Analysing Tweets: During natural disasters in Europe and Asia

Presenter: Kiran Zahra
6.2-magnitude earthquake crumbles central Italy

sbne.ws/r/vg6x
6.2-magnitude earthquake crumbles central Italy sbne.ws/r/vg6x
Article: 'It was so strong, so sudden,' says survivor of deadly #Italy #earthquake
Article: 'It was so strong, so sudden,' says survivor of deadly #Italy #earthquake
Strong

Sudden
P J Anthony @JMoota · 24 Aug 2016
Earthquake in Myanmar Kills 3 and Damages Over 170 Temples
nyti.ms/2bFtKyO
Earthquake in Myanmar Kills 3 and Damages Over 170 Temples
nyti.ms/2bFtKyO

3 Killed
Earthquake in Myanmar Kills 3 and Damages Over 170 Temples

3 Killed
170 Temples damaged
CJ Lopez @kickdblclj · May 1
That wasn't an #earthquake. That was my responsibilities crashing down around me
That wasn't an #earthquake. That was my responsibilities crashing down around me
Aim

• How geographic features are reported from two different continents during disasters

• The credibility of users reporting from two different continents

• Performance of machine learning algorithm to filter noise
Data Collection
Case-study

- Italy (Europe) 6.2 Magnitude
- Myanmar (Asia) 6.8 Magnitude
Research Question 1

• How reported geographic feature granularity vary from two continents?
Method

• 500 Tweets (Italy and Myanmar)

• Geographic feature extraction

• Count number of occurrences

• Geonames gazetteer

• Hierarchy of geographic features

• Geocode
Italy Earthquake Geographic Features Granularity

with number of times appeared in sample data

Italy 314
Perugia 23
Umbria 3
Norcia 2
Accumoli 12
Amatrice 42
L'Aquila 2
Marche 4
Vatican city 1
Rome 28
Epicenter

Legend
granularity_name
- Capital
- Independent state
- First order administrative division
- Third order administrative division
- Epicenter

0 15 30 60 90 120 Kilometers
Myanmar Earthquake Geographic Features Granularity

with number of times appeared in sample data

Legend
- Capital
- First order administrative division
- Second order administrative division
- Populated Place
- Epicenter

Bay of Bengal

India 62

Myanmar 95

Epicenter

Thailand 4

Bangkok 6
Research Question 2

• What is the quality of tweets in terms of credibility?
Twitter Metadata

- Retweet count
- Friends count
- Location
- Verified
- Screen name etc.
# User-based Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Account age (AG)</td>
<td>The time passed since the author registered their account</td>
</tr>
<tr>
<td>Statuses count (SC)</td>
<td>The number of tweet sent by the user</td>
</tr>
<tr>
<td>Followers count (FoC)</td>
<td>Number of people following this user</td>
</tr>
<tr>
<td>Friends count (FrC)</td>
<td>Number of people user is following</td>
</tr>
<tr>
<td>Verified (V)</td>
<td>If the account has been verified</td>
</tr>
<tr>
<td>Has description (D)</td>
<td>A non-empty bio</td>
</tr>
<tr>
<td>Has URL (U)</td>
<td>A non-empty homepage URL</td>
</tr>
</tbody>
</table>

(Castillo, Mendoza & Poblete 2011)
Credibility

- Credibility is the function of

\[ C = f (FrC, SC, FoC, AG, U, D, V) \]
Credibility

Friends Count

Status Count
Credibility

Followers Count

- Italy
- Myanmar

Account Age

- Italy
- Myanmar
Credibility

URL
- Italy
- Myanmar

Description
- Italy
- Myanmar
Classifiers trained for one event, work good for the same kind of event without training data

(Verma et al. 2011)
Research Question 3

• How well Naïve Bayes perform when the classifier is trained for one event and tested on another event of same nature?
Tweet Classification

• **Information**: Tweet text about disaster event and its location

• **Not Information**: Everything else falls in this category
Information

**UPDATE:** 73 people now known to have died in #Italy earthquake, Italy's civil protection unit says @BBCBreaking
Information

UPDATE: 73 people now known to have died in #Italy earthquake, Italy's civil protection unit says @BBCBreaking
I Love My Kids @MyKidsAreAholes · 24 Aug 2016
Told my wife I need to go to Italy to help the earthquake victims but she thinks I only want to go to drink red wine and eat pizza.
I Love My Kids @MyKidsAreAholes · 24 Aug 2016
Told my wife I need to go to Italy to help the earthquake victims but she thinks I only want to go to drink red wine and eat pizza.
Two major earthquakes. Death toll climbing. You want to have your shills flood our timelines with this crap? #BadTiming
Not Information

Mae ✨ @maepryor · May 4
Ok idk if that was an earthquake or I'm crazy but I've never been so scared 😂
Naïve Bayes Workflow

Reads Test and Training Data → Creates a Corpus → Classify based on Frequency of Terms
Precision & Recall for Italy

- **Information**: 98% Precision, 93% Recall
- **Not Information**: 92% Precision, 97% Recall
Naïve Bayes Workflow

1. Reads Test Data
2. Creates a Corpus
3. Classify based on Frequency of Terms
Precision & Recall for Myanmar

• **Information**: 88% Precision, 91% Recall

• **Not Information**: 92% Precision, 89% Recall
Key Points

• Granularity of geographic location is very important when reporting disasters

• Geographic locations can be differently reported from different regions of the world

• User-based credibility features reflect different user characteristics
Key Points

• Difficult to judge credibility based only on one set of features

• Naïve Bayes trained on one event performed well on the same kind of event in a different geographic region
Limitations

• Local administrative divisions do not match Geonames gazetteer

• Use of English language to collect data in non-native English speaking countries

• Twitter streaming API

• Sampling
Understanding the properties of social media are important before using this data during disasters