

A Sentiment Analysis on Twitter Data

Vaishnavi Charde¹ Tejeswy Chimnekar² Shital Badole³ Shubhangi Ghadinkar⁴ B. Kumbhare⁵

⁵Professor

^{1,2,3,4,5}Department of Computer Science & Engineering

^{1,2,3,4,5}N.I.E.T.M .Nagpur, India

Abstract— Sentiment analysis is a broad research area in academic as well as business field. The term sentiment refers to the feelings or opinion of the person towards some particular domain. Hence it is also known as opinion mining. It can be expressed in terms of polarity, reviews or previously by thumbs up and down to denote positive and negative sentiments respectively. Sentiments can be analyzed using NLP, statistics or machine learning techniques. Sentiment analysis may ask questions regarding “customer satisfaction and dissatisfaction, “public opinion towards new iPhone series launched” etc. In real world, public or consumer opinions about some product or brand are very important for its sell. Hence sentiment analysis is a very important research area for real life applications i.e. decision making. For this purpose, the methodology we use is as follows: access the twitter API to extract the tweets about elections. The extracted tweets are then processed so as to convert all letters in the lower case, to special characters etc. which would make the further tasks more efficient. We classify these processed tweets using a supervised classification approach. The classifier used is Naïve Bayes Classifier to classify the tweets as positive, negative or neutral. The classifier is trained using tweets which bear a distinctive polarity. The result can be used further to gain an insight into the views of the people using twitter about a particular topic that is being searched by the user. It can help corporate houses to devise strategies on the basis of the popularity of their product among the masses. It may help the consumers to make informed choices based on the general sentiment expressed by the Twitter users on a product.

Key words: Data Mining, Feature Extraction Naïve Bayes Classifier, Natural Language Processing, Twitter, Sentiment, Review, Aspect, Hashtag, Entity, Emotions

I. INTRODUCTION

Sentiment analysis is also called opinion mining, is the field of study that analyses people’s opinion, sentiments, evaluation, appraisals, attitudes and emotions towards entities such as product, service, organization, individuals, issues, events, topics and there attributes. There are also many names and slightly different tasks, e.g., sentiment analysis, opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining etc. Sentiment analysis is a type of natural language processing for tracking the mood of the people about the particular product or topic. It involves in building a system to collect and examining opinion about the product made in blog post, comments, reviews or tweets.

Sentiment Analysis is to detect the polarity of text in consideration in textual form. It is also known as opinion mining as it derives the opinion of the speaker or the user about some topic. In other words, it determines whether a piece of writing is positive, negative or neutral.

For Example:

- Positive-Weather →The weather is pretty good this morning!
- Negative-Work→ Dammn ...I hate this clerical work
- Neutral-Bus→The bus arrives at 8 in the evening.

II. DIFFERENT CLASSES OF SENTIMENT ANALYSIS

Sentiments can be classified into three classes. i.e. positive, negative and neutral sentiments.

A. Positive Sentiments

These are the good words about the target in consideration. If the positive sentiments are increased, it is referred to be good. In case of product reviews, if the positive reviews about the product are more, it is bought by many customers.

B. Negative Sentiments

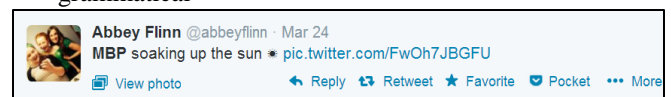
These are the bad words about the target in consideration. If the negative sentiments are increased, it is discarded from the preference list. In case of product reviews, if the negative reviews about the product are more, no one intend to buy it.

C. Neutral Sentiments

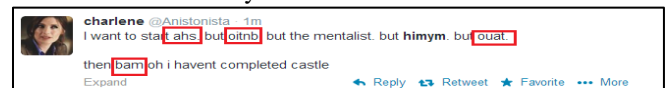
These are neither good nor bad words about the target. Hence it is neither preferred nor neglected.

III. CHALLENGES

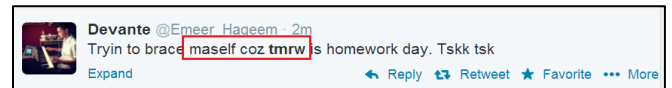
- Tweets are highly unstructured and also non-grammatical



- Out of Vocabulary Words



- Lexical Variation



- Extensive usage of acronyms like *asap*, *lol*, *afaik*

IV. LITERATURE SURVEY

In Twitter Sentiment Analysis: The Good, The bad and The OMG! “Paper, they have to investigate the utility of logistic feature for detecting the sentiment of Twitter messages. They have evaluated the usefulness of existing logical resources as well as features that capture information about the informal and creative language used in microblogging. A supervised approach has been introduced to solve the problem, but leverage existing hashtags in the Twitter data for building training data.

Sentiment analysis is the most important research area in business fields. Previously research was carried out for sentiment analysis in various domains like company product, movie reviews, politics etc. Previous research like Pang et al. has provided with the baseline for carrying out research in various domains. It uses star ratings as polarity signals in their training data. Even many authors have used the same concept provided by Pang et al.

The paper "Sentiment Analysis of Twitter Data" published in 2012 introduces a machine learning approach to implement sentiment analysis on data. They have performed sentiment classification of Twitter data where, the classes are "positive", "negative" and "neutral". Two kinds of models have been used: tree kernel and feature based model and both these models outperform the unigram baseline. For the feature-based approach, they performed feature analysis, which reveals that the most important feature are those that combine the prior polarity of words and their parts-of-speech tags.

V. PROPOSED SYSTEM

- Our proposed models were evaluated on real Twitter data.
- We have proposed a system that perform aspect level sentiment analysis on twitter data or tweets based on movies into three categories:
 - 1) Positive
 - 2) Negative
 - 3) Neutral

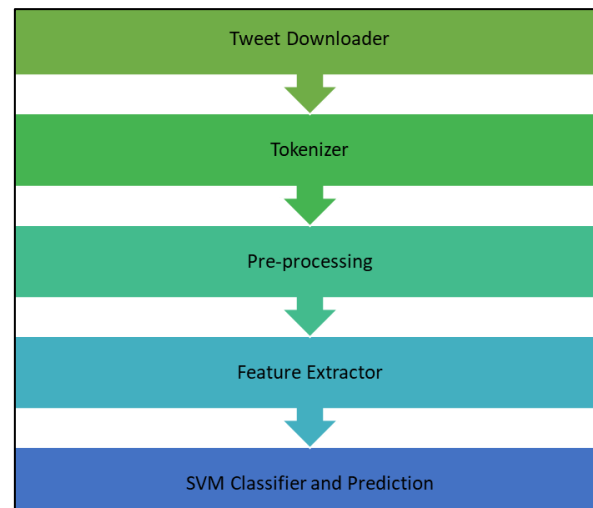
VI. WHY IS SENTIMENT ANALYSIS IMPORTANT?

- Microblogging has become popular communication tool
- Opinion of the mass is important
- Political party may want to know whether people support their program or not.
- Before investing into a company, one can leverage the sentiment of the people for the company to find out where it stands.
- A company might want find out the reviews of its products.

VII. USING TWITTER FOR SENTIMENT ANALYSIS

- Popular microblogging site
- Short Text Messages of 140 characters
- 240+ million active users
- 500 million tweets are generated everyday
- Twitter audience varies from common man to celebrities
- Users often discuss current affairs and share personal views on various subjects
- Tweets are small in length and hence unambiguous

VIII. FLOWCHART



IX. METHODOLOGY

A. Tweet Downloader

- Download the tweets using Twitter API

B. Tokenisation

- Twitter specific POS Tagger developed by Social Media Search

C. Pre-processing

- Removing non-English Tweets
- Replacing Emoticons by their polarity
- Remove URL, Target Mentions, Hashtags, Numbers.
- Replace Negative Mentions
- Replace Sequence of Repeated Characters eg. 'cooooooooool' by 'cool'
- Remove Nouns and Prepositions

D. Feature Extractor

- Number of Positive/Negative Words
- Number of Positive/Negative Hashtags
- Number of Positive/Negative/Extremely Positive/Extremely Negative Emoticons
- Number of Negation
- Number of special characters: ?,!,*

E. Classifier and Prediction

- The features extracted are next passed on to SVM classifier.
- The model built is used to predict the sentiment of the new tweets

X. SCREENSHOT OF THE SYSTEM

A. Tweet Downloader/ Extract Module

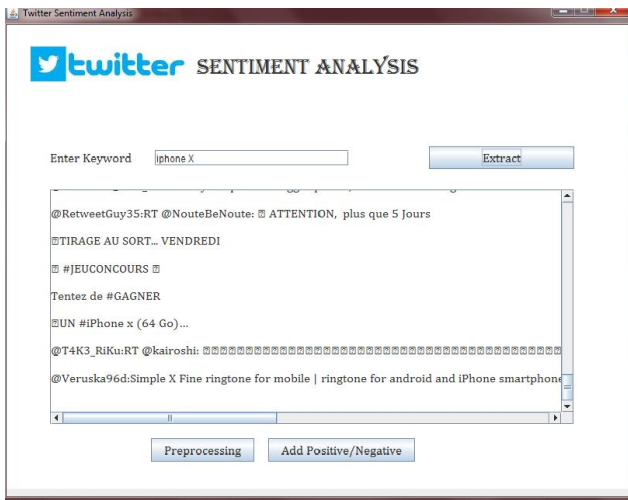


Fig. 1: Data Extraction

B. Add Positive/Negative Word Offline

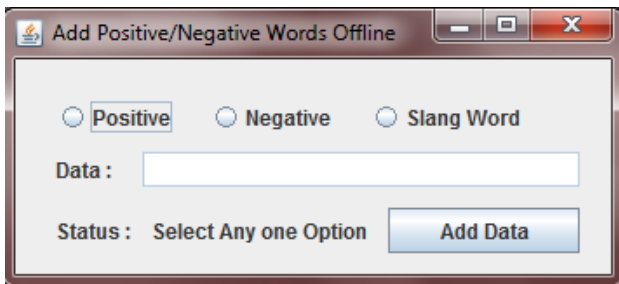


Fig. 2: Maintained Data Dictionary

C. Tweet Processing Module

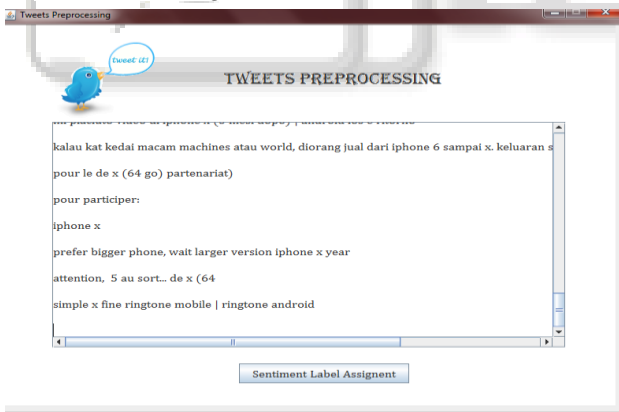


Fig. 3: Data Processing

D. Sentiment Label Assignment Module

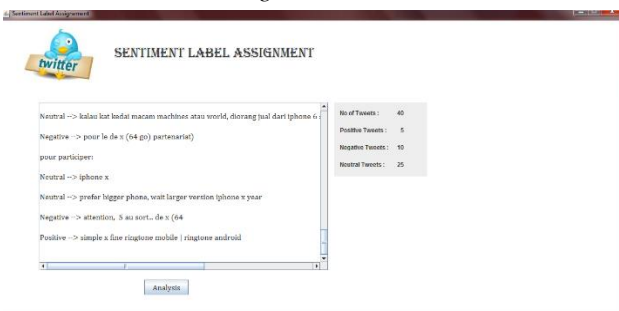


Fig. 4: Sentiment Label Assignment

E. Sentiment Analysis Module

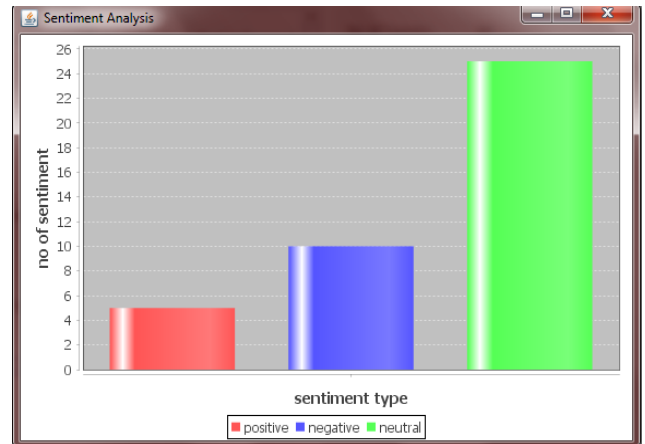


Fig. 5: Sentiment Analysis Result

XI. SOFTWARE & HARDWARE REQUIREMENTS

A. Software

- Operating System : Windows
- Technology : Java and J2EE
- Web Technologies : Html, JavaScript, CSS
- IDE : Net Beans IDE8.0.2
- Java Version : J2SDK1.5

B. Hardware

- Processor : Pentium
- Speed : 1GHz
- RAM : 1GB
- Hard Disk : 20GB
- Floppy Drive : 1.44 MB
- Key Board : Standard Windows Keyboard
- Mouse : Two or Three Button Mouse
- Monitor : SVGA

XII. APPLICATION

- Twitter
- Facebook
- Instagram
- Also mostly used in following domain: Politics, brands, products, celebrities etc.

XIII. FUTURE SCOPE

- It can be used for big data.
- A social analysis and marketing analysis platform.
- It should be possible to use our approach to classify sentiment in other languages with a language specific positive/negative keyword list.

XIV. CONCLUSION

- In this approach to perform aspect level sentiment classification for Twitter.
- Thus we have collect tweets using Twitter API, applied appropriate preprocessing on the tweets. As data receive from twitter is very dirty, it is difficult to perform aspect-level sentiment classification.

- Hence our classifier make use of POS tagger, dictionaries, aspect extraction and supervised machine learning algorithm.
- By the end of the project, we would understand the general sentiment around the iPhone X, which aspect of the iPhone X people like or dislike and again gain insights on how opinions on iPhone X change over a period of time.

REFERENCES

- [1] Kiruthica M.,Sanjana Woonaa,Priyanka Giri. Mumbai University, Vashi, Maharashtra, India "Sentiment Analysis of Twitter Data" from IJJET journal.
- [2] Pranav Waykar, Kailash Wadhvani,Pooja More. Pimpri, Pune "Sentiment analysis in twitter using Natural Language Processing (NLP) and classification algorithm" from IJARCET journal.
- [3] Bholane Savita Dattu, Prof.Deipali V. Gore. Pune, India "A Survey on Sentiment Analysis on Twitter Data Using Different Techniques" from IJCSIT journal.
- [4] T. Sakaki, M. Okazaki, and Y. Matsuo, Earthquake shakes twitter users: Real-time event detection by social sensors, in Proc. 19thInt. Conf. WWW, Raleigh, NC, USA, 2010.
- [5] Alexandre Trilla, Francesc Alias, Sentence-Based Sentiment Analysis for Expressive Text-to-Speech, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 21, NO. 2, FEBRUARY 2013.
- [6] D. Chakrabarti and K. Punera, Event summarization using tweets, in Proc. 5th Int. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.
- [7] Y. Hu, A. John, F. Wang, and D. D. Seligmann, Et-lda: Joint topic modeling for aligning events and their twitter feedback, in Proc.26th AAAI Conf. Artif. Intell. Vancouver, BC, Canada, 2012.
- [8] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Interpreting the Public Sentiment Variations on Twitter, IEEE Transactions on Knowledge and Data Engineering, VOL. 26, NO.5, MAY 2014.