

Polarity Detection using Twitter Sentiment Analysis

Namrata Jaiswal¹, Prajakta Borkar², Prajakta Pirdurkar³, Shivani Rahate⁴, S. V. Solanki⁵
^{1,2,3,4}Student, Department of Computer Technology, KDKCE, Nagpur, India
⁵Professor, Department of Information Technology, KDKCE, Nagpur, India

Abstract: Twitter is a social networking platform where people share different part and thoughts, where sentiment analysis is the process of understanding an opinion about particular post and getting its meaning. Covering the topic sentiment analysis is the application to show the feedback or the opinion or the post of the users. This application shows the positive, negative and neutral graph over the post. This paper focus on one of the research in the sentiment analysis with the help of python.

Keywords: Sentiment analysis, Natural Language Processing, Python.

1. Introduction

Nowadays social networking sites have gained so much popularity, which is not ignorable. Because social networks are serving people mood of sharing various kinds of information's, blogs, pictures and also millions of people post their views, comments and expressions on different topics in effective manner for number of times.

Such Opinion on various topics like discussing current issues, movies, products that are used I daily life make sense of general sentiment for company's study users reactions and reviews on their product for improvement in business. Different types of words which are classified to their semantic orientation used in dictionaries is one of the unsupervised semantics. In this paper, we use twitter for sentimental analysis since in social networks, mainly twitter consists of limited tweet size, use of slang, etc. This makes performing sentiment analysis challenging.

A. Sentiment Analysis

Sentiment Analysis is defined as the type of data mining which measures the inclination of people's opinions from natural language processing. Sentiment Analysis is classified into two ways: supervised and unsupervised which is based on semantics.

Supervised techniques are used in support vector machines and Naive Bayes classifier. Sentiment Analysis is generally applied for the voice of customer materials namely views and survey response, online and social media. Peoples can share their opinions through the tweets. Sentiment Analysis presents a characteristic which includes machine learning and lexicon based approaches. The Supervised method which is a part of sentiment analysis consisted such as decisions trees, support

vector machines, neutral networks also the methods which is based on probable, such as Naive Bayes, Bayesian networks and maximum entropy. Sentiment Analysis evaluates the expression which is positive, negative or neutral. It also evaluates the written/spoken language. Feelings and feedback of customer is very important to uncover the method of sentiment analysis which is used as a brands or any other tweets.

B. Twitter Sentiment Analysis

Sentiment Analysis is the process termed in the data mining where analysis is done on the people's comments, views or opinion about any particular topic that may be any business product, policies, movie review etc.

The best platform to implement this kind of analysis is the social media where people are show active all day, commenting on the topics of their interest. Thus, twitter is one of the biggest platform where we can carry out this analysis.

There are different people on twitter who tweets on different small topic in different ways using different emotions, abbreviations and some language slang which makes it sometimes difficult to analyses the sentiment. Natural Language Processing systems were unable to analyses this sentiment other researches using deep learning and other machine learning techniques to extract and mine the polarity.

2. Twitter Sentiment Analysis using Python

A. Python

Python is high level programming language developed by Guido Van Rossum is 1991. Python is designed such that it ensures code readability and also uses white space. Python has large standard library which is used for many tasks, applications.

B. Natural Language Toolkit(NLTK)

Natural Language Toolkit is a library and programs for symbolic and statistical Natural Language Processing. NLTK includes graphical representation and sample data. NLTK supports classification, Tokenization, tagging, and semantic reasoning functionalities. NLTK is a platform used for building python programs that work with human language data for applying in statistical Natural Language Processing. The NLTK library also embodies various classifier such as Naive Bayes

Classifier.

C. Natural Language Processing

Natural Language Processing is a method to translate between computer and human languages. It is a method of getting a computer to understandably read a line of text without the computer begin fed some sort of due or calculation. Natural Language Processing is about developing applications and services that are able to understand human languages. NLP are speech recognition such as understanding what the content is about or sentiment analysis etc.

D. Data Collection

There are two ways to collect data for Twitter Sentiment Analysis. First is to use Tweepy–Tweepy is the python client for the official Twitter Application Programming Interface (API). Tweepy can be installed using pip command: pip install Tweepy. To get tweets from the Twitter API one needs to create an account on Twitter. After that the following steps are performed:

- Open <https://apps.twitter.com/> and click button – ‘Create New App’.
- Fill the application details. You can leave the callback url field empty.
- Once the app is created, you will be redirected to the app page.
- Open the ‘Keys and Access Tokens’ tab.
- Copy ‘Consumer Key’, ‘Consumer Secret’, ‘Access Token’ and ‘Access Token Secret’.

The keys copy is then inserted into the code, which helps in dynamic collection of tweets every time we run the code. The other way is to collect data non dynamically using the existing data provided by websites and save the data into whatever format we require. The former method is quite slow in nature as it performs tweet collection every time we start the program. The later approach may not provide us with the quality of tweets we require.

3. Design and Implementation

This paper has implementation of Twitter sentiment analysis, by using the APIs provided by Twitter itself. There are many tools focusing on text mining and social networks. In this project many libraries are used. The approach to extract sentiment for tweets is as follows:

1. Start with downloading and storing the sentiment dictionary.
2. Download twitter testing data sets, input it in to the program.
3. Clean the tweets by removing the stop words.
4. Tokenize each word in the dataset and feed into the program.
5. For each word, compare it with positive sentiment and negative sentiment word in the dictionary. Then increment positive count or negative count.

6. Finally, based on the positive count and negative count, we get result in percentage about sentiment to decide the polarity.

A. Implementation

In this paper, we use python to implement twitter sentiment analysis. Some packages are used such as Tweepy and text blob. We can install these libraries by following commands:

- Pip install Tweepy
- Pip install text blob

The second step is to download the dictionary by using following commands: Python –m text blob. download-corpora. Corpora is nothing but a large and structured set of texts used for analyzing tweets. Text blob- text blob is the python library for processing textual data and it uses NLTK for Natural Language Processing.

B. Connect to Twitter using APIs

To connect to twitter and query latest tweets, we need to go to the apps.twitter.com/app/new and generate the api keys.

C. Sample Result

Following shows the sample output of the program for the ‘Cricket’ as a query based on the last 100 tweets from Twitter.

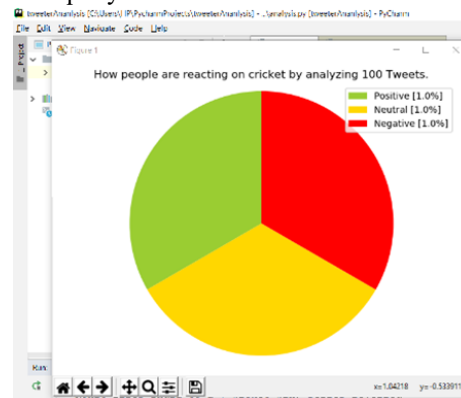


Fig. 1. How people are reacting on cricket by analyzing 100 tweets

Enter Keyword/hashtag to search about: cricket
 Enter how many tweets to analyze:100
 Hamilton #Waikato #New Zealand: 2019 Waikato Backyard Cricket Championships 16 Feb 2019 <https://t.co/xnEb2PbYFn>
<https://t.co/RUiu2c84ng>
 Rohit fires India to a 7-wicket win, series levelled
 Read More - <https://t.co/tmtvnFw17u>
 #INDvsNZ #RohitSharma... <https://t.co/QGX0K3Xquq>
 India crash to their worst T20I defeat after horror show against NZ
 Read More - <https://t.co/8RSojm5SMI>
 #NZvIND... <https://t.co/nyBmlsUtVf>
 How many People are reacting on cricket by analyzing 100 Tweets?

D. Applications

- **Business:** Every small or large business depend upon

the customers. What customer think, react whether they give good or bad reviews for the services provided by the particular business. sentiment analysis is one of the best mode to analyze the same. Businessman uses Twitter sentiment analysis for the growth of their businesses on the basis of the calculated polarity for their product or services, using this technique.

- *Politics*: There are large amount of tweets on Twitter related to politics. Politics is the topic wherein every person has his own opinion also they keep tweeting on such topics frequently. Those tweets of people are used in politics for taking some decisions they need to make their parties strong. People post their support or argument towards government policies, actions, elections, debates etc. Hence analyzing data from it can help is in determining public views.
- *Sports events*: Sports involve many events, championships, gatherings and some discussions related to sports too. Many people are interested in sports followers and follow their favorite players present on Twitter. These people frequently tweet about different sports related events. We can use the data to gather public view of a player's action, team's performance, official decisions etc.

4. Conclusion

Twitter is a platform where people are active all the day sharing their views which are then processed to locate some trends and patterns. Twitter sentiment analysis directly refers to opinion mining. It analysis's the sentiments from the tweets,

feed the data into machine learning model and checking its accuracy so that it can be used in future. Steps like data collection text preprocessing, sentiment detection, sentiment classification and testing are involved in twitter sentiment analysis. The accuracy determine is not always perfect but it is nearly to the point. There are many applications based on twitter sentiment analysis thus this project has a bright future scope.

References

- [1] Medhat, W., Hassan, A., Korashy, H. 2014. Sentiment analysis algorithms and applications: A survey," Ain Shams Eng.
- [2] M. Desai and M. Mehta, "Techniques for sentiment analysis of Twitter data: A comprehensive survey", 2016 International Conference on Computing, Communication and Automation (ICCCA), 2016.
- [3] Kristina Lerman, Luciano G. Marin, Megha Arora, Lucas H. Costa de Lima, Emilio Ferrara and David Garcia," Language, demographics, emotions, and the structure of online social networks", Journal of Computational Social Science, 1, (2018).
- [4] Chihli Hung," Word of mouth quality classification based on contextual sentiment lexicons", Information Processing & Management, 53, 4, (751), (2017).
- [5] Martin Haselmayer and Marcelo Jenny, "Sentiment analysis of political communication: combining a dictionary approach with crowd coding", Quality & Quantity, 51, 6, (2623), (2017).
- [6] David Zimbra, M. Ghiassi and Sean Lee, "Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", IEEE, 2016.
- [7] Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017).
- [8] Halima Banu S and S Chitrakala, "Trending Topic Analysis Using Novel Sub Topic Detection Model", (IEEE) ISBN-2016.
- [9] Boguslavsky, I. (2017). Semantic Descriptions for a Text Understanding System. In Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2017).
- [10] V. A. Kharde and S. S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques," Int. J. Compute. Appl., vol. 139, no. 11, 2016.