

Reviewing the Movies using Semantic Analysis Algorithm

Utkarsh Sharma¹, Shobhana Kashyap²

¹Under Graduate in Computer Science Department, ²Assistant Professor in
Computer Science Department

¹2014pgicsutkarsh@poornima.org, ²shobhana.kashyap@poornima.org

Abstract

The evolution of social web has made everyone capable to get globally connected with each-other and to easily create and share their ideas and opinion with people around the world. As the number of opinions and ideas from millions of computers, smart phones and cameras are increasing the amount of multimodal content on the web is also increasing for which the need of decoding such information into useful knowledge is also required. In this paper, sentiment analysis algorithm is used for extracting users' emotions and opinions from the text content.

Keywords: Social Media, YouTube, Artificial Intelligence, Machine Learning, Sentimental Approach

1. Introduction

This time we are at the age of big data. As the information and communication technologies are growing rapidly, people are also exhibiting totally different styles. Now a days, people from all over the world are enjoying the values that are being extracted from historical and real time data sets such as social interaction, weather forecast and so on. As per the survey report from International Business Machine(IBM), 2.5 quintillion bytes of data are being produced on the daily basis, among which 90 percent of data present today has been produced within the last two years. The huge datasets for which we searches through the search engines like Google, Yahoo and many others contains a big amount of information and knowledge which are for the advancement of different sectors of our society like medical, business, finance and so on. According to the leading researchers of the world, the main reason for the evolution of data mining and machine learning is the appearance of big data. The big data applications which are based on fancy mining and learning are depended on the efficient and effective support from the underneath layers even in the age of big data. Mainly there are two implicit conditions in the traditional mining algorithm, which are locality and homogenous. In the locality all the data are stored at a local place whereas in homogenous all the data that are to be used for mining must be homogenous. The process of sentimental analysis considers the datasets which contains the emotions, attitudes and so on based on human thinking. Sentiment analysis helps users to classify if the information about a particular object is up to the satisfaction level of customers or requires more information. This analysis is used by the companies to analyze their product and make them up to customers' satisfaction. In the process of semantic analysis two types of machine learning techniques are considered. These are supervised and unsupervised learning. Unsupervised learning conduct clustering as they do not provide the correct target. Since, supervised learning are based on labelled datasets, they provide correct output when considered in decision making. The given research paper gives a better understanding of sentiment analysis based on unsupervised machine learning.

2. Literature Survey

[1]Shi Yu concluded that we are at the doorstep of big data, and one of the most critical field for accommodating the need of big data application is the networking for big data. According to this paper, there are so many challenges and problems in the implementing big data applications. Also, huge number of problems will be faced during the enhancement of big data in the future both from the theoretical as well as practical view.

[2]SoujanyaProia proposed a novel multimodal affective data analysis framework, which includes the extraction of salient features, building features and decision-level fusion framework. Here, for the outer performance of the state of the art model's accuracy deep CNN-SVM based textual sentiment analysis component is given. In the case of fusion experiment a significant role is played by MKL. Also, one of the important contribution of the paper has been given to the novel decision level fusion architecture.

[3] GeetikaGautam proposed a set of techniques of machine learning with semantic analysis so as to classify the product reviews for the twitter dataset. For this purpose Naïve Byes technique is used here and the main aim of this is to analyze the large amount of datasets. Here the datasets used are labeled datasets. At last it is concluded that the Naïve Byes techniques gives the better visualization of the content which will be helpful to the users.

3. Dataset Description

The database used in the algorithm is real-time database. The real-time twitter events are passed as data to the algorithm. The tweets on a particular trending topic are stored are analyzed and are given as real-time database to the algorithm. The advantage of this type of database is that it fetches the data continuously so if any new data is added or we can say any new tweet is added it automatically analyze that tweet or data without making changes into database.

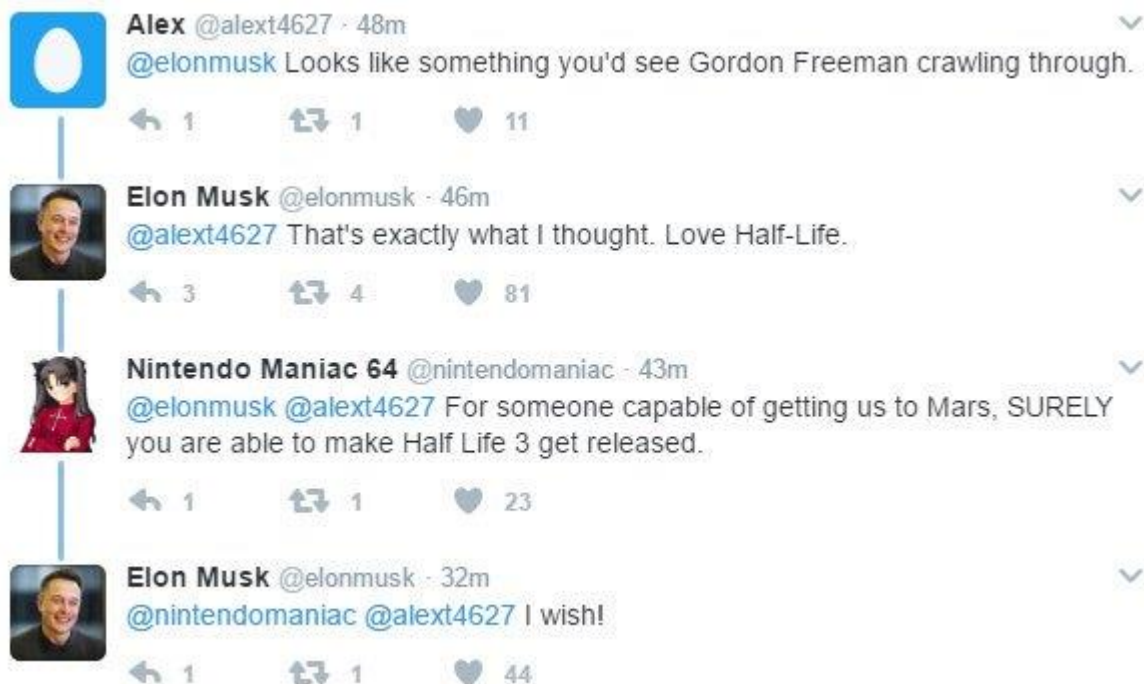


Figure-1 Live tweets database

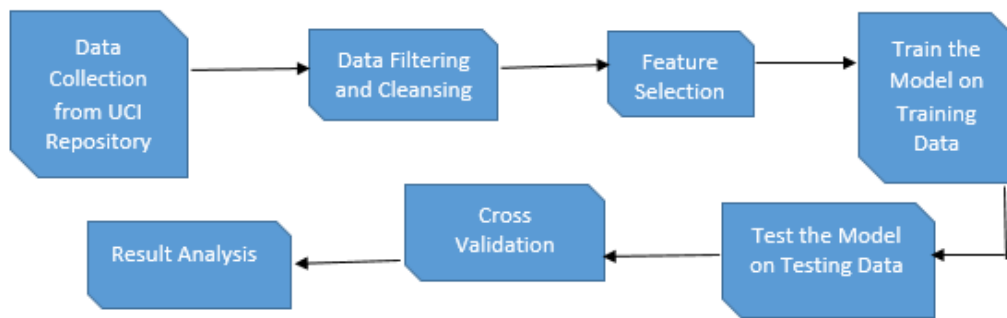


Figure 2: Data Flow

4. Methodology

1.1 R tool

R is a free and open source software for statistical computing and graphics available for Windows, MacOS and Windows platform. R provides a wide variety of statistical models like Linear and nonlinear modeling, Classical statistical tests, Time-series analysis and Classification/Clustering/Regression Models. It is a an effective data handling and storage facility, a suite of operators for calculations on arrays, in particular matrices, a large, coherent, integrated collection of intermediate tools for data analysis, graphical facilities for data analysis and display either on-screen or on hardcopy, and a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and I/O facilities. R provides a mixture of programming models. At its internal / ground, it is an imperative type of language where we can write a script which do one calculation at a time, but it also supports OOPs features where data and functions are encapsulated inside classes and also functional programming .

1.2 Algorithm

```

import time
from tweepy import Stream
from tweepy import OAuthHandler
from tweepy.streaming import StreamListener
import json
from textblob import TextBlob
import matplotlib.pyplot as plt
import re

"# -- coding: utf-8 --"
  
```

```
defcalctime(a):
    returntime.time()-a

positive=0
negative=0
compound=0

count=0
initime=time.time()
plt.ion()

ckey=""
csecret=""
atoken=""
asecret=""

class listener(StreamListener):

    defon_data(self,data):
        globalinitime
            t=int(calctime(initime))
        all_data=json.loads(data)
        tweet=all_data["text"].encode("utf-8")
            #username=all_data["user"]["screen_name"]
        tweet=" ".join(re.findall("[a-zA-Z]+", tweet))
        blob=TextBlob(tweet.strip())

        global positive
        global negative
        global compound
        global count

        count=count+1
        senti=0
        forsen in blob.sentences:
            senti=senti+sen.sentiment.polarity
        ifsen.sentiment.polarity>= 0:
```

```
positive=positive+sen.sentiment.polarity
else:
negative=negative+sen.sentiment.polarity
compound=compound+senti
print count
printtweet.strip()
printsenti
print t
printstr(positive) + ' ' + str(negative) + ' ' + str(compound)

plt.axis([ 0, 70, -20,20])
plt.xlabel('Time')
plt.ylabel('Sentiment')
plt.plot([t],[positive],'go',[t],[negative],'ro',[t],[compound],'bo')
plt.show()
plt.pause(0.0001)
    "if count==200:
return False
else:
return True"

defon_error(self,status):
print status

auth=OAuthHandler(ckey,csecret)
auth.set_access_token(accessToken,asecret)

twitterStream= Stream(auth, listener())
twitterStream.filter(track=["Elon Musk"])
```

5. Conclusion

In this paper we, we have proposed a set of technique of machine learning by using semantic analysis for dividing sentences and movie reviews based on twitter data. The main aim is to fetch and analyze the reviews by using twitter dataset. For this purpose we had developed an algorithm which is faster than previous algorithm developed for this purpose. This algorithm can fetch real time data by using twitter database. Thus the

algorithm can be used to rate the movies on the basis of live tweets which give feedback about the movie.

6. Acknowledgement

With immense pleasure I, Mr. Utkarsh Sharma presenting "Reviewing the movies using Semantic Analysis Algorithm" Seminar presentation as part of the curriculum of Bachelor Degree. I wish to thank all the people who gave me unending support. I express my profound thanks to seminar coordinator Mr. Bhanwar Veer Singh. I am thankful to my seminar guide Ms. ShobhanaKashyap for her kind support and providing me expertise of the domain to develop the research paper.

References

1.1 Journal Article

- [1.] Gautam, Geetika, and DivakarYadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014 seventh international conference on.IEEE, 2014.
- [2.] Cambria, E., Poria, S., Gelbukh, A., &Thelwall, M. (2017). Sentiment analysis is a big suitcase. IEEE Intelligent Systems, 32(6), 74-80.
- [3.] Al-Zaidy, R.A. and Giles, C.L., 2017, February. A Machine Learning Approach for Semantic Structuring of Scientific Charts in Scholarly Documents.In AAAI (pp. 4644-4649).
- [4.] R. Feldman, " Techniques and Applications for Sentiment Analysis," Communications of the ACM, Vol. 56 No. 4, pp. 82-89, 2013.
- [5.] Y.Luo,W.Huang," Product Review Information Extraction Based on Adjective Opinion Words," Fourth International Joint Conference on Computational Sciences and Optimization (CSO), pp.1309 –1313,2011.

1.2 Website

- [1.] <https://www.thebalance.com/what-is-social-media-2890301>
- [2.] <https://archive.ics.uci.edu/ml/datasets/YouTube+Spam+Collection>
- [3.] <https://en.wikipedia.org/wiki/AdaBoost>