MATTER: International Journal of Science and Technology
ISSN 2454-5880

Catanghal Jr et al., 2017

Volume 3 Issue 1, pp. 59-69

Date of Publication: 15th March, 2017

DOI- https://dx.doi.org/10.20319/Mijst.2017.31.5969

This paper can be cited as: Catanghal Jr. R. A., Palaoag T. D., & Malicdem A. R. (2017). Crowdsourcing Approach for Disaster Response Assessment. MATTER: International Journal of Science and Technology, 3(1), 59-69.

CROWDSOURCING APPROACH FOR DISASTER RESPONSE ASSESSMENT

Ricardo A. Catanghal Jr
College of Computer Studies, University of Antique, Sibalom, Antique, Philippines
catanghal.ric@gmail.com

Thelma D. Palaoag
College of Information Technology and Computer Science, University of the Cordilleras, Baguio City, Philippines
tpalaoag@gmail.com

Alvin R. Malicdem
Don Mariano Marcos Memorial State University, City of San Fernando, La Union, California, US
alvin.malicdem@dmmmsu-mlu.edu.ph

Abstract

Philippines are a country attuned to social media and a disaster prone country and recent research focused on the interesting use of Twitter. This work is motivated to provide information through crowdsourcing, which uses humans as sensors to observe and report events in the physical world. In this paper we propose that, Twitter feeds which consist of short messages to extract information as a tool in needs assessment for a disaster hit community. This information will serve as situation awareness through crowd sensing, in order to deliver the relevant basic needs to the disaster stricken community and humanitarian disaster response. The data were

obtain using the Twitters open search API, preliminary experiment is carried out, Naïve Bayes algorithm was used to classify disaster related tweets. The geo location feature in the tweets were extracted and translated into map for visualization and the information related to disaster. This study will be helpful in identifying, analyzing, monitoring and evaluating basic needs of the affected communities, in order for the decision makers to take necessary actions and respond to the needs of the people.

Keywords

Disaster Management, Data Analytics, Crowdsourcing, Data Mining

1. Introduction

The Philippines is a country that is much attuned to social media, is even nicknamed as the social networking capital of the world. It is also the country with the highest social networking penetration in the Asia-Pacific region. In fact, in the Philippines, even the government agencies have Twitter accounts to make dissemination of advisories or warnings to Filipinos faster and easier, such as the Department of Science and Technology (DOST)s Twitter account @dost_pagasa which tweets weather forecasts and updates (Beduya & Espinosa, 2014).

Philippine is prone to disasters, every year we are visited by around 19 tropical cyclones which enters our Philippine Area of Responsibility (PAR), and six to nine of it make landfall. In a report by German watch for the Global Climate Risk Index 2016, the Philippine is one the most affected extreme weather events, from the fifth place we are now on the fourth following Haiti, affecting millions of people (Kreft et al, 2011). The Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) have officially adopted the super typhoon (STY) category to classify a tropical cyclone, to adapt to this extreme weather change (Daluz, 2015).

Information for decision making even in the aftermath of every major disasters, whether it be man-made or by nature, for example the Typhoon Yolanda 2013, is very crucial for every decision makers and emergency response agencies. It is one of the key challenges that they may face, to collect relevant information for analysis. In emergencies it is vital to gain situational awareness what is happening in the ground in order to determine appropriately what is to be delivered (Ramchurn, Fischer & Simpson, 2015).

2. Related Work

2.1 Social Media

Twitter is one of the prevalent micro-blogging platforms today, as describe by Figure 1 (Statistica, 2016). One its major strength is its ability to inform users about what's happening in the world in real-time, be it news of natural disasters, breaking political stories, or the latest sports scores (Bahmani, Chowdhury, Goel, 2010). There are several ways Twitter message can be posted such as; emails, phones, other third party applications. Twitters popularity is attributed to its easiness of usage and portability among many others (Atefeh & Khreich, 2015).

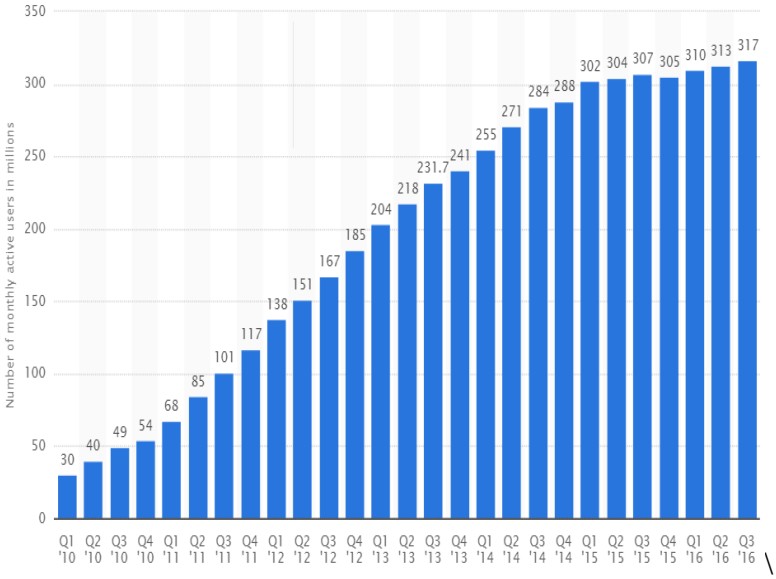

Figure 1: *Twitter Users Trend*

Twitter has altered the news by becoming the microphone of the masses (Murphy, 2011). Many examples in the real world have shown the effectiveness and the timely information during disasters. Some of its examples include the 2010 devastating earthquake in Haiti (Muralidharan, Rasmussen, Patterson & Shin, 2011), the real time earthquake detection in Japan, the detection of the outbreak of influenza (Aramaki, Maskawa, & Morita, 2011).

2.2 Crowdsourcing

Crowdsourcing has emerged as a new application paradigm for sensing the physical environment by using human as sensors (Wang, Marshall and Huang, 2016). This paradigm is

motivated by the wide spread use of digital sensors, ubiquitous wireless connectivity and massive data dissemination opportunities (e.g., Twitter) (Wang, Abdelzaher & Kaplan, 2015). Twitter as a micro blog one of its characteristics is real time in nature. It has various usages in real-time updates and notification such as those needing for help during a large-scale fire emergency and live traffic updates (Sakaki, Okazaki & Matsuo, 2010). For example, survivors may tweet to document the damage and outage in the aftermath of a disaster or emergency event. Citizens may report geotagged photos to document the potholes on city streets or regarding the status of the electricity.

Several studies demonstrated the use of Twitter for crowdsourcing or as a source of data for solving issues. Presenting several example, a study conducted by Aramaki et al in which they detect influenza epidemic by using Support Vector Machines (SVM) classifier on the collected tweets throughout a certain period of time (Ramchurn et al, 2015). Then another study by Sakaki et al. in which they predict earthquake in Japan by performing semantic analysis on real-time tweets (Sakaki et al, 2010). Another study done by Zin et al. in which they used knowledge-based approach to data from Youtube and Twitter for situation awareness in disasters specified such as earthquake and tsunami (Zin, Tin, Hama & Toriu, 2013). Osborne et al also used Twitter and Wikipedia for event detection via first story detection (Velev & Zlateva, 2012).

In crowdsourcing, each Twitter user is assumed as a sensor, the tweet that they make will be the sensory value information. This social sensor has a lot of noise or noisier compare to the physical sensors, such as heat sensors, kinetic sensors, light sensors. The user sensors are called virtual sensors, they have a different characteristics and a wide variety of characteristics: they are online most of the time, often others are seldom active. This virtual sensor are not functioning sometimes, they may be out of coverage or sleeping (Sakaki et al, 2010).

2.3 Tweets and Disaster

Twitter was originally used for social networking and interaction, it was not intended for use during the disaster response and emergencies, but the public and institutions are increasingly turning their attention for dissemination and gathering of vital information. The growing empirical work on Twitters functions in relation to disasters shows that it is a valuable channel of information for both official institutional sources such as government agencies and news outfits,

and witnesses on the ground close to the event who are able to post updates as texts and photos (David, Ong & Legara, 2016).

Its use has been studied in relation to the Red River Valley flood threat (Palen, Starbird, Vieweg & Hughes, 2010), the Pakistan floods (Murphy, 2011), the detection of the outbreak of influenza (Aramaki, 2011), the Australian floods (Cheong & Cheong 2011) and several others. Twitter contains features that are particularly useful for disaster reporting and monitoring: real-time posting, short-burst message style (140 characters), default public settings when posting, and easy retweeting or forwarding of messages to followers and the public Twitter space (David, Ong & Legara, 2016).

A number of studies have closely examined various aspects of Twitters use as an information source and channel for dissemination during disasters. In addition to being a place to find and share information, people also use it to share personal experiences, crack jokes, express concern, and ask questions. During emergencies Twitter functions as a backchannel source of information, a venue for information sourcing to enhance situational awareness, and as a channel for sharing opinions and experiences (David, Ong & Legara, 2016).

3. METHODOLOGY

3.1 Data Gathering

For this initial study, tweets will be gathered from twitter users about typhoon via Twitter Search API provided by Twitter in order to get public tweets that correspond to the given parameters. Specifically tweets regarding the Typhoon Melor (locally known Nona), from December 14 to 15, 2015 were gathered and used as sample. In gathering the said tweets, they were filtered to include messages that contain the following hashtags or keywords: Melor, nonaph, Nona, Typhoon Nona, Typhoon Melor, and TyphoonPH using a web scrapper. The Philippine government official hash tags and common hash tags were included. The local and the international name for the storm were included. In total, we collected 43, 648 tweets using the process.

3.2 Data Processing

For data cleaning, the proponents first converted the string tweets into lower case for easier comparison and then removed the stop-words, URL, hash symbol (#), emoticons, RTs,

mentions (@mention), expressions and other special/miscellaneous characters present in each tweet. This process was performed in order to produce a more compact "dictionary" which will help decrease the dimension of the data set that is going to be used. Filtering was done in order to remove repeated tweets or retweets, and filtering by length of 3. This will give us the unique tweets. After the process, 5780 unique values were returned.

For this initial study we categorize our outputs into 5: evacuation, roads and bridges, electricity, food and water, medical. The classification was done using the Naïve Bayes, several study demonstrated its use and performance in classification during disaster.

3.3 Location-Referencing

Geo-location defined (Vieweg, Hughes, Starbird & Palen, 2010) as clearly identifiable information which includes addresses of the streets and even its intersections, the city names, the province or regions or towns, roads and highways and name-places: such as establishments, universities, popular place, etc. Those tweets that include information regarding the whereabouts or hint of humans, flood, evacuees and the evacuation sites (among others) can help those who are relying on those information in assessing their personal situations and needs, this could also gain a much broader understanding of the situation as a whole. Location-referencing refers to "information that uses one place as a reference for another or the mention of location via a landmark, i.e. 'x miles from y,' where the reference point is ambiguous without knowledge of situational context" (Vieweg, Hughes, Starbird & Palen, 2010). For example, regarding power supply tweet sample we read:

No electricity Virac turned into a Ghost Town again.

Or from an evacuation tweets:

143,223 individuals (29,015 families) evacuated in Sorsogon

The Google Maps Javascript API V3 was used which is available in the Google Maps API family site to generate visual distributions on the map. It is assumed that the locations inputted by the user would not be the exact latitude and longitude coordinates of the place but the address of the place. Google Maps provided a way to convert the address string to geographic coordinates which is the latitude and longitude pair. That process of converting is called Geocoding.

Map Creation was done by initialization of the map. It used a fix center (latitude, longitude coordinates) and it is where the map will first project. It is set to a fixed width and height. Also, specific map options are specified like the zoom value, and the map type. These data were displayed in an Info Window (overlay that is like a speech bubble). These Info Windows show whenever Events are fired.

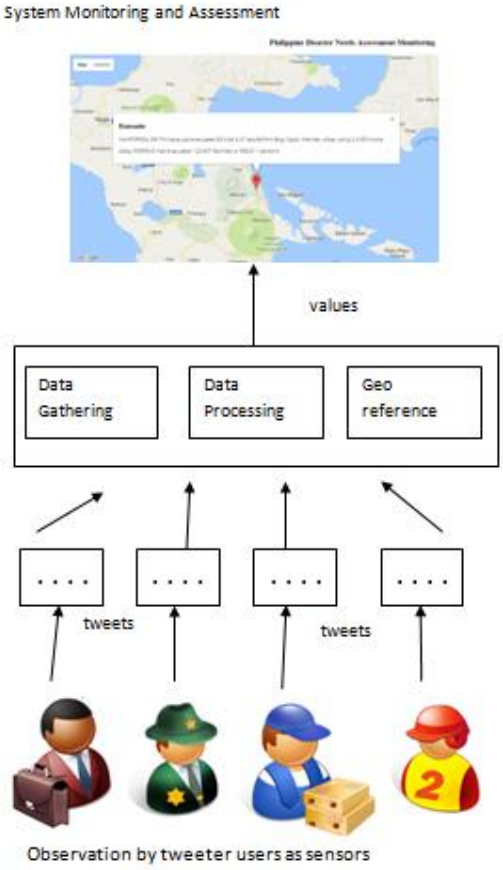

Figure 2: *Process for Crowdsourcing*

4. Discussion and Conclusion

This research try to extract information from social media specifically Twitter in order to identify, analyze, monitor and evaluate basic needs of the affected communities, in order for the decision makers to take necessary actions and respond to the needs of the people.

The proponents in this initial experimental set up gathered the tweets as a sensor values. The cleaning and filtering showed that out of 43, 648 tweets that were scrap using the Twitter Search API, 86.76% of it were noise or duplicates and only 13.24% are unique vales. It is

interesting that the word "evacuate" is among the top 10 words frequently mentioned in those unique documents, associated with some places. The top mentioned places mentioned in the tweets were the places where the path of the typhoon.

Table 1: *Top four places mentioned during the Typhoon Melor*

Word	Total Occurrence	Document Occurrence
Sorsogon	179	171
Samar	169	150
Romblon	136	100
Mindoro	131	123

Since this study is still ongoing, no definitive conclusion can be drawn yet. However, the proponents were able to establish that it is indeed possible to use crowdsourcing to identify disaster-related tweets in Twitter and plot them to Google Maps with marker and information as a monitoring and assessment tool. Crowdsourcing through social media is a useful tool in gathering data since almost everyone has access to the Internet and can supply information wherever they may be.

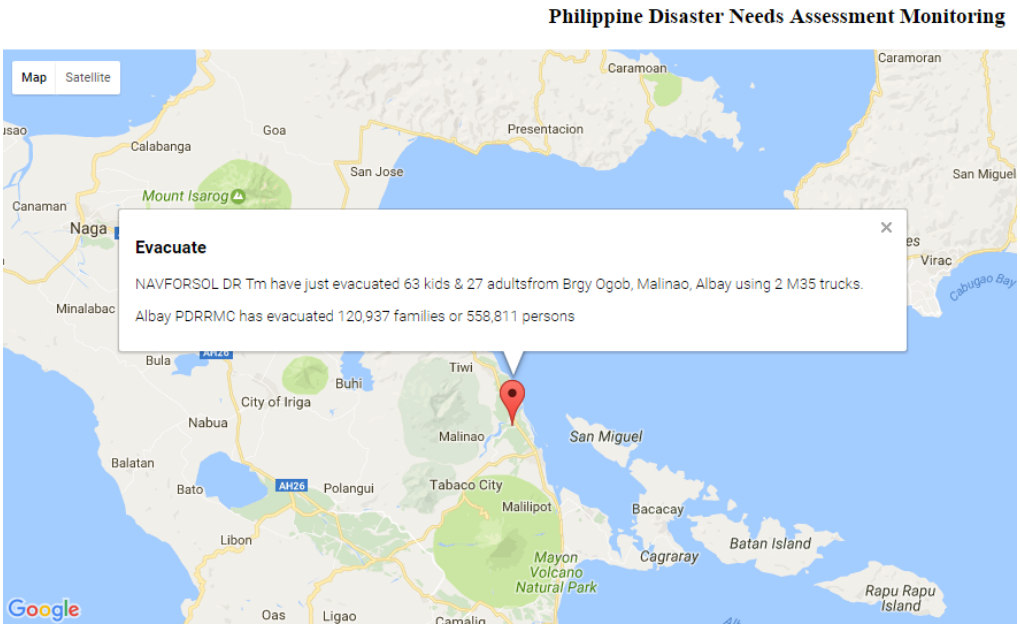

Figure 3: *Sample Generated screen shoots*

Future works of this study includes generating customized marker to suite or differentiate the type of disasters and the type of markers (for disaster or for class suspension). This will be a good visualization for monitoring and assessment of different area under disaster. Then a database will be setup to be linked to the Google Maps, then retrieve it and pin a marker. The system will also provide a mechanism to verify user post before saving in the database. It should also check if the place entered by the user is already in the database. Instead of pinning too many markers, the Info Window of the marker having the same location should accommodate the updates.

The use of some classification algorithm in classifying the disaster related tweets, by training it with a dataset. Then investigate how well they behave and how accurate in classifying the disaster related tweets would be interesting future topics.

References

Aggarwal, C., & Abdelzaher, T. (2013). Social sensing. In Managing and Mining Sensor Data, 237-297.

Aramaki, E., Maskawa, S., & Morita, M. (2011). Twitter catches the flu: Detecting influenza epidemics using twitter. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 1568-1576.

Atefeh, F., & Khreich, W. (2015). A Survey of Techniques for Event Detection in Twitter. Journal Computational Intelligence, 31(1), 132-164.

Beduya, L. J., & Espinosa, K. J. (2014). Flood-related disaster tweet classification using support vector machines. Proceedings of the 10th National Natural Language Processing Research Symposium, 76-81.

Cheong, F., & Cheong, C. (2011). Social media data mining: A social network analysis of tweets during the 2010 - 2011 Australian floods. Proceedings of the 2011 PACIS.

Daluz, C. (2015, April 2). PAGASA now has four cyclone categories. CNN Philippines. Retrieved from http://cnnphilippines.com/news/2015/03/24/pagasa-adopts-super-typhoon-category.html

David, C., Ong, J., & Legara, E. (2016). Tweeting Supertyphoon Haiyan: Evolving Functions of Twitter during and after a Disaster Even. PLOS ONE. http://dx.doi.org/10.1371/journal.pone.0150190

Gao, H., & Barbier, G. (n.d.). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. IEEE InTeLLIGenT SySTemS.

Kreft, S., Eckstein, D., Dorsch, L., & Fischer, L. (2016). "Global climate risk index 2016 who suffers most from extreme weather events? Weather related loss events in 2014 and 1995 to 2014 (Tech.). 72 D-10963 Berlin: Germanwatch, Stresemannstr.

Muralidharan, S., Rasmussen, L., Patterson, D., & Shin, J. (2011). Hope for Haiti: An analysis of facebook and twitter usage during the earthquake relief efforts. Public Relations Review, 37(2), 175-177.

Murphy, D. (2011). Twitter: Microphone for the masses? Media, Culture and Society, 779-789.Number of monthly active Twitter users worldwide from 1st quarter 2010 to 3rd quarter 2016 (in millions). (n.d.). Statistica. Retrieved from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Palen, L., Starbird, K., Vieweg, S., & Hughes, S. (2010). Twitter-based information distribution during the 2009 red river valley flood threat. Bulletin of the American Society for Information Science and Technology, 36(5), 13-17.

Ramchurn, S., Fischer, J., & Simpson, E. (2015). Hac-er: A disaster response system based on human-agent collectives. ," Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. Proceedings of International Conference on World Wide Web (WWW), 851-860.

Spiro, E., & Butts, C. (2013). When bad things happen to good people: Online communication dynamics during natural hazard events. Proceedings of Annual Meeting of the American Sociological Association.

Velev, D., & Zlateva, P. (2012). Use of social media in natural disaster management. International Proceedings of Economic Development and Research, 39, 41-45.

Vieweg, S., Hughes1, H., Starbird, K., & Palen, L. (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. CHI. Retrieved from http://cmci.colorado.edu/~palen/vieweg_1700_chi2010.pdf

Wang, D., Abdelzaher, T., & Kaplan, L. (2015). Social sensing: Building reliable systems on unreliable data.

Wang, D., Marshall, J., & Huang, C. (2016). Theme-Relevant Truth Discovery on Twitter: An Estimation Theoretic Approach. International AAAI Conference on Web and Social Media. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/ 13012

Zin, T., Tin, P., Hama, H., & Toriu, T. (2013). Knowledge based social network applications to disaster event analysis. Proceedings of the International MultiConference of Engineers and Computer Scientists, 1.