Using Big Data Analytics for Finding Natural Disasters

A.Susmitha, R.Vara Prasad

Computer Science and Engineering, G. Pullaiah College Of Engineering and Technology

Abstract -- *The source of information presented is social media, which gives the news of any major event including natural disasters. It also includes conversational data. As a result, the large massive amount of data on social media has a tremendous increase. During the time of natural disaster like earthquake, landslide, floods, tsunami, etc., people require information in those situations, so that the relief operations like help, medical facilities can save many lives from those disasters. A study on Geoparsing is done in this paper which will identify the places of disaster on a Map. Geoparsing is a process of converting text description of locations into the geographical identifier with the help of longitude and latitude. A real-time, example is the use of twitter messages which are the source of the information that can handle a large amount of data. collecting tweets at the real-time we can resolve them for the disaster situation and its location. This information will help to identify the exact location of the event. To know the information on the natural disaster, tweets are extracted from the twitter to the R-Studio environment. First, the derived tweets from twitter are parsed using R about "The Natural Disaster" that has occurred. Later the tweets are parsed and stored in CSV format in R database. Sentiment analysis is used on user tweets, which is useful for decision making.*

Key words: **Twitter, R-Studio, Natural disasters, Real-time systems, Geoparsing, Sentiment analysis**

I. INTRODUCTION

Big Data is an industry jargon that refers to the large amount of information currently being generated by websites, social media, physical network of sensors and other relatively new resources. Presently the frequency and intensity of natural disasters have increased significantly and this trend is expected to continue. Nowadays the major disasters are predicted earlier through the use of developments like Lazer beams, which is used to predict earthquakes. By using these technologies rescue organizations are trying ways in order to find the value in and perception from both structured and unstructured data from internal and external sources which are generated in large volumes.

Most of the methods reported in the literature so far are based on the statistical analysis of diverse geological indicators and certain precursory patterns. High-resolution satellite aerial images are the sources of valuable information. However, extraction of knowledge from such huge amount of massive data cannot always be performed by using this statistical analysis. Devices like Smartphones, personal home assistance usually have multiple built-in sensors that can combine, compute and store large data sets which can be used for processing also there is an estimate that nearly 6.4 billion of devices are connected to the internet of things. By using these Smartphones we are able to detect the natural disasters the moment they happened.

Big Data also allow the early tracking of floods. By combining data related to flooding from Twitter and satellite observations, a group of scientists (De Groeve, Kugler, and Brakenridge) built a real-time map of the location, timing, and hit of floods. The map, constantly revised, can be seen in online. Disaster-affected communities use social media as a evidence to the destruction around them, and scientists are looking to use that data to aid the affected communities. For example, during Hurricane Sandy, there were over 20 million tweets posted, which could have provided valuable information to emergency responders.

The United Nations (UN) has created an organization called Global Pulse which uses big data to understand the effects of storms or other disasters after they hit.The analysis of patterns within big data could reform the way we respond to events such as disease outbreaks, and natural disasters and so on.
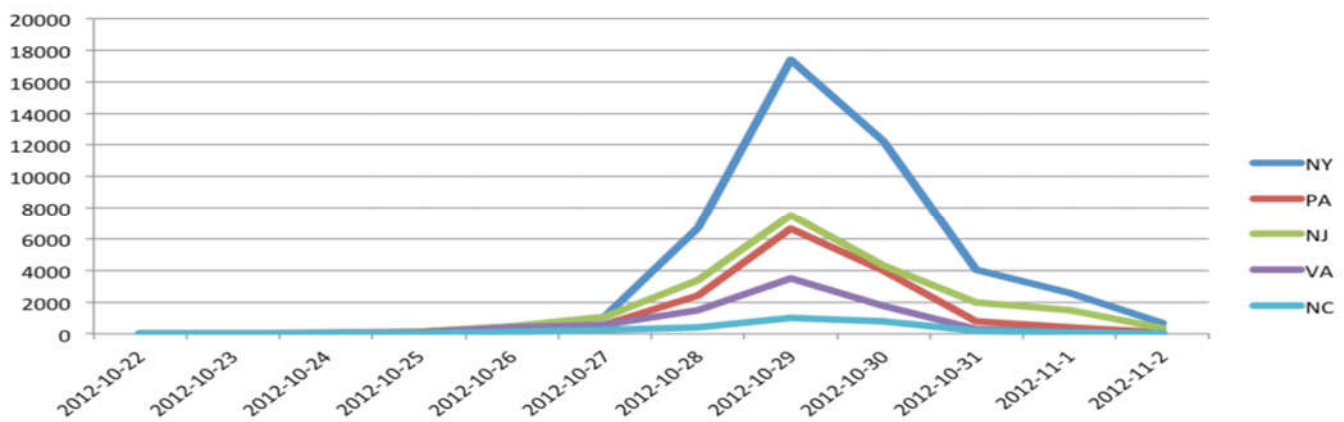
Fig 1: The number of tweets per day that contain the word "sandy" by state

Global Pulse was able to slide through thousands of tweets, and, using the metadata from related posts, they were able to determine and figure out the time when each post was uploaded, the GPS coordinates, and the types of damages which are present in the photos. With this metadata, they were able to create a precise map of the storm, which helped with relief efforts as they were happening.

In the future, UN Global Pulse desire is to create an algorithm that crawls all social media posts for text and pictures that could imply the initiation of a natural disaster. Big Data-powered visualization tools seem to be helpful for real-time predictions of weather system as well as for raising awareness of natural hazards among citizens.
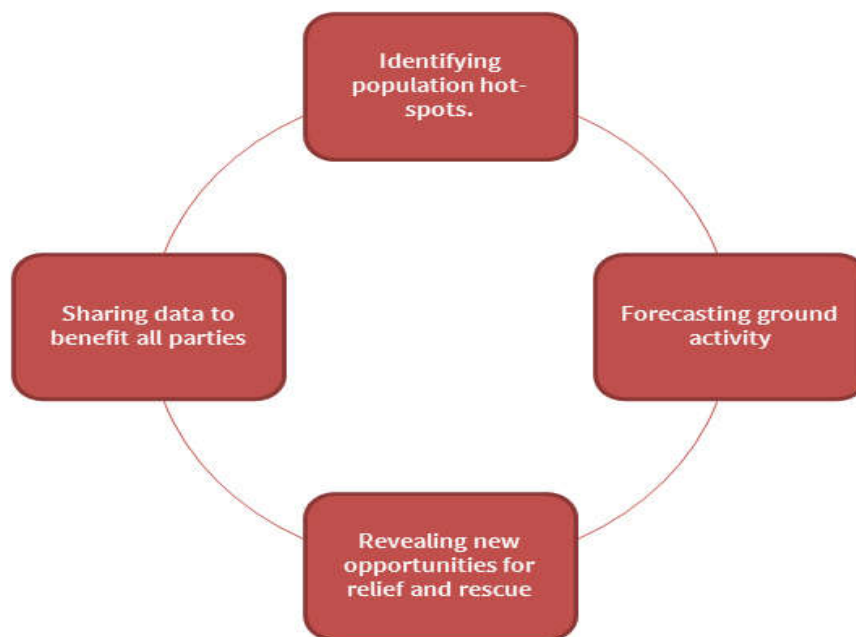

Fig:2 Use of Big Data in disaster management

With big data, documents or reports of important population subsets, such as areas with different communities like elderly communities or areas with infant and youth concentrations can be identified. Big data also helps to recognize the specific mobility support or resources that responders can apply to these population hot spots. It also gets easier, with big data, to send real-time information to residents when a disaster approaches.

There are two phases namely data streaming [1] from twitter and mining of the knowledge through R-Studio. For these two operations, the two approaches used are twitter API and sentiment analysis [2] through R. Twitter applications

requests for connection to the twitter database. When the connection is established authentication is generated, providing search key the natural disaster. A file will be data frame (DF) is generated with tweets and is converted into a comma-separated values (CSV) file which can be visualized in a Map. Twitter applications requests for connection twitter database. When the connection is established authentication is generated, providing search key the natural disaster.

II.   Big Data analytics

The process of evaluating and mining big data known as big data Analytics[3] can provide operation and knowledge of accuracy and uniqueness. One of the biggest tasks of big data is to analyze data collected by business houses for trends in advantage[4]. The inventive techniques of big data are majorly in data storage, it's processing and analyzing[5]. These methods ensure the reduction in storage cost and power consumption. All the Data centres will be profited by cost reduction techniques.

Further, the cloud computing increases the efficiency. Data tools like Clojure, Scala, Python, Hadoop and java for Natural Language Processing (NLP) and text mining and R, MAT can be used in data analytics. Processing the large volume of data in parallel mode is an important measure[6]. Also, the new development of packages the Hadoop for computation in distributed mode is a weaker tool. Tradition analytics, big data analytics has several differences.

III.   Sentimental analysis

Opinion mining (sometimes known as sentiment analysis or emotion AI) refers to the use of natural language processing, text analysis, computational linguistics, and so on. Often, sentiment analysis is done on the data that is collected from the internet and from various social media platforms.

IV.   Architecture

It consists of
1.Twitter environment.
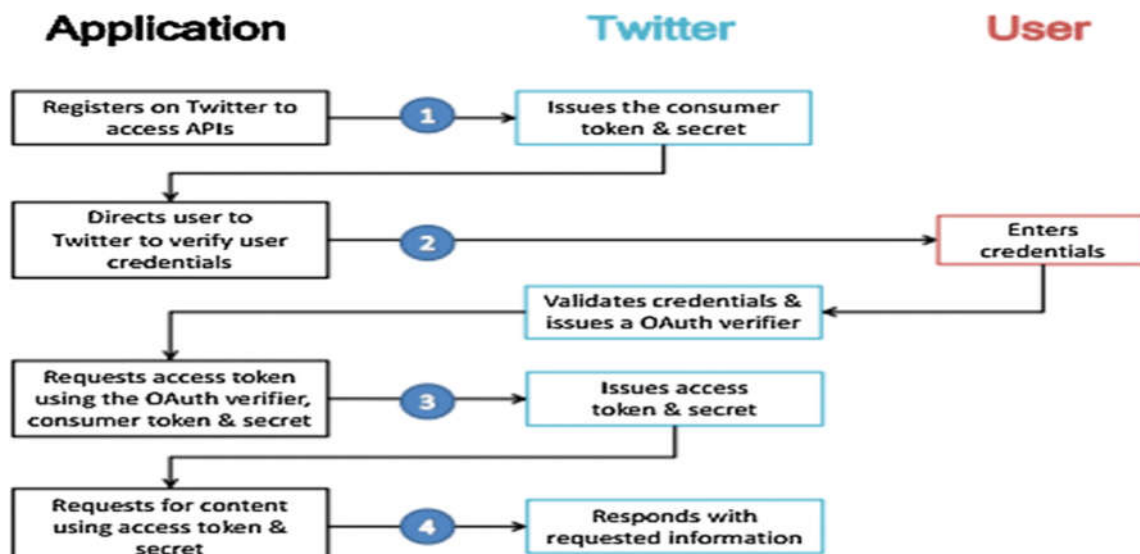2. R-Studio.
3. R-Packages.

Fig:3 Architecture

i.   *Environment of twitter:*

OAuth package of R is used for Twitter API confirmation procedure. Figure 3 shows the steps involved in the usage of OAuth to Access Twitter API.

1.Upon registration by the user in Twitter, a key and secret key are provided which are required for application authentication.

2.The authentication process is initiated with the help of these keys, which are used to create a twitter link. The twitter verifies the user's identity and issues PIN3 called verifier. This pin is required for the Twitter application.

3.This PIN is used for requesting an Access Token and Access Secret, which are exclusive to the particular individual from twitter API for continuing the next application process.

4.GetUserAccessKeySecret contains token and secret key, whose information is required for further use.

ii.   *R-studio:*

For graphical analysis, R-studio is used. It contains more than 8000 packages. Multiple window environments are provided in the R-studio nearly four. The work area called as R-console is present at the left bottom of the screen. The area called R-script area where R-scripts are written. The area called Global area where variables are defined and read. Display of charts and data is called a plot area, which is present at right bottom.

iii.   *R-packages*

For R functions compiling code on R-studio, we use the packages. These functions make the R library environment. Some of the packages are installed by default when we have installed the R studio. Other packages are loaded separately upon required specifications.

TwitterR: It is mainly used for providing an interface to Twitter API.

ROAuth: it provides the users it means of o authenticate to the server.

V. Conclusion

It is a challenging the task to analyze large data emanating from social media with the existing data mining tool. The aim is to access twitter and R-studio. Large data for decision making is done by twitter and R-studio analysis. From the retrieved "Natural Disaster in India" data from Twitter Sentiment Analysis has shown the opinion of the people. This concludes that R statistics tool is sufficiently used for the analysis of big data. The application of PYTHON for analysis of big data can be explored.

REFERENCES

1. Morstatter F, Pfeffer J. Is the sample good enough? Comparing data from twitter's streaming API with Twitter's Firehose. In: The 7th international AAAI conference on weblogs and social Media. 2013.
2. Agarwal A, Xie B, Vovsha I, Rambow O. Sentiment analysis of Twitter Data. In: Proceeding LSM '11, Proceedings of the workshop on languages in social media, p. 30–38, Portland, Oregon, 23 June 2011 Available from: dl.acm.org.
3. Sukhpal K, Rashid EM. Web news mining using back propagation neural network and clustering using K-Means algorithm in Big data. Indian Journal of Science and Technology, 9(41). doi:10.17485/ijst/2016/v9i41/95598.
4. Manyika J, Chui M, Brown B, Bughin J, Dobbs J, Roxburgh C, Byers AH. Big data: the next frontier for innovation, competition, and productivity. Chennai: McKinsey; 2011.
5. Wilder-James E. What is big data? An introduction to the big data landscape. 2012. https://www.oreilly.com/ideas/what-is-big-data
6. Prajapati V. Big data analytics with R and Hadoop. Birmingham: Packt Publishing; 2013.