Social Media Use During Humanitarian Emergencies and Disasters

Alexandra Olteanu
Social Computing + Computational Social Science

IBM Science for Social Good
Applied Data Science
IBM TJ Watson Research Center
IBM Science for Social Good Initiative

Directed by Aleksandra Mojsilovic & Kush Varshney
IBM Science for Social Good Initiative
Exploring a new approach in addressing world’s challenges

Partners IBM Research scientists with academic fellows and subject matter experts from a diverse range of non-governmental organizations (NGOs) to tackle emerging societal challenges using science and technology.
IBM Science for Social Good Initiative

Our projects demonstrate the realm of what’s possible

- What Works in Global Development
- Cognitive Disease Hunter
- Ask Nature for Design Inspiration
- Simpler Voice Overcoming Illiteracy
- Hunting Zika Virus with Machine Learning
- Changing Behaviors to Conserve Energy
- Combating the Opioid Crisis
- Demystify Social Entrepreneurship
- Disseminate the Best Treatment for Diarrhea
- Smarter Sustainable Development
- Accelerate Scientific Discovery
- Open Innovation Platforms
- Recognize Hate Speech
- Real-Time Understanding of Humanitarian Crises
- Emergency Food Best Practice
- The Digital Experience
- Cognitive Policy Advisor
- Neurology-as-a-Service
- Cognitive Financial Advisor
- For Low-Wage Workers
- How to Foster Innovation
Events & Social Media

High volumes of event-related tweets
- **25+ million** tweets for Sandy Hurricane
- **24+ million** tweets for 2016 Academy Awards
- **10+ million** tweets for Supreme Court ruling on marriage equality
- **4+ million** tweets about UK 2015 elections

Endorsed by governmental agencies

Social media are to enhance, not replace
MicroMappers, CrisisLex, CREES, AIDR, …
WORLD HUMANITARIAN
DATA AND TRENDS 2014

An annual OCHA publication that presents global and country-level data and trend analysis about humanitarian crises and assistance.
What kind of information people share during humanitarian crises?

with Carlos Castillo and Sarah Vieweg
✓ Twitter sample API
✓ Keyword-based searches
✓ 26 crisis events
✓ 1000 annotated tweets per crisis
Twitter sample API
  ✓ 2012 & 2013
  ✓ ~1% random sample of Twitter public stream
  ✓ ~130+ million tweets per month

Keyword-based searches
✓ 26 crisis events
✓ 1000 annotated tweets per crisis
Twitter sample API
✓ Keyword-based searches
 ✓ 26 crisis events
 ✓ 1000 annotated tweets per crisis
✓ Twitter sample API
✓ Keyword-based searches
  ✓ proper names of affected location
    ✓ manila floods, boston bombings, #newyork derailment
  ✓ proper names of meteorological phenomena
    ✓ sandy hurricane, typhoon yolanda
  ✓ promoted hashtags
    ✓ #SafeNow, #RescuePH, #ReliefPH
✓ 26 crisis events
✓ 1000 annotated tweets per crisis
Data Collection & Annotation

- Twitter sample API
- Keyword-based searches
- 26 crisis events
- 1000 annotated tweets per crisis
Data Collection & Annotation

✓ Twitter sample API
✓ Keyword-based searches
✓ 26 crisis events
  ✓ 14 countries and 8 languages
  ✓ 12 different hazard types
    ✓ earthquakes, wildfires, floods, bombings, shootings, etc.
  ✓ 15 instantaneous crises
✓ 1000 annotated tweets per crisis
✓ Twitter sample API
✓ Keyword-based searches
✓ 26 crisis events
✓ 1000 annotated tweets per crisis
✓ Twitter sample API
✓ Keyword-based searches
✓ 26 crisis events
✓ 1000 annotated tweets per crisis
  ✓ Content dimensions
    ✓ Informativeness
    ✓ Source of information
    ✓ Type of information
✓ Crowdsourcing workers from the affected countries
Content & Source Variations
Content & Source Variations

Eyewitness accounts
9% (min 0%, max 54%)

Outsiders
38% (min 3%, max 65%)

- Singapore Haze 2013
- Typhoon Pablo 2012
- Australia Wildfire 2013
- Guatemala Quake 2012
- Boston Bombings 2013
- Typhoon Yolanda 2013

Background
- Media Biases
- Collection Biases
- Data Collection
- Methods Sensitivity
- Conclusions
- Future Work
Content & Source Variations

Eyewitness accounts
9% (min 0%, max 54%)

Outsiders
38% (min 3%, max 65%)

Infrastructure & Utilities
7% (min 0%, max 22%)

Affected individuals
20% (min 5%, max 57%)

Savar Building Collapse 2013
Typhoon Pablo 2012
Alberta Floods 2013
Guatemala Quake 2012
Boston Bombings 2013
Typhoon Yolanda 2013
Content & Source Variations

Eyewitness accounts
9% (min 0%, max 54%)

Outsiders
38% (min 3%, max 65%)

Infrastructure & Utilities
7% (min 0%, max 22%)

Affected individuals
20% (min 5%, max 57%)

Typhoon Pablo 2012
Singapore Haze 2013
Australia Wildfire 2013
Guatemala Quake 2012
Boston Bombings 2013
Typhoon Yolanda 2013
Savar Building Collapse 2013
Alberta Floods 2013
Guatemala Quake 2012
Boston Bombings 2013
Typhoon Yolanda 2013

Background
Media Biases
Collection Biases
Data Collection
Methods Sensitivity
Conclusions
Future Work
Data Patterns: Types

[Tree diagram showing relationships between various events such as natural disasters and incidents.]

Typhoon_Pablo'12
Sardinia_floods'13
Colorado_floods'13
Queensland_floods'13
Australia_bushfire'13
Italy_earthquakes'12
Singapore_haze'13
Costa_Rica_earthquake'12
Typhoon_Yolanda'13
Philipinnes_floods'12
Alberta_floods'13
Manila_floods'13
Spain_train_crash'13
Brazil_nightclub_fire'13
Boston_bombings'13
Bohol_earthquake'13
Venezuela_refinery'12
Lac_Megantic_train_crash'13
West_Texas_explosion'13
Glasgow_helicopter_crash'13
Guatemala_earthquake'12
Colorado_wildfires'12
Russia_meteor'13
NY_train_crash'13
LA_airport_shootings'13
Savar_buildingCollapse'13

data patterns: lower similarity
Data Patterns: Types

Natural, diffused, and progressive

Human-induced, focalized, and instantaneous

lower similarity
Data Patterns: Sources

- Instantaneous, focalized, and human-induced
- Natural, diffused, and progressive
Similar events tend to have a similar distribution of message types & sources.

Data Patterns: Sources

Instantaneous, focalized, and human-induced

Natural, diffused, and progressive
Sources & Message Types

- Infrast. & Utilities
- Caution & Advice
- Donat. & Volun.
- Affected Ind.
- Sympathy
- Other Useful Info.

<table>
<thead>
<tr>
<th>Media</th>
<th>Outsiders</th>
<th>Eyewitness</th>
<th>Government</th>
<th>NGOs</th>
<th>Business</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="chart-data" alt="" /></td>
<td><img src="chart-data" alt="" /></td>
<td><img src="chart-data" alt="" /></td>
<td><img src="chart-data" alt="" /></td>
<td><img src="chart-data" alt="" /></td>
<td><img src="chart-data" alt="" /></td>
</tr>
</tbody>
</table>

Colors represent:
- 0% = no data
- 5% = light gray
- 10% = medium gray
- >15% = dark gray

Legend:
- Infrast. & Utilities
- Caution & Advice
- Donat. & Volun.
- Affected Ind.
- Sympathy
- Other Useful Info.
Sources & Message Types

- Infrast. & Utilities
- Caution & Advice
- Donat. & Volun.
- Affected Ind.
- Sympathy
- Other Useful Info.

Media
Outsiders
Eyewitness
Government
NGOs
Business

Sources & Message Types

- Media
- Outsiders
- Eyewitness
- Government
- NGOs
- Business

Sympathy
Infrast. & Utilities
Caution & Advice
Donat. & Volun.
Affected Ind.

0%
5%
10%
15%
>15%
Temporal Distribution: Information Types

Peak

12h  24h  36h  48h  ...  several days
Caution & Advice

Temporal Distribution: Information Types

- peak
- 12h
- 24h
- 36h
- 48h
- ...
- several days

Deeply touched by the blast at the Boston Marathon. Our thoughts with the affected people and their families. EH
Temporal Distribution: Information Types

- Caution & Advice
- Sympathy & Support
- Infrastructure & Utilities

COSTA RICA: @nacion In Nandayure, per mayor: are no electricity, no water in some parts. There are landslides in the highlands
Temporal Distribution: Information Types

- **Caution & Advice**
- **Sympathy & Support**
- **Affected Individuals**
- **Infrastructure & Utilities**

Timestamps:
- peak
- 12h
- 24h
- 36h
- 48h
- ...
- several days

Bangladesh factory building collapse kills over 70, injures hundreds [reut.rs/10csR0q]
If you are able to donate blood - Providence Hospital will have a blood drive in Waco from 11 a.m. to 5 p.m. #prayforwest
Temporal Distribution: Sources

Progressive peak

12h  24h  36h  48h  ...  several days
No way is the PSI 40 right now. Can barely see past Sheares Bridge from downtown. #sghaze.
Collated government advisories & situation reports on #YolandaPH, with crisis & relief map: gov.ph/crisis-respons...
Temporal Distribution: Sources

- **Progressive**
- **Government**
- **Eyewitness**

**Reuters**: Colorado wildfires worsen, 32000 flee homes | - goo.gl/news/shzz via news.google.com

Timeline:
- 12h
- 24h
- 36h
- 48h
- ... several days
Temporal Distribution: Sources

- Progressive
- Outsiders
- Media
- Government
- Eyewitness

My love and prayers go out to all the people involved with Aurora Colorado.
24 hrs after Super Typhoon Haiyan struck the Philippines, 1st shipment of health, med & shelter eqp left SD today expected to arrive Tues.
Temporal Distribution: Sources

During the early stage of the disaster, eyewitness reports dominated the immediate response, providing crucial information about the situation. As the scope of the disaster expanded, the government and official sources began to take center stage, issuing official statements and updates.

As the situation evolved, NGOs and media outlets became increasingly active, providing detailed reports and analysis. The peak of activity was reached by Outsiders, who contributed unique perspectives and insights.

Businesses and Progressive groups began to mobilize, offering support and resources. The timeline includes a mention of an event involving Tide washing clothes for free at Glenmore Coin Laundry. This statement is a specific example of how businesses can contribute during disasters, and it is accompanied by the hashtag #yycflood, indicating the location and possibly the nature of the event.

The timeline also indicates a peak in activity, with the text box stating:

"Until Sunday, Tide is washing clothes for free at Glenmore Coin Laundry. #yycflood"
How do we collect social media data during humanitarian crises?

with Carlos Castillo, Fernando Diaz, and Sarah Vieweg
Tweets are queried by

**Content**

#prayforwest  #abflood

**Location**

longitude: [-97.5, -96.5]  
& latitude: [31.5, 32]
Data Collection: How Is It Done?

Tweets are queried by:

**Content**

#prayforwest
#abflood

**Location**

longitude: [-97.5, -96.5]
& latitude: [31.5, 32]

Maximum 1% of all tweets

Low recall: 33%
Not everyone uses the keywords.
Maximum 400 terms.

Low precision: 12%
Not everyone on the ground talks about the event.
Maximum 25 geo-rectangles.
Key Insight: Distill a Crisis Lexicon

- damage
- affected people
- people displaced
- donate blood
- text redcross
- stay safe
- crisis deepens
- evacuated
- toll raises
- ... ...

Gov. McDonnell: Virginia 'Spared' in Hurricane Sandy
Damage [patch.com/A-zgoL]

Deeply touched by the blast at the Boston Marathon. Our thoughts with the affected people and their families. EH

NOTE: Oklahoma University is providing shelter in their dorm facilities for people displaced by the tornado in OK today - @KFOR

If you are able to donate blood, Providence Hospital will have a blood drive in Waco from 11 a.m. to 5 p.m. #prayforwest

Best way to help tornado victims is to donate to the Red Cross at redcross.org or text REDCROSS to 90999. #okwx

This flooding is crazy! Hoping my fellow Albertans and Calgarians stay safe! #abflood #yycflood
Precision vs. Recall

**Precision** is straightforward to measure. **Recall** requires a complete data collection. We use geo data as proxy.
**Precision vs. Recall**

**Precision** is straightforward to measure.

**Recall** requires a complete data collection. We use geo data as proxy.

- **Keyword based sampling**
  - #sandy, #bostonbombings, #qldflood

- **Location based sampling**
  - Tweets geo-tagged in area of the disaster

- **Desired**

![Graph showing precision vs. recall with Keyword based sampling and Location based sampling points](image-url)
Precision vs. Recall

**Precision** is straightforward to measure.

**Recall** requires a complete data collection. We use geo data as proxy.
Precision vs. Recall

**Precision** is straightforward to measure.

**Recall** requires a complete data collection. We use geo data as proxy.

---

**Desired**

- **Keyword based sampling**: #sandy, #bostonbombings, #qldflood
- **Location based sampling**: tweets geo-tagged in area of the disaster

---

High Precision

High Recall
Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.
Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.
Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.
Representativeness

Keyword-based samples appear to overrepresent media reports and underrepresent eyewitness accounts.

Lexicon-based samples better preserve the original distribution of messages types and sources.
KEEP CALM AND QUESTION EVERYTHING
Email: alexandra@aolteanu.com

Web: aolteanu.com

Twitter: @o_saja