Establishing effective communications in disaster affected areas and artificial intelligence based detection using socialmedia platform

Project tittle : Natural Disasters Intensity Analysis and Classification using Artificial Intelligence

Abstract:

Floods, earthquakes, storm surges and other natural disasters severely affect the communication infrastructure and thus compromise the effectiveness of communications dependent rescue and warning services. In this paper, a user centric approach is proposed to establish communications in disaster affected and communication outage areas. The proposed scheme forms ad- hoc clusters to facilitate emergency communications and connect end-users/ User Equipment (UE) to the core network. A novel cluster formation with single and multi-hop communication framework is proposed. The overall throughput in the formed clusters is maximized using convex optimization. In addition, an intelligent system is designed to label different clusters and their localities into affected and non-affected areas. As a proof of concept, the labeling is achieved on flooding dataset where region specific social media information is used in proposed machine learning techniques to classify the disaster-prone areas as flooded or unflooded. The suitable results of the proposed machine learning schemes suggest its use along with proposed clustering techniques to revive communications in disaster affected areas and to classify the impact of disaster for different locations in disaster-prone areas.

Keywords: Ad-hoc networks, heterogeneous networks (HetNets), social sensors, infrastructure less communications, machine learning. 5G, device to device (d2d), boosting classifiers

Introduction :

Natural hazards and catastrophes can significantly interfere and effect people's life, property and socioeconomic cycle. Natural hazards can be categorized into three main classes [1]: meteorological hazards, hydrological hazards and geological hazards. These hazards are depicted in Figure 1. Meteorological events are composed of tornados, hurricanes, thunderstorms, winter storms (ice storms) and summer storms (wildfire). Hydrological events consist of all flood types (fluvial, pluvial, and coastal), storms surges and tsunamis. Geological hazards comprise of earthquakes, volcanic eruptions and mass movements (land sliding, mudflows, avalanches etc.).

In the event of such hazards, disaster management play a vital role. Disaster management requires preparedness and timely responses to mitigate and recover normal state of living along with increasing resilience of society. However, limited capacity of the governmental institutes, non-governmental organizations (NGOs), first responders and rescue workers to effectively execute help and rescue services for the affected regions in natural disasters pose several challenges

The power failures in the affected regions result in inability of affected population to receive telecasts (one-way simplex communications for general instructions). The devastating impact of natural catastrophes on communication backbone networks disturb both wired and wireless links (duplex communications) alike. Lack of effective communications between the inhabitants of affected areas and the rescue workers/first responders, among various rescue teams and with the rescue operation coordinating.

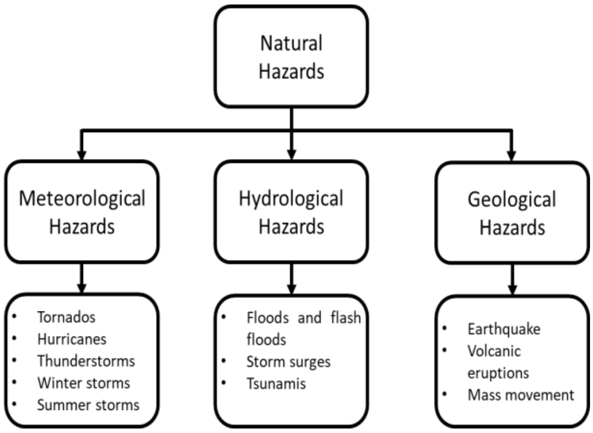


Figure 1 Classification of natural hazards

The importance of communications is evident in the event of a natural disasters [7, 8]. However, the lowresilience of traditional cellular infrastructure and inability to operate in severe circumstances makes it more vulnerable and fault intolerant. Besides, natural disasters also cause failure in operations of Base Station System(s) (BSS) and affects the communications of all connected Mobile Stations (MS) in the region. For instance, the 2011 earthquake in Japan caused over 16 thousand deaths while thousands were injured and missing [9]. The communications infrastructure including major coastal transmission lines and utility poles were heavily damaged, causing a shutdown of approximately 29,000 base stations [10].

The centralized hierarchy supported by the legacy cellular systems is especially affected in the areas in need of rescue services. Therefore, decentralized emergency communication frameworks are required to offer necessary feedback communication within the affected areas [37]. Most of the existing efforts in restoring disaster affected communications require deployment of remote BSS which restricts the suitability of such communication networks. Furthermore, the robustness and scalability in such systems is usually unattainable. Localization of the MS for accurate region-wise threat analysis and statistical feedback of trapped survivors is also a challenge. The bulk information generated in communication functional areas (where communication networks are restored or survived the disaster) is also hard to analyze using conventional techniques. Machine intelligence can be used to better read the situation in different disaster affected areas and characterize level of emergency in different regions based on the communications originating from that region.

This paper proposes a location-based ad-hoc network formation mechanism to restore necessary communications of UE/MS in case of failure in core communication infrastructure. The main contributions of the paper include

1) Single-hop and Multi-hop communications link establishment to the core network in the event ofcommunication infrastructure collapse
2) Efficient cluster formation and optimized cluster-head (CH) selection
3) Throughput optimization within the clusters
4) Machine learning based data analysis for identification of disaster situation in different regions
5) Accurate localization of the disaster affected areas with high performance

Related work:

The centralized architecture in cellular network make it vulnerable to large scale communications outage in the disaster affected areas. Under such circumstances, the device to device (d2d) and localized communications can play a vital role in restoring the necessary communications. Over the years, several schemes were proposed to offer sustainable communications in pre-disaster as well as post-disaster phases.

The inability of existing communication infrastructure to cope with extreme unprecedented events cause complete communication infrastructure breakdown. Gomes et al. [16] extensively reviewed the communication strategies in pre-disaster and post-disaster scenarios and highlighted the limitations of existing systems with possible improvement suggestions for robust communication infrastructure. It was suggested that the network solution requires added redundancy, content connectivity and traffic connection management to provide robust communication in pre-disaster scenarios. Added redundancy ensures accessibility of a UE to other UE within the network by introducing immunization strategies to provide extra protection to the network. Content connectivity allows content delivery to each node in disconnected network during a catastrophic event.

Over the years, several ad-hoc schemes are proposed to facilitate communications in disaster affected and communication outage areas. While some of the schemes lack scalability, others require dedicated equipment and additional resources, not readily available in disaster affected areas. In addition, almost all of the techniques lack the ability to identify region-wise vulnerability of disaster affected area. In the proposed work, a machine learning based vulnerability analysis of disaster prone/affected areas is evaluated along with a suitable ad-hoc communication infrastructure to cope with the adverse effects of communication failure in such regions. The proposed communication infrastructure enables instantaneous clustered network establishment to facilitate the communications where the cellular and other means of communications have failed. In addition, as a proof of concept, a machine learning technique is proposed to evaluate social media to assess vulnerability of people in disaster areas with potential extension to region wise susceptibility. The details of the proposed communication infrastructure and machine learning technique are covered in the following sections.

Proposed communication infrastructure in disastrous Situations:

The communications are essential for both first responders and civilians trapped in the disaster-stricken areas. In critical circumstances, failure in communications can affect the rescue services and limit the effectiveness of first responders in the affected areas. The rescue activities of the first responders and rescue workers are very important in minimizing further loss of life and injuries and therefore, the necessity of resilient communication networks cannot be denied.

In the event of a natural disaster, the electrical power and the communications infrastructure can be seriously affected. The inability to communicate affects coordination among the trapped survivors, rescue workers and outside world. In the event of a failure in cellular infrastructure, adaptive d2d and multi-hop communications

clustering and communications scheme is proposed to facilitate communications within the disaster affected and communications outage areas. The system parameters are presented in Table 1, whereas a detailed description of proposed scheme is as follows.

Table 1 System Parameters

Parameters	Variable(s)	Value(s)
Control channel Superframe duration	T_{sf}	4.83ms
Transmission slots	m	8
Receiver sensitivity	δ_r	-95 dBm
Desired SNR margin	SNR_m	10 dB
Received Power	P_r	-
Transmitted power	P_t	-
Transmitter gain	G_T	3dB
Receiver gain	G_R	3dB
Transmission power multiplying factor	g	1-3
Extended range of control channel	d_e	-
control channel band	f	900/1800/2100 MHz
Path-loss exponent	n	2-4
Average Smart Phone (SP) penetration	μ_1	
Expected deviation in average SP	σ_1	
Transmitted power of the satellite	P_s	
Noise Power	$\sigma_n{}^2$	
GPS coordinates: x-y plane mean location and expected deviation	$\mu_x, \sigma_x, \mu_y, \sigma_y$	
Bivariate gaussian distribution symmetry	ρ	0
Distance vector between j and k	d_{jk}	
Hop-count to the core network at time t	$H_{count}(t)$	

Extended coverage establishment:

In the event of a failure in one or more cellular BSS, the active BSS, near the communications outage area will initiate a high-power clustering request using the prespecified control channel. The proposed scheme uses synchronized communications from active BSS situated next to the disaster affected areas, allowing to extend its coverage to the communications dead zone. The extended coverage area for the control channel is presented in Figure 2.

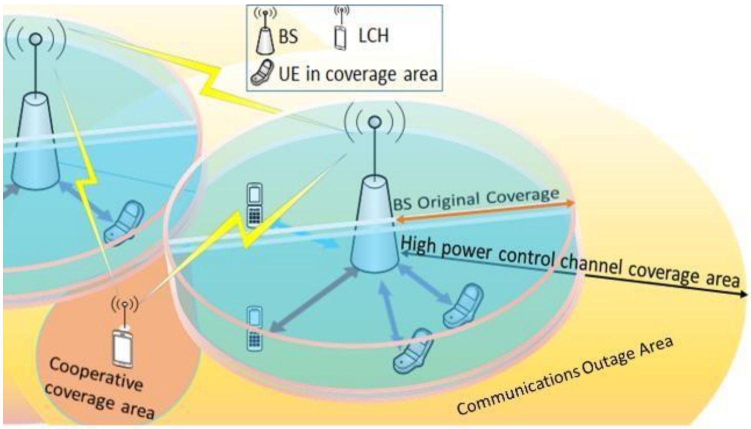


Figure 2 High power control channel coverage and cluster-head (CH) selection

The extended coverage range of the BSS is trigged in the event of a disaster. The extended coverage range is achieved by introducing two changes in control channel communications: 1) use of relatively low frequency and 2) high transmission power. The justification for the extended range can be validated from Friis' Equation which suggests the received power can be expressed as [38, 39]

$$P_r = \left(\frac{P_t G_T G_R \lambda^2}{(4\pi d_e)^2}\right) \tag{1}$$

$$d_e = \sqrt[n]{\left(\frac{gP_tG_TG_R\lambda^2}{(4\pi)^2\delta_r}\right)} \qquad (2)$$

here n = 2, and gP_t is g times original transmission power for extended coverage area.

Control communications and scheduling

Figure 2 highlights the cooperative coverage area (covered by more than one BSS). To avoid interference, time division multiple access (TDMA) is used to establish control channel communications, where the communications from different BSS are adaptively scheduled. An example scenario is presented in Figure 3, where BS1, BS2 and BS3 are the active BSS in the neighboring region of communications outage area. The hexagonal cells in Figure 3 represent original coverage area where extended coverage area is represented with circular coverage areas. The control channel communications schedule of these BSS is also presented in Figure 3 where these base stations cover the areas CA1 (Coverage area of BS1), CA2 (Coverage area of BS2), CA3 (Coverage area of BS3), CC12 (Cooperative coverage area of BS1 and BS2) and CC23 (Cooperative coverage area of BS2 and BS3). The slotted broadcast of the BSS uses a control channel frame of duration T_{sf} with m transmission slots, which can cover all possible aspects of collision and failure scenarios in hexagonal cellular structure.

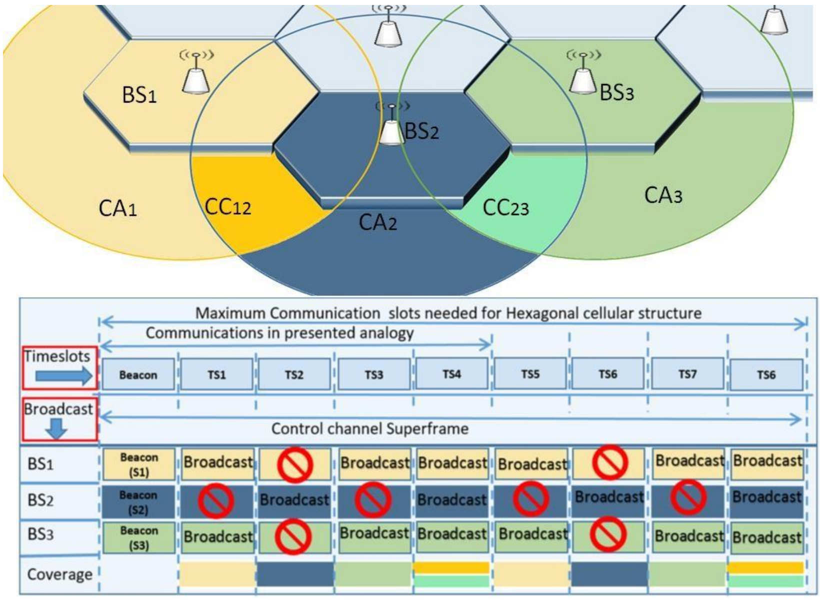
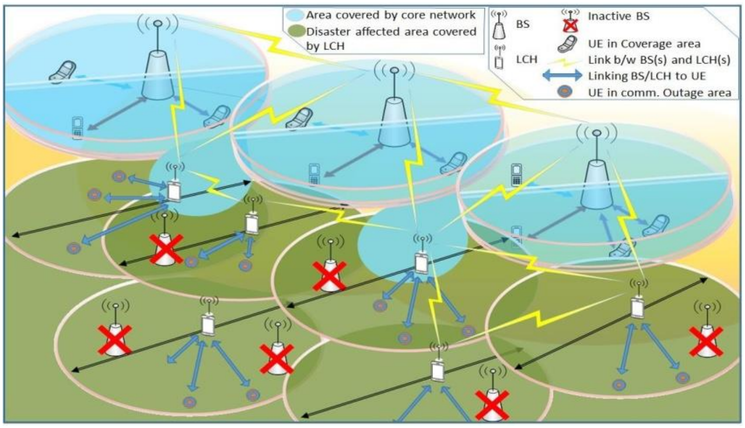


Figure 3 Control channel communication synchronization

Cluster-Head Selection and MS distribution:

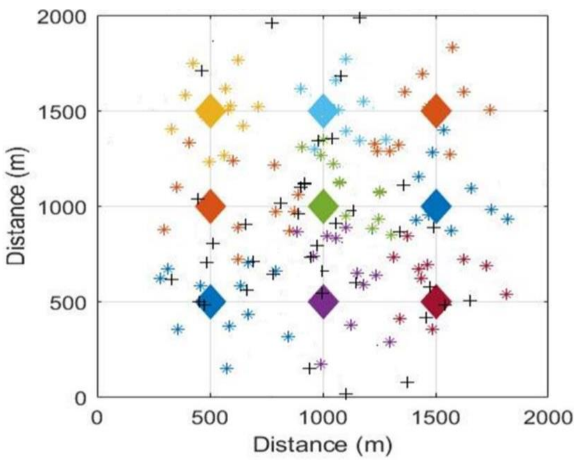
The overall disaster area can be presented as shown in Figure 4. As represented in the figure, only the smart phones can serve as a CH for obvious reasons of having extended features. Note that the initial criteria for being a local cluster-head (LCH) is to be in the extended transmission range of an active BSS in first instance. The potential CHs in suitable extended coverage location are then evaluated based on four parameters: 1) Hop count, 2) delay, 3) remaining battery and 4) received signal strength. A weighted voting of all these components is used to decide which CH is more suitable. In this study, equal voting isused to evaluate smart phone's suitability to become CH. Finally, the potential CHs are compared and most suitable of all is



. It is worth noticing that the threshold, (i), of i^{th} iteration of setup mode is given by

$$\psi(i) = \sum_{j=1}^{ci} C_{head}(j) \Big/ ci \tag{10}$$

here i is the number of times the setup mode is re-run. ci is the total number of MS in the network. Each MS designated as LCH serves as a BSS to penetrate further in the communications outage area and broadcast clustering beacon using designated control channel to allow further clustering as represented in Figure 5.



Machine learning based region wise disaster severity detection using social media platform:

A machine learning based disaster severity evaluation in different regions using social media platform is proposed along with the communication model for disaster affected areas. The fundamental objective of analyzing social media based platform in this study is to perform pre-disaster vulnerability prediction and localize disastrous areas until the communication infrastructure sustains. The role of proposed AI- empowered disaster detection system using social media is to assist government officials and emergency services to identify threats and overall vulnerability. The proposed machine learning scheme evaluates impact in different disaster affected areas where severity of underlying conditions is analyzed based on messages/pics/sentiments from the social media users within a certain geographical boundary. It is assumed that users' approximate location is known. Further details of the proposed machine learning technique and implemented system are presented in following subsections.

Dataset

The dataset used in this study in obtained from CrisisLex [34]. It is a data repository that uses social media platform (i.e. twitter) to understand the presence of disaster within a geographical boundary. CrisisLex contains the dataset of natural disasters (floods, hurricanes, tornados) as well as manmade disasters (explosions, bombings). This study utilizes only dataset of natural disasters and flooding in particular. The flooding dataset refers to 2013 flooding occurred in Alberta, Canada. The dataset was labeled using the crowd flower data annotation platform where each data sample (tweet) was labelled to either non- flood (irrelevant to Alberta flooding) or flood (relevant to Alberta flooding).

Pre-processing and feature-extraction of the dataset

The flood related social media messages obtained through twitter require extensive pre-processing before any feature extraction. It is essential in most of the natural language processing tasks to omit the unrelated information and to reduce the redundant information. For this purpose, regular expression (RE) operations and natural language toolkit (NLTK) [35] libraries have been used in python to remove the special characters, hashtags, URLs, punctuations, stop words. The resulted dataset is processed further through stemming and lemmatization operations, which helps to derive the root or basic form of words that can eventually ease the feature-extraction and further processing. Stemming is followed by feature extraction which is obtained through bag of words. This is a process to compute features from the text in such a way that the machine learning algorithm can be implemented to extract the meaningful patterns from the dataset. Bag on words consist of two steps:

Results and Discussion:

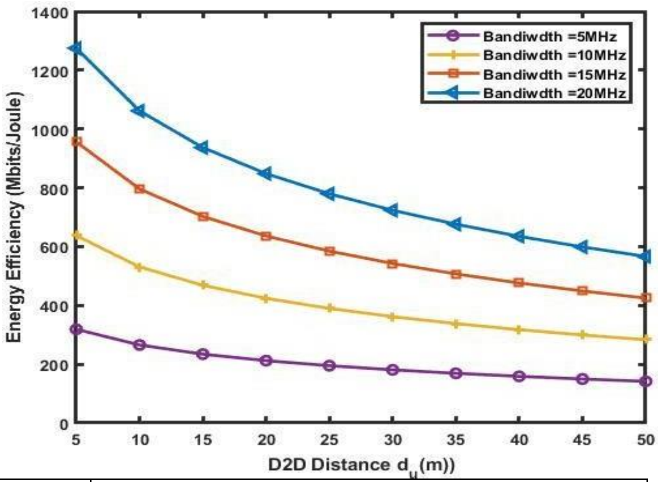
Performance evaluation of proposed communication infrastructure

The performance of the proposed clustering scheme is evaluated based on number of MS/UEs affiliated with the LCH, system throughput and energy efficiency. In the proposed scheme, the affiliation of UE is conditional to the throughput optimization, where if two UE can be affiliated to two LCHs, the selection will be made based on which UE's affiliation to which cluster maximizes the throughput. The results of the proposed throughput optimization scheme are also compared with the state-of-the-art distance-based affiliation scheme (D-Allocation) where affiliation to a cluster is dependent on the Euclidean distance.

Performance evaluation of machine learning based disaster detectionsystem:

The performance of proposed MNB and XGB is evaluated to check suitability of proposed methods in classifying flood relevant and flood irrelevant tweets. Both classifiers i.e. MNB and XGB performed wellin classifying the non-flood and flood related messages with an accuracy of above 90%. The achieved accuracy to differentiate flooding vs non-flooding events shows the strength of proposed system in localizing the flood within a geographical boundary. The XGB classifier performed better than MNB classifier with performance of 95.80%, while the MNB classifier achieved an accuracy of 91.22%. The confusion matrices obtained for MNB and XGB classifiers are presented in Table 3 and Table 4 respectively.

Table 3. Confusion matrix of MNB classifier



Accuracy 91.22%	Predicted class		
Actual class	**Classified as →**	**non-flood**	**flood**
	non-flood	1224	205
	flood	55	1476

95.80%			
Actual class	**Classified as →**	**non-flood**	**flood**
	non-flood	1365	64
	flood	60	1471

These findings show the strength of machine learning algorithm in detecting regional floods through social media platform. Although, this work only uses the twitter platform as a data source due to the unavailability of other means of data gathering mediums such as short messaging services (SMS), audio and video messages, yet the proposed method can easily be extended to mobile platforms using cellular networks, where flood affected inhabitants can inform the regional authorities and rescue services throughSMS and other means.

The SMS data source is much more realistic and effective than twitter in instances where communication infrastructure is already collapsed. In such conditions, proposed machine intelligence based data processing pipeline can also help in localizing the emergency cases in severely affected area. Nevertheless, twitter-based pre-disaster vulnerability detection and finding the epicenter of the disaster for emergency aid are key contributions of such systems, Moreover, the proposed data processing paradigm can easily be translated to other natural disasters such as earthquake, hurricanes,

Conclusion:

Natural disasters significantly impact natural habitat and socioeconomic system. The significance of communications in effective post-disaster support and rescue, is undeniable, therefore, in this work, an ad-hoc cluster formation framework was proposed to establish communications in disaster affected and communications outage areas. The work also incorporated convex optimization for throughput enhancement and localization of UE for novel machine learning based identification of disaster struck areas. The proposed work effectively establishes communication infrastructure to facilitate communications in the affected areas. In addition, the proposed machine learning scheme assists in identifying critical regions in the affected areas by analyzing bulk information through social messaging platform. The results have been very favorable and offered improvement in throughput and serviceable users in newly formed clusters in comparison to state-of-the-art. A relatively high accuracy of above 95% was also achieved on CrisisLex flooding dataset while using the proposed machine learning based approach. However, there are certain limitations of the work. The proposed cluster formation scheme suffers from low energy efficiency and in future can be further improved with the help of suitable transmission power control and energy optimization scheme.

References:

[1] E. Keller, D. DeVecchio, Natural hazards: earth's processes as hazards, disasters, andcatastrophes. in Pearson Higher Education AU, (2015).

[2] F.A. Turjman, F Cognitive routing protocol for disaster-inspired internet of things, Future Gener. Comput. Syst. 92 (2019)1103-1115. https://doi.org/10.1016/j.future.2017.03.014

[3] B. Varghese, R. Buyya, Next generation cloud computing: New trends and researchdirections. Future Gener. Comput. Syst. 79 (2018) 849-861. https://doi.org/10.1016/j.future.2017.09.020

[4] Y. Nakayama, K. Maruta, T. Tsutsumi, K. Sezaki, Wired and wireless network cooperation for wide-area quick disaster recovery, IEEE Access. 6 (2017) pp. 2410-2424. https://doi.org/10.1109/ACCESS.2017.2783050

[5] M. Erdelj, M. Król, E. Natalizio. Wireless sensor networks and multi-UAV systems for natural disaster management. Comput. Netw., 124 (2017) 72-86. https://doi.org/10.1016/j.comnet.2017.05.021

[6] A. Izaddoost, S. S. Heydari, Risk-adaptive strategic network protection in disaster scenarios, J. Commun. and Netw., vol. 19 (2017) 509-520. https://doi.org/10.1109/JCN.2017.000082

[7] A. Merwaday, A. Tuncer, A. Kumbhar, I. Guvenc. Improved throughput coverage in natural disasters: Unmanned aerial base stations for public-safety communications. IEEE Vehic.ular Tech. Mag., 11(2016), 53-60. https://doi.org/10.1109/MVT.2016.2589970