



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 5, Issue 2)

Available online at: www.ijariit.com

Twitter sentimental analysis

Mandar Menjoge

mandarmenjoge@gmail.com

MIT Polytechnic, Pune, Maharashtra

Mazhar Sayyad

msayyad1155@gmail.com

MIT Polytechnic, Pune, Maharashtra

Vedant Bhawalkar

vedantbhl@gmail.com

MIT Polytechnic, Pune, Maharashtra

Jainam Gosaliya

jainamg48@gmail.com

MIT Polytechnic, Pune, Maharashtra

ABSTRACT

This paper presents the effectiveness of linguistic features to identify the sentiment of Twitter messages using the apache storm framework. We calculate the effectiveness of present lexical resources and features that capture information about the informal and creative language used in microblogging. In the past few years, there has been a huge growth in the use of microblogging platforms such as Twitter. Influenced by intensification, companies, and media organizations are increasingly seeking ways to excavate Twitter information about what people think and feel about their products and services. Here we download Twitter messages for a particular hashtag and carry out sentiment analysis i.e. to find a positive, negative or neutral sense of that tweet using apache storm framework. Each hashtag may have 1000 of comments and new comments are added every minute, in order to handle so many live tweets we are using apache storm framework.

Keywords— Sentiment analysis, Apache storm framework, Preprocessing, NLP, K-means clustering, Porter stemming algorithm

1. INTRODUCTION

Online social networks (OSNs), such as Twitter, Facebook, and some enterprise social network, have become extremely popular in the last few years. A person expends huge time in OSNs for making friends with people who they are familiar with or interested in. Twitter, which was founded in 2006, has become one of the most popular microblogging service sites. Currently, 200 million Twitter users generate over 400 million new tweets per day. Twitter is one such popular network where the short message communication (called tweets) has enticed a large number of users.

However, tweets are retrieved in a streaming manner, and Twitter provides the Streaming API for developers and researchers to access public tweets in real time.

With the rapid growth of social media platforms such as microblogging services, social networking sites, and short messaging services, people increasingly share their views and

opinions online. As such, sentiment analysis has attracted much attention since opinions or sentiments detected from the text are potentially useful for downstream applications including recommender systems, social network analysis, market forecasting and the prediction of political topics [1].

2. LITERATURE SURVEY

Ali Hasan et al.[1] introduces growth in the area of opinion mining and sentiment analysis has been rapid and aims to explore the opinions or text present on different platforms of social media by using machine-learning techniques with the sentiment, subjectivity analysis or polarity calculations. Even with the use of a variety of machine-learning techniques and tools for sentiment analysis during elections, there is a direct need for a state-of-the-art approach. A comparison of techniques of sentiment analysis of political views by applying supervised machine-learning algorithms such as Naïve Bayes and support vector machines (SVM) is given.

Mathura Tasnim et al.[2] gives depression detection by analyzing the Psychological Review. The paper gives a correlation between depression levels and behavioral trends of individuals. The papers propose a survey outcome under the psychological viewpoint and implementing the results, system design has been proposed that identifies who is in need of psychological support around us and take appropriate action.

Chae Won Park et al.[3] give an analysis of the user's opinions about three artificial intelligence assistants. For that Twitter, data were collected and classified into positive, negative, neutral opinions by a lexicon named Valence Aware Dictionary and Entiment Reasoner (VADER). The tweets are analyzed through independent samples T-test, Kruskal-Wallis test, and Mann-Whitney test to show the statistical significance among groups. The results suggested the highest rank of three artificial intelligence assistants by using statistical analysis.

M. Trupthi et al.[4] provide an interactive automatic system which predicts the sentiment of the review/tweets of the people posted in social media using Hadoop. The exact technique is

used for predicting sentiment polarity, which helps to get better marketing strategies. This paper deals with the challenges that come out in the Sentiment Analysis process, where real-time tweets are considered as they are rich sources of data for opinion mining and sentiment analysis. This paper focuses on Sentiment analysis, Feature-based Sentiment classification and Opinion Summarization.

Huma Parveen et al. [5] discuss the extraction of sentiment from a famous micro-blogging website, Twitter where the user posts their views and opinion. The sentiment analysis is performed on tweets which help to give some prediction on business intelligence by using Hadoop Framework for processing movie data set that is available on the Twitter website in the form of reviews, feedback, and comments. Results of sentiment analysis on Twitter data will be classified as positive, negative and neutral sentiments.

Lakshmi Kaushik et al. [6] proposed a new method for recognizing sentiment in audio using keyword spotting (KWS) for sentiment detection with the use of audio obtained from videos in youtube.com and UT-Opinion corpus. The system uses the iterative methodology to automatically extract sentiment behavior keywords from the text. The Maximum Entropy-based approach is used to enlarge the sentiment classifier.

3. PROPOSED SYSTEM

3.1 System Architecture

Comments, reviews and opinion of the people on social sites play an important role in predicting the sentiment of a wide variety of people on a particular event of interest. So to discover the overall sentiment of population i.e. positive, negative or neutral based on comments given on social networking we are developing a system that can use Twitter data Perform Sentiment Analysis using apache storm framework. Figure 1 gives the structural design of the proposed system followed by the detailed working of the system.

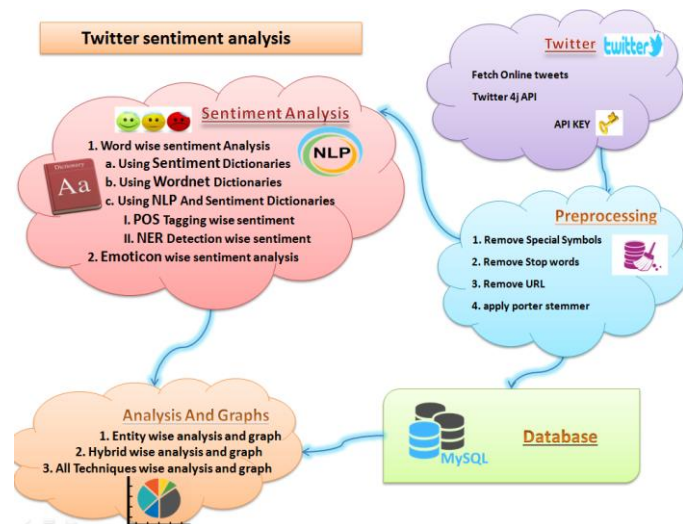


Fig. 1: System Architecture

3.1.1 Fetch Entity Wise Twitter tweets and Comments: The user collects Twitter tweets using the Twitter4j graph API. The entities are politicians, TV stars, places, actors etc. The collected data includes twitter tweets and hash tag comments on it.

3.1.2 Preprocessing: The user can apply to preprocess on the collected data. Preprocessing includes

(a) Tokenization

It splits sentences into words.

(b) Normalization

- Removes stop words.
- Removes #@.
- Removes any URL links

(c) Part-of-speech (POS) tagging

Identify if the word token is a noun, verb, adjective

3.1.3 Sentiment Analysis Using Different techniques

Sentiment analysis performed using the following techniques,

(a) Word Wise Sentiment Analysis

- **Using SentiWord Dictionaries:** SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to every synset of Word Net three sentiment scores: positivity, negativity, objectivity. SentiWordNet means A general thought, feeling, or sense. Feelings, especially tender feelings, as apart from reason or judgment. Gentle or tender feelings, sometimes of a weak or foolish kind.
- **Using Word Net Dictionaries:** Word Net is a combination of dictionary and thesaurus. It groups English words into sets of synonyms called synsets, provides short definitions and usage examples, and records a number of relations among these synonym sets or their members.
- **Using NLP and Sentiment Dictionaries:** Natural language processing (NLP), a branch of artificial intelligence concerned with automated interpretation and generation of human language. It is the ontology -assisted method of programming in terms of natural language sentences.
 - **POS Tagging wise:** Part-Of-Speech (POS) detects if the word token is a noun, a verb, an adjective. The word is assigned in accordance with its syntactic functions. In English, the main parts of speech are noun, pronoun, adjective, determiner, verb, preposition, adverb, conjunction, and interjection.
 - **NER Detection Wise:** Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as person and company names, or gene and protein names.

(b) Emoticon Wise Sentiment Analysis: The Emoticon data set is created by collecting status and comments with positive ‘:)’ and negative ‘:(’ emoticons. In this approach, messages are classified based on positive and negative emotion.

4. ANALYSIS AND GRAPHS

The system gives the final outcome on the following points, Entity and techniques wise analysis and graph. Comparison of used techniques and algorithms

4.1 Algorithm used

4.1.1 K- means Clustering Algorithm: [5]K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. K-means clustering aims to partition observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

Given a set of observations (x_1, x_2, \dots, x_n) , where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets $(k \leq n)$

$S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS):

$$\arg \min_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

Where μ_i is the mean of points in S_i .

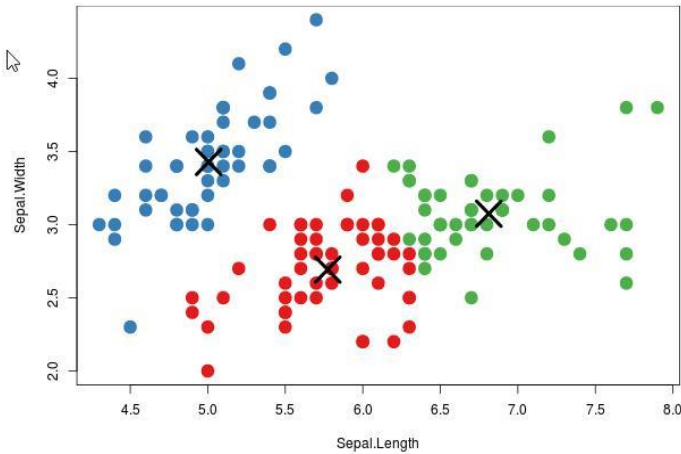


Fig. 2: K- means Clustering Algorithm

4.1.2 Porter Stemmer Algorithm: [10] Porter stemming algorithm (or 'Porter stemmer') is a process for removing the commoner morphological and inflexional endings from words in English. Following are the steps of this algorithm:

- Gets rid of plurals and -ed or -ing suffixes Turns terminal y to i when there is another vowel in the stem
- Maps double suffixes to single ones: -ization, -ational, etc.
- Deals with suffixes, -full, -ness etc.
- Takes off -ant, -ence, etc.
- Removes a final -e

5. IMPLEMENTATION DETAILS

This system can perform overall sentiment of population i.e. positive, negative or neutral based on comments given on social networking we are developing a system that can use Twitter data Perform Sentiment Analysis using apache storm framework. Below figure show the GUI of the proposed system.

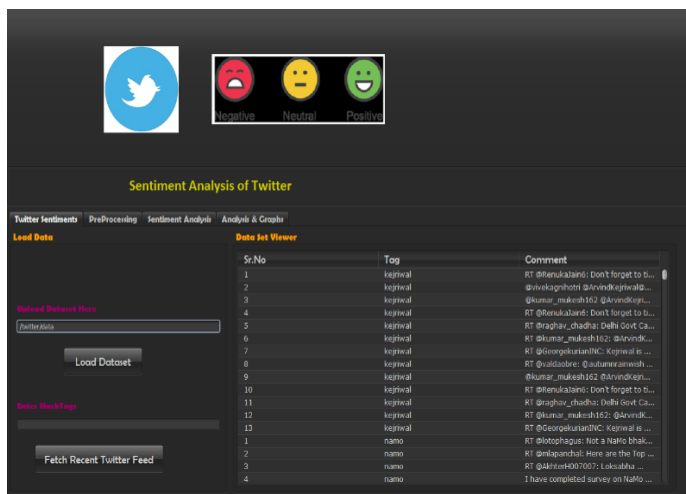


Fig. 3: System GUI

In this system, we can fetch the live Twitter messages for a particular hashtag and perform sentiment analysis i.e. to find a positive, negative or neutral sense of that tweet.

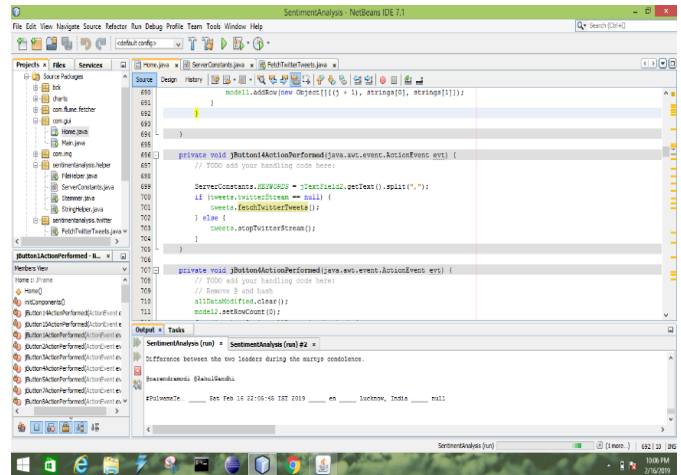


Fig. 4: Sentiment Analysis

6. CONCLUSION

Nowadays users have much interested in social sites as they prefer to share their opinion about particular things on social media instead of sharing with one another. The core technology Sentiment Analysis helps you to detect and analyze the sentiment of the comments based on the user response it receives. In addition, this type of analysis helps the businessmen and administrator to examine the issue discussed to enhance their service. This project helps to find the positive, negative or neutral sense of that tweet using an apache storm framework.

7. REFERENCES

- [1] Ali Hasan, Sana Moin, Ahmad Karim and Shahabuddin Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts" Math. Comput. Appl. 2018, 23, 11; doi:10.3390/mca23010011 www.mdpi.com/journal/mca.
- [2] Mashrura Tasnim, Nowshin Nahar, Hossain Mahmud "Intelligent Depression Detection and Support System: Statistical Analysis, Psychological Review and Design Implication", 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom).
- [3] Chae Won Park; Dae Ryong Seo , "Sentiment analysis of Twitter corpus related to artificial intelligence assistants", v 2018 5th International Conference on Industrial Engineering and Applications (ICIEA).
- [4] M. Trupthi; Suresh Pabboju; G. Narasimha "Sentiment Analysis on Twitter Using Streaming API", 2017 IEEE 7th International Advance Computing Conference (IACC).
- [5] Shruti Kapil; Meenu Chawla; Mohd Dilshad Ansari 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC).
- [6] Huma Parveen; Shikha Pandey , "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm", 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT).
- [7] M. Rambocas, J. Gama, "MarketingResearch: The Role of Sentiment. Analysis", the 5th SNA-KDD Workshop '11, 2013.
- [8] A. K. Jose, N. Bhatia, S. Krishna, "Twitter Sentiment Analysis", National Institute of Technology Calicut, 2010
- [9] P. Lai, "Extracting Strong Sentiment Trend from Twitter" in, Stanford University, 2012.
- [10] Biju Issac ; Wendy J. Jap TENCON 2009 - 2009 IEEE Region 10 Conference.