Identification of Natural Disaster Affected Area Precise Location Based on Tweets

B. Premamayudu, P. Subbarao, Koduganti Venkata Rao

Abstract:Twitter is an "in the moment" platform designed for customers to post tweets about any event, which reports any major event including natural disasters. Hence, social media creates volumes of data on an event. Therefore, during the time of natural disaster like tsunami, earthquake, floods, landside etc., people of that area require information in those situations to enable relief operations to save many lives. This paper presents the identification of natural disaster affected area based on twitter tweets using Geoparsing to mark the places of disaster on a world map. In the proposed mechanism, longitude and latitude location of twitter message can extracted to map geographical coordinates in GoogleMapPlotter. The source of the geographical coordinates in real time is twitter messages collected based on the keyword and timeline. We can parse real time collected twitter messages for the natural disaster effected areas and locations. The collected tweets and their location information will help us to identify the exact place of disaster event. These tweets location information is stored in database or saved in CSV format to create the dataframe in python pandas. Further, the visualization is performed on the prepared dataframe using GoogleMapPlotter. This visualization is helpful for the disaster relief operations and estimates the severity of the natural disaster. The truthiness of the user tweets is evaluated using sentiment analysis for decision making.

Keywords: twitter, tweets, dataframe, pandas, python, googlemapplotter, sentiment analysis.

I. INTRODUCTION

Interactive computer mediated technology is a boon for data. Social media is a real time mechanism which provides a gateway for effective data[1][2]. Though the data is huge, unstructured and noisy the current focus is on extracting the best out it. Twitter is a micro blogger site to broadcasting information to the people. The well-known focal point for the information is twitter, where people can converge and share during the natural disaster. Information from broadcast media will be gathered directly or indirectly. Generally, information can be obtained directly from the people involved in the disaster[3].

B. Premamayudu,Department of Information Technology, Vignan Foundation for Science Technology and Research Deemed be University, Vadlamudi, Guntur,Andhra Pradesh India

P.Subbarao,Department of Information Technology, Vignan Foundation for Science Technology and Research Deemed be University, Vadlamudi, Guntur,Andhra Pradesh India

KodugantiVenkataRao,Department of Computer Science and Engineering, Vignan's Institute of Information Technology,Gajuwaka, Visakhaptnam (Dt), Andhra Pradesh, India

Some time, it can be gathered from various broadcasting blogger and reports or pointers of the information available from public sites. The natural disaster information in social media can be directly clouded by conversation and incomplete[4].

Data analysis on obtained tweets will help to know the current trend on the disaster. This analysis will help to identify the various patterns such as focal point of natural calamity or logistics at the located area[5]. To build decision making system from the obtained patterns by combining geo locations with sentimental analysis. Those decisions will help to take correct decision that help the victims and also save the lives of humanity in disaster. This article attempts to find the precise location of the natural calamity. Thereby, it will be informed to the rescue teams and send them to those locations to save the lives of humanity which are affected in natural disasters[4].

The proposed system has two major phases to find the precise location of disaster. In the first phase crawling the twitter to obtain data streaming and mining the knowledge using python pandas library. For these two phases, twitter API and social media informative sentiment analysis through data science and analytics. Crawling data from the twitter required to establish connection with twitter database. This connection needs authentication. There is a procedure to create an authentic connection with twitter database. The authentication connection and search key for the natural disaster will create data stream of tweets form the twitter database. In this process all the tweets are stored in data frame and data frame is converted into a comma separated values (CSV) file for identifying geo location of disaster through second phase. In second phase, sentimental analysis will be applied. Based on the result of sentimental analysis, the tweets and geo locations to extracted from the data frame and then visualized in a world map[5].

A. BigDataAnalytics

Big data analytics provides uniqueness of mining knowledge from the huge data frames. The analytics methods are more suitable for trends in leverage[6]. There are several methods available to improve the performance of trends analysis by ensuring the less resource utilization like storage cost, power consumption, and processor cycle etc. The new development in big data analytics, many modern tools are available to build analytic model with more accuracy. Big data analytical tools namely python, Hadoop, R studio, MAT lab etc are more popular to performance multidimensional analysis to mine the required knowledge from the given data set[7].

B. Why Twitter?

Twitter is a "day long broadcasting session" and widely used social media platform across the world. With around 319 million monthly active users are in across the world for twitter and 23.2 million monthly active users in India alone.Twitter is described as "the SMS of the internet", which restrict the number of characters in a tweet to 280 characters. It makes the tweet short and precise, making the text processing fast and easy. User of the twitter can see the updates when a message is posted in application. The messages of twitter are made public. The user has to follow the submitter to receive the updates. The tweets are identified using topics or keywords. Twitter follows its own concession to make distinct from other textual messages.

Terminologies used by twitter:

a. The maximum of 280 characters in a message posted in twitter is known as twit
b. "@" is the starting character for twitter account username
c. "#" symbol is used to mark topics or keywords
d. Videos, photos, links and text of tweet are mapped

C. Natural Disasters in India:

Natural disasters are regular in India, which cause massive loss of life and property. There are several causes to get the disasters like flooding, volcanic eruption, landslides, hurricanes, earthquakes, etc. Every year there is a lot of damage and loss in property and human lives due to natural disaster. The technological advanced in internet and way of sharing information can be used to some extent to reduce the damage created due to disasters. If the damage identified in shorter notice, then there will be change to reduce the suffering of the tragedy victims. More importantly, it is possible to get back some lives of suffered people by identifying exact or precise geo location of the disaster. The main objective of this attempt is that tweets are the real-time and quick source of information for any event happened at a particular area in the world.

II. ARCHITECTURE

It consists of

i. Build an environment to crawl the tweets from twitter application
ii. Jupyter Note Book to create python interactive environment for programming and for visualization
iii. Pandas package from the python

i. Build an environment to crawl the tweets from twitter application:

Python language has tweepy package to establish connection with twitter API. The step by step procedure to use tweepy package to access twitter API.

A. User has to create registration in twitter API. This registration provides authentication to access twitter API from any process. The authentication is compressed with key and secret key. These two keys are used in tweepy package methods to establish authentic connection with twitter application

B. The authentic connection creates a twitter link. This step creates the PIN also called verifier with the help of user's identity.

C. Access Token and Access Secret values provide exclusive access with twitter API to download tweets based on the search keyword or topic. The PIN is used to obtain an access token and access secret values.

ii. Jupyter Note Book to create python interactive environment for programming and for visualization:

Jupyter note book is open source software (or) Integrated Development Environment(IDE) which supports python and almost all the packages of python. Jupyter note book supports interactive data science and scientific computing across all programming languages.

iii. Pandas package from the python:

Python has a large number of packages which can be used for a large scale of applications. Five python packages are used for the implementation of this paper.

A. Pandas: It is a package for data analysis and data computations. This tool is used to create data frames from the CSV files which are obtained twitter API

B. NumPy: It is mathematical python library. It consists of multidimensional array objects and various mathematical function to operate on arrays and data frames created with pandas.

C. Tweepy: Tweepy supports accessing Tweeter via a method, OAuth.

D. Geocoder: Geocoder is a simple and consistent geocoding library written in python.

E. Gmplot: Gmplot is a Matplot-lib like interface to generate the HTML and javascript to render all the data users would like on top of Google Maps.

III. METHODOLOGY

Steps for tweet analysis to build disaster tweets mapped with geo location:

A. Build the application or process model for the system.
B. Collect the authentication requirements form twitter API.
C. Extract the keyword or topic based tweets.
D. Parse the tweet data which is in json format and get the required fields.
E. Store the tweet data into data set (as CSV file).
F. Find the location coordinates of the tweet.
G. Plot the coordinates on a map

A. Build the Application or Process Model for the System

There should be an application model to perform data clearing and conversion of data from obtained format to python data frame and also in CSV format. In addition, apply various filtering techniques on the data and remove all the NaN values and tweets without geo locations from the data frame for further processing. The total procedure is show in the figure 3 that present the pseudo code of the proposed model.

B. Collect the Authentication Requirements form Twitter API:

Use the python interface code to establish connection with twitter API. The following python packages required to access twitter API.

- tweepy
- twitter_credentials
- numpy
- pandas
- geocoder
- gmplot

C. Extract the Keyword or Topic based Tweets:

Once tweets are crawled from the twitter API database based on the keyword or topic search, they are analysed based on the area mapped with tweets. The area of real tweets is fixed and conformed based on the truthiness of the tweets using sentiment analysis. When the correct area is conformed, tweets will be extracted using the area in addition to keyword or topic search.

D. Parsing of Tweet Data:

A tweet is downloaded in JSON format containtweet objects, the tweet is parsed to extract only the required objects from the tweet.

E. Store the Data into Database:

The extracted data from the tweets is stored into data frames for the processing same values are then stored in the database for future use and cross verification of the results.

F. Find Location Coordinates from Tweet:

Location of the tweet (where it is posted from) is extracted from the tweet data using tweet.user.location, this method returns the coordinates of the location from where the tweet is posted from.

G. Plot the Coordinates on a Map:

The coordinates which are extracted are plotted on a map using gmplot, to visualize the coordinates.

IV. ALGORITHM

1. Start
2. Load Twitter API into memory.
3. Set CONSUMER_KEY, CONSUMER_KEY_SECRET.
4. Get Instances using Twitter API.
5. Initialize AccessToken, AccessTokenSecurity, CosumerKey, ConsumerSecret.
6. Create Query()
7. Store all the tweets info into database and CSV file.
8. If(Location_Co-ordinates):
   Extract co-ordinates.
   else:
   Convert the location into coordinates.
9. Get the coordinates
10. Plot the coordinates on the map.
11. End.

The crawled tweets from the twitter API are in JSON format. We have loaded the JSON formatted tweets into python data frame object. Fig 1 shows the tweets displayed from the python data frame to understand the tweet attributes.

	Tweets	id	len	date	source	likes	retweets	location
0	RT @Kalaari: Interesting stats from @SharmaShr...	1095329584315158528	140	2019-02-12 14:31:37	Twitter for Android	0	33	New Delhi
1	RT @jPraveenKrSingh: #MeraParivarBhajapaPariva...	1095329579965636608	99	2019-02-12 14:31:36	Twitter Web Client	0	43	New Delhi
2	RT @sidmtweets: Narendra Modi he is right , he...	1095329578145337345	112	2019-02-12 14:31:36	Twitter for Android	0	261	New Delhi
3	RT @RuchiraC: PM Modi = Prime #MiddlemanModlin...	1095329577998503936	68	2019-02-12 14:31:36	Twitter for Android	0	50	India
4	RT @mohitbharatiya_: Presenting the glimpse of...	1095329577306451968	140	2019-02-12 14:31:36	Twitter for Android	0	34	India

Fig 1 the python data frame of tweets and their attributes

V. RESULTS

Figure 2 show the natural disasters in India from disaster tweets. Visualization shows the sentiment of twitter users by creating visual maps. The locations of processed tweets are mapped to disasters can be marked through gmap.heatmap() function. The gmap.draw() function creates the world map shape file in html format.

Fig. 2 natural disaster in India tweets

```
from tweepy import API
from tweepy import Cursor
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
import twitter_credentials
import numpy as np
import pandas as pd
import geocoder
import gmplot
class TwitterClient():
    def __init__(self, twitter_user=None):
        self.auth = TwitterAuthenticator().authenticate_twitter_app()
        self.twitter_client = API(self.auth)

        self.twitter_user = twitter_user

    def get_twitter_client_api(self):
        return self.twitter_client

    def get_user_timeline_tweets(self, num_tweets):
        tweets = []
        for tweet in Cursor(self.twitter_client.user_timeline,
        id=self.twitter_user).items(num_tweets):
            tweets.append(tweet)
        return tweets

    def get_friend_list(self, num_friends):
        friend_list = []
        for friend in Cursor(self.twitter_client.friends,
        id=self.twitter_user).items(num_friends):
            friend_list.append(friend)
        return friend_list

    def get_home_timeline_tweets(self, num_tweets):
        home_timeline_tweets = []
        for tweet in Cursor(self.twitter_client.home_timeline,
        id=self.twitter_user).items(num_tweets):
            home_timeline_tweets.append(tweet)
        return home_timeline_tweets
    def get_keyworded_tweets(self,num_tweets):
        keyword_tweets=[]
        for tweet in Cursor(self.twitter_client.search,q="modi").items(num_tweets):
            keyword_tweets.append(tweet)
        return keyword_tweets
class TwitterAuthenticator():

    def authenticate_twitter_app(self):
        auth = OAuthHandler(twitter_credentials.CONSUMER_KEY,
        twitter_credentials.CONSUMER_SECRET)
        auth.set_access_token(twitter_credentials.ACCESS_TOKEN,
        twitter_credentials.ACCESS_TOKEN_SECRET)
        return auth
class TweetAnalyzer():
    def tweets_to_data_frame(self, tweets):
        df = pd.DataFrame(data=[tweet.text for tweet in tweets], columns=['Tweets'])
        df['id'] = np.array([tweet.id for tweet in tweets])
        df['len'] = np.array([len(tweet.text) for tweet in tweets])
        df['date'] = np.array([tweet.created_at for tweet in tweets])
        df['source'] = np.array([tweet.source for tweet in tweets])
        df['likes'] = np.array([tweet.favorite_count for tweet in tweets])
        df['retweets'] = np.array([tweet.retweet_count for tweet in tweets])
        df['location']=np.array([tweet.user.location for tweet in tweets])
        return df
twitter_client = TwitterClient()
tweet_analyzer = TweetAnalyzer()
tweets=twitter_client.get_keyworded_tweets(20)
```

Fig 3 pseudo code of proposed model

There is a knowledge base tweets list consists event related words. This list has a set of standard words related to natural disasters. The list might be framed from the National Disaster Management Authority (NDMA). This can be updated in regular intervals to improve the

performance of analytics. From the disaster related tweets, we can extract the geo location values (latitude and longitude). We have created two lists, one for storing latitude values and other one for storing longitude list. Figure 3 shows the pseudo code for given results in this paper.

To get the exact location of user, the constraint is that, user has to enable the option "Precise-Location" while tweeting in their mobile phone settings. Figure 4 shows the enable procedure of precise location in the settings app of android operating system.
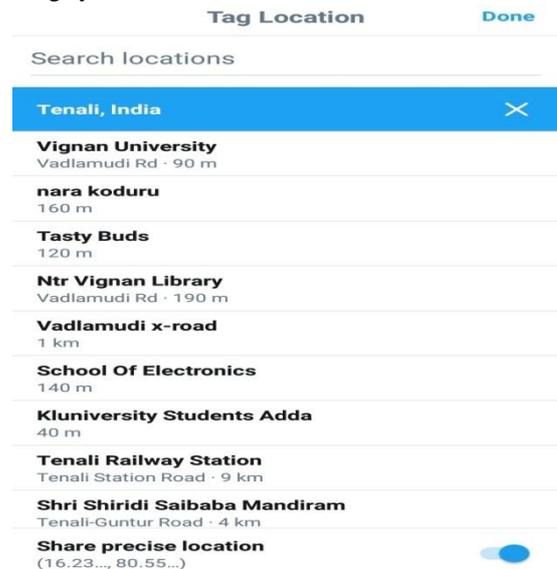

Fig 4 android operating system settings app to enable share precise location

VI. CONCLUSION AND FEATURE WORK

It is a challenging task to analyse data emitting from social media using existing tools. Therefore, during the time of natural disaster like tsunami, earthquake, floods, landside etc., people of that area require information in those situations to enable relief operations to save many lives. This idea can be enhanced by applying Natural Language Processing (NLP) to the text of the tweet to know the requirement of those people in critical situation and locating the precise location for the relief operations in more accurately.

REFERENCES

1. X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting Social Relations for Sentiment Analysis in Microblogging Categories and Subject Descriptors," vol. 1.
2. X. Hu, J. Tang, H. Gao, and H. Liu, "Unsupervised sentiment analysis with emotional signals," pp. 607–618, 2016.
3. N. P. Rao, S. N. Srinivas, and C. M. Prashanth, "Real Time Opinion Mining of Twitter Data," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 3, pp. 2923–2927, 2015.
4. F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley, "Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose," 2013.
5. A. Bruns and Y. Liang, "Tools and methods for capturing Twitter data during natural disasters," *First Monday*, vol. 17, no. 4, 2012.
6. M. F. Abdullah, M. Ibrahim, and H. Zulkifli, "Big Data Analytics Framework for Natural Disaster Management in Malaysia," no. IoTBDS, pp. 406–411, 2017.

7. M. Arslan, A. Roxin, C. Cruz, and D. Ginhac, "A Review on Applications of Big Data for Disaster Management To cite this version : HAL Id : hal-01678238 A Review on Applications of Big Data for Disaster Management," *13th Int. Conf. SIGNAL IMAGE TECH- Nol.*, 2017.

AUTHORS PROFILE

Dr.B. Premamayudureceived Ph.D at JNT University, Kakinada in Computer Science and Engineering in 2018. He is currently associate professor in Department of Information Technology at Vignan Foundation for Science, Technology and Research University, Guntur, India. His research interest focuses on Security in wireless sensor networks, IoT and future networking.

Dr. P. Subbarao received Ph.D at JNT University, Kakinada in Computer Science and Engineering in 2017. He is currently associate professor in Department of Information Technology at Vignan Foundation for Science, Technology and Research University, Guntur, India. His research interest focuses on software engineering, IoT and future networking.

Dr.KodugantiVenkataRao received the Ph.D degree from Andhra University inComputer Science and Systems Engineering in 2008. He is currently Professor in thedepartment of Computer Science and Engineering and Dean IQAC at Vignan's Institute of Information Technology, Visakhaptnam. His major research interests include keymanagement, authentication protocols and light weight protocol analysis and design innetworks.