International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Situation Awareness Extraction and Rehabilitation System for Disaster Events

Neha Pandey[1], Dr. Natarajan S[2]

[1]M.Tech, Department of ISE, PES Institute of Technology, Karnataka, India
[2]Professor and Key Resource Person, Department of ISE, PES Institute of Technology, Karnataka, India

---

Abstract - *Millions of tweets are generated during a normal day. When a disaster occurs, the population residing in that place is adversely affected. When there is an emergency situation like a disaster enormous numbers of tweets are generated. Some tweets can provide useful information to the general public during the disaster event. This type of information is called situation awareness information. Extraction of situation awareness information from tweets is a challenging task since the vocabulary used by the users differs as well as the writing short hand notations reduces the readability of the tweets. Generally tweets collected are not in the form of a labelled data so implementing supervised learning approach can be tedious and time consuming task. Hence, we followed a method that uses the semi-supervised machine learning approach which can be used to obtain such useful information .This semi-supervised approach is based on the supervised learning algorithm i.e. support vector machine. Another aspect that has to be considered after a disaster strikes is the rehabilitation activity that focuses on bringing back the state of affected people to stability. Providing information about relief operation post disaster can help speed up rehabilitation activity for the disaster affected population. This can account to extraction of relief related urls that provides information such as volunteering, donation, relief camps etc. Also, use of crowdsourcing system to obtain relief related areas can help general public become aware of various locations where relief operations are carried out. An interactive map based on open street map can be used to develop a crowdsourcing system.*

Key Words: **disaster management, semi-supervised learning, situation awareness, crowdsourcing, interactive maps**

1. INTRODUCTION

Disaster, whenever strikes leave the affected population in an unstable situation. Since disasters are usually unanticipated events, they are unavoidable. Hence, preparedness strategies can become very important as they ensure to lower the panic among the general public [5]. The impact of a disaster is the extent to which it damages the lives of the population it affects. If the population is aware about the situation during a disaster, the impact of disaster can be reduced [9]. Data generated from twitter during a disaster can be very in analyzing the content posted by the users. Relevant information from twitter can be used in preparedness of a disaster [1]. Study shows that only 8% to 20% data from microblogs like twitter account to situation awareness data during disaster. This creates a challenging task of extraction of relevant information from twitter.

Crowdsourcing is a method of gathering information from the crowd on the internet. The authenticity of the information obtained by crowdsourcing cannot be known as it is accessed by general public. But during a disaster event it is assumed that the information generated from this is relevant to the emergency situation. Crowdsourcing mechanism can be used to develop a software system that can provide relief, response and volunteering related information [10].

The main aim of this research work is to develop a prototype system that can demonstrate the use of semi-supervised learning approach to extract relevant twitter data. The twitter data used in this study is related to Chennai Floods 2015 [13]. This system also integrates crowdsourcing technique [2] to create awareness during and post disaster using interactive maps and extraction of useful relief related urls about various efforts carried out by different government and NGOs.

2. RELATED WORK

A study was conducted to compare the bag of word model with proposed features used to extract situation awareness tweets using the supervised learning method i.e. support vector machine by Anirban Sen, Koustav Rudrat and Saptarshi Ghosh [3]. The results demonstrated that the proposed low level linguistic approach provided better results in-domain than cross-domain. Aibek Musaev, De Wang and Calton Pu developed a prototype system called LITMUS which was a landslide detection service developed using physical as well as social media data [7]. The system used the spatiotemporal feature of the data collected from these sources and resulted in detection of 25 out of 27 landslides reported by USGC in December 2013.

As the twitter data set is unlabeled using supervised learning method would not be a desirable approach. There was a study conducted by Isaac Triguero, Salvador Garcia and

Francisco Herrera [4] provided an intensive survey of self-labelling methods of semi-supervised learning to be able to identify apt technique while working with semi-supervised classification.

Mapbox developed a tool based on recent Chennai floods to locate flooded streets and locations on the map [12]. This was a crowdsourcing technique to collect threat locations during the floods. This was used to visualize the flood pattern in the city and the vulnerable areas that during the disaster. This tool was used by the general public to create awareness about the flood affected regions in the city.

Many other research and studies [6] are carried out in this domain and various tools have been developed to help the process of disaster preparedness, mitigation and response.

3. PROPOSED SYSTEM

In this paper, we utilize the ideas of the previously carried out work in the disaster management domain to build a system that can be used to provide relevant disaster related information through analyzing social media [11] content such as twitter. We also integrate this system with crowdsourcing technique by generating interactive maps that can be used by the general public to post disaster event.

3.1 Tweet Collection and Pre-processing

We collected around 2 lakh Chennai floods related tweets. The table below shows the popular hashtags used during the disaster to retrieve tweets.

Table -1: Hashtags used to collect tweets

Hashtags
#chennaifloods,#chennairainhelp,#chennairainsrelief,#chennaivolunteer,#donate Chennai, #support Chennai, #volunteer Chennai

Tweet collection was followed by pre-processing based on the subjective/ objective nature, numerals, and emoticons. Also, features from [3] were used to during pre-processing.

3.2 Classification of Twitter data

Tweet classification was based on semi-supervised learning algorithm. Since the tweets collected were not labelled, we initially labelled a small amount of tweets and rests of the tweets were kept unlabeled. Then an initial supervised learning using support vector machine [8] was performed on this dataset to generate initial labels. For the next iterations, the model used the resulting labels from the previous iteration as a part of training data to label the tweets. Until all the rows in the dataset were labeled, the self-training of the model was done. Once all the unlabeled data was labeled, the training of the model reached stopping condition.

3.3 Identification of Relief Locations and Crowdsourcing

Using the OSM (Open Street Map) as base layer, an interactive map was created to locate relief locations on the map and all the locations identified were stored on the google sheet that was published on the web for crowdsourcing. The algorithm used to create Crowdsourcing map is given below.

Table -2: Algorithm-Relief_Location_Crowdsourcing

```
use google sheet key to establish connection with map
 run the map session
 enters the relief related information
 if location marked on map
                reverse geocode longitude and latitude to address
                store address and user information on google sheet
                display relief location with information and address
                of marked location
 end if
```

3.4 Extraction of relief related URLs

Various NGOs and other agencies help the disaster affected public by collecting donation, food, clothing and other relief related materials from other people across the globe during a disaster event. Various websites that provide such information are tweeted in the twitter. Extraction of such urls can provide significant help to user that intent to contribute/ help the disaster affected victims. The algorithm below is used to extract urls from tweets.

Table -3: Algorithm-Extract_URLs_Hashtags

```
create a parser
 parse each line of the file
 for each parsed line
    if URLs exists
          then extract URLs and hashtags
    else
          read next line
 end for
```

4. RESULTS AND DISCUSSIONS

A 10-fold cross-validation was performed which resulted in an average accuracy of 90%. The result obtained by using this technique also had some miss-classified tweets like" real problem mentality people they always want require #chennaimicro #chennairains flood relief"," just pray chennai.. #chennairains pray chennai", etc. These tweets were classified as situation awareness tweets. The reason behind this was the initial training size i.e. the number of labelled data that was initially used was not very large. Hence, the model was able to self-train itself based on this initial training size. A better approach can be identifying other features for classification as well as providing an optimum initial labelled training size so that the model can correctly differentiate between situation awareness and non- situation awareness data. Thus, reducing the number of

false positive and false negative data. Figure 1 displays the cross-validation result and figure 2 shows the classified situation awareness twitter data.

```
> svm <- cross_validate(container,10,algorithm="SVM")
Fold 1 Out of Sample Accuracy = 0.9144737
Fold 2 Out of Sample Accuracy = 0.9007444
Fold 3 Out of Sample Accuracy = 0.9146067
Fold 4 Out of Sample Accuracy = 0.9262673
Fold 5 Out of Sample Accuracy = 0.9
Fold 6 Out of Sample Accuracy = 0.9176201
Fold 7 Out of Sample Accuracy = 0.9116162
Fold 8 Out of Sample Accuracy = 0.8974359
Fold 9 Out of Sample Accuracy = 0.8767507
Fold 10 Out of Sample Accuracy = 0.9159292
> svm
[[1]]
 [1] 0.9144737 0.9007444 0.9146067 0.9262673 0.9000000 0.9176201 0.9116162 0.8974359 0.8767507 0.9159292

$meanAccuracy
[1] 0.9075444
```

Fig -1: Cross-Validation Result

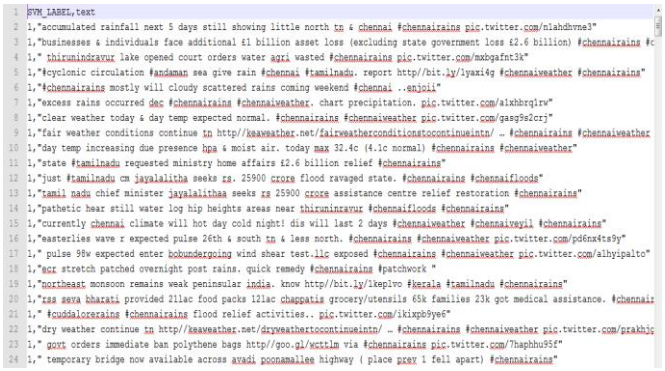

Fig -2: Extracted Situation Awareness Data

The relief location and crowdsourcing module was successfully able to locate user marked locations along with the information on the map and this data was stored on the google sheet that could be viewed by the general public to utilize this information for other use. Also, the urls along with their hashtags were extracted from the twitter data related to relief operations. Figure 3 displays the relief location on map, figure 4 shows the crowdsourced relief information on google sheet, figure 5-6 shows the extracted urls and hashtags from the relief related tweets and displays it on browser.

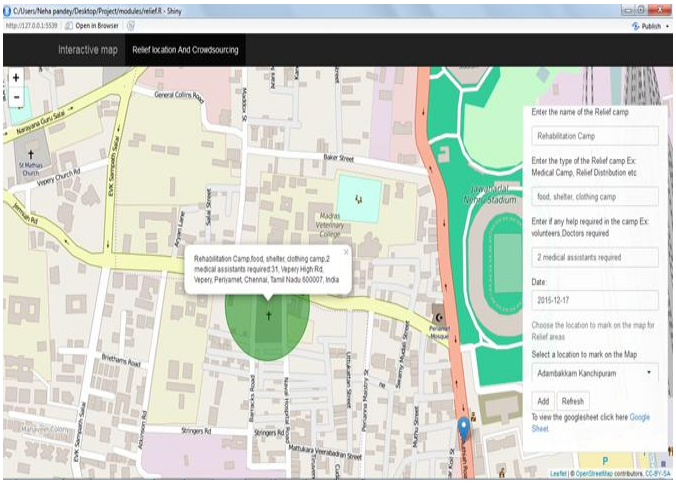

Fig -3: Visualizing Relief Location on Map

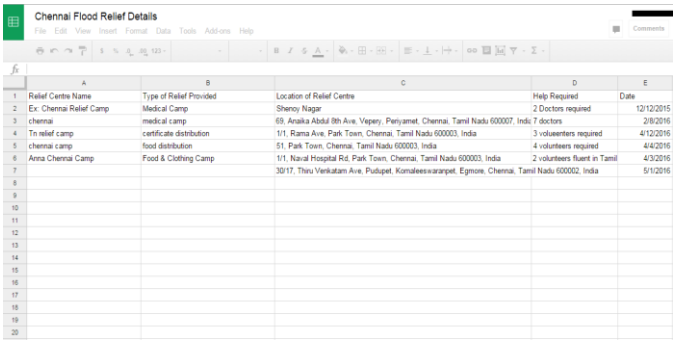

Fig -4: Crowdsourced data on Google Sheet

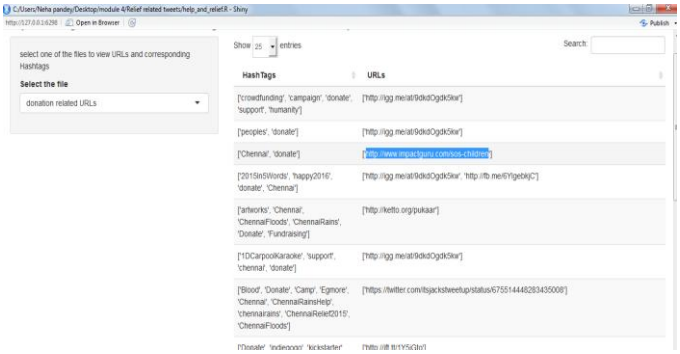

Fig -5: Extracted URLs and Hashtags

Fig -6: Selected URL displayed on browser

5. CONCLUSION

Disasters are unanticipated, unavoidable events. One way to handle this situation is to create awareness among people about the disaster incidents. Disaster preparedness, mitigation and response mechanism helps to speed up the process of stabilization of the affected area and population. Thus, a system that can help create such awareness can aid in the task of decision making during disaster.

This study provided a prototype system that can be used during the disaster event to provide situation awareness information to the public. The semi-supervised approach used in this study provided an average accuracy of 90% with classification of tweets as situation awareness and non-situation awareness. The system provided a crowdsourcing

solution that could be used during the relief and response operation and also a method to extract useful relief related urls links from the twitter data.

This study was carried out to provide a novel system that can be used by the general public during and post a disaster event. This study was based on the twitter data related Chennai floods 2015. Based on this prototype system, a real time system can be deployed during a disaster which can provide useful information to the users of the system.

REFERENCES

[1] Dennis Thom Robert Kruger Thomas Ertl, Ulrike Bechstedt Axel Platz, Julia Zisgen, Bernd Volland,"Can Twitter Really Save Your Life? A Case Study of Visual Social Media Analytics for Situation Awareness", IEEE Pacific Visualization Symposium 2015

[2] Jens Ortmann, Minu Limbu, Dong Wang1, and Tomi Kauppinen, "Crowdsourcing Linked Open Data for Disaster Management",Institute for Geoinformatics, University of Muenster, Germany,2011

[3] Anirban Sen, Koustav Rudrat and Saptarshi Ghosh, "Extracting Situational Awareness from Microblogs during Disaster Events", Social Networking Workshop, COMSNETS 2015

[4] Isaac Triguero, Salvador Garcia and Francisco Herrera,"Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study", Springer-Verlag London, 2013

[5] J. Rogstadius, M. Vukovic ,C. A. Teixeira ,V. Kostakos ,E. Karapanos and J. A. Laredo, "CrisisTracker: Crowdsourced social media curation for disaster awareness",IBM J. Res. & Dev. Vol. 57 No. 5 Paper 4 September/October 2013

[6] TweetComP Stuart E. Middleton, Andrea Zielinski, Ocal Necmioglu, and Martin Hammitzsch,"Spatio-Temporal Decision Support System for Natural Crisis Management", Springer International Publishing, Switzerland 2014

[7] Aibek Musaev, De Wang and Calton Pu, "LITMUS: a Multi-Service Composition System for Landslide Detection", IEEE Transactions on Services Computing 2013

[8] Thorsten Joachims,"Text Categorization with Support Vector Machines: Learning with Many Relevant Features", University at Dortmund Informatik LS-8, Baroper Str. 30144221Dortmund, Germany, April 1998

[9] Ken Moule,"Situation Awareness for Disaster Management in the Information Age", Global GBM, 2012

[10] Kate Starbird,"Digital Volunteerism During Disaster: Crowdsourcing Information Processing", May 7-12, 2011, Vancouver, BC, Canada,ACM 978-1-4503-0268-5/11/05, CHI 2011

[11] Jie Yin,Andrew Lampert, Mark Cameron, Bella Robinson, and Robert Power," Using Social Media to Enhance Emergency Situation Awareness", 1541-1672/12,IEEE Intelligent Systems, 2012

[12] https://www.mapbox.com/blog/chennai-flood-map/

[13] http://indianexpress.com/article/india/india-news-india/chennai-floods-rains-jayalalithaa-imd-reasons-rescue-news-updates/