



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 7, July 2023



Impact Factor: 8.423



# Social Media-Driven Big Data Analysis for Disaster Situation Awareness

Anushree KR, Prof Sowmya M S

Dept. of MCA, Bangalore Institute of Technology, Bangalore, India

Asst. Professor, Dept. of MCA, Bangalore Institute of Technology, Bangalore, India

**ABSTRACT:** Disaster response heavily relies on electronic communications to assess the situation and dispatch aid and rescue resources effectively. In recent times, focus of these communications have been shifted in favor of social media platforms, particularly Twitter. Nevertheless, deriving valuable insights from an extensive pool of accessible data presents numerous hurdles, all of which are comprehensively explored in this research. These challenges encompass (a) the necessity of sieving out irrelevant data, (b) the integration of diverse data types from social media and alternative sources, and (c) the handling of partially geo-tagged social media data to grasp the requirements of affected individuals.

To tackle these obstacles effectively, the utilization of big data techniques becomes imperative, primarily due to the vast volume of data, a substantial proportion of which might not be directly pertinent. Spatial analytics plays a chief role in comprehending the disaster situation, although its availability is limited due to the reluctance of many users to be tracked. Additionally, this paper explores the role of edge computing in handling analytics in scalable manner.

In conclusion, this paper highlights the critical part of social-media in disaster assessment and response, and emphasizes the necessity of advanced technologies such as big-data analysis and spatial analytics to effectively utilize this invaluable resource during times of crisis

Keywords: Communication, Social Media, Big Data, Edge Computing, Disaster

## I. INTRODUCTION

Situational scenario plays highest importance in disaster situation, yet obtaining and organizing the obligatory information in a lucid manner proves to be challenging. Difficulties stem from potential mutilation to and overload of communications networks, but the lack of clarity about the relevant information and its optimal gathering also contributes significantly. As disasters continue to develop over prolonged periods, monitoring situational awareness becomes increasingly difficult.

Lately, social media has emerged as a predominant channel for acquiring real-time information and capturing the necessities of disaster-affected individuals. Given that numerous people use smartphones to access social media during such events, there exists a theoretical possibility of identifying the geographical origins of the shared data.. Nevertheless, in practice, location information is frequently scarce due to privacy considerations. Among various social media platforms, Twitter has emerged as the preferred means of disaster communication because of its low networking demands, user-friendly interface, and concise format. For example, in the wake of the 2011 Japanese earthquake, Twitter experienced an astonishing influx of over 5,500 tweets per second.

## II. LITERATURE SURVEY

Twitter has been extensively utilized in various disaster scenarios, encompassing significant events such as three hurricanes of 2017 (Harvey, Maria, and Irma) impacting the Caribbean and the US East Coast [1], the 2019 Pan-European floods [2], and the 2019 US Midwestern floods [3]. Additionally, it played a pivotal role in upshot of the Kumamoto Earthquake, which struck Kumamoto City in Kyushu Region, Japan, in 2016. Notably, the density of relevant keywords in these tweets displayed a strong correlation with the shake map observed to the east of Kumamoto City, as reported by the Geospatial Information Authority of Japan and the National Disaster Institute for Earth Science and Disaster Suppleness [5].

During the occurrence of Hurricane Sandy in 2012, Twitter proved instrumental in tracking power outages in New York. Storm's impact on the city was severe, leading to a widespread blackout on the night of October 29th, affecting

hundreds of thousands of people [6]. Geo-tagged tweets played a crucial role in highlighting and documenting the power outage-related information during this disaster.

### III. EXISTING SYSTEM

In the framework of the existing system, media has proven to be a powerful tool, offering a involved and collective erection that facilitates collective knowledge-building for public info and caution approaches. Considering this perspective, the author proposes an intelligent public information and warning system for disasters, leveraging social media, which comprises three crucial functions: (1) the proficient and effective acquisition of disaster situational awareness information, (2) facilitation of self-organized peer-to-peer assistance initiatives, and (3) empowering disaster management agencies to interact with the public and gather invaluable insights.

To realize this idea, authors conducted an extensive examination of 304 revisions led between 2008 and 2018. This systematic evaluation of the current literature aimed to understand The analysis explores the communication patterns observed on social media during disasters, as well as the current advancements in social media informatics in such scenarios. The review highlights the challenges encountered by previous research and presents a detailed research roadmap to address these issues and realize the envisioned intelligent public information and warning system.

This comprehensive review acts as a valuable connection for researchers engaged in studying social media within disaster scenarios, providing valuable insights into the latest advancements in related fields. The review's discoveries underscore the significance of specific research areas, such as the creation of a precise disaster social media ontology with semantic interoperability, the examination of network patterns involving trending information and influential users, in-depth evaluations of societal impacts arising from infrastructure disruptions, and the establishment of best practices for utilizing social media during disasters.

### IV. PROPOSED SYSTEM

The proposed system incorporates Spatial clustering, a method that Cluster analysis, within the context of unsupervised learning in machine learning and concept hierarchies in statistics, groups similar objects together using diverse measures like distance, connectivity, or their relative density in space. The primary goal is to detect significant structures or clusters within data, minimizing the dependence on substantial background knowledge.

Spatial clustering could be classified into three categories: partitional clustering, density based clustering, and grid based clustering. Partitional clustering involves dividing the entire dataset into a group of disjoint clusters. For instance, the widely used K-means method partitions n data points into k groups, where individual data point belongs to the collection with the closest mean.

Density-based clustering, Conversely, it identifies groups of closely packed points based on neighborhood relationships while also pinpointing outliers situated in isolated low-density regions. Prominent method for density-based clustering is DBSCAN, which identifies clusters based on certain conditions, such as a specified radius (Eps) and a min number of items (MinPts) within that radius, forming dense regions. Nonetheless, DBSCAN exhibits relatively high computational complexity since it processes each data point individually.

In contrast, grid-based clustering takes a different approach by assigning a value to each cell of a grid that covers several data points. This method proves to be more efficient in processing big data, especially when the grid cell size is appropriately chosen to avoid excessive granularity.

Overall, the use of Spatial clustering offers an effective way to group data and uncover meaningful structures, and the specific clustering method chosen depends on the characteristics of the data and the desired computational efficiency.

### V. IMPLEMENTATION

#### Admin

Within this module, the Service Provider is required to log in using a valid username and password. Once logging in is successful, they gain access to various operations, enabling them to perform certain tasks.

#### Login

The module provides the following functionalities to the logged-in Service Provider:

- All users can be view and authorize them.
- View all friend requests and responses.





- Access details of all datasets.
- Access details of datasets through blockchain technology.
- Access details of datasets through clustering.
- View all reviews.
- View disaster status results.
- 

### View and Authorize Users

This module grants the admin the ability to examine the roster of registered users. The admin has the option to access individual user information, such as their username, email, and address. Furthermore, the admin holds the power to authorize users when necessary.

### End User

Within this module, a certain number (n) of users are accommodated. To perform any operations, users must complete a registration process. Upon successful registration, their information will be stored in the database. After registration, users are required to log in using their authorized username and password. Once logged in, users can engage in various activities, including registration and login, viewing their profile, searching for friends and sending friend requests, accessing a list of all friends, uploading datasets, viewing all uploaded datasets, checking disaster status, obtaining disaster type results via a hash code, and viewing reviews from all friends.

## VI. CONCLUSION

During disasters, relevant data for situational assessment originates from diverse sources, counting social media platforms like Twitter, straight communications with unaffected users who share information on media, and observations from deployed monitoring infrastructure, among others. Numerous type of disaster-related data may stem from user-end interactions, social media platforms, or network operators. However, the data collected from these sources often contains a significant amount of immaterial or weakly related info, necessitating the use of big-data techniques to extract valuable insights.

Spatial information and context play a critical role in this process. The paper at hand focuses on exploring prospects and challenges related to spatial big data analytics, especially when dealing with partial spatial information. The ultimate goal is to extract situational awareness of the disaster. I believe that the above article will inspire many others for future enhancement using bigdata analysis and disaster situation awareness.

## REFERENCES

- [1] F. Alam, F. Ofli, M. Imran, and M. Aupetit, "A twitter tale of three hurricanes: Harvey, irma, and maria," in Proc. 15th Intl. Conf. on Information Systems for Crisis Response and Management, ISCRAM 2018 (B. Tomaszewski and K. Boersma, eds.), vol. 2018-May, pp. 553– 572, 1 2018.
- [2] V. Lorini, C. Castillo, F. Dottori, M. Kalas, D. Nappo, and P. Salamon, "Integrating social media into a pan-european flood awareness system: A multilingual approach," ArXiv, vol. abs/1904.10876, 2019.
- [3] "Midwest farmers take to twitter to document flood disaster," May 2019.
- [4] J. Yee, "Hurricane sandy: A chance to identify vulnerabilities, learn from the past, and increase future resilience." [https://research.library.fordham.edu/envirom\\_theses/5](https://research.library.fordham.edu/envirom_theses/5), 2013.
- [5] Asian Disaster Reduction Center, "2016 Kumamoto Earthquake Survey Report (Preliminary)." <http://www.adrc.asia/publications/201604> KumamotoEQ/ADRC 2016KumamotoEQ Report 1.pdf. Accessed: March, 2017.
- [6] "Hurricane Sandy 2012." [https://www.hu\\_ngetonpost.com/2012/10/31/hurricane-sandy-new-york-city-power-outage-map](https://www.hu_ngetonpost.com/2012/10/31/hurricane-sandy-new-york-city-power-outage-map) n 2050380.html. Accessed: January, 2018.
- [7] S. Shekhar, P. Zhang, Y. Huang, and R. R. Vatsavai, "Trends in Spatial Data Mining," Data mining: Next generation challenges and future directions, pp. 357–380, 2003.
- [8] J. Wang, Y. Wu, N. Yen, S. Guo, and Z. Cheng, "Big data analytics for emergency communication networks: A survey," IEEE Communications Surveys & Tutorials, vol. 18, no. 3, pp. 1758–1778, 2016.
- [9] Z. Jiang, "A survey on spatial prediction methods," IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 9, pp. 1645–1664, 2018.
- [10] A. McGovern, N. Troutman, R. A. Brown, J. K. Williams, and J. Abernethy, "Enhanced spatiotemporal relational probability trees and forests," Data Mining and Knowledge Discovery, vol. 26, no. 2, pp. 398–433, 2013.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.423**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

**9940 572 462** **6381 907 438** **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details