

Big Data Implementation of Natural Disaster Monitoring and Alerting System in Real Time Social Network using Hadoop Technology

S. Dhamodaran*, K. R. Sachin and Rahul Kumar

Faculty of Computing, Computer Science and Engineering, Sathyabama University,
Chennai - 600119, Tamil Nadu, India; s.dhamodaran07@gmail.com, krsachin001@gmail.com,
k.s.rahul.singh@gmail.com

Abstract

Background/Objectives: The information generated by the social networks is exponentially higher and demand effective systems to yield effective results. In conventional techniques stay unqualified because they ignore the social related data. **Methods/Statistical Analysis:** The existing system doesn't provide any proper methodology to report people about any natural disasters rapidly or to notify the rescue agencies for taking immediate actions to carry out the rescue process. The only existing methodology of reporting people is the media, i.e. news, radio, etc. The use of a particle filter fetches the necessary keywords from tweets by the use of Stemming along with the location and time. When the system encounters keywords related to natural disasters, an auto alert is sent to the people in the nearby locations and the rescue teams based on a proper verification algorithm. **Findings:** The method finding the two data set representations: one is considering the two directional social relations, and the other considering the one directional social relation. It is seen that the performance of the recommender system can be greatly boosted by the mentioned contextual factors. **Application/Improvements:** On validation of the authenticity user, the user is allowed to enter the application. The user is allowed to tweet using the same application.

Keywords: Bigdata, Disaster, Hadoop Technology, Social Network

1. Introduction

The users of the social network tend to generate a large amount of information, which could help in generating useful results by the use of highly accurate recommender systems. A recent study has proposed a framework that utilizes the social relation data to regularize user latent space by exploiting the friendship information. However, this proposed work doesn't consider the social contextual information to the fullest. It is a challenging task to discover and integrate the social contextual factors into a unified framework.

The Figure 1 describes how links from various social networks can be used to derive entire social contextual information. A novel framework understands the mechanism of user behavior on social networks and summarizes into two factors.

The two factors summarized here are:

- Individual preference.
- Interpersonal influence.

The decision of users on information adoption is affected by the individual preference and interpersonal influence. It is noticed that individuals are influenced by others behaviour to some extent, rather than taking independent decisions, i.e. purely preference driven. The present work here demonstrates on how the incorporation of the preference of individuals and interpersonal influence into recommendation could be reduced and quality improved.

The usage of a particle filter extracts the necessary keywords from tweets by the use of Stemming along with the location and time. When the system encounters key-

*Author for correspondence

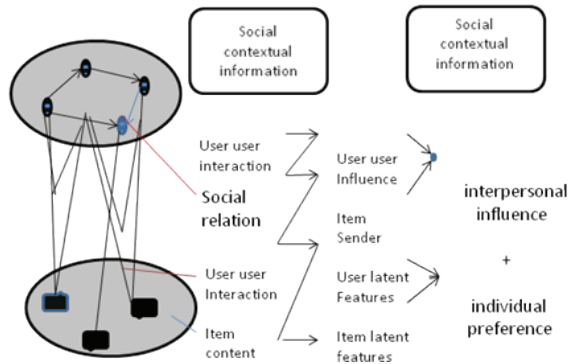


Figure 1. Derivation of social contextual information.

words related to natural disasters, an auto alert is sent to the people in the nearby locations and the rescue teams based on a proper verification algorithm.

The study is based on the social networking site, Twitter to consider the two dataset representations: one considering the two directional social relations, and the other considering the one directional social relation. It is seen that the performance of the recommender system can be greatly boosted by the mentioned contextual factors.

The paper has been organized into various sections considering the domain specification, applications of the existing system with their disadvantages, the proposed methodologies with their advantages over the existing system and the module descriptions supported by the architecture diagrams.

2. Domain Specification – Data Mining

Data mining is often known as data discovery. It is a process of summarizing it into some useful information after analyzing the data from various perspectives. The summarized information can be used for various purposes like increasing revenues, cutting costs and many more.

Our study demonstrated here deals with the purpose of summarizing the disaster related information from various social networks into an emergency alert system to be used in case of natural disasters at any location. The software used for mining data is one of the most important analytical tools among the various analytical tools available for various purposes. This software effectively allows users to analyze data from multiple angles and

dimensions and further categorizing them to have a useful summary of the relationships identified. Technically, data mining is defined as the process of finding patterns and correlations or patterns from various and large relational databases.

3. Technology used – Hadoop Technology

Hadoop technology is an open-source software framework developed by the apache software foundation. It provides a set of algorithms for storage and processing of distributed data sets. The Hadoop comprises of a storage part named as Hadoop Distributed File System (HDFS) along with a processing part called - map reduce. The mapreduce functionality is used for large scale data processing. In our paper this function is used to find the exact location of any disaster and then reducing the complexity of the problem. Initially the mapreduce function searches of all the locations for a particular disaster and then further reduces the complexity of the large amounts of data involved by specific identification of the location with the corresponding disaster identified. So, basically the Hadoop technology helps us in managing the large amounts of data and then deriving a solution by effective and accurate processing methods.

According to the Existing System, there is no proper automatic alert system implemented to report people during natural disasters. Hence, there is no way to take immediate actions for rescuing people. The only existing methodology of reporting people is the media, i.e. news, radio etc. Such methodologies are time consuming and fail to report people immediately, resulting in human and economic losses.

In the social recommendation problem is investigated based on sociology and psychology studies, which further present the importance of various factors for prediction of online behavior^{1,2}.

The concept in^{3,4} explains the revolution in micro messaging with special attention to Twitter and its various related features like tweets, followers etc. respectively.

The studies reveal the information related to the need of twitter and other micro blogging sites, further enabling us to cite event detection as a critical need of twitter⁵.

The concept cited in helps us in knowing the concept of twitter as a social network in detail. It clearly states the difference between a social network and social media. Having concluded it as a social network, we use

this concept to accumulate data from various users in the form of tweets⁶.

The reference is the key to detecting events in our paper. Its concept has already succeeded in predicting key events like elections, which further helps us in predicting events like disasters⁷.

The existing process is very time consuming as it does not provide any methods of fast communication in case of emergency. Delayed communication leads to various physical and economic losses.

4. Proposed System

We are implementing a particle filter mechanism which is used to extract the keywords from the tweets using Stemming Algorithm along with location and time of the tweets. By using this information, the system can analyze the peaks of the keywords like “Earthquake”, “Tsunami” etc. at a specific time and location. Based on the authentication of tweets the peak will trigger automatic alerts to people who are present in the nearby locations and the rescue organizations, which can carry out the rescue process as soon as possible. This would help in saving lives and prevent further losses. The alerts are sent via SMS and Emails to the registered tweet users and the nearby rescue organizations.

4.1 Advantages

- The system proposed by us helps in saving time as the automatic alert is sent to the registered users present in locations close proximity of the disaster affected areas.
- It increases the probability of saving various losses by alerting the rescue teams immediately and enabling them to conduct rescue operations immediately.

4.2 Architecture Diagram

Figure 2 explains the proposed system:

Initially, the system receives multiple tweets with specific keywords related to disasters. On performing the process of extracting the required keywords using the stemming algorithm, the selected words are stored in two classes based on valid criteria. The required keywords are stored in the positive class and processed further while the unnecessary keywords are stored in the negative class.

The important keywords from the positive class are further processed to detect the event type and the loca-

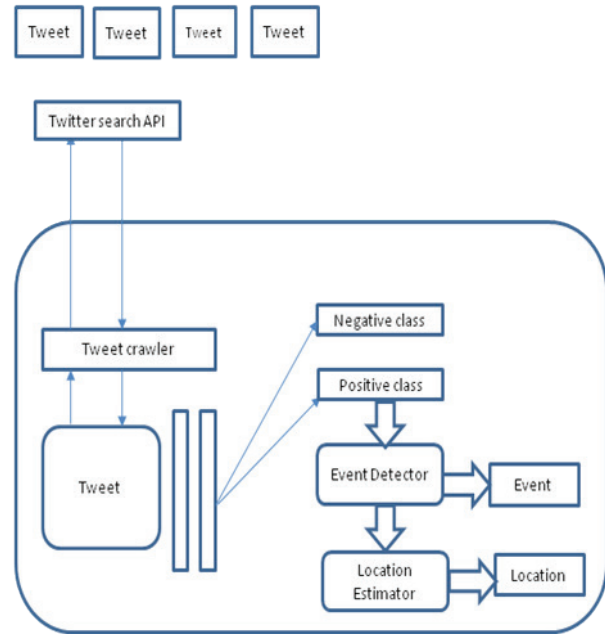


Figure 2. Architecture of disaster management system.

tion of its occurrence. The system is set to have three features namely:

- Counting the number of words from the tweets.
- Displaying the individual words.
- Identifying the vulnerable and location keywords.

The example shown below explains the functionality of the system using the features specified:

“I am in Chennai earthquake right now”

The above tweet will be processed by the 3 features as:

- Seven words.
- Chennai, earthquake.

This procedure helps in accurate identification of the disasters and their corresponding location.

5. Module Description

The methodology used by us has been categorized into the following four modules for effective functioning:

5.1 Application Creation

This part deals with creating an application to tweet our friends and followers on Twitter. For creating the

application, we will be using Advanced Java Concepts like JSP and Servlets. For the application to be user friendly and secure, we design fields like Username, Password, Phone and other personal information. For creation of the application, the user is allowed to enter the data which will be stored in the server. On validation of the authenticity user, the user is allowed to enter the application. The user is allowed to tweet using the same application.

5.2 Database Setup

By a database, we refer to a location for storing all the contents required for the effective implementation of any work or project. We use MySQL as the software for designing and maintaining the database. The database of this project deals with the storage and retrieval of various entities like registration details, login details, tweets, followers, etc. The Server will analyze the contents of the user in order to extract the required keywords. Also the Server will retrieve the user information like access time and location which would be used to find the user's location to provide any necessary help when needed.

5.3 Extraction of Keywords

The purpose of a server is to analyze the tweets between the users and then extract the keywords using particle filters. The particle filter extracts the keywords and filters the other words, using the Stemming Algorithm. By using this algorithm we can filter the unwanted words from the tweets in order to calculate the extracted keyword counts. When the count value of the keywords reaches the peak value (set by default), an automatic alert message is sent via SMS and Emails to the registered users and the rescue organizations.

5.4 Invocation of Automatic Alert System

This part of our work is responsible for sending alert messages via SMS and Emails to the registered users and the rescue organizations, once the keyword count reaches the peak value. To generate an SMS alert, we include the Java archive file called "JSMS". The contact information of the rescue organizations will be specified in the codes being used. To implement this concept on a small scale, we will connect a PC suite configured mobile device via a data cable. This PC suite configured mobile will transmit the SMS to the rescue organizations and registered users.

6. Conclusion and Future Enhancement

Our paper has successfully filtered the essential tweets posted in our application by comparing them with various other tweets on the social networking site, twitter and also by comparing the number of tweets with a fixed count for proper authentication. By means of these tweets, we collect the required keywords related to any disasters and store them in a database. Further the collected information is used to generate and spread alerts to the rescue agencies and the users registered with our application using the Hadoop technology which enables management of big data over widespread networks. These alerts can help in saving lives and other entities on a short notice as compared to other media of communication which require more time for necessary actions. Hence, our application can be effectively used for generating emergency alerts in case of any disasters and other natural calamities.

7. References

1. Dhamodaran S, Kunapareddy S. Density dynamics using GIS. *Int J Appl Eng Res.* 2015; 10(5):12229–36.
2. Dhamodaran S, Raja A. Prediction of weather condition using probability function. *Int J Appl Eng Res.* 2015; 10(4):10665–70.
3. Surendren SK, Dhamodaran S. Multi-resolution pruning based co-location identification in spatial data. *IOSR-JCE.* 2014 Mar-Apr; 16(2):1–5.
4. Java A, Song X, Finin T, Tseng B. Why we Twitter: Understanding microblogging usage and communities. *Proceedings of 9th WebKDD and First SNA-KDD Workshop Web Mining and Social Network Analysis (WebKDD/SNA-KDD);* 2007. p. 56–65.
5. Huberman B, Romero D, Wu F. Social networks that matter: Twitter under the microscope. 2008 Dec. Available from: <http://arxiv.org/abs/0812.1045>
6. Kwak H, Lee C, Park H, Moon S. What is Twitter, a social network or a news media? *Proceedings of 19th International Conference on World Wide Web;* 2010. p. 591–600.
7. Tumasjan A, Sprenger TO, Sandner PG, Welpe IM. Predicting elections with Twitter: What 140 characters reveal about political sentiment. *Proceedings of 4th International AAAI Conference on Weblogs and Social Media (ICWSM);* 2010. p. 178–85.