

SYNOPSIS

Sentiment Analysis of Twitter Data Using Python

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE AWARD OF

Degree of Bachelor of Technology in Computer Science & Engineering



SUBMITTED BY

Name: Anmol Singh Rajput

UID:- 17BCS2898

Name:- Arjun Shrivastva

UID:- 17BCS2910

Name:- Satyam Priyadarshi

UID:- 17BCS2903

SUBMITTED TO:

RAJANI MAM

**Department of Computer Science &
Engineering CHANDIGARH UNIVERSITY**
Gharuan, Mohali

1. Introduction

Social networks are the main resources to gather information about people's opinion and sentiments towards different topics as they spend hours daily on social medias and share their opinion. In this project, we show the application of sentimental analysis and how to connect to Twitter and run sentimental analysis queries. We run experiments on different queries from politics to humanity and show the interesting results. We realized that the neutral sentiment for tweets are significantly high which clearly shows the limitations of the current works. It will be helpful to political party for reviewing about the program that they are going to do or the program that they have performed. Similarly, companies also can get review about their new product on newly released hardware's or software's. Also, the movie maker can take review on the currently running movie. By analysing the tweets analyser can get result on how positive or negative or neutral are peoples about it.

2. Feasibility: -

A feasibility study is a preliminary study which investigates the information of prospective users and determines the resources requirements, costs, benefits and feasibility of proposed system. A feasibility study takes into account various constraints within which the system should be implemented and operated. In this stage, the resource needed for the implementation such as computing equipment manpower and costs are estimated. The estimated are compared with available resources and a cost benefit analysis of the system is made. The feasibility analysis activity involves the analysis of the problem and collection of all relevant information relating to the project. The main objectives of the feasibility study are to determine whether the project would be feasible in terms of economic feasibility, technical feasibility and operational feasibility and schedule feasible or not. It is to make sure that the input data which are required for the project are available.

3. Methodology

- The first phase will include greeting the user and getting sample commands in the form of twitter from the user.
- In the second phase, we need to test the inputs against the hardcoded commands. We'll be testing the command against each if-else statement, which will instruct it to do a certain task.
- At the third phase, semantic will get its checked through test software that we will develop for the sake of simplicity and efficiency.
- Now it's time to store the data (tasks). They will be stored in data files, so that Twitter can access them anytime.
- Next phase is to set up alarms/reminders. This is a bit tricky, so for starters we'll be using alert signals after execution of a task to set as the reminder for the next task.
- After all the functions and modules are coded and thoroughly tested, Twitter will be available in a UI environment.
- The program will go through beta testing in the last phase before launching.

- Lastly, the project will be launched. Hopefully it proves useful for the work and time taken to complete it.

4. Module and work distribution

4.1Data: -

To gather the data many options are possible. In some previous paper researches, they built a program to collect automatically a corpus of tweets based on two classes, “positive” and “negative”, by querying Twitter with two type of emoticons:

- Happy emoticons, such as “:.)”, “:P”, “:-)” etc.
- Sad emoticons, such as “:(“, “:’(“, “=(“.

Others make their own dataset of tweets by collecting and annotating them manually which is very long and fastidious. Additionally, to find a way of getting a corpus of tweets, we need to take care of having a balanced data set, meaning we should have an equal number of positive and negative tweets, but it needs also to be large enough. Indeed, more the data we have, more we can train our classifier and more the accuracy will be. After many researches, I found a dataset of 1578612 tweets

	ItemID	Sentiment	SentimentSource	SentimentText
0	1	0	Sentiment140	is so sad for my APL friend.....
1	2	0	Sentiment140	I missed the New Moon trailer...
2	3	1	Sentiment140	omg its already 7:30 :O
3	4	0	Sentiment140	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)..
4	5	0	Sentiment140	i think mi bf is cheating on me!!! T_T
5	6	0	Sentiment140	or i just worry too much?
6	7	1	Sentiment140	Juuuuuuuuuuuuuuuuuusssst Chillin!!
7	8	0	Sentiment140	Sunny Again Work Tomorrow :- TV Tonight
8	9	1	Sentiment140	handed in my uniform today . i miss you already
9	10	1	Sentiment140	hmmmm.... i wonder how she my number @-)

4.2 Pre-processing

Now that we have the corpus of tweets and all the resources that could be useful, we can pre-process the tweets. It is a very important since all the modifications that we are going to during this process will directly impact the classifier's performance. The pre-processing includes cleaning, normalization, transformation, feature extraction and selection, etc. The result of pre-processing will be consistent and uniform data that are workable to maximize the classifier's performance. All of the tweets are pre-processed by passing through the following steps in the same order.

	ItemID	Sentiment	SentimentSource	SentimentText
0	1	0	Sentiment140	is so sad for my APL friend.....
1	2	0	Sentiment140	I missed the New Moon trailer...
2	3	1	Sentiment140	omg its already 7:30 :O
3	4	0	Sentiment140	.. Omgaga. Im sooo im gunna CRy. I've been at this dentist since 11.. I was suposed 2 just get a crown put on (30mins)..
4	5	0	Sentiment140	i think mi bf is cheating on me!!! T_T
5	6	0	Sentiment140	or i just worry too much?
6	7	1	Sentiment140	Juuuuuuuuuuuuuuuusssst Chillin!!
7	8	0	Sentiment140	Sunny Again Work Tomorrow :- TV Tonight
8	9	1	Sentiment140	handed in my uniform today . i miss you already
9	10	1	Sentiment140	hmmmm.... i wonder how she my number @-)

Table 2.3.1.1: Before processing emoticons, list of tweets where some of them contain emoticons.

5.Work Distribution:

- Anmol Singh Rajput: -Twitter data analysis and software development for semantic analysis of tweets.
- Arjun: -To apply effective algorithms
- Satyam: -To analysis and test output produce by develop software

6.Software and Hardware Requirements

- Software: Python3, mpg123 and any Debian based Linux Distro (Although Ubuntu is recommended).
- Hardware: PC or laptop with a fast Internet connection.

7.Conclusion: -

Nowadays, sentiment analysis or opinion mining is a hot topic in machine learning. We are still far to detect the sentiments of s corpus of texts very accurately because of the complexity in the English language and even more if we consider other languages such as Chinese. In this project we tried to show the basic way of classifying tweets into positive or negative category using Naive Bayes as baseline and how language models are related to the Naive Bayes and can produce better results. We could further improve our classifier by trying to extract more features from the tweets, trying different kinds of features, tuning the parameters of the naïve Bayes classifier, or trying another classifier all together.

Bibliography

- <https://gtts.readthedocs.io/en/latest/>
- <https://realpython.com/python-speech-recognition/>
- <https://docs.python.org/3/>
- <https://stackoverflow.com/>