

SUMMARIZATION APPROACH OF SITUATIONAL INFORMATION FROM MICROBLOGS DURING DISASTER EVENTS: A SURVEY

Pooja B.Kawade¹ and Dr.Nitin N.Pise²

Abstract — The status and events updates posted to social networks, such as Twitter and Facebook, contain a myriad of information about what people are doing and looking. During events, such as disaster events, many updates are sent describing and expressing opinions about the event. As soon as natural disaster events happen, users are willing to know more about them. Twitter is a great source that can be exploited for obtaining such fine-grained structured information for fresh natural disaster events. Such events are often reported on Twitter much earlier than on other news media. This paper covers (i)An overall survey of different methods, techniques and systems using for tweet summarization and tweets extraction. (ii)Flow diagram. (iii)Proposed research direction.

Keywords – Disaster events, Twitter, situational information, classification, summarization

I. INTRODUCTION

Crisis situations such as disasters brought on by natural hazards present unique challenges to those who study them, creating conditions that call for particular research methods. In this paper, we survey methods for studying disasters from the perspective of information processing and management. We know that information posted to social media platforms in time and safety critical circumstances can be of great value to those tasked with making decisions in these fraught situations. Microblogging sites like Twitter have become imperative sources of real time information during disaster events. A natural disaster is a main inimical event resulting from natural processes of the Earth; examples include floods, hurricanes, tornadoes, volcanic eruptions, earthquakes, tsunamis, and other geologic processes. A natural disaster can cause loss of life or properties damages, and typically leaves some economic damage in its wake, the severity of which depends on the affected populace's resilience, or ability to recover and also on the infrastructure available[16]. A significant amount of valued situational information is available in these sites; however, this information is immersed among hundreds of thousands of tweets, generally containing opinion of the masses, that are posted during such events. To effectively exploit microblogging sites during disaster events, it is necessary to (i) Extract the situational information from among the large amounts of opinion, and (ii) Summarize the situational information. To help decision-making processes when time is critical. In recent years, microblogging sites such as Twitter have become important sources of real-time information,

¹ Department of Computer Engineering MIT, Pune, Maharashtra, India

² Department of Computer Engineering MIT, Pune, Maharashtra, India

especially during disaster events. Several recent research studies [13,14,15] have shown the importance of microblogging sites in improving situational awareness during such events[1]. During the course of several natural disasters in recent years, Twitter has been initiate to play an important role as an additional medium for many-to-many disaster communication. Emergency services are successfully using Twitter to inform the public about current developments, and are increasingly also trying to source first-hand situational information from Twitter feeds (such as relevant hashtag). The additional study of the uses of Twitter during natural disasters relies on the development of flexible and reliable research infrastructure for analysing Twitter feeds at scale and in close to real time[3]. Some agencies have even begun to formally incorporate social media monitoring and communication during mass emergency situations. The American Red Cross (ARC), in a survey, reported the effectiveness of social media and mobile apps. ARC recently opened their Social Media Digital Operations Center for Humanitarian Relief. The goals of this center are to “source additional information from affected areas during emergencies to better serve those who need help; spot trends and better anticipate the public’s needs; and connect people with the resources they need, like food, water, shelter or even emotional support”[9].

II. LITERATURE SURVEY

We have done a comprehensive literature survey. We have focused on work related to tweet summarization. Aforementioned research has shown that information which contributes to situational awareness is reported via Twitter. Table 1 gives the list of work on disaster events-

Table – 1 The List Of Work On Disaster Events

Paper Name, Year	Methods and Techniques	Conclusion
Tools and Methods for capturing twitter data during natural disaster 2013[3]	Tracking twitter through yourtwrapper, Advanced system for analyzing tweets.	They presented an approach to the tracking and analysis of Twitter user activities.
A Participant-based Approach for Event Summarization Using Twitter Streams 2013[8]	Hybrid TF-IDF used to extract the representative sentence from tweets	They generate event summaries using Twitter data streams.
On Summarization and Timeline Generation for Evolutionary Tweet Streams 2013[6]	Develop a TCV-Rank summarization technique for generating online summaries and historical summaries of arbitrary time durations.	Online tweet stream clustering algorithm to cluster tweets and maintain distilled statistics in a data structure called Tweet Cluster Vector (TCV)
Structured Information Extraction from Natural Disaster Events on Twitter 2014[5]	Three algorithms are used here.	In this paper, the problem of extracting structured information for natural disaster events from Twitter.
Efficient Online Summarization of Microblogging Streams 2014[4]	TOWGS, A word graph, along with optimization techniques	Improve the efficiency of this process by designing an online abstractive algorithm.
Extracting Situational Information from Microblogs during Disaster Events: a	Classification-summarization framework for disaster-specific situational information on twitter.	The authors have recommended working with tweet fragments rather than entire tweets. Distinct lexical and syntactic features present in

Classification-Summarization Approach 2015[1]		tweets can be used to separate out situational and non-situational tweets, which leads to significantly better summarization.
Open Domain Event Extraction from Twitter[12]	Twical	Open domain approach to extracting and categorizing events from status messages.
Summarizing Situational Tweets in Crisis Scenario2016[2]	An Integer-linear programming (ILP) based optimization technique and word based abstractive summarization technique to produce the final summary	They developed a complete system to generate summaries in real time from the incoming stream of tweets.

In [3], Axel Brun-2013, the authors have proposed a solution for tracking hashtags. YourTwrapperkeeper is the open-source tool. Building on PHP and MySQL, it draws mainly on the Twitter streaming API to track a number of keywords selected by its user, using the search API to fill any gaps which may exist in the data received from the streaming API.

In [8], Chao Shen-2013, propose a participant-based event summarization approach that “zooms-in” the Twitter event streams to the participant level TF-IDF approach to extract the representative sentences from a collection of tweets In this approach, each tweet was considered as a sentence. The sentences were ranked according to the average TF-IDF score of the consisting words; top weighted sentences were iteratively extracted, while excluding those that have high cosine similarity with the existing summary sentences.

In [6], Zhenhua Wang-2013, in the tweet stream clustering module, they design an efficient tweet stream clustering algorithm, an online algorithm allowing for effective clustering of tweets with only one pass over the data. This algorithm employs two data structures to keep important tweet information in clusters. The first one is a novel compressed structure called the Tweet Cluster Vector (TCV). TCVs are considered as potential sub-topic delegates and maintained dynamically in memory during stream processing The high-level summarization module supports generation of two kinds of summaries: online and historical summaries. (1) To generate online summaries, we propose a TCV-Rank summarization algorithm by referring to the current clusters maintained in memory.

In [5], Sandeep Panem-2014, three algorithms are used, namely: 1. Extracting Complete Attribute Names 2.

Extraction Of Fact Triplets 3. Extraction of Fact Triplets. These three algorithm are used here.

In [4], Olariu-2014, the Olariu introduces us to TOWGS, a highly efficient algorithm capable of online abstractive microblog summarization. A word graph, along with optimization techniques such as decaying windows and pruning is introduced.

In [1], Koustav Rudra-2015, a novel content-word based summarization approach (COWTS) to summarize the situational tweet stream by optimizing the coverage of important content words in the summary, using an Integer Linear Programming (ILP) framework. The authors have recommended working with tweet fragments rather than entire tweets. Distinct lexical and syntactic features present in tweets can be used to separate out situational and non-situational tweets, which leads to significantly better summarization.

In his work [12], Alan Ritter has introduced the first open-domain event-extraction and categorization system for twitter, named TwiCal. A scalable and open-domain approach to extracting and categorizing events from status messages.

In [2], Koustav Rudra 2016, an Integer-linear programming (ILP) based optimization technique and content word based abstractive summarization technique to produce the final summary. They have developed a complete system to generate summaries in real time from the incoming stream of tweets.

In this survey we have seen different methods and techniques. They have developed different systems that are used to extract relevant information from twitter, in Ritter et al.[2013] TwiCal system, in Ashktorab et al.[2014] tweetdr system, in Imran et al.[2014] AIDR system. These all systems are used to extract crisis relevant information from twitter.

III. FLOW DIAGRAM

This paper covers the survey of summarization approach of Situational Information from Microblogs during Disaster Events. Proposed Flow Diagram from the survey has shown in fig1.

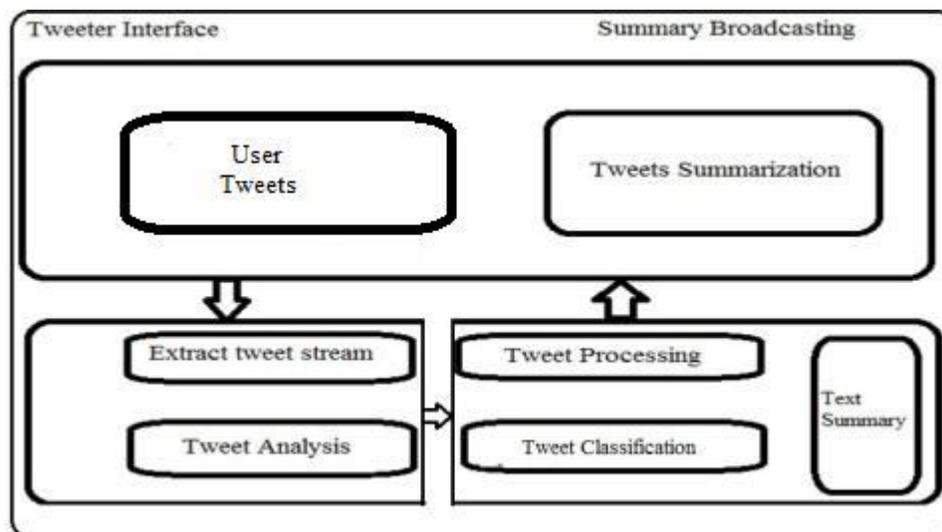


Figure1. Proposed Flow diagram from survey

- User Interface-Collect the user tweets along with the all user details.
- Tweet Extraction-Extracting tweets from tweet set.
- Tweet Analysis-Analyzing Extracted tweets to normalize.
- Tweet Processing- Applying Stemming tweets as per the analyzed log to further sub Classification
- Tweet classification-User tweet are classified on the basis of tweet events. For example: Bomb blast, Floods.
- Text summary-Generating text summery from stored tweets to all tweets.
- Tweet Summarization-Generating tweet summary in the form of summarized statement.
- Broadcasting Tweet.

Fig1. shows a flow diagram with three major steps-(1) Tweet extraction (2) Analysis (3) Summarization. The Input will take as a tweeter user tweets and Output will be the Summary from tweets

IV.CONCLUSION

This paper covered 1.The survey of different methods, techniques and systems 2.Proposed Flow diagram from survey. From this survey, research direction can be proposed- (i)A new framework in which first extract tweets from different situational tweet data. In two stage summarization framework, first getting a set of important tweets from the large set of information and then follows a word graph and content word based abstractive summarization technique to create the final summary. Additionally for security, use URL spamming detection to remove the fake URL tweets. Because work on fake tweets will give fake or false summarization result. (ii)Extract situational tweets with the help of AIDR system and use Hybrid TF-IDF to generate final summary. This research is useful for the public or formal response organizations, that it has the prospective to save lives or property during an emergency, remains to be seen.

REFERENCES

- [1] Koustav Rudra, Subham Ghosh, Niloy Ganguly, Extracting Situational Information from Microblogs during Disaster Events: a Classification- Summarization Approach. CIKM15, Melbourne, VIC, Australia,2015.
- [2] Koustav Rudra, Siddhartha Banerjee, Niloy Ganguly, Summarizing Situational Tweets in Crisis Scenario. HT 16, Halifax, NS, Canada,2016.
- [3] Axel Bruns, Yxian Liang, Tools and Methods for capturing twitter data during natural disaster. In First Monday, Volume 17, Number 4-2April 2013.
- [4] Andrei Olariu, Efficient Online Summarization of Microblogging Streams. In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, pages 236-240, April 26-30 2014 Gothenburg, Sweden.
- [5] Sandeep Panem, Manish Gupta, Vasudeva Varma, Structured Information Extraction from Natural Disaster Events on Twitter. KDD'14, xian, china, 2014.
- [6] Zhenhua Wang, Lidan Shou, Ke Chen, Gang Chen, On Summarization and Timeline Generation for Evolutionary Tweet Streams. In IEEE Transactions on Knowledge and Data Engineering, 2013, DOI 10.1109/TKDE.234-5379,2014.
- [7] Muhammad Imran, Fernando Diaz, Carlos Castillo, Processing Social Media Messages in Mass Emergency: A Survey. ACM Computing Surveys, Vol. 47, No. 4, Article 67,2015.
- [8] Chao Shen, Fei Liu, Fuliang Weng, Tao Li, A Participant-based Approach for Event Summarization Using Twitter Streams,2014.
- [9] Muhammad Imran, Carlos Castillo, Ji Lucas, AIDR: Artificial Intelligence for Disaster Response. In WWW14 Companion, April 7-11, 2014, Seoul, Korea.
- [10] Sarah Vieweg, Carlos Castillo, and Muhammad Imran, Integrating Social Media Communications into the Rapid Assessment of Sudden Onset Disasters. In Springer LNCS 8851, pp. 444-461 2014.
- [11] Miles Osborne, Elizabeth Cano, Craig Macdonald, Real-Time Detection, Tracking, and Monitoring of Automatically Discovered Events in Social Media,2014.
- [12] Alan Ritter, Mausam, Open Domain Event Extraction from Twitter. In KDD12, August 12-16, 2012, Beijing, China.
- [13] Robert Power, Bella Robinson, John Colton, Emergency Situation Awareness. Twitter Case Studies. In Springer ISCRAM-med, LNBP 196, pp. 218-231,2014.
- [14] Pengyi Zhang, Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake. In CSCW 2011, March 19-23, Hangzhou, China.
- [15] https://en.wikipedia.org/wiki/Natural_disaster (Visited On 23/10/2016).