“Eye of the Storm”: Social Sensing of Extreme Weather Events Using Social Media

Michelle Spruce
PhD Computer Science
PI – Dr Hywel T. P. Williams

Research undertaken as part of University of Exeter SEDA Lab (Social & Environmental Data Analysis @ University of Exeter)

RMetS NCAS Atmospheric Science Conference - July 2018
Project Context

• **Extreme weather events** cause disruption to communities and economies.

• However the **specific impacts** can be hard to forecast and observe.

• ‘**Social Sensing**’ provides an opportunity to improve understanding of impacts of extreme weather events.

• With **Social Media**, public can comment on and respond to their experiences of events such as extreme weather events.

• **Improved understanding** of impacts would allow better verification of meteorological forecast models and **aid impact based forecasting**
Social Sensing

• The systematic analysis of unsolicited social media data (*user generated content shared socially via the web*) to observe real-world events

• Involves event detection, location and characterisation
Aim:
• To use ‘Social Sensing’ to map extreme weather events and understand the social impacts

Questions:
• What social impact information can we determine from social media?
• Can we assess the impact severity of an extreme weather event using Social Media?
Wildfires  
*Boulton et al, 2016*

Hurricanes

Floods / Heavy Rain  
*Arthur et al, 2018*

Earthquakes  
*Sakaki et al, 2010*

Heatwaves
Number of social media users worldwide from 2010 to 2021 (in billions)

https://www.statista.com
Tue 3rd July 11:30am
British Summer Time (BST)
Searching for floods on Twitter...

Connor @ConnerFlood30 • 5h
damn this weather is hella nice

Joyce Peterson @MemphoNewsLady • 6h
#metsunset pics 4 days apart at Greenbelt Park in Harbor Town.
#MississippiRiver at Memphis is about to hit 34 ft flood stage. (33.6 ft now.)

Fawks @OhFawker • May 7
This area is normally a field... Water is almost up to the bench seats
#OttawaGatineauFlood https://instagram.com/p/B7zHiEkhWzq/

Natchez Events @NatchezEvents • 8m
Officials keep eyes on Mississippi River as it rises above flood stage
.l/tt/12q2a2G

Ottawa Citizen @Ottawa Citizen • 6h
Photos: The view from the sky of the massive Ottawa-Gatineau flood
.l/ly/N04W30bybxV
…or for news about Storm Hector...

Storm Hector has already delivered me a few presents. So far a plant pot, a garden glove, couple empty bags of compost and a cat bed. Have it on good authority that the trampoline may be on its way 😞engo #stormhector

Good morning, #mayo. It’s a wild one out there today with some debris on the road. Storm Hector has arrived and has brought strong gale force winds, so be careful. Take care and remember to tie those trampolines down. @MayoDotIE .@MayoCoCo

Thank you storm hector. It appears I have a helipad now instead of a trampoline 😂😂

Give over... it’s not ‘Storm Hector’ it’s just a blowy

Weather is just weather.
How to locate tweets?
• Only 1% of tweets contain a ‘geotag’ – specific location co-ordinates
• Another 1-2% contain a ‘place’ element
• Therefore need to infer location using user location, place name mentioned in text or timezone
Social sensing of floods in the UK

Rudy Arthur\textsuperscript{1,*}, Chris A. Boulton\textsuperscript{2#}, Humphrey Shotton\textsuperscript{3#}, Hywel T. P. Williams\textsuperscript{1}

\textsuperscript{1} Computer Science, CEMPS, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4OE, United Kingdom. \textsuperscript{2} Earth System Science, CLES, University of Exeter, Laver Building, North Park Road, Exeter, EX4 4OE, United Kingdom

\* These authors contributed equally to this work.
\# R.Arthur@exeter.ac.uk

Abstract

“Social sensing” is a form of crowd-sourcing that involves systematic analysis of digital communications to detect real-world events. Here we consider the use of social sensing for observing natural hazards. In particular, we present a case study that uses data from a popular social media platform (Twitter) to detect and locate flood events in the UK. In order to improve data quality we apply a number of filters (timezone, simple text filters and a naïve Bayes ‘relevance’ filter) to the data. We then use place names in the user profile and message text to infer the location of the tweets. These two steps remove most of the irrelevant tweets and yield orders of magnitude more located tweets than we have by relying on geotagged data. We demonstrate that high resolution social sensing of floods is feasible and we can produce high-quality historical and real-time maps of floods using Twitter.

Introduction

Natural hazards such as floods, wildfires, storms and other extreme weather events cause substantial disruption to human activity and are predicted to increase in frequency and severity as
## UK Storm Names 2017/18

<table>
<thead>
<tr>
<th>Name</th>
<th>Date named</th>
<th>Date of impact on UK and/or Ireland</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aileen</strong></td>
<td>12 September 2017</td>
<td>12 - 13 September 2017</td>
</tr>
<tr>
<td><strong>Ex-Hurricane Ophelia</strong></td>
<td>11 October 2017 (Named by NHC)</td>
<td>16 - 17 October 2017</td>
</tr>
<tr>
<td><strong>Brian</strong></td>
<td>19 October 2017</td>
<td>21 October 2017</td>
</tr>
<tr>
<td><strong>Caroline</strong></td>
<td>5 December 2017</td>
<td>7 December 2017</td>
</tr>
<tr>
<td><strong>Dylan</strong></td>
<td>29 December 2017</td>
<td>30 - 31 December 2017</td>
</tr>
<tr>
<td><strong>Eleanor</strong></td>
<td>1 January 2018</td>
<td>2 - 3 January 2018</td>
</tr>
<tr>
<td><strong>Fionn (F-yunn)</strong></td>
<td>16 January 2018</td>
<td>16 January 2018</td>
</tr>
<tr>
<td><strong>David</strong></td>
<td>17 January 2018 (Named by Meteo France)</td>
<td>18 January 2018</td>
</tr>
<tr>
<td><strong>Georgina</strong></td>
<td>23 January 2018</td>
<td>24 January 2018</td>
</tr>
<tr>
<td><strong>Hector</strong></td>
<td>13 June 2018</td>
<td>13 - 14 June 2018</td>
</tr>
</tbody>
</table>

[https://www.metoffice.gov.uk/barometer/uk-storm-centre](https://www.metoffice.gov.uk/barometer/uk-storm-centre)
Collecting Twitter Data (Tweets)

**Twitter collections** – Using Twitter API¹ collect tweets containing natural hazard/weather related key words – tweets received in JSON format

- Storm Name Collection: storm, Ophelia, Ofelia, Opelia, Ohpelia, ophelia, Ophelia, opheliaireland, Brian, brian, Caroline, caroline, Dylan, dylan, Eleanor, eleanor, Fionn, fionn, Fion, fion, Georgina, georgina, Emma, emma

- Wind Collection: wind, gale, windstorm, hurricane

**Extract tweets** from collections for a specific time period/keywords

https://developer.twitter.com/content/developer-twitter/en.html
- Tweets collected 16th October 2017 – 10th March 2018
- Over 35 million individual Tweets collected
Method – relevance filter

Bot filter
- Remove usernames known to be automated ‘Twitter bots’

Weather station filter
- Remove tweets with known automated weather station structure

Irrelevant Term filter
- Exclude tweets containing terms such as: ‘cook up a storm’, ‘wind up’, ‘throw caution to the wind’, etc

Machine learning
- Created training corpus using 5000 labelled tweet examples
- Apply Naïve Bayes algorithm to remove tweets based on training corpus
"ophelia" tweets: 16/10/2017 - 31/10/2017

All Ophelia Tweets (Unfiltered)

Filtered for relevance

Tweet Count

Hour of Date_Hour [October 2017]

Hour of Date_Hour [2017]
"Brian" tweets - 20/10/2017 - 01/112017

All Brian Tweets - Unfiltered

Filtered

Hour of Date Hour [October 2017]
Wind tweets (filtered for relevance) versus Named storm tweets (filtered for relevance)

- Peaks in wind tweets coincide with peaks in named storm tweets
- Also peaks in wind tweet activity at other times to named storm events – are these ‘extreme weather events’? Or just windy days?
Method – Location Inference

• Filter for **timezone** (UK/Ireland only)

• Locate tweets using **Geotag** (GPS cords)
  • **Place** (polygon cords)

• **User location** (*GPS cords, if not lookup text with Geonames*¹ *db*)

• **Place names** mentioned in tweet text (*dbpedia*² *spotlight lookup identifies placenames – then Geonames lookup to get coords*)

¹ [http://www.geonames.org/](http://www.geonames.org/)
² [https://wiki.dbpedia.org/](https://wiki.dbpedia.org/)
<table>
<thead>
<tr>
<th>Tweet Collection</th>
<th>All Tweets (unfiltered)</th>
<th>Filtered for relevance</th>
<th>% of All Tweets</th>
<th>Filtered for relevance AND located</th>
<th>% of All Tweets</th>
<th>% of Filtered for relevance Tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wind</td>
<td>26,298,449</td>
<td>831,076</td>
<td>3.2%</td>
<td>472,586</td>
<td>1.8%</td>
<td>56.9%</td>
</tr>
<tr>
<td>2. All Storm names</td>
<td>8,101901</td>
<td>278412</td>
<td>3.4%</td>
<td>214,220</td>
<td>2.6%</td>
<td>76.9%</td>
</tr>
<tr>
<td>ophelia</td>
<td>897,054</td>
<td>214,730</td>
<td>23.9%</td>
<td>167,369</td>
<td>18.7%</td>
<td>77.9%</td>
</tr>
<tr>
<td>brian</td>
<td>2,037,045</td>
<td>12,970</td>
<td>0.6%</td>
<td>9,439</td>
<td>0.5%</td>
<td>72.8%</td>
</tr>
<tr>
<td>caroline</td>
<td>1,199,149</td>
<td>8,552</td>
<td>0.7%</td>
<td>4,993</td>
<td>0.4%</td>
<td>58.4%</td>
</tr>
<tr>
<td>dylan</td>
<td>2,504,264</td>
<td>3,907</td>
<td>0.2%</td>
<td>2,410</td>
<td>0.1%</td>
<td>61.7%</td>
</tr>
<tr>
<td>eleanor</td>
<td>555,433</td>
<td>11,872</td>
<td>2.1%</td>
<td>9,761</td>
<td>1.8%</td>
<td>82.2%</td>
</tr>
<tr>
<td>fionn</td>
<td>43,936</td>
<td>1,260</td>
<td>2.9%</td>
<td>878</td>
<td>2.0%</td>
<td>69.7%</td>
</tr>
<tr>
<td>georgina</td>
<td>104,327</td>
<td>894</td>
<td>0.9%</td>
<td>650</td>
<td>0.6%</td>
<td>72.7%</td>
</tr>
<tr>
<td>emma</td>
<td>760,693</td>
<td>24,227</td>
<td>3.2%</td>
<td>18,720</td>
<td>2.5%</td>
<td>77.3%</td>
</tr>
</tbody>
</table>
## Locating Tweets

<table>
<thead>
<tr>
<th>Tweet Collection</th>
<th>All Tweets (unfiltered)</th>
<th>Filtered for relevance AND located</th>
<th>Geo co-ords</th>
<th>Place co-ords</th>
<th>User location (co-ords)</th>
<th>User location (resolvable place name)</th>
<th>Place name mentioned in text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wind</td>
<td>26,298,449</td>
<td>473,740</td>
<td>7,351</td>
<td>18,539</td>
<td>21,871</td>
<td>361,156</td>
<td>64,823</td>
</tr>
<tr>
<td>2. All Storm names</td>
<td>8,101901</td>
<td>214,220</td>
<td>1349</td>
<td>7169</td>
<td>933</td>
<td>159207</td>
<td>45562</td>
</tr>
</tbody>
</table>
Social Impact

Using all filtered tweets:
- Count tweets by date/hour
- Sentiment analysis
- Categorisation (disruption, damage, warnings, news, humour, other)

Using located tweets:
- Map tweet activity (absolute, percentile, sentiment)
- Percentile normalises for variations in population, tweet activity, etc
Locating tweets during Ex-hurricane Ophelia – 16/10/2017
Sentiment Analysis

• Can we use social media posts to infer how a person feels about an event?

Machine Learning / Natural Language processing – use tweets as training data, assign tweet to sentiment

Lexicon based – look up to dictionary of sentiment words

Measuring the Sentiment of ‘Ophelia’
Tracking the Sentiment of ‘Ophelia’
Categorisation of Tweets

Filtered tweets are manually labelled into categories:

![Graph showing categories and percentages of tweets]

- Humour: 26%
- Warnings: 18%
- Other: 18%
- News: 17%
- Disruption: 17%
- Damage: 4%
- Information: 1%
### Tweet Content – Storm Brian

*(Example tweet text from each category, not actual Twitter posts)*

<table>
<thead>
<tr>
<th>Humour</th>
<th>Disruption</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Brian? What kind of name is that for a storm? Everyone knows Brian is a snail.’</td>
<td>“Train delay: National Rail have warned of delays due to high winds from Storm Brian”</td>
</tr>
<tr>
<td>‘Am I the only one to find it really hard to take a storm called #Brian seriously?’</td>
<td>“Storm Brian latest - tree blocks railway lines and hovercraft suspended”</td>
</tr>
<tr>
<td>‘And Brian? Really? Storm Rambo or Terminator would be far better than #StormBrian’</td>
<td>“Major motorway was CLOSED after Storm Brian floods carriageway”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Warnings</th>
<th>Damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘#StormBrian could lead to travel disruption this weekend.’</td>
<td>“This is the scene this morning as the waves have damaged the Harbour Office during Storm Brian.”</td>
</tr>
<tr>
<td>‘Storm Brian set to batter UK with heavy rain and 70mph winds.’</td>
<td>“Storm Brian damage causes floodlight damage. Revised home game vs @ChesterCityFC”</td>
</tr>
<tr>
<td>‘Take care on the coast folks. Waves are quite high with #StormBrian’</td>
<td>“Scaffolding in Helsby High Street BLOWN OVER by #StormBrian high winds”</td>
</tr>
</tbody>
</table>
Next Steps...

• *(in progress)* Use of Met Office observation / WOW data to compare ‘Twitter talk’ against observed weather conditions

• *(in progress)* Apply UK Storm model to extreme weather events outside of the UK

• Categorisation of tweets over time – how does this change over the period of the storm?

• Impact terminology – can we categorise the impact of an event using tweet text?
Applying model outside the UK

Searching for heavy rainfall and flooding...

Tweets about rain/flooding above the 70th percentile of these tweets for each grid square on 29th April 2017
Blue indicates unusually high number of tweets about rainfall/flooding

Rainfall radar for the USA at 7am on 29th April 2017
Areas shaded red were most affected by heavy rainfall

http://www.wpc.ncep.noaa.gov
Summary

• Social Sensing allows us to filter the ‘noise’ from Twitter and create a dataset of social media posts relating to extreme weather events

• 60-80% of tweets can be located using location inference method (as opposed to only 1% if just use geotag)

• Using percentile of tweets at a given time and place allows us to account for population size, prevalence of Twitter use, etc

• Sentiment analysis shows us that emotion becomes less positive during the period of a named storm

• Categorisation is a work in progress, however we find that about a quarter of tweets fall into the humour category

• Overall social sensing provides a data source which can be used to help in understanding of the impact of weather events in addition to traditional methods
Thank you

Michelle Spruce: m.spruce@exeter.ac.uk
Hywel Williams: h.t.p.williams@exeter.ac.uk
SEDA Lab: http://sites.exeter.ac.uk/seda-lab/