International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 6 ǁ 2022 ǁ PP. 1337-1342

Data Modelling for Effective Disaster Rehabilitation using Tweets

Prakash C O, Sujay Vittal, R Lohit, Shylaja S S
Department of Computer Science and Engineering
PES University
Bangalore, India

Abstract— *Twitter's popularity as an information source has led to the development of applications and research in various domains. Humanitarian Assistance and Disaster Relief is one domain where information from Twitter is used to provide situational awareness to a crisis situation. Researchers have used Twitter to predict the occurrence of earthquakes and identify relevant users to follow to obtain disaster related information. We present a study of different tweets obtained from twitter – how these could possibly be used in Disaster rehabilitation. Our analysis reveals interesting inferences and provides a good visual representation of the same.*

Keywords—*Empirical Analysis; Twitter; Tweets; Tweepy; Predictive Analysis; Python; Disaster Rehabilitation; ChennaiFloods; South Indian Floods*

Date of Submission: 06-06-2022 Date of acceptance: 21-06-2022

I. INTRODUCTION

During the recent years Chennai floods, as part of the aftermath, over 1.5 lakh (150,000) street vendors sustained losses of over □300 crores (US$45 million). The death toll in the Chennai region was alone close to 700. There were people who were stuck in the middle of the disaster and because of help not reaching them. Not that there wasn't enough effort that was made to save people, the problem was with the approach. The disaster was so unexpected that not that the priority was to deploy the army and help as many people as possible. Timely help couldn't be extended to many people and hence that lead to the death toll to reach close to 700.

Disaster management involves a wide spectrum of problems that can be addressed. Through our project, we are looking to narrow the existing gap between the victims and the relief associations by providing good communication mechanisms and better reach to relief materials for the victims of the disaster using a combined setup. This setup involves two channels - one focusing on the victim and the other on the relief organization [2].

The victim can always be in talking terms with the relieforganizations and other people stuck in the disaster using our tool 'Crisiss'. On the other side, relief organizations can use our heat maps setup to better locate the people who are requiring immediate help (based on the degree of help required - we colour code the indication on the map of Chennai pointing to the location of the victim). The relief organizations can also continuously look into the other Heatmap that indicates the possible dynamic weather changes and hence decide which places require immediate assistance and which don't.

The major aim of this project is to narrow the gap between the relief organizations and the disaster victims. Through this project we want to show that using our proposed solution we can better help people in disaster situations by implementing better communication and geo-locating the people in danger torescue them or provide relief materials to them. We are considering the cases where a lot of relatives of the people who are under direct problems took the issues to social media platforms such as Twitter, Facebook and blogs to inform others about how helpless they are and what sort of an urgent help do these guys need. Here is where we thought we can make a difference. With the completion of this project, we hope that our solution acts as a better connect between the different organizations that are ready to provide relief and the disaster victims themselves.

II. IMPLEMENTATION

A. *Twitter and Tweepy:*

Twitter is a massive social networking site tuned towards fast communication. More than 140 million active users publish over 400 million 140- character "Tweets" every day. Twitter'sspeed and ease of publication have made it an important communication medium for people from all walks of life. Twitter has played a prominent role in socio-political events, such as the Arab Spring and the Occupy Wall Street movement.

Twitter has also been used to post damage reports and disaster preparedness information during large natural disasters, suchas the Hurricane Sandy. Similarly, we are using the Chennai floods as the use case to our project and have tried to tackle as many challenges as we could during. This disaster so that we can be better prepared for the next one

Users on Twitter generate over 500 million Tweets every day. Some of these Tweets are available to researchers and practitioners through public APIs at no cost.

APIs to access Twitter data can be classified into two types based on their design and access method:

- REST APIs are based on the REST architecture now popularly used for designing web APIs. These APIs use the pull strategy for data retrieval. To collectinformation a user must explicitly request it.
- Streaming APIs provides a continuous stream of public information from Twitter. These APIs use the push strategy for data retrieval. Once a request for information is made, the Streaming APIs provide a continuous stream of updates with no further input from the user. They have different capabilities and limitations with respect to what and how much information can be retrieved. The Streaming API has three types of endpoints:
    - **Public streams**: These are streams containing the public tweets on Twitter.
    - **User streams**: These are single-user streams, with to all the Tweets of a user.
    - **Site streams**: These are multi-user streams and intended for applications which access Tweets from multiple users.
- As the Public streams API is the most versatile Streaming API, we will use it.
- Tweepy is open-sourced, hosted on GitHub and enables Python to communicate with Twitter platform and use its API. Tweepy supports accessingTwitter via OAuth. OAuth is now the only way to usethe Twitter API.

B. OAuth:

Twitter APIs can be accessed only via authenticated requests. Twitter uses Open Authentication and each request must be signed with valid Twitter user credentials. Access to Twitter APIs is also limited to a specific number of requests within a time window called the rate limit. These limits are applied both at individual user level as well as at the application level. A rate limit window is used to renew the quota of permitted API calls periodically. The size of this window is currently 15 minutes.

Open Authentication (OAuth) is an open standard for authentication, adopted by Twitter to provide access to protected information. Passwords are highly vulnerable to theft and OAuth provides a safer alternative to traditional authentication approaches using a three-way handshake. It also improves the confidence of the user in the application as the user's password for his Twitter account is never shared with third-party applications

The authentication of API requests on Twitter is carried out using OAuth. Twitter APIs can only be accessed by applications. Below we detail the steps for making an API call from a Twitter application using OAuth:

- Applications are also known as consumers and all applications are required to register themselves with Twitter. Through this process the application is issued a consumer key and secret which the application must use to authenticate itself to Twitter.
- The application uses the consumer key and secret to create a unique Twitter link to which a user is directed for authentication. The user authorizes the application by authenticating himself to Twitter. Twitter verifies the user's identity and issues a OAuthverifier also called a PIN.
- The user provides this PIN to the application. The application uses the PIN to request an "Access Token" and "Access Secret" unique to the user.
- Using the "Access Token" and "Access Secret", the application authenticates the user on Twitter and issues API calls on behalf of the user. The "Access Token" and "Access Secret" for a user do not change and can be cached by the application for future requests. Thus, this process only needs to be performed once, and it can be easily accomplished using the method.

C. Collecting Search Results:

Searching on Twitter is facilitated through the use of parameters. Acceptable parameter values for search include keywords, hashtags, phrases, geographic regions, and usernames or userids. Twitter search is quite powerfuland is accessible by both the REST and the Streaming APIs. There are certain subtle differences when usingeach API to retrieve search results. We use Streaming API.

Streaming API:

Using the Streaming API, we can search for keywords, hashtags, userids, and geographic bounding boxes simultaneously. The filter API facilitates this search and provides a continuous stream of Tweets matching the search criteria. POST method is preferred while creating this request because when using the GET method to retrieve the results, long URLs might be truncated.

Using the Streaming API, we can search for keywords, hashtags, userids, and geographic bounding boxes simultaneously. The filter API facilitates this search and provides a continuous stream of Tweets matching the search criteria. POST method is preferred while creating this request because when using the GET method to retrieve the results, long URLs might be truncated.

The key parameters include:

- Follow: A comma-separated list of userids to follow. Twitter returns all of their public Tweets in the stream.
- Track: A comma-separated list of keywords to track. Multiple keywords are provided as a comma separated list.
- Locations: A comma-separated list of geographic bounding box containing the coordinates of the southwest point and the northeast point as(longitude, latitude) pairs.

Rate Limit:

Streaming APIs limit the number of parameters whichcan be supplied in one request. Up to 400 keywords, 25 geographic bounding boxes and 5,000 *userids* can be provided in one request. In addition, the API returns allmatching documents up to a volume equal to the streamingcap. This cap is currently set to 1% of the total current volumeof Tweets published on Twitter.

D. Strategies to identify the location of a Tweet:

Location information on Twitter is available from two different sources:

- **Geotagging Information**: Users can optionally choose to provide location information for the Tweets they publish. This information can behighly accurate if the Tweet was published usinga smartphone with GPS capabilities.
- **User Profile**: User location can be extracted from the location field in the user's profile. The information in the location field itself can beextracted using the APIs discussed above. Approximately 1% of all Tweets published on Twitter are geo-located. This is a very small portion of the Tweets, and it is often necessary to use the profile information to determine the Tweet's location. This information can be used in different visualizations. The location string obtained from the user's profile must first be translated into geographic coordinates. Typically, a gazetteer is used to perform thistask. A gazetteer takes a location string as input, and returns the coordinates of the location that best correspond to the string. The granularity of the location is generally coarse. The response is provided in JSON, from which the coordinates can be easily extracted. If the service is unable tofind a match, it will return (0,0) as the coordinates.

E. Google maps API, Heatmaps and Visualizing Geo-SpatialInformation:

Geo-spatial visualization can help us answer thefollowing two questions:

- **Where are events occurring? and,**
- **Where are new events likely to occur?**

The location information is typically used to gain insightinto the prominent locations discussing an event. Maps are an obvious choice to visualize location information. In this section we use Google maps API to generate maps to effectively summarize location information and aid in the analysis of Tweets.

In a geo-spatial visualization, we want to quickly identify regions of interest or regions of high density of Twitter users. This information for example could be used for targeted advertising as well as customer base estimation.

Kernel Density Estimation is one approach to estimating the density of Tweets and creating such heatmaps, whichhighlight regions of high density.

Kernel Density Estimation (KDE): Kernel DensityEstimation is a non-parametric approach to estimating the probability density function of the distribution from the observations, which in this case are Tweet locations [1]. KDE attempts to place a kernel on each point and then sums them up to discover the overall distribution. Appropriate kernel functions can be chosen based on the task and the expected distribution of the points. A smoothing parameter called bandwidth is used to decide if the learned kernel will be smooth or bumpy.

F. "Crisiss" Tool with Firebase:

The Crisiss tool is an android application that uses very less resources on the phone. It is a mode of communication where the disaster victim can talk to the relief organization and ask them for reliefs. It ensures that every message that goes into the server has an embedded latitude/longitude in it so that the admin can easily locate from where the messages are coming from.

The Crisiss tool used firebase which is a real-time database which can power our app's backend, including data storage, user authentication, static hosting, and more. It provides first class security features and works offline too.

The tool is optimized to use less than 0.5% of theKernel, about ~2.63 Kb/s for network calls. Hence, it would use very less battery and is suitable for a disaster struck situation.

III. TECHNICAL OVERVIEW

The project is hosted using a Firebase server, the communications happen over it. The disaster victims have a mobile application which can be used for communication. There are separate features where the victim can ask for relief materials from food to clothing or simply ask the relief organization for miscellaneous help. The messages are encoded with the latitudinal and longitudinal details so that the relief organizations can directly locate the victim. Also, other victims can offer to help as the message is made available to everyone in that demography.

The tweets obtained from Twitter are filtered and analyzed using ***TextBlob*** that assigns weights to the textual part of the tweets and the weights along with the latitude and longitude is used to construct the ***Heatmaps*** which can be monitored by the relief organizations and they can facilitate in extending more assistance.

The messages asking for help sent by the victims can be accessed by hosts sitting in any part of the world provided they have the authorization credentials to the ***Firebase*** server.

Disaster management involves a wide spectrum of problems that can be addressed. Through our project, we are looking to narrow the existing gap between the victims and the relief associations by providing good communicationmechanisms and better reach to relief materials for the victims of the disaster using a combined setup. This setup involves two channels - one focusing on the victim and the other on the relief organization.

The performance of the proposed solution is tabulated as follows:

Properties	Performance (Approx.)
Memory	< 7.862 MB
CPU	< 0.5% of Kernel
Network	~2.63 KB/s

The victim can always be in talking terms with the relief organizations and other people stuck in the disaster using our tool 'Crisiss'. On the other side, relief organizations can use our heat maps setup to better locate the people who are requiring immediate help (based on the degree of help required - we color code the indication on the map of Chennai pointing to the location of the victim). The relief organizations can also continuously look into the other heatmap that indicates the possible dynamic weather changes and hencedecide which places require immediate assistance and which don't.

The following depicts the workflow:

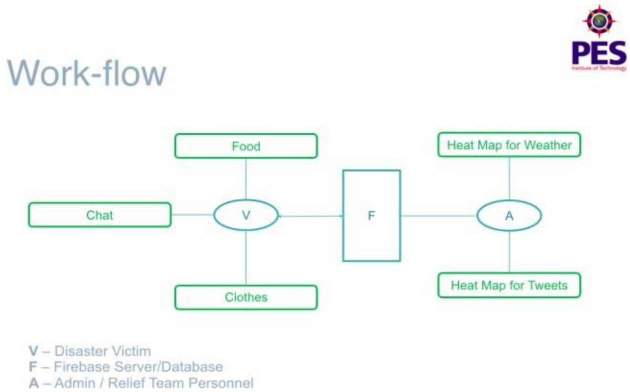

Figure A: Workflow

IV. OBSERVATIONS

The major observations that were made as part of this paper was that there exists a strong co-relation between the tweets that originate from certain groups and their cumulative reactions towards a certain happening.
The data was taken over a wide spectrum of topics/events.
Here, it's disasters. Analysis was done over a very small duration of time. However small the duration was, since it was live twitter data, we had about five thousand entries within a very short duration of time. Figure A shows the work flow that has been followed.

The relation between tweets and associated problems proved very helpful in building the bridge that is crucial in helping people rehabilitate during disasters. The major aim of this project is to enhance and strengthen the bridge between the relief organizations and the disaster victims. Considering the cases where a lot of relatives of the people who are under direct problems took the issues to social media platforms such as Twitter, Facebook and blogs to inform others about how helpless they are and what sort of an urgent help do these guys need. Here is where we thought we can make a difference. With the completion of this project, we hope that our solution acts as a better connect between the different organizations that are ready to provide relief and the disaster victims themselves.

Disaster management involves a wide spectrum of problems that can be addressed. Through our project, we are looking to narrow the existing gap between the victims and the relief associations by providing good communication mechanisms and better reach to relief materials for the victims of the disaster using a combined setup. This setup involves two channels - one focusing on the victim and the other on the relief organization [3].

The victim can always be in talking terms with the relief organizations and other people stuck in the disaster using our tool 'Crisiss'. On the other side, relief organizations can use our heat maps setup to better locate the people who are requiring immediate help (based on the degree of help required - we color code the indication on the map of Chennai pointing to the location of the victim).

V. CHALLENGES

The analysis of social media comes with a set of extremely challenging situations which demand specific actions. The major challenges we faced as part of this paper was the mining of data relevant to our study [4].

Twitter produces data in extremely large quantities. The major challenges were finding answers to questions such as – How effective are the predictions that are being made? Can issues be predicted on the basis of Empiricism? And can predictive models be held accountable for predicting imperfections/problems?

- Data obtained from Twitter

{
"created_at":"Wed Nov 25 17:05:55 +0000 2015","id":647457013530759168,"id_str":"647457013530759168","text":"RT http:\/\/t.co\/bcj7jNy11R Help required, save us! #ChennaiRains #ChennaiRainsHelp #\u2026 http:\/\/t.co\/ScfcBpEhNN","source":"\u003ca href=\"http:\/\/ifttt.com\" rel=\"nofollow\"\u003eIFTTT\u003c\/a\u003e","truncated":false,
"Latitude": 13.0827, "Longitude":80.2707
}

Figure B.1: Unsorted Data

- Data obtained from Twitter is then processed by TextBlob which assigns weights to the text based on certain test data

{
"created_at":"Wed Nov 25 17:05:55 +0000 2015","id":647457013530759168,"id_str":"647457013530759168","text":"RT http:\/\/t.co\/bcj7jNy11R **Help required, save us! #ChennaiRains #ChennaiRainsHelp** #\u2026 http:\/\/t.co\/ScfcBpEhNN","source":"\u003ca href=\"http:\/\/ifttt.com\" rel=\"nofollow\"\u003eIFTTT\u003c\/a\u003e","truncated":false,
"Latitude": 13.0827, "Longitude":80.2707
}

Figure B.2: Sorted Data

The major challenge that we faced while developing this project was extraction of text data/relevant text data. Figure
B.1 gives an illustration of the unsorted data we came across during data extraction. Figure B.2 shows the extracted statement done using TextBlob.

VI. RESULT AND ANALYSIS

The Goal of this paper is to infer hypothesis from the patterns - relate different Tweets with the same inferred pattern (testing). We deduced relationships between patterns and tweets. Say, for problem/situation X, there are Y types of typical tweets. The model would be able to predict the type of problems a particular pattern may indicate and this is the primary goal of this research.

Visualizations shown in the conclusion section give a better picture of the same scenario. The following was the flowdiagram that mainly highlights the work that has been done.

Using Tweepy, TextBlob predictive analysis on twitter data was done to build a hypothesis. Tweets were used as data and also were used as benchmarks to compare other tweets.

VII. CONCLUSION

Disaster Management when considered as a topic has a wide spectrum of activities that can be performed. This work looks to tackle one aspect of the whole spectrum. We are trying to simplify and at the same time make the communication between the victims and the relief organizations better. We have been able to prove that ourapproach is possible and have a POC to back our claims.

When we consider building a mobile project for the victims to use, we face problems such as battery, network related and other phone related issues as it will be continuously subjected to disaster prone conditions. However, we have worked quite a bit on the optimization fronts and are happy to share that our mobile application uses very less amount of data (~2.63 KB/s), uses less than 0.5% of the Kernel. The only worry is that the application requires minimum of 7.862 MB of RAM to run. We will be working on optimizing the memory which the application uses in the next phase of this work.

Overall, the basic goals of this work have been achieved and a working prototype of the same has been developed.

REFERENCES

[1]. Aydinoglu, A. C., and M. S. Bilgin. "Developing an open geographic data model and analysis tools for disaster management: landslide case." Natural Hazards and Earth System Sciences 15.2 (2015): 335-347.

[2]. Dilo, Arta, and Sisi Zlatanova. "Spatiotemporal data modeling for disaster management in the Netherlands." Information Systems for Crisis Response and Management (2008): 517-528.

[3]. FEMA – US Department of Homeland Security

[4]. Sujay Vittal, R Lohit and C O Prakash, "Empirical Analysis of Social Networks: Predictive Analysis of Tweets – Visualization",17th International Conference on Computer Science and Information Technology (ICCIT 2016), pp. 27-29.