BIG CRISIS DATA

Twitter: @BigCrisisData

CARLOS CASTILLO
Director of Research for Data Science at Eurecat

Data for Good Workshop – Turin, October 2016
TOPIC OF THIS SHORT TALK: SOCIAL MEDIA IN DISASTERS

Patrick Meier
QCRI →

Alexandra Olteanu
EPFL →

Muhammad Imran
QCRI

Aditi Gupta
IIIT Delhi →

Sarah Vieweg
QCRI

P. K. Kumaraguru
IIIT Delhi

Irina Temnikova
QCRI

Hemant Purohit
Wright State → George Mason

Ji Lucas
QCRI

Fernando Diaz
Microsoft

Ferda Ofli
QCRI

Work done (2012-2015) by Crisis Computing team at QCRI

Book (July 2016) from Cambridge University Press
KEY MESSAGES

● Social media is a huge force during crises, emergencies, and disasters
  – *You cannot control the wind, but you can harness its power*

● Social media contains highly relevant information surrounded by tons of noise
  – *Mining is an industrial process, or at the very least one that requires special tools*

● Digital volunteers will help you if you let them
  – *Crisis mapping is just the first of many paradigms for participatory data mining*
Social Media is a huge force during crises, emergencies, and disasters.
TWO MONTHS AGO (MARCH 22, 2016)

Attacks in the airport and a metro station in Brussels kill 35 and injure 340

Police asks the public:

1) to use social media, not phone;
2) to reduce video/audio streaming;
3) to avoid sharing real-time information about police actions
In the first 8 hours after the attacks...

- 2,800-words Wikipedia article
- Reddit post with 17,000 comments
- 700+ YouTube videos per hour
- Tweets and photos
- Facebook pages and Safety Check
A COMMON PATTERN

Disaster or mass-convergence event

People have increased communication needs

People are familiar with social media

Internet is not bullet-proof but fairly resilient

Emergency agencies encourage social media usage

Intensive usage of social media by the public for emergency communications
EXAMPLES FROM #SMEM PAPERS

“OMG! The fire seems out of control: It’s running down the hills!”

“Red River at East Grand Forks is 48.70 feet, +20.7 feet of flood stage... #flood09”
Red River Valley floods in 2009 [Starbird et al. 2010]

“My moms backyard in Hatteras. That dock is usually about 3 feet above water [photo]”
Hurricane Sandy 2013 [Leavitt and Clark 2014]

“Sirens going off now!! Take cover...be safe!”
Moore Tornado 2013 [Blanford et al. 2014].

“There is shooting at Utøya, my little sister is there and just called home!”
2011 attacks in Norway [Perng et al. 2013]
Problem 1: Scale

- 2012 Hurricane Sandy
- 2013 Boston bombings
- 2011 Japan earthquake
- 2013 Pope elected
- 2015 Super Bowl
- 2014 World Cup final
- 2013 Castle in the Sky on Japanese TV

Tweets Per Second (TPS)
PROBLEM 2: NOISE

- The Internet is strong because it is decentralized
- So, as they say ...
SOCIAL MEDIA CONTAINS RELEVANT INFORMATION SURROUNDED BY NOISE
THE TENSION: FAMILIAR TOOLS VS NEW USES

- “In times of disaster and crisis, people tend to gravitate towards the systems and networks that are most familiar to them. They shun the specialized sites set up by aid workers for the comfort of already-established systems they know well from prior experience.”

- “Familiar sites for sharing photos of weddings become locations for sharing breaking news ... Participants are using social web tools in ways the designers of such systems had neither anticipated nor considered. As a result, the vast majority of today's technologies are woefully ill-equipped for crisis situations”
MINING IS NOT THIS
MINING IS MORE LIKE THIS
DATA MINING IS NOT THIS
DATA MINING IS MORE LIKE THIS

```scala
val postIDTags = postsXML.flatMap { line =>
  // Matches Id="..." ... Tags="..." in line
  val idTagRegex = "Id="\((\d+)\)".+Tags="\([^"\]+\)"\".r
  // // Finds tags like <TAG> value from above
  val tagRegex = "&lt;\([^&]+\)&gt;".r

  // Yields 0 or 1 matches:
  idTagRegex.findFirstMatchIn(line) match {
    // No match -- not a line
    case None => None
    // Match, and can extract ID and tags from m
    case Some(m) => {
      val postID = m.group(1).toInt
      val tagsString = m.group(2)
      // Pick out just TAG matching group
      val tags = tagRegex.findAllMatchIn(tagsString).map(_.group(1)).toList
      // Keep only question with at least 4 tags, and map to (post,tag) tuples
      if (tags.size >= 4) tags.map((postID, *)) else None
    }
  }
  // Because of flatMap, individual lists will concatenate
  // into one collection of tuples
}
```

Click on map icons to see local reports
DON'T DO IT ENTIRELY BY HAND

- Many tools and systems, some of them free, are available for:
  - Organizing your communications with the public through social media
  - Monitoring the response of people to social media activity
  - Monitoring social media activity
- You probably don't need “big systems,” but most likely a collection of specialized tools
  - Each tool does one job well
  - Tools may overlap a bit, sometimes a lot
  - If a tool ceases to be relevant, stop using it and find a new one
- Big tightly-integrated systems bring some comfort but you pay a lot and loose flexibility
DATA MINER REFLEX: TO CLASSIFY AND CLUSTER

Caution & Advice | Information Sources | Damage & Casualties | Donations
---|---|---|---
Gov | | | 
NGO | | | 
Media | | | 
Eyewitness | | | 
Outsider | | | 
... | | |
A STUDY OF TWITTER ON 26 CRISIS

Results from Olteanu et al. CSCW 2015. Data available at http://crisislex.org/
Results from Olteanu et al. CSCW 2015. Data available at http://crisislex.org/
TEMPORAL PROGRESSION

Peak

- Caution and advice
- Sympathy and support
- Affected individuals
- Infrastructure and utilities
- Other specific information
- Donations and volunteering

12 hr 24 hr 36 hr 48 hr ... several days

Results from Olteanu et al. CSCW 2015. Data available at http://crisislex.org/
INFORMATION EXTRACTION + DONATIONS MATCHING

See Purohit et al. 2013 for automatic donations matching.
Digital volunteers will help you, if you let them.
Physical and Digital Convergence

- Disasters involve the convergence of people, resources, and information.
  - Some of the people, some of the resources, and some of the information are actually helpful in the disaster response [Fritz and Mathewson 1975]

- Social media and mobile phone cameras evaporate traditional scene control [Crowe 2012], yellow tape does not work so well online
  - Note: rescue personnel, including firefighters, police officers, and others who physically converge on the scene have responsibility and legal codes to respect. As the public also enters the scene (physically or virtually), they must adhere to ethical standards they may be unfamiliar with. Technology can be used to weaken or to strengthen those standards.
THE GEOGRAPHY OF AN EVENT  

2015 NEPAL EARTHQUAKE

Facebook: Visualizing Crisis Relief in Nepal (September 2015).

A map.”

Patrick Meier – http://irevolution.net/
COMMUNITY-POWERED CRISIS MAPPING

Social media is critical for humanitarian work & now you can see why. Crisis Map of Libya is now public: http://bit.ly/g8xCtm #UN #OCHA

© 2011 Twitter About Us Contact Blog Status Resources API Business Help Jobs Terms Privacy
CROWDSOURCED MAPPING OF WILDLIFE IN NAMIBIA

See Ofli et al. 2016 for details.
AUTOMATIC MAPPING

Floods in Germany
De Albuquerque et al. 2015

Dengue in Brazil
Gomide et al. 2015

Earthquakes in Italy
Cresci et al. 2014
HYBRID MAPPING: AIDR + MICROMAPPERS

Automatic processing: machine learning

Manual processing: crowdsourcing

See Imran et al. 2014 for details on AIDR. Find out more at http://aidr.qcri.org/
See Imran et al. 2014 for details on AIDR. Find out more at http://aidr.qcri.org/
CONTENT CREDIBILITY

- Internet content is considered as less truthful or valid than other contents
  - Reality is more nuanced, otherwise nobody would use Wikipedia or Google
  - Are emergency managers used to work with perfect information?

- Learn to deal with imperfect information, e.g., the Admiralty code:

<table>
<thead>
<tr>
<th>A</th>
<th>Reliable</th>
<th>1</th>
<th>Confirmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>Usually reliable</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>Fairly reliable</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>Not usually reliable</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>Unreliable</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>Cannot be judged</td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>
THERE ARE NO EASY ANSWERS, BUT TOOLS CAN HELP A BIT

http://twitdigest.iiitd.edu.in/TweetCred/
CROWDSOURCED TOOLS CAN ALSO HELP

One can create tools that encourage a behavior, e.g., we don't want people to upvote/downvote content, but instead to provide evidence.

https://veri.ly/
THE FUTURE: REAL-TIME CROWDSOURCED MINING

- Mapping disaster-affected areas using UAVs
- Is crowdsourced stream mining possible?

See Patrick Meier's blog post from Nov. 2015 for details.
Visit the UAViators community for more information on video clickers.
TO RE-CAP

• Social media is a huge force during crises, emergencies, and disasters
  – *You cannot control the wind, but you can harness its power*

• Social media contains highly relevant information surrounded by tons of noise
  – *Mining is an industrial process, or at the very least one that requires special tools*

• Digital volunteers will help you if you let them
  – *Crisis mapping is just the first of many paradigms for participatory data mining*
Good projects in this space

- Computationally feasible
- Supported by data
- Useful

Poorly planned projects :-(

AI-complete problems

Temptation! Danger!
THANK YOU!

Patrick Meier
QCRI →

Muhammad Imran
QCRI

Sarah Vieweg
QCRI

Irina Temnikova
QCRI

Ji Lucas
QCRI

Ferda Ofli
QCRI

Alexandra Olteanu
EPFL →

Aditi Gupta
IIIT Delhi →

P. K. Kumaraguru
IIIT Delhi

Hemant Purohit
Wright State → George Mason

Fernando Diaz
Microsoft

chato@acm.org

BIG CRISIS DATA
Social Media in Disasters and Time-Critical Situations
CARLOS CASTILLO

BIGCRISISDATA.ORG