ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

A PRAGMATIC EVALUATION OF USER SENTIMENT EXTRACTION FROM SOCIAL MEDIA FOR ASSORTED APPLICATIONS

Yatin Kumar Manocha

M.Tech. Research Scholar

Department of Computer Science and Engineering

Universal Group of Institutes

Lalru, Punjab, India

Priyanka

Head of Department, Assistant Professor

Department of Computer Science and Engineering

Universal Group of Institutes

Lalru, Punjab, India

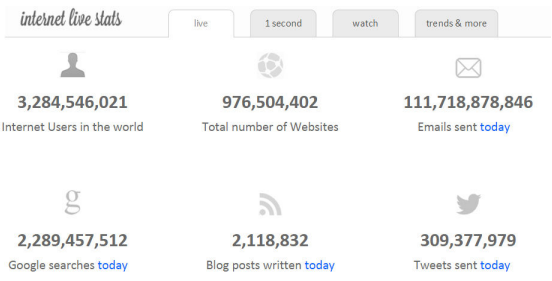
Abstract

Sentiment Data Analysis is one of the prominent and widely used domains by the research scholars and practitioners. In this approach, there are number of tools and technologies available for fetching the live datasets, tweets, emotional attributes. Using these tools, the real time tweets and messages can be extracted from Twitter, Facebook, WhatAapp and many other social medial portals. In this paper, the implementation of fetching live tweets from Twitter using python programming is presented. By this approach, the emotional attributes of the Internet users on social medial portals can be analyzed and prediction is done. Suppose we want to evaluate the overall cumulative score of a celebrity. For this, the python or PHP based programming scripts can fetch the live tweets about that celebrity or renowned person from Twitter. After that, using natural language processing toolkits, the fetched data in form of tweets or messages can be analyzed and prediction is done whether that particular person or movie or celebrity is getting fame or not.

Keywords: User Belief Mining, User Sentiment Mining, Sentiment Extraction

Introduction

Following is the statistical reports from InternetLiveStats.com and Statista.com about the real time data on social media and related web portals.



ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

Figure 1 - Screenshot from InternetLiveStats.com

On Twitter, 350 Million Tweets are flowing daily by more than 500 Million Accounts. Around 571 New Websites are hosted every minute on World Wide Web. There are more than 5 billion users on mobile phones concurrently.

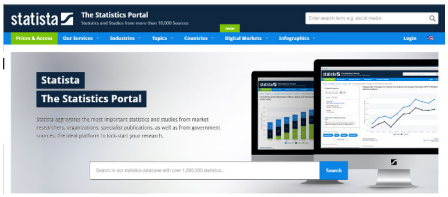


Figure 2 - Screenshot from Statista.com

On WhatAapp, there are 700 Million Active Users. There are more than 1 Million New Users Registration. Around 30 Billion Messages Sent and 34 Billion Messages Received everyday on WhatAapp. On Facebook, 5 New Profiles are created every second. There are around 83 Billion fake Profiles. Around 300 Billion Photos are Uploaded everyday by 890 Billion Daily Active Users. The data of 320 TB is processed daily with 21 Minutes per User Average Time spent. All these statistical reports are amazing and growing in huge volume every time instance.

Now, the question is how to do research work on these datasets. By which technologies the real time datasets can be fetched. The live streaming data can be fetched using Python, PHP, PERL, Java and many other used for network programming.

FETCHING LIVE STREAMING DATA FROM TWITTER USING PYTHON CODE

For fetching live tweets from Twitter, the specific packages named tweepy and Twitter is required with python. After installation of these packages, the python code will be able to fetch live data from Twitter.

These can be installed using pip command as follows -

$ python -m pip install tweepy

$ python -m pip install Twitter

CODE TO FETCH LIVE TWEETS FROM TWITTER

```
from tweepy import Stream
from tweepy import OAuthHandler
from tweepy.streaming import StreamListener
my_app_consumerkey = 'XXXXXXXXXXXXXXXXXXXXXXXXX'
my_app_consumersecret = 'XXXXXXXXXXXXXXXXXXXXXXXXX'
my_app_accesstoken = 'XXXXXXXXXXXXXXXXXXXXXXXXX'
```

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

```
my_app_accesssecret = '
XXXXXXXXXXXXXXXXXXXXXXXXXX
'
class TweetListener(StreamListener):
  def on_data(self, mydata):
    print mydata
    return True
  def on_error(self, status):
    print status
auth = OAuthHandler(my_app_consumerkey, my_app_consumersecret)
auth.set_my_app_accesstoken(my_app_accesstoken, my_app_accesssecret)
stream = Stream(auth, TweetListener())
stream.filter(track=[Name of the Celebrity or Movie or Person'])
```

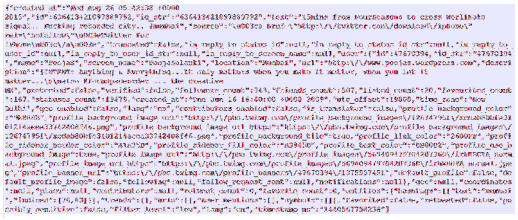


Figure 3 - Live Tweets fetched from Twitter in JSON Format

After execution of this script, the output dataset is fetched in JSON file format. The JSON file can be parsed using OpenRefine tool in the XML, CSV or any other readable format by the data mining and machine learning tools.

Openrefine is one of the powerful and effective tools that is used for processing the BigData and JSON file format.

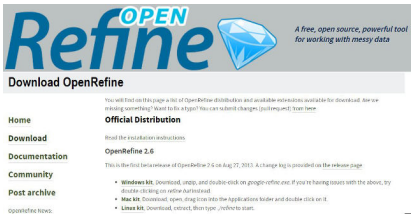


Figure 4 - OpenRefine Tool for Processing of DataSets

In the similar way, the timeline of any person or Twitter id can be fetched using the following code -

```
import tweepy
import time
my_app_consumerkey = 'XXXXXXXXXXXXX'
my_app_consumersecret = 'XXXXXXXXXXXXX'
my_app_accesstoken = 'XXXXXXXXXXXXX'
my_app_accesssecret = 'XXXXXXXXXXXXX'
auth = tweepy.auth.OAuthHandler(my_app_consumerkey, my_app_consumersecret)
auth.set_my_app_accesstoken(my_app_accesstoken, my_app_accesssecret)
api = tweepy.API(auth)
list= open('Twitter.txt','w')
if(api.verify_credentials):
  print 'Connected to Twitter Server'
```

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

```
currentuser = tweepy.Cursor(api.followers, screen_name="gauravkumarin").item()
while True:
    try:
        u = next(currentuser)
        list.write(u.screen_name +' \n')
    except:
        time.sleep(15*60)
        u = next(currentuser)
        list.write(u.screen_name +' \n')
list.close()
```

The following script of Python can be used to parse the JSON to CSV format

```
JSON - CSV Parser
import fileinput
import json
import csv
import sys
l = []
for currentline in fileinput.input():
    l.append(currentline)
currentjson = json.loads(''.join(l))
keys = {}
for i in currentjson:
    for k in i.keys():
        keys[k] = 1
mycsv = csv.DictWriter(sys.stdout, fieldnames=keys.keys(),

quoting=csv.QUOTE_MINIMAL)
mycsv.writeheader()
for row in currentjson:
    mycsv.writerow(row)
```

FETCHING DATA FROM TWITTER USING PHP CODE

For fetching the live tweets using PHP code, the API *TwitterAPIExchange* is required. After including this API in this PHP code, the script will directly interact with the Twitter Servers and live streaming data.

```
<?php
error_reporting(0);

define('CURRENTDBHOST','localhost');

define('CURRENTDBUSERNAME','root');

define('CURRENTCURRENTDBPASSWORD','');
    define(' CURRENTDBPASSWORD','Twitter');

define('CURRENTTWEETTABLE','Twittertable');

require_once('TwitterAPIExchange.php');
    $settings = array(
    'oauth_my_app_accesstoken' => "XXXXXXXXXXXXXXXXXX",
```

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

```
'oauth_my_app_accesstoken_secret' => " XXXXXXXXXXXXXXXXX ",
    'my_app_consumerkey' => " XXXXXXXXXXXXXXXXX ",
    'my_app_consumersecret' => " XXXXXXXXXXXXXXXXX "
    );
    $url = "https://api.Twitter.com/1.1/statuses/user_timeline.json";
    $myrequestMethod = "GET";
    $getfield = '?screen_name=gauravkumarin&count=20';
    $Twitter = new TwitterAPIExchange($settings);
    $string = json_decode($Twitter->setGetfield($getfield)
    ->buildOauth($url, $requestMethod)
    ->performRequest(),$assoc = TRUE);
    if($string["errors"][0]["message"] != "") {echo "<h3>Sorry, there was a problem.</h3><p>Twitter returned the following error message:</p><p><em>".$string[errors][0]["message"]."</em></p>";exit();}
    foreach($string as $items)
    {
        echo "Tweeted by: ". $items['currentuser']['name']."<br />";
        echo "Screen name: ". $items['currentuser']['screen_name']."<br />";
        echo "Tweet: ". $items['text']."<br />";
        echo "Time and Date of Tweet: ".$items['timestamp']."<br />";
        echo "Tweet ID: ".$items['id_str']."<br />";
        echo "Followers: ". $items['currentuser']['followers']."<br /><hr />";
        echo insertTweetsDB($items['currentuser']['name'],$items['currentuser']['screen_name'],$items['text'],$items['timestamp'],$items['id_str'],$items['currentuser']['followers']);
    }
    function insertTweetsDB($name,$screen_name,$text,$timestamp,$id_str,$followers){
        $mysqli = new mysqli(CURRENTDBHOST, CURRENTDBUSERNAME, CURRENTCURRENTDBPASSWORD, MYDBNAME);
        if ($mysqli->connect_errno) {
```

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

```
        return 'Failed to connect to Database: (' . $mysqli->connect_errno . ') ' . $mysqli->connect_error;
    }
    $QueryStmt='INSERT INTO '.MYDBNAME.'.'.CURRENTTWEETTABLE.' (name, screen_name, text, timestamp, id_str, followers) VALUES (?,?,?,?,?,?);';
    if ($insert_stmt = $mysqli->prepare($QueryStmt)){
        $insert_stmt->bind_param('ssssid', $name,$screen_name,$text,$timestamp, $id_str,$followers);
        if (!$insert_stmt->execute()) {
            $insert_stmt->close();
            return 'Tweet Creation cannot be done at this moment.';
        }elseif($insert_stmt->affected_rows>0){
            $insert_stmt->close();
            return 'Tweet Added.';
        }else{
            $insert_stmt->close();
            return 'No Tweet were Added.';
        }
    }else{
        return 'Prepare failed: (' . $mysqli->errno . ') ' . $mysqli->error;
    }
}
```

Conclusion

Using these technologies, the parsing, processing and prediction on the real time tweets and their association with the particular event can be mapped. The news channels adopt this type of implementation for exit polls which predict the probability of winning by any political party or candidate. In similar aspect, the prediction on the business of any movie can be done after careful analysis on the live streaming data. The research scholars can work on such real life topics related to BigData Analytics so that effective and presentable research work can be accomplished.

References

[1] Bifet, A., & Frank, E. (2010, October). Sentiment knowledge discovery in twitter streaming data. In International Conference on Discovery Science (pp. 1-15). Springer Berlin Heidelberg.

[2] Bollen, J., Mao, H., & Pepe, A. (2009). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. ICWSM, 11, 450-453.

[3] Bollen, J., Mao, H., & Pepe, A. (2010). Determining the Public Mood State by Analysis of Microblogging Posts. In ALIFE (pp. 667-668).

[4] Asur, S., & Huberman, B. A. (2010, August). Predicting the future with social media. In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

IEEE/WIC/ACM International Conference on (Vol. 1, pp. 492-499). IEEE.

[5] Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., & Li, P. (2011). User-level sentiment analysis incorporating social networks. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1397-1405). ACM.

[6] Saif, H., He, Y., & Alani, H. (2012, November). Semantic sentiment analysis of twitter. In International Semantic Web Conference (pp. 508-524). Springer Berlin Heidelberg.

[7] Leong, C. K., Lee, Y. H., & Mak, W. K. (2012). Mining sentiments in SMS texts for teaching evaluation. Expert Systems with Applications, 39(3), 2584-2589.

[8] Wang, H., Cambria, E., Schuller, B., Liu, B., & Havasi, C. (2013). Knowledge-based approaches to concept-level sentiment analysis. IEEE Intelligent Systems, 28(2), 12-14.

[9] Dong, H., Shahheidari, S., & Daud, M. N. R. B. (2013, July). Twitter sentiment mining: A multi domain analysis. In Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on (pp. 144-149). IEEE.

[10] Cambria, E., Fu, J., Bisio, F., & Poria, S. (2014). AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis. In AAAI (pp. 508-514).

[11] Kotwal, A., Fulari, P., Jadhav, D., & Kad, R. (2014). Improvement in Sentiment Analysis of Twitter Data Using Hadoop. Imperial Journal of Interdisciplinary Research, 2(7).

[12] Poria, S. Cambria, E., Fu, J., Bisio, F., & (2015). AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis. In AAAI (pp. 508-514).

[13] Guo, Y., Rao, J., Cheng, D., & Zhou, X. (2016). ishuffle: Improving hadoop performance with shuffle-on-write. IEEE Transactions on Parallel and Distributed Systems.

[14] Davis, B., Hürlimann, M., Cortis, K., Freitas, A., Handschuh, S., & Fernández, S. (2016, September). A Twitter Sentiment Gold Standard for the Brexit Referendum. In Proceedings of the 12th International Conference on Semantic Systems (pp. 193-196). ACM.

[15] Martínez-Cámara, E., Martín-Valdivia, M. T., Urena-López, L. A., & Montejo-Ráez, A. R. (2014). Sentiment analysis in Twitter. Natural Language Engineering, 20(01), 1-28.

[16] Abdul-Mageed, M., Diab, M., & Kübler, S. (2014). SAMAR: Subjectivity and sentiment analysis for Arabic social

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 4 July 2017

International Manuscript ID : 23482001V5I4072017-01

(Approved and Registered with Govt. of India)

media. Computer Speech & Language, 28(1), 20-37.

[17] Hassan, S., Yulan, H., Miriam, F., & Harith A.(2015, June). Contextual Semantics for Sentiment Analysis of Twitter. Elsevier.