Third International Conference on
CONTEMPORARY COMPUTING AND INFORMATICS

10 - 12 October 2018 | Delhi, India | Venue: Amity University Gurgaon

Finding out the needy one from Tweets : An analysis using #keralafloods

Navanath Saharia, Himangshu Sarma, Rahul Ranjan

Indian Institute of Information Technology Senapati (Manipur)

Outline

Introduction

Introductions

- It is very difficult to predict Natural disasters and when it happens its very difficult for government and different aid agencies to get the real time information about the disaster effected people or properties.

- This System helps analysing the situations during disasters such as Keralafloods and collects needed information through tweets as twitter is one of the best way of spreading awareness about the worsening situation in affected areas.

Related Work

- A SMS system designed for communication in between disaster effected people and aid agencies named as Trilogy Emergency Relief Application (TERA).

- NASA launched a device named as Finding Individuals for Disaster and Emergency Response (FINDER) to detect human's who are buried under buildings, roads etc.

Objectives

Objectives

- In this system, We used tweets of Kerala floods as an input to analyze the different scenario to help the people in the realtime.
- Using twitter data, We have analysis the whole crisis properly by processing the data.
- We have extracted the hashtages from tweets and applied clustering algorithms (e. g K-Means) to group similar messages and information like rescue,actions,supplies,emergency calls, together which took place on twitter.

Proposed System

Proposed System

- In our system, we used tweets of Kerala floods as an input to analyze the different scenario to help the people in the realtime.
- Using twitter data we have analysis the whole crisis properly by processing the data.

Proposed System (Contd.)

- We have extracted the hashtages from tweets and applied clustering algorithms (e. g K-Means) to group similar messages and information like rescue,actions,supplies,emergency calls, together which took place on twitter.

Implementations & Results

1. **Data**

   - The analysis is done from the tweets which are extracted during the early August 2018 when the flood happens. The analysis is restricted to 15000 tweets extracted by looking for the hashtag #keralaflood.

2. **Data Processing**

   - The hashtags were extracted from the tweets stored in (. csv) file and on proper analysis it gives various information about the situation in Kerala like weather forecasts (#rain), relief (#help), support(#chiefministerfund), location (#saveErnakulam) , etc.
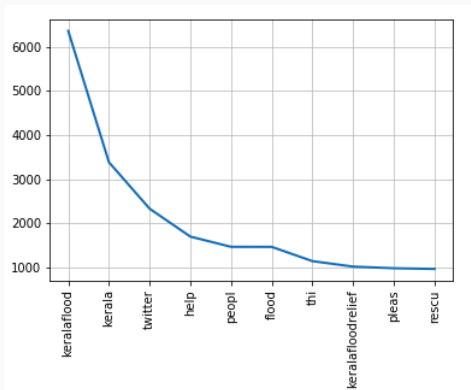
Implementations & Results (Contd.)

Processing of Data involves -

- Removal of punctuations.
- Removal of stop-words, emoticons, numbers and slangs.
- Removal of URLs.
- Extract meaning bearing words.

- After preprocessing, we have counted, which hashtag was used most frequently, counted its frequency and plotted using matplotlib (Figure 1).

Implementations & Results(Contd.)

- The TF-IDF vectorizer converts the tweets into vectors when each tweets are processed and appended to a list and this list was provided to the TF-IDF vectorizer.

- .Further we used K-Means clustering algorithm to group tweets into chosen number (say, here five) of groups. The Output on using five groups cluster we obtained is shown in Table I.

0	keralaflood, twitter, kerala, rescue, flood	115
1	help, donate, people, please, need	3925
2	chiefminister, minister, relief fund, crore	72
3	children, feet long bridge, bridge rescue, citizen malampuzha, senior citizen malampuzha	257
4	twitterkeralaflood, status, keralafloodrelief, keralaflood, kerala	1824

Table 1: Five different groups with their respective hashtags

3. Cluster-based & topic modeling

- K-means generates the following clusters for the K-value 5.
    - Cluster 0: Words: alert district rain affect these district
    - Cluster 1: Words: announce person maharashtra serious injury decrease
    - Cluster 2: Words: keralaflood kerala twitter help people
    - Cluster 3: Words: will keralaflood keralafloodrelief train kerala
    - Cluster 4: Words: chief minister chief minister announce announce crore

4. Topic Modeling Using LDA

- A topic model is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents and used for text-processing as it deduces the theme of texts .
- We implemented LDA to identify the topic of the tweets, as shown in Table II.

Topics	Words(Clustered After Topic Modeling)
0	keralaflood, kerala, nation, need, flood, relief
1	keralaflood, twitter, family, rescue, flood, people
2	keralaflood, twitter, kerala, rescue, water, operation
3	keralaflood, road, twitter, ernakulam, near, please
4	keralaflood, twitter, train, kerala, govern, keralafloodrelief

Table 2: LDA output for five different topics

- The words are associated with each and every topic and they represent certain meanings. Topic 0 is about Help, Topic 1 is about Action, Topic 2 is somewhat similar to Topic 1,Topic 3 is about Affected Location or Geography, Topic 4 is about Support and so on other can be identified , etc.

Conclusions & Future Works

Conclusions & Future Works

- It is clear from the above analysis that how powerful a social media is and social media can be harnessed to great effect in times of crisis.
- Some of the steps which are taken in this article are also adopted by twitter itself to help surrounding community in fighting against disaster, crisis, etc. & Facebook also introduced a feature 'Mark-safe' to fight against these disasters.
- We will focus on making it more accurate and useful by applying some Neural Network concepts in this project in future.
- To make this system more powerful and useful, by implementing some technique that can detect non-hashtag words that are relevant for analysis.(E.g.- e 'People' as 'ppl').
- Machine learning techniques can be employed to check the veracity of social media by comparing contents from actual news reports.

References

References - I

[1]. N. Saharia, "Detecting emotion from short messages on nepal earthquake," in *Speech Technology and Human-Computer Dialogue (SpeD), 2015 International Conference on.* IEEE, 2015, pp. 1–5.

[2]. S. Kedar, S. Owen, C. Jones, A. Donnelan, M. Glasscoe, and R. Duren, "Select technologies and capabilities to improve earthquake resiliency in california," 2016.

[3]. D. M. Blei, A. Y. Ng, and M. I. Jordan, "*Latent dirichlet allocation,*" *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[4]. J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[5]. D. Hardt, "The oauth 2.0 authorization framework," Tech. Rep., 2012.

References - II

[6]. H.-N. Teodorescu and N. Saharia, "An internet slang annotated dictionary and its use in assessing message attitude and sentiments," in *Speech Technology and Human-Computer Dialogue (SpeD), 2015 International Conference on*. IEEE, 2015, pp. 1–8.

[7]. G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, no. 5, pp. 513– 523, 1988.

[8]. H. M. Wallach, "Topic modeling: beyond bag-of-words," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 977–984.

Source data Repository

- Source code and Dataset for this Project can be found at - Here (Github)

Thank You