Utilizing Online Social Media for Post-Disaster Relief

Saptarshi Ghosh
Department of CSE, IIT Kharagpur
http://cse.iitkgp.ac.in/~saptarshi/

ACM IKDD CODS-COMAD, January 2018
Disasters: man-made & natural

Paris terrorist attacks  Nepal earthquake  Chennai floods
Lack of situational information, mismanagement

No coordination in relief efforts

“Often there are enough resources, but no one knows what is needed where”

Discussions with victims, volunteers, NGOs related to Nepal earthquake, Chennai floods

Need mechanisms for gathering situational information

Automated systems for coordinating relief efforts
From where to get information?

Utilizing social intelligence of people at disaster site (social sensors)

Ubiquity of smartphones, mobile Internet
Real-time dissemination of information
WHAT IS POSTED ON OSM DURING DISASTERS?
From the point of view of agencies responding to a disaster

- The good stuff
- The bad stuff
- The ugly stuff
From the point of view of agencies responding to a disaster

- The good stuff
- The bad stuff
- The ugly stuff
BBC Breaking News @BBCBreaking
Update: 152 people killed and 2,600 injured in #Sichuan earthquake, Chinese officials say bbc.in/14Dd1EL pic.twitter.com/WaeQIS2yl7

Retweeted 1846 times

Expand  Reply  Retweet  Favorite  More
Tweets to organize donations

**CHIRANJEEVI @mega_fanz**
Mumbai Blasts 2011: Emergency Numbers: Mumbai, Jul 13: Maharashtra Chief Minister Prithvir... [link]

**Paula Jones @tensilica**
RT @bfuller9: Good place to start out to help victims of #haitiearthquake is Google: [link] Complete with satellite imagery

**Haiti Earthquake @haitiquake2010**
Help for #Haiti after 7.0 Earthquake near Port-Au-Prince, donations needed [link] #haitiearthquake
Hi A2 Mahalakshmi Flats, 64 65 Jones Road Saidapet <mobile number>. Kids don’t have drinking water. Please help.

HELP 5 adults+ 80 yrs old & 1.5 yr old stuck in the first floor No 12, first cross st, Vijayaraghava road, T Nagar. Contact: <mobile number>

#ChennaiRainsHelp A friends wife, 4 wk baby, family on terrace at 10/1 Janakiraman st, muthurangamblock, jafferkhanpet <mobile number>
Tweets providing help / resources

Drinking water in Chennai (thiruvanmiyur) available. 60 1 litre bottles, 20 packets. Contact @samyu_d <mobile number>

Willing to do recharge for chennai numbers. Anyone needed please message me <mobile number> #ChennaiRainsHelp

Volunteers are needed to distribute food packets at North Chennai pls contact aravind <mobile number>
From the point of view of agencies responding to a disaster

- The good stuff
- The bad stuff
- The ugly stuff
Opinion and sentiment

arjun bhadana
@arjunbhadana

#HyderabadBlasts terror has no religion, sick ppl who worship violence n chaos, who wanna rule by fear and division of ppl, don't let them

10:31 पूर्वोत्तर - 22 फरवरी 2013

Kiran Bedi
@thekiranbedi

@sardesairajdeep #Hyderabadblasts VIPs must avoid going to site immed after. Their visits divert attention and focus!

KitesRunner
@Raj_saiifi

Prayers n only prayers required for d victims n the ppl who lost loved ones. hyderabad blasts. May Allah Bless All
#Hyderabadblasts peace

10:32 पूर्वोत्तर - 22 फरवरी 2013
Opinion and sentiment

Bad from the perspective of agencies responding to a disaster:

Constitutes a vast majority of the posts

Of limited use for responding to the disaster
From the point of view of agencies responding to a disaster

- The good stuff
- The bad stuff
- The ugly stuff
More than 40 crocodile escaped from park at Chennai due to overflow of water. On the roads of ecr side. Vellachery
#ChennaiFloods
Communal tweets

if u hear terrorist shouting, allah hu akbar, & killing innocents den u must not retaliate coz u'll b termed as islamophobe
#gurdaspurattack

If u think bringing any “persecuted Christians” into America from Syria and no terrorists will slip through, you’re a f**ing idiot. #ParisAttack

Who made kashmir the heaven? they r Kashmiri Pandits, Muslims made it Hell. #SaveKashmiriPandits #kashmirFloods
Our motivation

- Extracting the good stuff and the ugly stuff from among the huge amounts of bad stuff

- Utilizing the good stuff (e.g., generating summarized reports, identifying needs and availabilities of resources)

- Countering the ugly stuff
OUR WORKS

... ON THE GOOD STUFF
Collaborators

- Professor Niloy Ganguly, IIT Kharagpur
- Professor Somprakash Bandyopadhyay, IIM Calcutta
- Dr. Pawan Goyal, IIT Kharagpur
- Dr. Arnab Jana, IIT Bombay
- Dr. Kripabandhu Ghosh, IIT Kanpur
- Students: Koustav, Moumita, Anurag, Prannay, ...

Acknowledgement: partial funding from
  - ITRA, Media Labs Asia and DeITY, GOI
  - Microsoft Research India
EXTRACTION AND SUMMARIZATION OF SITUATIONAL AWARENESS (SA)
## SA vs. Non-SA classification

<table>
<thead>
<tr>
<th>Features</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Count of numerals</strong></td>
<td><strong>Count of Intensifiers</strong></td>
</tr>
<tr>
<td>urgent blood required at omni hospital please contact 8096369999 040 67369999</td>
<td>Blasts - sadly, one of the very, very few things that put us on world news!</td>
</tr>
<tr>
<td><strong>Count of wh-words</strong></td>
<td><strong>Presence of Emoticons and Exclamations</strong></td>
</tr>
<tr>
<td>Why don’t you submit your coalgate scam money to disaster</td>
<td>RT @DavaS12: RIP all you Angels! #sandyhook <a href="http://t.co/yrGA39Pc">http://t.co/yrGA39Pc</a></td>
</tr>
<tr>
<td><strong>Count of personal pronouns</strong></td>
<td><strong>Presence of Question Marks</strong></td>
</tr>
<tr>
<td>its just few km frm my place n i cant reach my people</td>
<td>another school shooting? what kind of sick mentality is this?? #blast</td>
</tr>
<tr>
<td><strong>Presence of Modal Verbs</strong></td>
<td><strong>Fraction of Subjective Words</strong></td>
</tr>
<tr>
<td>#HyderabadBlasts we're talking abt it we will forget this in few days.it should not end here.we should find out</td>
<td><strong>Count of non-situational Words</strong></td>
</tr>
<tr>
<td>(condemn, donate, sympathize)</td>
<td>(condemn, donate, sympathize)</td>
</tr>
<tr>
<td>Pm condemns hyderabad blast</td>
<td>Pm condemns hyderabad blast</td>
</tr>
</tbody>
</table>

**Classifier achieved around 80% accuracy**
Summarizing situational updates

- Content words: particular types of informative words important during disaster
  - Numerals (e.g., number of casualties, helpline nos.)
  - Nouns (e.g., names of landmarks, people, hospital)
  - Main Verbs (e.g., killed, injured, stranded)

- Proposed **ILP-based framework** that maximizes the coverage of important content words in the summary
Baseline summarization methods

- **RTS**: Outlier based event detection and generating summary of tweets with high weightage [Hypertext 2012]
- **Sumblr**: Clustering-based tweet stream summarization algorithm [SIGIR 2013]
- **DIS**: Affinity clustering based supervised disaster related news articles summarizer [KDD workshop 2014]

- Gold standard summaries generated by human volunteers
- All methods used to generate summary of same number of words
- Summarization evaluated using ROUGE scores

Proposed methodology performed better than all baseline approaches, for tweet streams posted during several recent disasters
Extracting & Summarizing SA

Initial work on English tweets only:

Extracting Situational Information from Microblogs during Disaster Events: A Classification-Summarization Approach
ACM Conference on Information and Knowledge Management (CIKM) 2015

Extended for English and Hindi tweets:

Extracting and Summarizing Situational Information from the Twitter Social Media during Disasters
Accepted in ACM Transactions on the Web 2018
Are the summaries practically useful?

- Summaries are better than those produced by other summarization algorithms
- But, are the summaries practically useful?

- Feedback on the summaries from NGOs
  - Summaries lack actionable information
  - Need summaries of specific types of information

- What is most frequent is not the most useful
Typical information requirements during disaster relief operations

- What resources are required, and where?
- What resources are presently available (or will soon be available), and where?
- What infrastructure damages, casualties are being reported?
- What are the on-ground activities of different Government organizations / NGOs?
Attempts to extract actionable info

- Shared task in FIRE 2016
- Data challenge as part of ECIR 2017 Workshop on Exploiting Social Media for Emergency Relief and Preparedness (SMERP 2017)
- Participating teams submitted results, which are evaluated using manually generated gold standard
Attempts to develop IR methods

- Different methods tried
  - Traditional NLP, e.g., Named Entity Recognition, WordNet
  - IR methods – query expansion, Indri, Lucene
  - Neural network methods – Word2vec, Glove
  - Semi-automatic methods involving manual selection of query terms / training on manually extracted data
Attempts to develop IR methods

- Best performing automatic method:
  - MAP: 0.18, Precision@20: 0.44
  - Word2vec, query expansion, Named Entity Recognition

- Why such poor performance?
  - Variable vocabulary of tweets – expected keywords often missing (different ways of framing a tweet)
  - Variations in spelling
  - Arbitrary shortenings (140 character limit)
Variable vocabulary of tweets

Sample tweets reporting infrastructure damage:

- Kathmandu-Lamjung road cut off after earthquake. Follow live updates

- Historic Dharahara Tower in #Kathmandu, has fallen #quake

- Building Collapsed along with fallen electric pole in Golmadhi, Bhaktapur-7
Spelling variations

#Sindhupalchowk 1100+deaths and 99% Houses are Down

Indian national Azhar 23, missing. Last location Sindhupalchok. Plz help.

Food Distribution in sindupalchowk, sufficient for 7 days for 500 victims
Foreign Secy & Defence Secy giving latest updates on earthquake relief [url]

4 planes to leave for #Nepal tmrw carry meds, med team, 30-Bed Hospital

Nepal quake stresses importance of earthquake resistant bldg designs in entire NCR.
Addressing the challenges

Variable vocabulary of social media

Spelling variations & arbitrarily shortened words

Use neural network models
- Contextual stemming to unify variants of words
- Capture semantic context
UNIFYING WORD VARIATIONS
(CONTEXTUAL STEMMING)
Measuring similarity of two words

- **String similarity**
  - Length of common prefix (has to be at least $p \sim 2$
  - Longest Common Subsequence of the words

- **Contextual similarity**
  - Word2vec to get local word vectors which are expected to capture the local context
  - GloVe pre-trained model on billions of tweets to capture global context of words
  - Cosine similarity of the word vectors
For a word $w$

- Identify a group of words $G_w$ having sufficient string similarity and contextual similarity – candidate stems
- The word in $G_w$ having minimum length is chosen as the stem of the set $\{w \cup G_w\}$
Contextual stemming algorithm

- Experiments over tweets posted during disasters as well as standard microblog datasets
- Proposed stemming enables significantly better retrieval compared to several baseline stemmers

A Novel Word Embedding based Stemming Approach for Microblog Retrieval during Disasters, ECIR 2017

Combining Local and Global Word Embeddings for Microblog Stemming, ACM CIKM 2017
IDENTIFYING NEEDS AND AVAILABILITIES OF RESOURCES
Focus on two specific types of info

- Two types of information which can critically help coordinate post-disaster relief operations
  - What resources are needed
  - What resources are available, or likely to be available soon
Mobile phones are not working, no electricity, no water in #Thamel, #Nepal, #earthquake

@canvassss @skyasesh they are in search of blood donors for the people who are injured in earthquake
Availability-tweets

Please contact for drinking free service water specially for Earthquake Victim. Sanjay Limbu [mobile num]

If blood required then please contact representative of Nepal Voluntary Blood Donors Society [url]
Needs / availabilities very sparse

- 2015 Nepal earthquake dataset
  - Total tweets: 50,068
  - Need-tweets: ~500
  - Availability-tweets: ~1,300

- 2016 Italy earthquake dataset
  - Total tweets: 70,487
  - Need-tweets: ~180
  - Availability-tweets: ~230

Less than 3% tweets are about needs / availabilities
Needs / availabilities very sparse

- 2015 Nepal earthquake dataset
  - Total tweets: 50,068
  - Need-tweets: ~500
  - Availability-tweets: ~1,300

- 2016 Italy earthquake dataset
  - Total tweets: 70,487
  - Need-tweets: ~180
  - Availability-tweets: ~230

Less than 3% tweets are about needs / availabilities

Challenge 1

Automatically identifying need-tweets and availability-tweets

- Availability-tweets: ~230
Matching needs and availabilities

Mobile phones are not working, no electricity, no water in #Thamel, #Nepal, #earthquake

Please contact for drinking free service water specially for Earthquake Victim. Sanjay Limbu [mobile num]

@canvassss @skyasesh they are in search of blood donors for the people who are injured in earthquake

If blood required than please contact representative of Nepal Voluntary Blood Donors Society [url]
Matching can be challenging

Nepal quake: thousands in need of shelter in Sindhupalchowk; little able to cope

Can anyone we know pick the 2000 second hand tents from Sunauli and distribute it to the people in need in Nepal?
Matching can be challenging

भूकंप पीड़ितों को खाना-टेंट चाहिये

Call [cell number]. Dhading Naubise with medicine, tent, rice for victims. #NepalQuake

Nepalis, r w/o water & electricity. Water is essential to be supplied to the affected people in Nepal

भरतीय रेलवे ने १ लाख पानी की बोतले भेजी हैं। धन्यवाद @sureshprabhu जी #NepalEarthquake
Matching can be challenging

Challenge 2
Automatically matching need-tweets and appropriate availability-tweets

भूकंप पीड़ितों को खाना-टेंट चाहिये

भरतीय रेलवे ने 1 लाख पानी की बोतले भेजी हैं। धन्यवाद @sureshpprabhu जी #NepalEarthquake
Identifying needs and availabilities

- Compared different methodologies for identification or need-tweets and availability-tweets
  - Pattern matching based techniques [Temnikova, ISCRAM15]
  - Retrieval using language models
  - Retrieval using neural network models
Prior works used pattern matching: Baseline1: EMTerms [Temnikova, ISCRAM 2015]

<table>
<thead>
<tr>
<th>Category</th>
<th>Example Patterns</th>
<th>Matching tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Need of / offered supplies</td>
<td>{Number} bags, aid, aids, bottled water, donate any supplies</td>
<td>We need tents, dry foodstuffs, bottled water, blankets, medicine for nepal earthquake victims raincoats, rainboots</td>
</tr>
<tr>
<td></td>
<td>{Number} {medical aid}</td>
<td>India sends mobile hospital with doctors &amp; medical aid, as hospitals in Nepal don't have enough rooms</td>
</tr>
<tr>
<td>Volunteer / professional service offered</td>
<td>Volunteer, heads, relief aid, help victims</td>
<td>#MSGHelpEarthquakeVictims 2000 Volunteers of @derasachasauda will go to Nepal with Food, Medicines, Clothes…</td>
</tr>
<tr>
<td></td>
<td>Blood bank</td>
<td>#Nepal Pls Donate blood at blood banks and Hospitals. #earthquake [url]</td>
</tr>
<tr>
<td>Logistics and transportation</td>
<td>{Number} trucks, helicopters, rescue boats</td>
<td>#Russia: We are ready to send field hospital, fresh water plant and heavy transport helicopters to #Nepal</td>
</tr>
</tbody>
</table>

Total: 953 patterns  Precision ~ 0.03  Recall ~ 0.70
Identifying needs and availabilities

- Compared different methodologies for identification or need-tweets and availability-tweets
  - Pattern matching based techniques [Temnikova, ISCRAM15]
  - Retrieval using language models
  - Retrieval using neural network models

- Neural network models perform significantly better, especially in terms of precision

*Microblog Retrieval for Post-Disaster Relief: Applying and Comparing Neural IR Models, ACM SIGIR 2017 Workshop on Neural Information Retrieval*
Future challenges

- Develop better algorithms and tools
  - Reduce training data of NN models (models pre-trained on past events?)
  - Retrieval algorithms need to be tuned over time
  - Optimal match between needs and availabilities (resource optimization problem)
  - Multi-view and multi-modal information extraction

- Involve active crowdsourcing with algorithms
SMERP 2018 workshop with WWW

Second workshop on:
Exploitation of Social Media for Emergency Relief and Preparedness (SMERP)

23rd or 24th April, 2018

Co-located with:
The Web Conference 2018 (formerly WWW),
23rd - 27th April, 2018,
Lyon, France

https://www.cse.iitk.ac.in/users/kripa/smerp2018/
Summary

- Use of OSM during disasters
  - Good, bad, and ugly information
  - Extract, summarize the good information
  - Challenges due to the noisy vocabulary of OSM

- Future Challenges
  - Need better and faster IR and NLP algorithms
  - Need to combine with resource optimization
  - Multi-modal and multi-view info extraction

Thank You!
Questions / Suggestions?