Social media mining for post-disaster management - A case study on *Twitter* and news

*K. Banujan[1], B. T. G. S. Kumara[1] and Incheon Paik[2]

[1]*Department of Computing & Information Systems, Faculty of Applied Sciences,*
Sabaragamuwa University of Sri Lanka, Sri Lanka
[2]*School of Computer Science and Engineering,*
University of Aizu, Aizu-Wakamatsu, Fukushima, Japan
*bhakuha@gmail.com

Abstract

A natural disaster is a natural event which can cause damage to both lives and properties. Social media are capable of sharing information on a real-time basis. Post disaster management can be improved to a great extent if we mine the social media properly. After identifying the need and the possibility of solving that through social media, we chose *Twitter* to mine and News for validating the *Twitter* Posts. As a first stage, we fetch the *Twitter* posts and news posts from *Twitter* API and News API respectively, using predefined keywords relating to the disaster. Those posts were cleaned and the noise was reduced at the second stage. Then in the third stage, we get the disaster type and geolocation of the posts by using Named Entity Recognizer library API. As a final stage, we compared the *Twitter* datum with news datum to give the rating for the trueness of each *Twitter* post. Final integrated results show that the 80% of the Twitter posts obtained the rating of "3" and 15% obtained the rating of "2". We believe that by using our model we can alert the organizations to do their disaster management activities. Our future development consists mainly of two folds. Firstly, we are planning to integrate the other social media to fetch the data, i.e. Instagram, YouTube, etc. Secondly, we are planning to integrate the weather data into the system in order to improve the precision and accuracy for finding the trueness of the disaster and location.

Keywords: API, Data mining, Disaster management, Social media, Twitter

Introduction

A natural disaster is an event such as a flood, landslide, earthquake, tsunami or hurricane which is capable of causing great damage to properties or loss of lives. The loss of lives and property damage will increases globally (Alexander, 1993). Even though some disasters can be predicted earlier it very hard to stop the impact fully. Therefore, post-disaster management plays an important role to get back to the routine life. Data Mining is a process of extracting knowledge hidden, from large volumes of raw data. The knowledge must be able to use for various purposes. According to Frawley et al. (1992), data mining is defined as "nontrivial extraction of implicit, previously unknown, and potentially useful information from data". Data mining is one of the tasks in the process of knowledge discovery from the raw data.

Twitter is a promising channel through which we can explode to explore for situation awareness because of the availability of open Application Programming Interfaces (APIs) that give access to almost all of the communications in real time. Unlike traditional communication technologies such as cell phones, *Twitter* communication is largely public and can be monitored, and members of the disaster-affected population can be employed as a sensor network (Rogstadius et al. 2013). Further, during disasters, all the conventional communications generally stop functioning. At