

Report on the Second Workshop on Exploitation of Social Media for Emergency Relief and Preparedness (SMERP 2018) at the Web Conference (WWW) 2018

Saptarshi Ghosh

Indian Institute of Technology, Kharagpur, India
saptarshi@cse.iitkgp.ac.in

Kripabandhu Ghosh

Indian Institute of Technology, Kanpur, India
kripa@cse.iitk.ac.in

Debasis Ganguly

IBM Research Labs, Dublin, Ireland
debasis.ganguly1@ie.ibm.com

Tanmoy Chakraborty

IIT Delhi
tanmoy@iiitd.ac.in

Gareth J. F. Jones

Dublin City University, Ireland
Gareth.Jones@dcu.ie

Marie-Francine Moens

Katholieke Universiteit Leuven, Belgium
marie-francine.moens@cs.kuleuven.be

Abstract

The Second Workshop on Exploitation of Social Media for Emergency Relief and Preparedness (SMERP) was held in conjunction with The Web Conference (WWW) 2018 at Lyon, France. A primary aim of the workshop was to promote multi-modal and multi-view information retrieval from the social media content in disaster situations. The workshop programme included keynote talks, a peer-reviewed paper track, and a panel discussion on the relevant research problems in the scope of the workshop.

1 Introduction and Motivation

User-generated content on online social media (OSM) platforms such as *Twitter*, *Facebook*, *WhatsApp*, has become an important source of real-time information during emergency events (e.g. natural disasters such as earthquakes, cyclones, floods, fire, epidemics or man-made disasters such as terror attacks, riots) [4, 8]. During such events information posted on OSM platforms can contribute significantly to relief operations. Additionally, crowdsourced content created using OSM platforms can also be utilised for emergency preparedness, such as for identifying disaster-prone regions and infrastructures, developing early warning systems, developing emergency-resilient communities, and so on.¹

Given the huge volume and the rapid rates at which content is posted on OSM platforms, automated techniques and information systems need to be developed for extracting, summarizing and presenting critical information in a timely and useful way. The Exploitation of Social Media for Emergency Relief and Preparedness (SMERP) workshop series aims to provide a forum for researchers working in fields related to this topic to gather, to present and discuss their findings, and to explore the directions for further developments in this area. Our workshops are targeted at bringing together researchers from diverse fields – Information Retrieval, Data Mining and Machine Learning, Natural Language

¹Note: the terms “emergency” and “disaster” are used interchangeably in this note.

Processing, Machine Translation, Social Network Analysis, Computational Social Science, Human Computer Interaction – who can potentially contribute to utilising social media for emergency relief and preparedness. The SMERP workshop series also aims to provide the research community with real-life datasets to foster research on the theme.

The first SMERP workshop was held in conjunction with ECIR 2017 [3]. SMERP 2018 was the second edition of the workshop was held in conjunction with the Web Conference (WWW) 2018 in Lyon, France (<https://www.cse.iitk.ac.in/users/kripa/smerp2018/>). SMERP 2018 included two keynote talks, a peer-reviewed research paper track, and a panel discussion. Additionally, the SMERP organizers released a dataset for use by the research community.

2 Keynote Talks

The workshop had two keynote talks by eminent researchers working on topics at the centre of the interests of the SMERP workshop.

(1) Utility of Social Media Information in Response to Natural Disasters

Speaker: Prasenjit Mitra, Professor of Information Sciences and Technology and Associate Dean for Research, The Pennsylvania State University, U.S.A.

Abstract: In this talk Professor Mitra explored the use social media information to improve responses to natural disasters. Drawing from his own experiences using AIDR, a tool that enables information gathering for disaster response, he showed how social media information can be used effectively to help in disaster response. His work in this area addresses questions of how tweets can be classified in real-time using volunteers who assist in training a machine learning based system. He has explored whether and how his existing datasets from previous disasters can be utilized to effectively enable disaster relief. He then showed how reports can be generated in natural language utilizing information from the tweets; these reports must be generated to satisfy the different perspectives and information needs of different first responders and volunteers who help orchestrate and co-ordinate disaster relief efforts. Finally, he outlined how images can be processed automatically and to help improve the response. He concluded by examining issues related to the veracity of the information obtained during disasters and the limits of using information from social media with an eye towards the future.

(2) Transforming Social Big Data into Timely Decisions and Actions for Crisis Mitigation and Coordination

Speaker: Amit Sheth, Executive Director of Kno.e.sis, Wright State University, U.S.A.

Abstract: Crises are imposing massive costs to economies worldwide. Natural disasters caused record \$306 billion in damage to the U.S. in 2017! Real-time gathering of relevant data through the now ubiquitously present mobile technologies, and the ability to disseminate it through social media has forever changed how disaster and health crisis monitoring and response are carried out. Both traditional crisis response organizations as well as temporary, informal, self-organized and community-based organizations have come to rely increasingly on social media. Furthermore, the ability to collect, re-purpose and reuse data from past events is helping with preparedness and planning for future events. In this talk, Professor Sheth reviewed his extensive experience of: (a) interactions with a variety of stakeholders involved in emergency response at city, county, country and international levels, (b) research on real-time social media analysis spanning spatio-temporal-thematic; people-content-network; linguistic-sentiment-emotion-intent analysis dimensions, (c) development and use of crisis response specific tools (location identification, demand-supply match) and the comprehensive Twitris semantic social intelligence system (which is also commercialized as *Cognovi Labs*), and (d) a variety of real-world evaluations and real-time uses (e.g., supplying data for Google Crisis map during Uttarakhand Floods, rescue during Kashmir Floods, neighborhood image map during Chennai Floods, providing information to FEMA during Oklahoma tornados), spread of disease and epidemiology (e.g., Zika spread), metro-level multi-agency disaster preparedness exercise, etc.

3 Peer-reviewed Research Paper Track

The workshop solicited original contributions related to relevant topics, which included (but was not limited to):

- Multi-modal and multi-view Information Retrieval
- Addressing the code-mixed and noisy, informal vocabulary of OSM content
- Transfer learning - applying models trained on prior emergency event(s) on a future event
- Detection of events and emerging themes
- Real-time management and summarization of dynamic content streams
- Detection of rumours and false news, verification of news and identification of trustworthy sources
- Geo-tagging and geo-localisation of content and sources
- Social network models for information diffusion in emergency situations
- Identifying disaster-prone or accident-prone regions and infrastructures
- Designing crowdsourcing systems for emergency preparedness and post-disaster relief
- Building knowledge bases for automatically mining social media posts during emergency

Nine (9) papers were submitted, out of which five (5) papers were accepted for presentation at the workshop, through a rigorous peer-review process (the Technical Program Committee is detailed at the workshop website).

We now briefly describe the contributions of the accepted papers. Kumar et al. [5] modelled the evacuation behaviour of the residents of *New York City* in the aftermath of Hurricane Sandy. Mountzidou et al. [6] combined textual and visual information to determine the relevance of social media content with respect to flooding. Ghosh et al. [2] reported that class specific TF-IDF boosting improves classifier performance on microblogs posted during disaster events. Soni et al. [7] studied the incompleteness of the gold standard created for a shared task relevant to the workshop theme (the FIRE 2017 IRMiDis track), and provided suggestions for the creation of more robust gold standard for such tasks. Finally, Dutt et al. [1] developed a real-time system that identifies locations using natural language processing techniques on the microblog text during disaster situation. The details of the papers are as below.

Title: Utilizing Geo-tagged Tweets to Understand Evacuation Dynamics during Emergencies: A case study of Hurricane Sandy

Abstract: Hurricane evacuation is a complex process and a better understanding of the evacuation behavior of the coastal residents could be helpful in planning better evacuation policy. Traditionally, various aspects of the household evacuation decisions have been determined by post-evacuation questionnaire surveys, which are usually time-consuming and expensive. Increased activity of users on social media, especially during emergencies, along with the geo-tagging of the posts, provides an opportunity to gain insights into user's decision-making process, as well as to gauge public opinion and activities using the social media data as a supplement to the traditional survey data. This paper leverages the geo-tagged Tweets posted in the New York City (NYC) in wake of Hurricane Sandy to understand the evacuation behavior of the residents. Based on the geo-tagged Tweet locations, we classify the NYC Twitter users into one of the three categories: outside evacuation zone, evacuees, and non-evacuees and examine the types of Tweets posted by each group during different phases of the hurricane. We establish a strong link between the social connectivity with the decision of the users to evacuate or stay. We analyze the geo-tagged Tweets to understand evacuation and return time and evacuation location patterns of evacuees. The analysis presented in this paper could be useful for authorities to plan a better evacuation campaign to minimize the risk to the life of the residents of the emergency hit areas.

Authors: Dheeraj Kumar and Satish Ukkusuri

Affiliation: Purdue University, United States

Title: Flood relevance estimation from visual and textual content in social media streams

Abstract: Disaster monitoring based on social media posts has raised a lot of interest in the domain of computer science the last decade, mainly due to the wide area of applications in public safety and security and due to the pervasiveness not solely on daily communication but also in life-threatening situations. Social media can be used as a valuable source for producing early warnings of eminent disasters. This paper presents a framework to analyse social media multimodal content, in order to decide if the content is relevant to flooding. This is very important since it enhances the crisis situational awareness and supports various crisis management procedures such as preparedness. Evaluation on a benchmark dataset shows very good performance in both text and image classification modules.

Authors: Anastasia Mourtzidou, Stelios Andreadis, Ilias Gialampoukidis, Anastasios Karakostas, Stefanos Vrochidis and Yiannis Kompatsiaris

Affiliation: Information Technologies Institute, CERTH, Greece

Title: Class Specific TF-IDF Boosting for Short-text Classification

Abstract: Proper formulation of features plays an important role in short-text classification tasks as the amount of text available is very little. In literature, Term Frequency - Inverse Document Frequency (TF-IDF) is commonly used to create feature vectors for such tasks. However, TF-IDF formulation does not utilize the class information available in supervised learning. For classification problems, if it is possible to identify terms that can strongly distinguish among classes, then more weight can be given to those terms during feature construction phase. This may result in improved classifier performance with the incorporation of extra class label related information. We propose a supervised feature construction method to classify tweets, based on the actionable information that might be present, posted during different disaster scenarios. Improved classifier performance for such classification tasks can be helpful in the rescue and relief operations. We used three benchmark datasets containing tweets posted during Nepal and Italy earthquakes in 2015 and 2016 respectively. Experimental results show that the proposed method obtains better classification performance on these benchmark datasets.

Authors: Samujjwal Ghosh and Maunendra Sankar Desarkar

Affiliation: Indian Institute of Technology, Hyderabad, India

Title: Gold Standard Creation for Microblog Retrieval: Challenges of Completeness in IRMiDis 2017

Abstract: Microblogging sites like Twitter, Facebook, etc., are important sources of first-hand accounts during disaster situations, and have the potential to significantly aid disaster relief efforts. The IRMiDis track at FIRE 2017 focused on developing and comparing IR approaches to automatically identify and match tweets that indicate the need or availability of a resource, leading to the creation of a benchmark dataset for future improvements in this task. However, based on our experiments, we argue that the gold standard data obtained in the track is substantially incomplete. We also discuss some reasons why it may have been so, and provide some suggestions for making more robust ground truth data in such tasks.

Authors: Ribhav Soni and Sukomal Pal

Affiliation: Indian Institute of Technology (BHU), Varanasi, India

Title: SAVITR: A System for Real-time Location Extraction from Microblogs during Emergencies

Abstract: We present SAVITR, a system that leverages the information posted on the Twitter microblogging site to monitor and analyse emergency situations. Given that only a very small percentage of microblogs are geo-tagged, it is essential for such a system to extract locations from the text of the microblogs. We employ natural language processing techniques to infer the locations mentioned in the microblog text, in an unsupervised fashion and display it on a map-based interface. The system is designed for efficient performance, achieving an F-score of 0.81, and is approximately two orders of magnitude faster than other available tools for location extraction.

Authors: Ritam Dutt, Kaustubh Hiware, Avijit Ghosh and Rameshwar Bhaskaran

Affiliation: Indian Institute of Technology, Kharagpur, India

It is evident that the papers presented at the workshop included a mixture of works that developed algorithms, works that modelled human behaviour during emergencies, and works that developed systems for use during emergencies.

4 Dataset Released

As stated earlier, the SMERP workshop series aims to provide the research community with real-life datasets to foster research on the theme. To this end, we made available (upon request) a large dataset and invited interested participants to develop methods to solve some practical challenges over the dataset. Specifically, we made available the following resources related to a particular disaster event – the Nepal earthquake in April 2015: (1) A set of microblogs (tweets) posted during the event, (2) A set of images posted along with the tweets, and (3) A set of news media articles about the event, posted on various mainstream news media sites.

Interested participants could use this dataset to develop methods for tasks such as developing algorithms for multi-modal information extraction, e.g., identifying information about infrastructure damage from all the data sources.

We invite the research community to use the dataset for developing improved algorithms for utilizing social media at the times of emergency. Those who are interested to get the dataset should follow the procedure stated at the workshop website.

5 Panel Discussion

The workshop ended with a panel discussion on the upcoming practical challenges related to use of social media during emergencies. The panelists were our two invited speakers – Prof. Mitra and Prof. Sheth, Prof. Niloy Ganguly (IIT Kharagpur, India) and Prof. Moens (KU Leuven, Belgium; also a co-organizer of SMERP). Some of the questions discussed were: (1) How effective can models trained on prior events be during a future event?, (2) Can social media be combined with other sources of information for more effective extraction of critical information?, (3) How effective will be the use of geographical knowledge bases?

6 Conclusion and Future Directions

Further details of the SMERP 2018 workshop can be found at the site (<https://www.cse.iitk.ac.in/users/kripa/smerp2018/>). It is clear to us through the process of organizing this workshop that the topics related to the SMERP workshops are of significant interest to the research community. Hence, we plan to continue organizing SMERP workshops in the coming years.

Acknowledgements

We are thankful to the Web Conference (WWW) for providing the opportunity to organize the workshop. The peer-review track would not have been possible without the support of the Program Committee members – the full list is available on the workshop website. We are also thankful to the authors of the papers submitted to the workshop for considering SMERP.

References

- [1] Ritam Dutt, Kaustubh Hiware, Avijit Ghosh, and Rameshwar Bhaskaran. SAVITR: A system for real-time location extraction from microblogs during emergencies. In *Proc. of SMERP 2018*, 2018.
- [2] Samujjwal Ghosh and Maunendra Sankar Desarkar. Class specific tf-idf boosting for short-text classification. In *Proc. of SMERP 2018*, 2018.
- [3] Saptarshi Ghosh, Kripabandhu Ghosh, Debasis Ganguly, Tanmoy Chakraborty, Gareth J. F. Jones, and Marie-Francine Moens. ECIR 2017 Workshop on Exploitation of Social Media for Emergency Relief and Preparedness (SMERP). *SIGIR Forum*, 51(1):36–41, 2017.

-
- [4] Muhammad Imran, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. Processing Social Media Messages in Mass Emergency: A Survey. *ACM Computing Surveys*, 47(4):67:1–67:38, June 2015.
 - [5] Dheeraj Kumar and Satish Ukkusuri. Modeling evacuation behavior of nyc twitter users during hurricane sandy. In *Proc. of SMERP 2018*, 2018.
 - [6] Anastasia MOUNTZIDOU, Stelios Andreadis, Ilias Gialampoukidis, Anastasios Karakostas, Stefanos Vrochidis, and Yiannis Kompatsiaris. Flood relevance estimation from visual and textual content in social media streams. In *Proc. of SMERP 2018*, 2018.
 - [7] Ribhav Soni and Sukomal Pal. Gold standard creation for microblog retrieval: Challenges of completeness in irmidis 2017. In *Proc. of SMERP 2018*, 2018.
 - [8] Sarah Vieweg, Amanda L. Hughes, Kate Starbird, and Leysia Palen. Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proc. ACM SIGCHI*, 2010.