Separating Fact from Fear:
Tracking Flu Infections on Twitter
Alex Lamb, Michael Paul, Mark Dredze
Johns Hopkins University
Flu Monitoring via Twitter

- Intuition: people tweet when they get sick

  I’ve been struck down by some sort of evil flu.

- Can aggregate flu tweets to estimate population rate

- Why? Potential for early outbreak detection

- A number of papers have shown influenza can be accurately tracked in Twitter

  Culotta 2010; Lampos et al 2010; Paul and Dredze 2011; Aramaki et al 2011; Sadilek et al 2012; Doan et al 2012
Calls to censor details of potential killer flu

Alison Caldwell

Updated December 22, 2011 08:43:08

The suppression of breakthrough research into deadly bird flu strains has been labelled scientific censorship by some, but others say it is a necessary step to prevent a possible biological attack.

Last month researchers in the Netherlands discovered that the H5N1 influenza virus, or bird flu, could develop into a dangerous virus that can spread between humans.

The H5N1 strain of bird flu is fatal in 60 per cent of human cases but only 350 people have so far died from the disease largely because it cannot be spread by sneezing or coughing.

But by using ferrets in a lab, the researchers proved it was possible to change H5N1 into an aerosol-transmissible virus that can be easily spread rapidly through the air.

The genetic mutations could trigger deadly epidemics in humans, and the scientists behind the research have now agreed to remove key details of their work from publication.

The research - known as the Erasmus study - alarmed the National Science Advisory Board for Biosecurity (NSABB), a US government science committee.

Related Story: Flu scare sparks mass Hong Kong chicken cull
Map: United States
Challenges

- People tweet about the flu in different contexts
  - “Michael Jordan's ‘flu game’ is on NBATV right now.”
  - “Starting to get worried about swine flu…”
  - “going over to a friends house to check on her son. he has the flu and i am worried about him”

- Hypothesis: some contexts will predict flu rates better than others

- Approach: use rich features to differentiate these types of tweets
Categorization of Flu Tweets

- 3 classifiers:
  - Related vs Unrelated
  - Infection vs Awareness
  - Self vs Others

- Annotations using Mechanical Turk:
  - 11,900 tweets each labeled by 3 annotators
  - We discarded low quality annotators (on gold standard)
  - We hand-corrected 14% of the labels after majority vote
Learning and Classification

- MaxEnt model, L2 regularization
  - Features: n-grams + more (next slide)
- Two-stage classification:
  - 1\textsuperscript{st} stage: related vs unrelated
  - 2\textsuperscript{nd} stage (‘related’ tweets only):
    - infection vs awareness
    - self vs others
**Features**

- **8 manually created word classes**

<table>
<thead>
<tr>
<th>Category</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infection</td>
<td>getting, got, recovered, have, having, had, has, catching, catch, cured, infected</td>
</tr>
<tr>
<td>Disease</td>
<td>bird, the flu, flu, sick, epidemic</td>
</tr>
<tr>
<td>Concern</td>
<td>afraid, worried, scared, fear, worry, nervous, dread, dreaded, terrified</td>
</tr>
<tr>
<td>Treatment/Prevention</td>
<td>vaccine, vaccines, shot, shots, mist, tamiflu, jab, nasal spray</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Features

- **Stylometry**
  - retweets, user mentions
  - links/URLs
  - emoticons (positive or negative)
Features

- Part of Speech templates
  - (subject, verb, object) tuples
  - pronoun/noun pairings ("my son has the flu")
  - whether phrases begin with a verb ("getting the flu")
  - numeric references
  - whether "flu" is a noun or adjective
  - and more – see paper!
Results - Classification

- **F1-score:**
  - Related/Unrelated: 0.77
  - Awareness/Infection: 0.80
  - Self/Others: 0.86

- **Feature ablation experiments**
  - Word class features helped the most for aware./inf.
  - Stylometry features helped the most for self/others
Results - Classification

Related vs Unrelated

F1 = 0.7562
F1 = 0.7665

Related vs Unrelated

- **N-Grams**
- **All Features**
Results - Classification of Infection vs Awareness

Infection vs Awareness

Precision

Recall

F1 = 0.7985

F1 = 0.7891

N-Grams

All Features
Results - Classification

Self vs Others

F1 = 0.8550
F1 = 0.8499

Recall vs Precision

N-Grams
All Features

Self vs Others
Influenza Surveillance

- 2 data sets from 2009-2010 and 2011-2012
  - each ~2 billion tweets
- Weekly flu rate:
  - (# flu tweets from US in week) / (# all tweets in week)
- Compare to US government weekly estimates
  - Normalized # hospital admissions for flu symptoms
- source: Centers for Disease Control and Prevention (CDC)
- Metric: Pearson correlation coefficient
## Results - Influenza

<table>
<thead>
<tr>
<th>Data</th>
<th>System</th>
<th>2009-2010</th>
<th>2011-2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>Flu Trends</td>
<td>.9929</td>
<td>.8829</td>
</tr>
<tr>
<td>Twitter</td>
<td>ATAM</td>
<td>.9698</td>
<td>.5131</td>
</tr>
<tr>
<td></td>
<td>Keywords</td>
<td>.9771</td>
<td>.6597</td>
</tr>
<tr>
<td></td>
<td>All Flu</td>
<td>.9833</td>
<td>.7247</td>
</tr>
<tr>
<td></td>
<td>Infection</td>
<td>.9897</td>
<td>.7987</td>
</tr>
<tr>
<td></td>
<td>Infection+Self</td>
<td>.9752</td>
<td>.6662</td>
</tr>
</tbody>
</table>
Flu Rate

Date

08/30/09  11/08/09  01/17/10  03/28/10  06/06/10  08/15/10

CDC
Twitter
Conclusion

- We showed that we can mine real-world trends through social media feeds
- Even though we are aggregating many tweets, we gained improvements by modeling individual tweets with rich features
- Takeaway: deeper content analysis matters!
- Our annotated data will be available soon
Thank You

- Johns Hopkins HLT COE is hiring!
- Research scientist / Postdoc positions
- [http://hltcoe.jhu.edu](http://hltcoe.jhu.edu)