary unstructured that can affect the results of analysis. Tool performance also affected through trendiness of subject. If subject is trendy data extraction and analysis will be fast and if subject is not form current issues than analysis will take time.

B. Experimental Results

Table 4 Overall Opinion Value and Accuracy of Proposed Algorithm

Analysis Text	Opinion (Textblob)	Opinion (Sentdex)	Accuracy(%)
Bitcoin	10.39	134	85%
Bitcoin	4.64	49	88%
Bitcoin	6.5	163	81%
Bitcoin	-2.3182	218	85.33%
cricket	5.40756	105	82.00%

General Purpose Opinion Mining System for Social Media Platform Using Machine Learning Based Api

cricket	3.39413	63	93.00%
cricket	-2.3329	70	89.00%
AQUAMAN	2.33641	22	94.00%
iphone x	3.52534	47	96.77%
Mortal Engines	-1.2746	-16	89.74%

VII CONCLUSION

Opinion Mining can play an important role in understanding different issues and various other subjects. Opinion mining is an easy and fast way of doing this. Many opinion mining techniques are studied and explored. After analysis all of them, one efficient tool is developed. This tool works on Tweeter data. It extracts the tweeter data and analyses its text contents. It classifies it into three different categories positive, negative and neutral. The three main properties of this tool are; it works on real time tweets, it is generalized tool for any type of subject, and this tool This proposed opinion mining solution works on live streaming. It extracts live tweets form the Tweeter and processes them for opinion calculation. So that it gives real time analysis and we get to know current public opinion about selected topic. It has been observed during study of literature that there are some tools available for opinion mining.

But they are specific to any particular field. This proposed tool is generalized. As result analysis section shows that we have taken different categories for analysis and for each category proposed toll executed successfully.

REFERENCES

1. B.Pang,L.Lillian,"Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, 2008,Vol. 2, Nos. 1–2, pp. 1–135.
2. K.Sarvakar and U.Kuchara, "Sentiment Analysis of movie reviews: A new feature-based sentiment classification", International Journal of Scientific Research in Computer Science and Engineering, 2018,Vol.6, Issue.3, pp. 8-12.
3. Shrija Madhu, "An approach to analyze suicidal tendency in blogs and tweets using Sentiment Analysis", International Journal of Scientific Research in Computer Science and Engineering,2018, Vol.6, Issue.4, pp.34-36.
4. Soujanya Poria, Erik Cambria and Alexander Gelbukh, "Aspect Extraction for Opinion Mining with a Deep Convolutional Neural Network", Knowledge-Based Systems 2016
5. Shiliang Sun, Chen Luo, Junyu Chen, A Review of Natural Language Processing Techniques for Opinion Mining Systems, Information Fusion 2016.
6. M. Rathan, Vishwanath R. Hulipalled, K.R. Venugopal and L.M. Patnaik, "Consumer Insight Mining: Aspect Based Twitter Opinion Mining of Mobile Phone Reviews", Applied Soft Computing Journal, 2017.
7. NazanÖztürka and SerkanAyvazb, "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis", Telematics and Informatics 35 ,2018,pp136–147.
8. ThienHai Nguyen and KiyoakiShirai, "Topic Modeling based Sentiment Analysis on Social Media for Stock Market Prediction", Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on NaturalLanguage Processing, pages 1354–1364, Beijing, China,2015, pp 26-31.
9. Farhan Hassan Khan, Saba Bashir and UsmanQamar, "TOM: Twitter opinion mining framework using hybrid classification scheme", Decision Support Systems, 57, 2014, pp. 245–257.
10. AmitGupte, Sourabh Joshi, Pratik Gadgul and AkshayKadam, "Comparative Study of Classification Algorithms used in Sentiment Analysis", International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6261-6264

AUTHORS PROFILE

UrmitaMehtapursed Bachelor of Science from Devi Ahilya University of Indore, in 1998 and Master of Computer Management from ShriVaishnav Institute ofManagement and Science, DAVV University in 2001. She is currently pursuing Ph.D. from Dr A.P.J Abdul Kalam University and currently working as Assistant Professor in Department of Computer Science at S.J.H.S Gujarati Innovative College of Commerce and Science, since 2001. Her main research work focuses on Machine Learning Algorithms, Opinion Mining. She has 18 years of teaching experience and 4 years of Research Experience

DhanrajVerma Bachelor of Science from Vikram //university of Ujjain, 1997 and Master of Technology from Devi Ahilya University indore in 2007. PhD (CSE) from BU University in 2013.Currently working as Professor in Department of Computer Science & Engineering, Dr APJ Abdul Kalam University, indore.He is member of IEEE & IEEE computer society since 2012, Life member of CSI since 2012. He has published more than 24 research paper National/ international journals . His focus on network security, cloud security & privacy, Data mining, IoT& Computational intelligence based education He has 18 years of teaching experience and 6 years of Research Experience