Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 115 (2017) 350–358

Procedia Computer Science

www.elsevier.com/locate/procedia

7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

Usage and analysis of Twitter during 2015 Chennai flood towards disaster management

Meera.R.Nair[1], G.R.Ramya[1,*], P.Bagavathi Sivakumar[1]

[a]*Dept. of Computer Science and Engineering*
Amrita School of Engineering, Coimbatore
Amrita Vishwa Vidyapeetham
Amrita University, India

Abstract

Social media plays a major role in the propagation of information during disasters. This paper mainly contains a study regarding how people of Chennai used social media especially twitter, in response to the country's worst flood that had occurred recently. The tweets collected were analysed using machine learning algorithms such as Random Forests, Naive Bayes and Decision Tree. By comparing the performances of all the three it was found that Random Forests is the best algorithm that can be relied on, during a disaster. This paper also targeted the sources of the Twitter messages to explore the most influential users of Chennai flood.

Peer-review under responsibility of the Organizing Committee of ICACC 2017.

Keywords: Twitter Analysis ; Random Forests ; Naive Bayes ; Decision tree ; Influential users ;

1. INTRODUCTION

Communication means sharing or exchange of thoughts, messages, visuals or signals. It requires a sender and a receiver. Traditional communication techniques are uncomplicated and are elementary in nature. However, these methodologies were limited to speed and distance. This was overwhelmed by modern communication technologies. Social media or social networking sites play a major role in the modern communication techniques [**?**]. These techniques seem to be much faster and are used more frequently than traditional technologies. Modern technology succeeded in creating a more global society by enabling the people to communicate or interact directly with one another from all over the world. Popular social media includes Twitter, Facebook, LinkedIn, blogs, YouTube, wikis etc. They provide a platform for the users to share or post their views, information etc. In this modern era Facebook is the biggest social medium network on the internet in terms of popularity. It is one of the best tool for connecting people all over the world. Twitter is a real time microblogging network. LinkedIn is a network that is dedicated to professionals. YouTube provides a platform for uploading or sharing videos and music among friends and family.

* Corresponding author. Tel.:+91 04222685000
E-mail address: gr_ramya@cb.amrita.edu

1877-0509
Peer-review under responsibility of the scientific committee of the 7th International Conference on Advances in Computing & Communications
10.1016/j.procs.2017.09.089

For the past few years, Twitter is one of the standard medium for news and communication. Twitter, a micro-blogging service in which registered users can post messages called tweets [**?**]. Registered users can broadcast tweets, follow other users tweets etc. Twitter messages are only 140 characters long and they are called tweets. Tweets can be published from multiple platforms and devices. The advantage of Twitter is that anyone can follow anyone on public Twitter. Tweets are delivered to users in real time. In order to connect to a general topic, users can add hash tag as keywords to their posts. Hash tag is a meta character which is expressed as #keyword or #hash tag. Hash tag helps people to pursue their interested topics very easily and quickly. Hash tag will provide the tweets related to a common or particular topic. For eg the keyword #flood will retrieve all the tweets that contain the keyword flood.

During the time of natural calamities or natural disasters like earthquake, flood or hurricane, where all the wired connections have been universally down, these social media tools appear to be more beneficial[**?**]. Twitter plays a major role in the rapid propagation of information during disasters [**? ?**]. It allows accessing or dispersing crucial information or breaking news directly from the affected areas.

Social media analysis is the process of collecting information or data from social media sites such as Twitter, Facebook, LinkedIn, YouTube etc and an analysis is done to get useful or meaningful results from it. Twitter analysis includes scrutinizing the tweets of a particular event or matter [**?**]. Several machine learning algorithms such as Decision Tree, Support Vector Machine, Random Forests, Naive Bayes, Logistic Regression etc can be applied to the data for analysis. These algorithms help in obtaining useful outcomes from the data and also help in visualising the data in a more precise manner.

For the past few years several analysis of tweets are being carried out on the disaster affected areas across different countries, from this, it is evident that, tweets generated during the time of natural disasters are considered, predominantly, to get vital information regarding the affected areas[**?**]. This work focused on tweets regarding the 2015 Chennai Flood. The tweets have been analysed in order to get meaningful knowledge from it. It was found that the tweets obtained during the time of flood mainly fall under five categories. It includes Need for help, Relief Measures, Express Gratitude, Complaints and Other. Tweets collected were given as input to Weka. Weka is a free data mining analytical tool[**?**]. The tweets were classified using three machine learning algorithms such as Random Forests, Decision trees and Naive Bayes. Performance was carried out in terms of Precision, Recall and F-measure. It also targeted the sources of the Twitter messages to explore the most influential users. Top ten influential users of the dataset were identified on the basis of the count of tweets, retweets and the followers that each user has. The rest of the paper is structured as follows. Section 2 focuses on Literature Survey and Related works. A background study of Chennai flood is presented in Section 3. Methodology of machine learning algorithms such as Random Forests, Naive Bayes and Decision tree is discussed in Section 4. It also includes the analysis of the tweets generated from Twitter. Section 5 contains the results of analysis. This section mainly contains the performance comparison of all the three algorithms based on Precision, Recall and F-measure. It also identifies the most influential users of Chennai flood on the basis of count of tweets, retweets and the followers that each user has. Conclusion is given in Section 6.

2. LITERATURE SURVEY AND RELATED WORKS

Twitter plays a major role in the dissemination of information during natural disasters. The various techniques that have been applied to extract data from Twitter are described below.

The work of Starbird et.al reveals the use of Twitter during the Red River Flood that occurred in the Red River valley in central North America[**?**]. They have collected all the tweets regarding the Red River Flood using the keyword #redriver. Each tweet contains attributes such as originator of the tweet, retweet, geographical location, affiliation and approximate distance from the event. Tweets analysed fall under one of the following category which includes hopeful, fear, support, and humor.

The work of Aisha et.al[**?**] mainly focused on the 2014 Malaysia Flood, that had devastated Malaysia fully. It was the worst flood they had witnessed in the recent decades. The number of Twitter messages during the flood period was higher when compared to the normal period. Twitter messages were analysed using a 5-point Likert scale, were 1 corresponds to never and 5 corresponds to always. This mainly focuses on the satisfaction gained by the Twitter user from sharing of information during the flood.

Case Study of 2011 Thai Flood [**?**] studied the use of Twitter during the Thai Flood. Twitter played a major role in providing first-hand information during this disaster. #thaiflood tweets were collected and analysed using the keyword