WHITE PAPER

latentview
Actionable Insights • Accurate Decisions

Social media listening during disasters: 2018 Kerala floods in focus

Abstract

Social media as a news source has grown as the mainstream news outlet for millennials. In the occurrence of a natural disaster, social media is redefining communication and broadcast for public safety. For example, during the 2015 (india) Chennai rains & the 2017 Houston floods, Twitter was extensively used by local communities to relay information about flooded areas, rescue agencies, and relief centers. This widespread use of social media as information-sharing platforms can be leveraged by public agencies for their disaster management policies. When disaster strikes, the capacity of telecommunication networks to cope with the surge in voice call volumes is severely limited, thus overwhelming and jamming phone lines. However, data networks like LTE remain operational. In this context, social media ends up playing a crucial role, for the public to contact emergency response teams.

Digital data, thus, can act as a supplemental listening channel for government agencies, apart from the traditional sources of information. It is a source of real-time, geographic-based information, provided directly by social media users from the affected community, thereby providing enhanced situational awareness of disaster-hit areas. Essentially, it is a repository of valuable information which can help in coordinating relief and rescue operations.

This paper discusses the standardized and well-tested framework that could be employed when a disaster strikes. The framework has been utilized for the 2015 Chennai floods and the 2018 Kerala floods.

Digital data, thus, can act as a supplemental listening channel for government agencies, apart from the traditional sources of information. It is a source of real-time, geographic-based information, provided directly by social media users from the affected community, thereby providing enhanced situational awareness of disaster-hit areas.

Introduction

In recent years, the world has witnessed a series of natural disasters. From earthquakes to tsunamis to the recent polar vortexes, the risk of natural disasters is on the rise. A recent report by the United Nations Office for Disaster Risk Reduction (UNISDR) states a 151 percent increase in climate-related disaster costs in the last 20 years. The economic and human loss due to natural disasters is staggering.

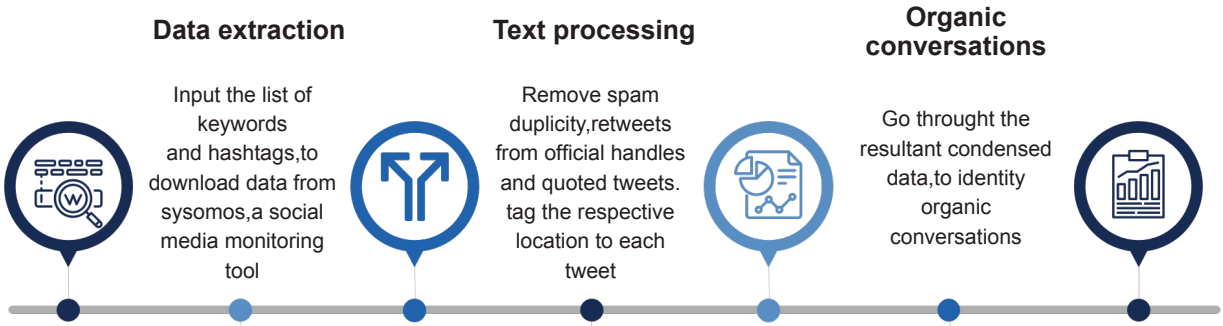
Disaster preparedness coupled with quick relief efforts can significantly minimize the losses suffered. In this environment, social media has been a key communication platform for providing information, broadcasting real-time alerts and warnings, and mobilize relief efforts among many other uses. Social media platforms are addressing the critical gap in communication by creating new products for the masses. For example, Facebook introduced Disaster Maps – a product to improve disaster response and disaster management. The product aggregates de-identified Facebook data to provide critical insights and information to trusted organizations during a natural disaster.

Using social media for disaster management

Social media is not only pertinent during a disaster but can be utilized in every phase of disaster planning and management.

I. **Mitigation:** Impact of natural disasters can be mitigated in-part, by strategically reducing the vulnerabilities of communities. Govt. agencies can use social media to routinely engage with the public and raise general awareness of natural disasters and safety measures.

II. **Preparedness:** During preparations for an imminent disaster, social media can act as a medium for information dissemination, to issue warnings and alerts.



III. **Response:** Social Media can serve as a crucial mode of communication even if the traditional voice-based infrastructure is hit. Using it as a listening tool will not only ensure that everyone is heard but also help the emergency response teams manage the massive inflow of information and rapidly understand the needs of a community – in order to deliver a quick and prioritized response.

IV. **Recovery:** Social media can be employed as a platform to discuss key problems facing the community and the necessary actions required to aid recovery.

The following table gives an overview of processes that could be adopted by government agencies, for using social media in disaster management.

Brand building	1 Establish a social media presence over various platforms 2 Get the twitter handle officially verified, to enhance credibility 3 Build-up a follower count through active campaigning and promotion
Social engagement	4 Frequently engage with stakeholders by regularly posting content 5 Send out timely and frequent updates, while responding to questions and clarifications from public
Social media listening	6 Establish a **Media Monitoring Team** 7 Create and popularize official hashtags, specific to the disaster 8 Develop a list of keywords and identify trending hashtags, to monitor -> use them to extract relevant conversations from Twitter 9 Apply text mining techniques to remove spam and bucket conversations into respective localities and categories -> e.g. Health, Relief & Rescue 10 Prioritize conversations based on urgency, time of posting and numbers affected 11 Have the information verified – the authenticity of posts – contact listed phone number
On-field action & recovery	12 Relay information to a dedicated control room, post verification 13 Coordinate relief and rescue with other organizations through social media posts/WhatsApp groups 14 Monitory conversations around **health and disease** to capture any potential outbreaks

A dedicated control centre and resource person would be required, to actively monitor and handle media platforms. Frequent and consistent pre-event engagement with the public is necessary in order to be able to play a prominent role during the time of crisis. Government agencies could also look to engage influencers and celebrities to reach out to a wider audience and lead on-field response efforts.

Key challenges

In the event of any disaster, social media witnesses high volumes and velocity of unstructured data. The data generated can be classified under two major categories of informational and conversational posts. Informational posts are those which convey facts, examples would include news articles, political chatter, posts of sympathy and those urging for charity & help. These are largely irrelevant to rescue & relief measures and contribute to the digital noise. Conversational posts, on the other hand, originate from the affected area and tend to be more insightful. They are either requests for help or post detailing available relief. Such organic posts are valuable but are vastly overshadowed by a plethora of other irrelevant posts around the disaster.

The primary challenge rests in extracting useful insights from massive amounts of raw, unstructured data. The rapid influx of information and the critical nature of the situation also necessitate a quick response time, in order to have any meaningful impact.

The primary challenge rests in extracting useful insights from massive amounts of raw, unstructured data.

1 Limitations in accessing data across multiple channels

States like Kerala have a more active user base on Facebook when compared to Twitter. Thus, a high proportion of social data remains inaccessible due to the privacy policy of Facebook. We would have to primarily rely on Twitter data for analyses.

2 Dispelling rumors, misinformation, and re-circulation of old messages

Posts uncorroborated with videos and photos have the possibility of containing misinformation. However, the current scope of our framework does not extend toward verification of false information.

3 Analyzing in local languages

Use of multiple languages for communication would pose a challenge for consolidation of data since information gets split across languages. In this scenario, it would become necessary to first identify the dominant language(s) in use and subsequently apply appropriate text mining techniques. Alternatively, local language tweets can also be translated using third-party translation tools. However, this would necessitate additional cost and time. For Kerala floods, our team manually classified the Malayalam language tweets.

4 Quick response time

During disasters quick turnaround time is necessary to facilitate rapid movement of resources. Information can be become outdated within a span of hours. It is thus necessary to keep a constant track of tweets and prepare social reports at hourly intervals, to have maximum impact. For Kerala floods, due to an inherent lag in the availability of the data in Sysomos and the processing time of tweets, our social reports were limited to a daily basis.

A standardized protocol should be able to address the challenges of parsing informal messages, handling information overload and prioritizing the different types of data and information. The focus of our framework lies in extracting the user-generated comments requesting for help, while also giving a high-level summarized overview of the social media conversations. This would be valuable in offering prioritized and actionable insights to agencies working on the ground.

A standardized framework

The steps below detail a standardized methodology for handling unstructured data. The process outlined is applicable for any unstructured textual data, however, they can be customized for disaster management scenarios.

1 Framing keywords

Hashtags are often used to categorize content and track specific topics or trends on social media platforms. Identifying and following relevant hashtags of a disaster helps in tracking and listening to disaster-related conversations. For the Kerala floods, we identified a list of 30 trending hashtags to track and monitor during the period. A few of the top trending hashtags included **#Keralafloods, #StandWithKerala, #opmadad, #KeralaSOS and #KeralaFloodRelief.** Apart from incorporating hashtags, framing a list of disaster-specific keywords will help filter and extract only the conversations of interest.

2 Data extraction

There are various social media tools available in the market for data extraction. For the purposes of Kerala floods, we used ***Sysomos*** for collecting our data. Also, feeds obtained directly from Twitter were streamed to a live dashboard for real-time monitoring. Data collected from Sysomos was used for further analyses.

3 Language split

In countries like India, where a multitude of languages are spoken the first logical step in breaking language barriers would be to split the dataset on the basis of language. For Kerala floods, our team consisted of native speakers who manually categorized the local language conversations.

4 Text processing

The scope of this framework is for capturing tweets requesting for help. Such critical posts tend to get lost in the sea of irrelevant tweets. Thus, spam removal is a primary step to condense the dataset extract only useful tweets. This process eliminates official handle posts, factual content, and duplicity of data which appear in the form of retweets & quoted tweets. Location tagging is another important procedure that captures the location mentioned in the content of the tweet. We used 'Parts of Speech' tagging to identify localized needs and help officials divert the required resources to the right areas.

5 Theme analysis

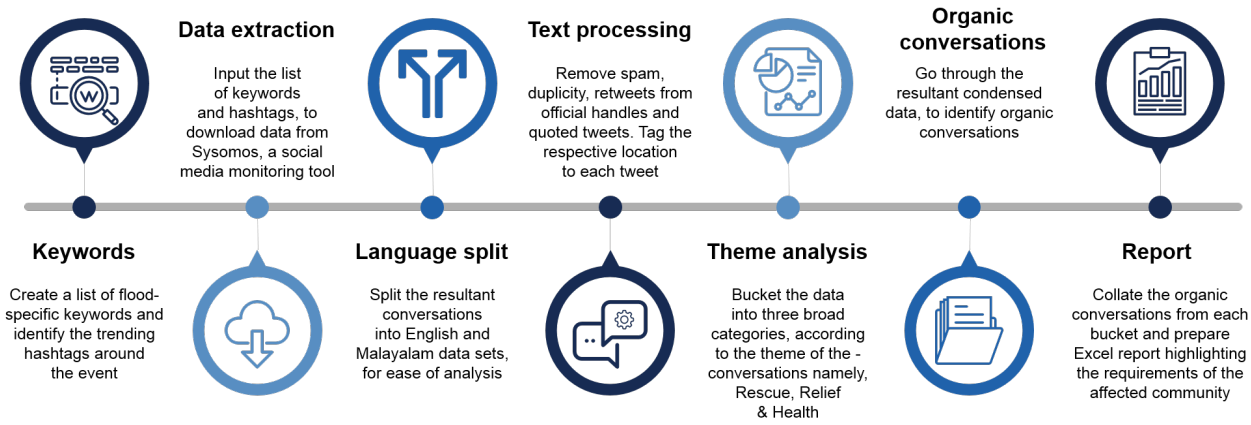
During a disaster, the top three priorities for civic administration are Rescue, Relief, and Health. Splitting data into these three buckets would construct a coherent approach to address calls for help. A combination of standard NLP techniques like N-grams and Word2Vec can be used.

N-Grams provide us with a broad idea of the frequently occurring words in the dataset, which can be helpful in framing keywords for categorizing themes. However, N-grams are not exhaustive, and only capture high-frequency words whereas word embedding approaches like Word2Vec offer a more comprehensive solution.

Word2Vec works on the premise of grouping together closely associated words occurring in similar contexts. We can use the high-frequency words identified by the N-grams approach, as inputs for the model. As an example, for the word 'trapped', our Word2Vec model would point us to other closely associated words like **'stuck', 'stranded'** and **'evacuate'**. Another example would be to discern all the words used in association with **'medicine':** during Kerala floods, words like Doxycycline and Paracetamol found high mentions in relation with **'medicine'**. This approach enables us to identify a wide gamut of important words being used by the public to convey their requirements.

6 Organic conversations

Though the spam removal process eliminates much of the irrelevant data, our dataset may still be left with unwanted factual tweets. The final step entails manual intervention to differentiate purely user-generated tweets from residual spam tweets.



The above figure illustrates the methodology adopted for processing data during the 2018 Kerala floods.

2018 Kerala floods – Our approach

In the first few weeks of August 2018, Kerala witnessed record levels of rainfall resulting in unprecedented floods. Reported as the largest flood in a century, the Kerala floods led to severe economic losses, death, and displacement of lives. Red alert was declared across 12 districts as death tolls rose.

Having experience in using social media to mitigate the 2015 Chennai flood, LatentView Analytics reached out to several administration officials involved in rescue and relief efforts, and offered:

i. **A real-time, interactive dashboard** – Kerala floods related conversations were collected from Twitter using the Twitter API. Elasticsearch was used for the storage and retrieval of conversations. The dashboard itself was built in Kibana and hosted in AWS cloud. It allowed us to monitor the trends, popular hashtags, key influencers, number of mentions etc. It also had options to filter for specific words and contained functionalities for viewing district-wise metrics. The dashboard was refreshed with the latest data for every 15 minutes.

ii. **Daily social listening reports** – These reports consisted of lists of organic tweets, sorted based on time, and organized into different themes. Overall volumes of the daily social media conversations and key health-related conversations were also highlighted.

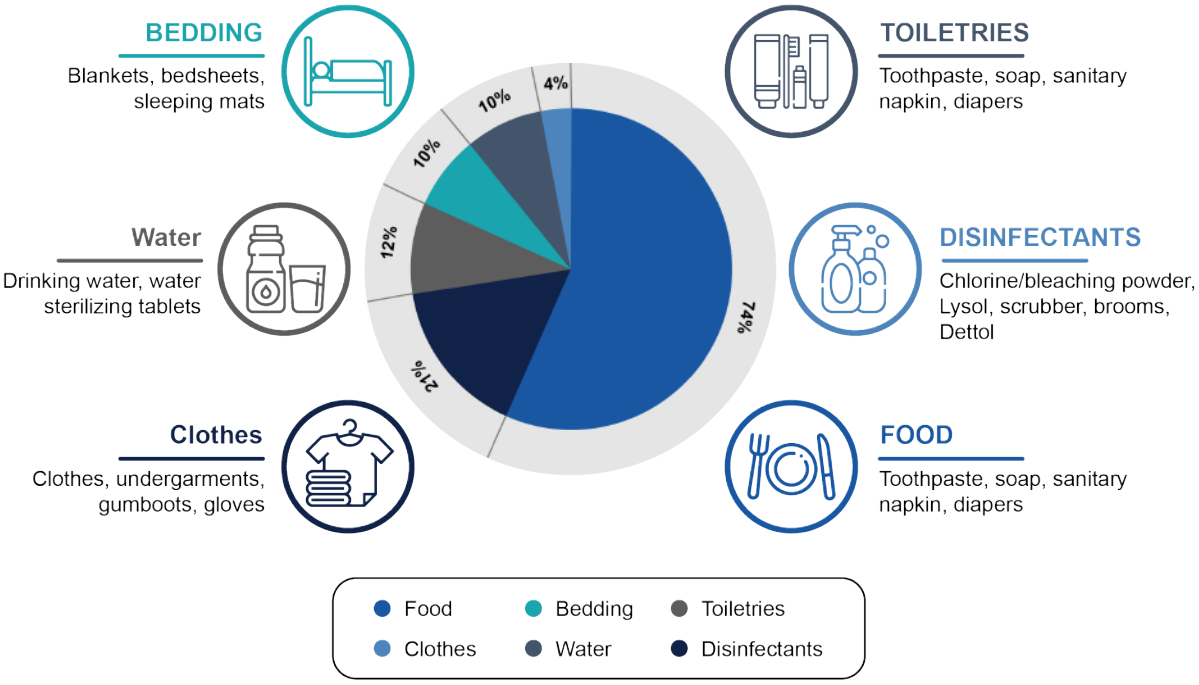
Subsequently, the social media scenario surrounding Kerala floods was actively monitored by the LatentView Analytics team for a period of two weeks. Our social reports were sent to an email distribution list on a daily basis. The reports broadly covered information pertaining to the topics of Rescue, Relief, and Health. The Kerala Health Control team specifically wanted to monitor health-related conversations for any signs of a potential epidemic breakout.

The figure below shows a glimpse of the content observed in each of the categories:

- Mentions of being trapped or stranded on the terrace
- Requests for boats/helicopters to rescue stuck relatives
- Requests for rice, milk powder, bread, pulses and wheat
- Requests for blankets, chlorine, clothes, sanitary pads, diapers and drinking water
- Requests for blood donation
- Concerns about snake bites, dengue, leptospirosis
- Requests for paracetamol and doxycycline

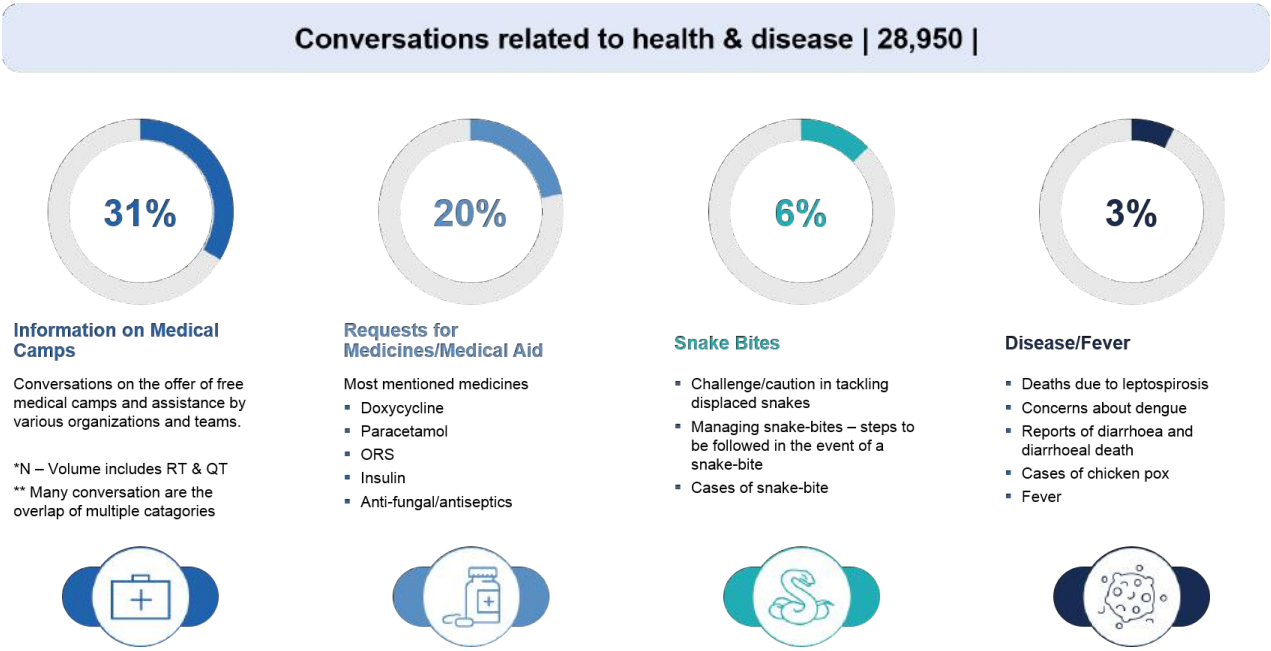
N-grams proved useful in finding the most frequently occurring words within each of the three buckets of Relief, Rescue, and Health. The illustration below provides a high-level overview of the top requirements and split of different relief items (in terms of number of mentions) from each district.

Conversations related to supply of relief materials |65,932|



*N-volume includes RT & QT
**Most conversations are the overlap of multiple categories

Conversations around health were dominated by offers of free medical aid by volunteers. A significant number of conversations also requested general medical assistance and medicines.



Impact of social media listening

Our daily social listening reports served as an additional source of information for a media monitoring team coordinating relief and rescue. It ensured that the burst of activity in the social media was observed and reported. The reports served to capture any major blind spots missed by the field staff.
LatentView Analytics developed and tested a framework which has proven to work well for natural disasters like floods. This framework can easily be scaled for handling other types of disaster as well.
We make use of free, open source tools and technologies like Elasticsearch, Logstash, Kibana, and Python, for our analyses. For building the real-time dashboard, the entire software setup can be completed within a couple of hours, including open source software installation and requesting servers in AWS cloud platform. Using Twitter feed data, the dashboard allows us to monitor high volumes and velocity of data in real-time. Our framework for text processing helps extract meaningful entities from vast amounts of unstructured data. The insights generated were considered as a value-addition for the media monitoring team, active during Kerala floods.

> ***LatentView Analytics developed and tested a framework which has proven to work well for mitigating natural disasters like floods. This framework can easily be scaled for handling other types of disaster as well.***

Future course of action

A key area to improve upon would be the processing time since few aspects of the framework continue to require manual effort. Automatic categorization of tweets into Rescue, Relief and Health, using supervised machine learning algorithms like Naïve Bayes, Support Vector Machines & Neural Networks could serve to speed up the process. In the future, we could also use ML/AI techniques to automatically extract contact details (name, phone number) of the person requesting rescue or relief operations.

Data analytics can also be used in other capacities to aid disaster management. Other options to explore would be to apply satellite imagery for flood mapping. Hue transformation can be used to locate the water in images, and analyzing pixel-to-pixel comparison of watershed masks pre-and post-disaster can tell us the change in water levels. This would help assess areas most severely impacted by floods.

In the future, we could also use ML/ AI techniques to automatically extract contact details (name, phone number) of the person requesting rescue or relief operations.

Conclusion

Social Media can be used in varying capacities, to aid in each phase of disaster management. In the event of a disaster itself, it becomes a valuable repository of real-time, geographic based information, provided directly by the affected community. Thus, leveraging Social Media as a supplemental listening channel can help coordinate rescue and relief operations.

About the author

Kiruthika Rajendiran

Kiruthika is Senior Business Analyst at LatentView Analytics. As a core member of the Social Intelligence team, she works extensively in text mining and NLP techniques, delivering critical insights for numerous Social Media Analytics driven projects.

About: LatentView Analytics

LatentView Analytics is a leading global data analytics firm, helping Fortune 500 companies harness the power of data to drive digital transformation. LatentView Analytics's solutions provide a 360-degree view of the digital consumer, fuel machine learning capabilities and support artificial intelligence initiatives. Companies in the retail, CPG, BFSI, high tech, healthcare and other sectors use LatentView Analytics's solutions to predict new revenue streams, anticipate product trends, improve customer retention, optimize investment decisions and turn data into a valuable business asset. LatentView Analytics has offices in Princeton, N.J., San Jose, Calif., London, Singapore and Chennai, globally.

For more information, please visit **www.latentview.com** **or** write into: **sales@latentview.com**