

A Review on Application of Data Mining Techniques to Implement Natural Disaster Alert System

Yash Pratap Singh¹ Mrs. Shalu Bhasambu²

^{1,2}Department of Information Technology

^{1,2}Maharaja Agrasen Institute of Technology, India

Abstract— Thousands of human lives are lost every year around the globe, apart from significant damage on property, animal life, etc., due to natural disasters (e.g., earthquake, flood, tsunami, hurricane and other storms, landslides, cloudburst, heat wave, forest fire). In this paper, we focus on reviewing the application of data mining and analytical techniques designed so far for (i) prediction, (ii) detection, and (iii) development of appropriate disaster management strategy based on the collected data from disasters. A detailed description of availability of data from geological observatories (seismological, hydrological), satellites, remote sensing and newer sources like social networking sites as twitter is presented. An extensive and in-depth literature study on current techniques for disaster prediction, detection and management has been done and the results are summarized according to various types of disasters.

Key words: Data Mining, Natural Disaster

I. INTRODUCTION

In this section, broad categories of tasks that can be solved using different types of data have been discussed. We can classify the objectives of the tasks, in the following three major categories

A. Prediction:

These sets of tasks involve prediction of the natural disaster, disaster prone area and different attributes of a natural disaster that can occur. Basically, these tasks involve prediction or forecasting of time, place and magnitude of the disaster.

B. Detection:

These sets of tasks involve detection of the natural disaster promptly after it has occurred. Literature studies indicate that the social sensors in terms of tweets, other social media websites report a natural disaster much faster than the observatories.

C. Disaster Management Strategies:

These methods deal with identification of different entities that are taking part in combating a disaster so that communication is enhanced, appropriate concern of the affected people is identified and distribution of relief items are optimized. In many cases, the classification stated above is overlapping.

As an example, it can be surely argued that, detecting the natural disaster helps in disaster management strategies. Even the psychological studies can give lot of insight to disaster management strategies. So, the classification is based on the direct objective of the task involved.

D. Prediction:

There is no doubt that this would be the most “ideal” problem to solve. But very often, this is not a problem that can be solved with available data and techniques. However, it is possible to predict, the areas which are susceptible to a particular type of disaster, let’s say, landslide or flood. The prediction techniques have been seen to be of more use for predicting various characteristics of a natural disaster, which has occurred. As an example, the techniques can be used to predict the magnitude of an earthquake, track and intensity of a cyclone etc. Analysis of various spatial and temporal data is often needed for such tasks. Though handful, another branch of research has focused on using unusual animal behavior to predict a natural disaster.

E. Detection:

Often the meteorological observatories detect the natural disaster, but the news of the detection takes a long time to be communicated to proper authorities with the exact location of the detection.

F. Disaster Management Strategies:

These sets of tasks are involved in forming appropriate disaster management strategies. An example of such tasks is identifying critical entities for disaster management; identify proper communication study, identifying the needs of the disaster affected area. Social media data is very important in these types of tasks.

The aim of disaster management should be the following-

- Minimize casualties
- Rescue victims on time
- Offer first aid instantly
- Evacuate people and animals to safe places
- Reconstruct the damages immediately

II. DATA SOURCES & TYPES

The objective of this section is to give researchers and practitioners a high level overview of the type of data that are useful for analysis and prediction of a natural disaster. Much of the data will qualify to be called ‘Big Data’, because of all or some of the dimensions of volume, variety and velocity as listed below:

- Volume (GIS data, meteorological data, social media data).
- Variety (text, time series, spatial data, GIS images).
- Velocity (because of the rate in which data are generated as well as because of the speed in which a decision needs to be taken).

The following questions arise thus-

- What are the different types of data that are useful for each type of disaster?
- What are the sources and format of such data, at national and international level?

- How the data can be accessed? Whether it is freely available or not?

III. TWITTER & SOCIAL MEDIA AS SOURCES

Twitter as a data source has gained lot of prominence in recent years. It is ranked as one of the top 10 popular websites, having 400 million registered users and over 500 million tweets generated every day. Additionally, information about disasters can be extracted from news channels & blogs through APIs, RSS feed or web scraping. Sentiment Analysis, Stock Market Public Health, General public mood and finding political alignments are some of the areas where twitter data have been used, Some of the advantages of tweets are-

- Though it is unstructured, it has some structure by the limitation of 140 characters.
- It can use hashtags, which gives semantic annotations of the tweets.
- The tweets have geocoded, which can help us spatially map the sources of the tweets.

In many countries, Twitter has been used to effectively manage disasters; however, in Indian context we have not seen a lot of referential work. Twitter has been effectively deployed to: -

- Detect disasters faster than observatories.
- Identifying key entities in disaster management & relief organization.
- Temporal study of needs and concerns after the disaster

IV. PROPOSED SYSTEM

I intend to build a database for natural disasters happening in India. In the project I want to concentrate on sources like Twitter. Like any standard system, I will need the following components as shown in the figure.

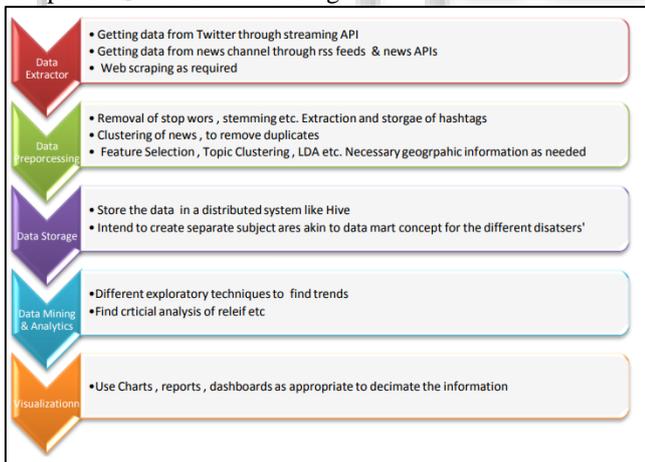


Fig. 1: Components involved in the system

The focus would be, on understanding peoples need during the disaster and evaluate the social impact and changes due to natural hazards.

Task	Detailed objective	Model techniques used	Data source, type	Country
Disaster management	Build a tool to extract important information from tweets for relief workers	Text processing techniques	Twitter	USA
	Build a system, for disaster discovery and humanitarian relief based on tweets. The system consists of a stream reader, a data storage and a visualization module.	SVM, LDA, topic clustering	Twitter	USA
Prediction	Build a geo hazard database for early prediction system, by using the Google news service. Geo tagging is done for geo referencing. The purpose is using this database extensively for disaster management	Text processing	Google news service, RSS feed	Italy

Table 1: General data, model and task summary

In the table referred above, different research directions in combating natural disaster have been discussed. This is a multidisciplinary activity needing experts from environmental science, geology, meteorology, social science, computer science, etc. The above list is not exhaustive, but an effort has been made to cover last 10–12 years data in this section. Here are few of our observations:

Twitter as a source has become important for real time detection and understanding of the need and concern of the affected people. Ten out of the forty papers we reviewed above use twitter as the data source. Interestingly, we did not find any referential work, where twitter has been used in an Indian context.

It is also observed, though India has a much higher loss in terms of human life and property, and adequate research as in countries such as USA has not been done in India

- Data Sources: I am focusing on Twitter at this point. The data extraction scripts will be written in Python, Javascript or R as applicable.
- Data Storage: The data storage is envisaged to be on geoJSON.

V. CONCLUSION

Natural disasters in forms of earthquake, floods, landslides, storms claim numerous lives, cause significant damage to property. The effects have been much more severe in a developing country like India compared to developed countries. India is a victim of natural disasters every year and the loss of lives and properties adds up to millions of rupees which this country cannot afford to lose. There are certain reasons for such poor disaster management procedures followed in this country.

- Inadequate early warning system.
- Poor preparation before the disaster occurs.
- Inadequate and slow relief operation.
- Lack of proper administration.
- Slow process of rehabilitation and reconstruction.
- Poor management of finances for relief work.
- Lack of effective help to victims.

There have been many efforts to predict the disasters based on various sources of data. In my literature survey, I explore the multidisciplinary nature of the task, where data mining models are being applied on various types of data, requiring deep subject matter expertise.

Recently Social Media and Internet have also emerged as an important source of information. These sources may not be used in prediction of the disasters, but they have contributed significantly to early detection and adoption for appropriate disaster response. I observe in my study that, there have not been enough works done in this area to tap the potential of these sources especially in context of India. I propose to build an alert system for natural disasters from these sources.

REFERENCES

- [1] Liu, S. B., B. Bouchard, D. C. Bowden, M. Guy, and P. Earle. "USGS Tweet Earthquake Dispatch (@ Suggested): Using Twitter for Earthquake Detection and Characterization." In AGU Fall Meeting Abstracts, vol.1, p. 2431. 2012.
- [2] Zhang, Xiao Yu, Xiang Li, and Xiao Lin. "The Data Mining Technology of Particle Swarm Optimization algorithm in Earthquake Prediction." Advanced Materials Research. Vol. 989. 2014.
- [3] Cradler, URLs. "Calandra's: A rapid overview about felt earthquakes through data mining of web- accesses." EGU General Assembly Conference Abstracts. Vol. 15. 2013.
- [4] Takako Hashimoto, Asabi Chakraborty, Mesuji Kolyma and Yukari Shirota, "Temporal Awareness of Needs after East Japan Great Earthquake using Latent Semantic Analysis", Information Modelling and Knowledge Bases XXV, Edited by Kokoda et.al. pp. 200-212, IOS press, 2014.
- [5] Bhargava, Neeta, V. K. Katihar, M. L. Sharma, and P. Pradhan. "Earthquake prediction through animal behavior: A review." *Indian J. Biotech* (2009): 7-8.
- [6] 6. Adele, Hajjam, and Asif Palakkad. "A probabilistic neural network for earthquake magnitude prediction." *Neural Networks* 22.7 (2009): 1018-1024.
- [7] Namawanda, Shastri L., and Heinz Dreher. "Ontology based data warehouse modeling and mining of earthquake data: prediction analysis along Eurasian-Australian continental plates." *Industrial Informatics*, 2007
- [8] 5th IEEE International Conference on. Vol. 1. IEEE, 2007.
- [9] Aydin, Ishan, Mehmet Karaoke, and Erhan Akin. "The prediction algorithm based on fuzzy logic using time series data mining method." *World Academy of Science, Engineering and Technology* 51.27(2009):91-98.