Social Media as Sensors: Microblogs as Sensors
Microblogging during two natural hazards events: what twitter may contribute to situational awareness
  • http://doi.acm.org/10.1145/1753326.1753486

Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake
  • http://doi.acm.org/10.1145/1958824.1958830

Twitter use during an emergency event: the case of the UT Austin shooting
  • http://doi.acm.org/10.1145/2037556.2037613
Situational Awareness

- Idealized Understanding
  - Command/Control operations
  - Groups
  - Communities

- Requires Communication
  - ICT – Information and Communication Technologies
    - Microblogging!
We use Twitter as a Sensor Array

- Important news items
- Weather reports
- Emergencies
- Updates on health/well-being
- Etc.
Complex Sensors

- Keyword search
- Geolocation
- Specific Users
- Hashtagging
Red River Floods
  • Spring 2009

Lots of advance warning

49 Twitter authors, 19k tweets
  • Twitter API, keywords
  • E-Data Viewer, coded tags
Wildfire

- Oklahoma Grassfires
  - Spring 2009

- Little if any warning

- 46 Twitter authors, 2.8k tweets
Information from Tweets

GEO-LOCATION INFO

SITUATIONAL UPDATES

<table>
<thead>
<tr>
<th>Coding Category</th>
<th>OK</th>
<th>RR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warning</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>Preparatory Activity*</td>
<td>N/A</td>
<td>7%</td>
</tr>
<tr>
<td>Fire Line/Hazard Location*</td>
<td>21%</td>
<td>1%</td>
</tr>
<tr>
<td>Flood Level*</td>
<td>N/A</td>
<td>17%</td>
</tr>
<tr>
<td>Weather*</td>
<td>6%</td>
<td>11%</td>
</tr>
<tr>
<td>Wind*</td>
<td>10%</td>
<td>1%</td>
</tr>
<tr>
<td>Visibility*</td>
<td>1%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Road Conditions</td>
<td>2%</td>
<td>3%</td>
</tr>
<tr>
<td>Advice (emergency)</td>
<td>1%</td>
<td>2%</td>
</tr>
<tr>
<td>Advice (information space)*</td>
<td>0.3%</td>
<td>2%</td>
</tr>
<tr>
<td>Evacuation Information*</td>
<td>12%</td>
<td>4%</td>
</tr>
<tr>
<td>Volunteer Information*</td>
<td>2%</td>
<td>6%</td>
</tr>
<tr>
<td>Animal Management</td>
<td>1%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Damage/Injury reports*</td>
<td>15%</td>
<td>2%</td>
</tr>
</tbody>
</table>
High- and low-level situational features

- Could be used to better develop a system that would be useful in emergencies
Earthquake

- Yushu Earthquake
  - April 2010
- No advance warning
- 94k Sina-Weibo posts and nearly 42k re-posts
  - Search interface, keywords
Peaks occur
- The day of the earthquake
- National day of mourning

Use of microblogging for:
- Situational updates
- Emotional support
- Opinions
- Calls for action

Figure 4. Category Trends (5% Sample)
University of Texas at Austin
  • September 2010

2.8k Twitter authors
  • Peak of 15k tweets/day
  • Followers of @UTAustin with public accounts
Day before, day of, day after
Event happened around 8AM, Sept. 28
"shooting", "utshooting", "suspect", and "university"—besides "campus", "UT", "Austin", and "RT" (which stands for retweets)—were the most frequent words. We presented this result in a word cloud (Figure 2), which gave the user a quick snapshot of frequent words. For more detailed word counts histograms could be displayed along with word clouds.

Comparing the distribution of number of posts over time for Sept. 28 to the day before and after it (Figure 3), the Twitter posts nearly doubled to 900 per hour at 8 AM as compared to other days when by this time it would be below 500 posts. The peak of Sept. 28 was at 9 AM with 2,623 Twitter posts, when other days it would be around 600 by that time, an increase of over 400%.

Users mostly tweeted about the shooting event for 7 hours after it happened. At 7 AM there was nothing related to it, but by 8 AM (Figure 4), around when the event happened, the words "UT", "campus", "gunman" and "shooter" were among the most used on tweet posts. The UT shooting dominated the Twitter posting of this community until 3 PM.

The 'collocation' feature in NLTK provides the top 20 frequently appearing word pairs in a data file. After separating tweets from the dataset by hour into different files, we ran the NLTK toolkit on them. At 7 AM on the day of the incident, no word pairs were marked as related, as we can expect. But it began to change radically from 8 AM. People were tweeting about the incidents frequently. Example pairs include 'active shooter', 'shot himself', 'armed suspect', 'Castaneda Library' (the library where the suspect finally went and committed suicide), and 'emergency text'.

From this study we observed that during crises people used Twitter to share and to comment on information about the event. We find that a spike in the number of tweets and changes in the ideas in the tweets signal that an event is occurring or occurred. Our content analysis method in this study was based on word frequencies; however, other methods such as semantic analysis [3] can further help distinguish events of interest to government emergency teams [4], for example. Future work will include other crisis related events and comparisons of results, adding also retweet and location analysis.

4. ACKNOWLEDGMENTS

We thank our supporters: NSF grant (0916733) and CAPES scholarship (BEX 1385/10-0).

5. REFERENCES


Word clouds over time

8AM

2PM
Challenges/Opportunities

- Determining importance of individual microblog posts
- Determining category of posts
  - And therefore relevance?
- Gathering accurate and timely news from tens to thousands (millions?) of users
- Presenting information in a useful manner
  - Audience-based?
Questions?
References