

Harnessing Multiple Data from Twitter and Performing Sentiment Analysis on Tweets

Hajeera Nikhat Khanam¹, Farhana Kausar²

¹Student, Department of Computer Science and Engineering, Atria Institute of Technology, Bangalore, India

²Assistant Professor, Dept. of Computer Science and Engg., Atria Institute of Technology, Bangalore, India

Abstract: Social media websites have emerged as one of the platforms for raising the opinions of users and influencing the marketing of any business. People's opinion is very important in analyzing how information propagation affects life in a large-scale network like Twitter. Twitter sentiment analysis determines the polarity of the vast population towards a particular subject, item or entity. During public elections, film promotions, endorsements of mobile phones and many other things, the applications of such sentiment analysis can be easily observed. In our project, we exploited the fast and efficient framework 'Tweepy' to extract the tweets and we are using framework 'TextBlob' to perform sentiment analysis in a less complex and efficient way. Using these frameworks, it helps us to extract relevant tweets and to extract the required amount of data from the twitter.

Keywords: Multisource Data Processing, Social Network Data Analysis, Sentimental Analysis.

1. Introduction

A. Overview

As internet is growing bigger, its horizons are becoming wider. Social Media platforms like Facebook, Twitter, Instagram dominate in spreading encapsulated news and current trending topics across the globe at a rapid pace. A topic becomes more popular if multiple users are contributing their opinion and feedbacks, thereby making it a valuable source of online perception. From these topics generally we can spread awareness or promote public figures, political campaigns during elections, product endorsements and entertainment like movies, award shows or launch of new smart phones. Many organizations take advantage of people's opinions by their feedback to improve their products and services which further help in enhancing marketing strategies.

B. Sentiment analysis

It is the prediction of different kind of emotions in a word, sentence or corpus of documents. It is intended to serve as an application to understand the attitudes, opinions and emotions expressed within an online mention. The main intention is to gain an overview of how people are reacting on certain topics. Precisely, it is a paradigm of categorizing conversations into positive, negative or neutral sentiments. Nowadays people use social media sites for networking with other people and to be updated with all information. These sites (Twitter, Facebook,

Instagram, google+) offer a platform to people to voice their opinions. These information forms a basis for people to evaluate, rate about the performance of not only any movie but about other products and to know about whether it will be a success or not. This type of vast information on these sites can be used for marketing and social studies. Therefore, Sentiment Analysis has many applications like emotion mining, polarity, classification analysis etc.

C. Twitter

Twitter is an online networking site driven by tweets which are 140-character limited messages. We have limitations on the characters that we post and using these characters we can express our sentiments. So generally most of the tweets that we send carries some value or they can be neutral, but in most of the cases we find people actually picturing out the sentiments so that is where the whole world got the idea that why not use the data to make business sense. Tweet is the data which can be executed from twitter. Twitter sentiment analysis involves the use of natural language processing to extract and identify the sentiments.

2. Connecting to twitter

A. Introduction

- In order to connect to Twitter, first we need to create an account in Twitter by entering valid name and phone number. The steps below outline how to create an application.
- Open the Twitters Developers site and sign in
- Go to my applications and create a new application
- Go to my applications and create a new application
- Fill in the details required
- When the application is created we will be back to the app page
- Then we need to click on the 'Keys & tokens' tab
- Generate the 'Access Token', 'Access Token Secret', 'Consumer Key' and 'Consumer Secret'

B. Tweepy

It provides access to Twitter Application Programming Interface. Tweepy API is an API that we can access in Python

package which will allow us to very easily stream tweets in real time directly from Twitter. We need to install tweepy. Tweepy provides the Cursor Interface to iterate through different type of objects. Maximum of 3200 tweets are allowed by Twitter for extraction.

C. OAuth (Open Authentication)

It is an open standard for authentication. In order to secure the information, we use this. Nowadays hackers are stealing our passwords, for this reason we should use OAUTH which will provide safer authentication approaches. Twitter verifies the identity of the user and issues a pin called an OAuth verifier.

D. Collecting end user’s information and tweets

End Users will make profiles on Twitter to expose themselves to others. A user’s profile consists of all the information needed to profile users according to their tweets and information. Using the OAuth, we can access the public API of the Twitter, which in turn gives us a stream of tweets that are just a sample of total tweets posted at a time. The important information we are storing from a user’s profile with the help of his tweets are the user’s screen name, user’s twitter Id etc.

In Twitter, the status messages are known as tweets. A tweet can have a maximum length of 140 characters. Using Twitter API, tweets can be published. Retweet is a special type of tweet that is generated when one user posts back again on someone’s tweets.

E. TextBlob

It is a python library that provides a simple API for accessing its methods and carrying out basic NLP tasks such as tokenization, parts of speech tagging, extraction of noun phrases, sentiment analysis, classification, translation etc.

This gateway basically performs authentication from users using a third-party application. The users must authenticate themselves. The developer of the project must have an account in the developer platform of twitter. The developer uses set of keys in order to perform challenges & response type of key exchange in order to authenticate themselves.

C. Twitter authorization gateway

This module represents the authorization gateway of twitter. This gateway basically performs authorization of users using a third-party application. The system checks whether the user has privileges to fetch the data and perform operation on them. The developer of the project must have an account in the developer platform of twitter. The developer uses set of access token and access token secret key in order to perform the challenge and response type of key exchange in order to perform authorization. In this process the rights of the developer is usually checked by the twitter back end in order to allow the users to fetch the tweets and analyze them.

D. Twitter database

This module represents a secure storage of tweets or user’s data. The database generally is located at remote geographical location. These databases consist of tweets according to the hashtags used. Twitter database consists a huge amount of data which are getting appended every second. The project helps in retrieving those data and analyzing them.

E. Sentiment analysis module

This module performs the entire analysis process. Using TextBlob, we process those data. During this process, a value is calculated based on different type of tagging methodology (POS Tagging, etc.). That value is called polarity. The developer then sets certain parameters in order to use that polarity value to analyze whether the specific sentence of the corpus is a positive, negative or neutral statement. This entire process is called sentiment analysis. This can also be achieved using natural language toolkit (NLTK). TextBlob is an enhanced tool of NLTK which can be used to simplify the complexity in the algorithm.

3. System architecture

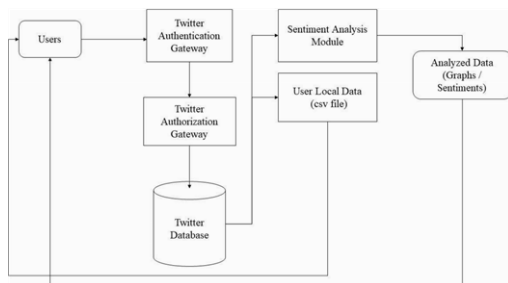


Fig. 1. Design

A. Users

This module represents the activities that are performed by the user. Here, the user performs operations like giving input to the next module and the system must contain the required packages and a python compiler. This module represents the client end of the entire system architecture.

B. Twitter authentication gateway

This module represents the authentication gateway of twitter.

4. Conclusion

We need to analyze the sentiments of tweets and visualize the result in the form of graph. We used TextBlob which is an open source text processing library to calculate the polarity of the tweets. Polarity was classified into positive, negative and neutral sentiments. Demo is designed and implemented.

References

[1] L. You, B. Tunçer and H. Xing, "Harnessing Multi-Source Data about Public Sentiments and Activities for Informed Design," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 343-356, 1 Feb. 2019.
 [2] L. You and B. Tunçer, "Informed Design Platform: Interpreting “Big Data” to Adaptive Place Designs," *2016 IEEE 16th International*

Conference on Data Mining Workshops (ICDMW), Barcelona, 2016, pp. 1332-1335.

- [3] M. Vogler, J. M. Schleicher, C. Inzinger, S. Dustdar, and R. Ranjan, "Migrating smart city applications to the cloud," *IEEE Cloud Computing*, vol. 3, no. 2, pp. 72–79, May 2016.
- [4] L. You and B. Tunçer, "Exploring public sentiments for livable places based on a crowd-calibrated sentiment analysis mechanism," *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, CA, 2016, pp. 693-700.
- [5] L. You, G. Motta, K. Liu, and T. Ma, "City feed: A pilot system of citizen-sourcing for city issue management," *ACM Transactions on Intelligent Systems and Technology*, vol. 7, no 4, pp. 53, July 2016.