CHALLENGES IN INFLUENZA FORECASTING AND OPPORTUNITIES FOR SOCIAL MEDIA

MICHAEL J. PAUL, MARK DREDZE, DAVID BRONIATOWSKI
NOVEL DATA STREAMS FOR INFLUENZA SURVEILLANCE

• New technology allows us to analyze new types of data to infer influenza prevalence
  • Especially data from the Web

• Most (in)famously Google Flu Trends

• Many other promising sources:
  • Social media (especially Twitter)
  • Mobile apps
  • Wikipedia
WHAT ABOUT FORECASTING?

• Detection is easy and utility is limited
  • Reliable forecasting is important for making preparations and allocating resources

• Google Flu has been shown to improve forecasting
  • Shaman and Karspeck (2012)
  • Nsoesie, Marathe, Brownstein (2013)
  • Dugas et al. (2013)

• Social media hasn’t been evaluated yet
**CDC PREDICT THE FLU CHALLENGE**

**CDC Competition Encourages Use of Social Media to Predict Flu**

November 25, 2013 — CDC has launched the “Predict the Influenza Season Challenge,” a competition designed to foster innovation in flu activity modeling and prediction. The registrant who most successfully predicts the timing, peak and intensity of the 2013-2014 flu season using social media data (e.g., Twitter, internet search data, web surveys) will receive an award of $75,000 and CDC recognition. Full details of the contest requirements—including eligibility rules, how to enter the contest, and scoring—are available via the official contest announcement at [https://federalregister.gov/a/2013-28198](https://federalregister.gov/a/2013-28198).

- Contest to forecast the 2013-14 flu season by augmenting existing surveillance with Web data
- Three metrics:
  - Start of season
  - Peak of season
  - Intensity of season (peak rate and duration)
So what exactly are we trying to predict?

- **ILINet**
  - CDC-run network of thousands of US providers
  - Hospitals report % of outpatients seen for influenza-like illness
  - Weekly reports of estimated ILI prevalence
  - Most commonly used flu metric

- Data is lagged by a week
  - Real time surveillance doesn’t exist through traditional means
  - This is why novel data streams can help
FORECASTING MODEL

- Forecasts and current-week nowcasts can be produced using standard time series models with the lagged ILINet data

- Basic autoregressive model:

\[ y_{w+k} = \alpha_1 \tilde{y}_{w-1} + \alpha_2 \tilde{y}_{w-2} + \alpha_3 \tilde{y}_{w-3} \]

- This works quite well
  - Especially for nowcasting
Can we improve this with social media data?

- Twitter can give estimates for the current week
- These estimates can be included in the model

\[ y_{w+k} = \gamma z_w + \alpha_1 \tilde{y}_{w-1} + \alpha_2 \tilde{y}_{w-2} + \alpha_3 \tilde{y}_{w-3} \]
TWITTER FLU DETECTION

• We used our state-of-the-art Twitter system
  • Lamb et al (2013) and Broniatowski et al (2013)
• Two streams downloading data since Nov 2011
  • 1% sample and stream filtered for health keywords
  • About 4 million per day
• Cascade of tweet classifiers:
  • Relevant to health
  • Relevant to flu
  • Indicates flu infection (vs general awareness)
• Can produce daily or weekly prevalence estimates
  
  \[
  \frac{\text{# of tweets classified as flu infection}}{\text{# of tweets from full sample}}
  \]
RESULTS

- Mean absolute error when *nowcasting*:

<table>
<thead>
<tr>
<th>Data</th>
<th>Average error per season</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2011-12</td>
</tr>
<tr>
<td>ILINet</td>
<td>.19</td>
</tr>
<tr>
<td>Twitter</td>
<td>.34</td>
</tr>
<tr>
<td>ILINet+Twitter</td>
<td>.15</td>
</tr>
</tbody>
</table>
RESULTS

• Mean absolute error when forecasting:

<table>
<thead>
<tr>
<th>$k$</th>
<th>ILI Only</th>
<th>ILI+Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.28 ± .07</td>
<td>.19 ± .03</td>
</tr>
<tr>
<td>1</td>
<td>.41 ± .14</td>
<td>.29 ± .06</td>
</tr>
<tr>
<td>2</td>
<td>.51 ± .20</td>
<td>.37 ± .09</td>
</tr>
<tr>
<td>3</td>
<td>.62 ± .26</td>
<td>.47 ± .12</td>
</tr>
<tr>
<td>4</td>
<td>.75 ± .32</td>
<td>.56 ± .15</td>
</tr>
<tr>
<td>5</td>
<td>.88 ± .39</td>
<td>.65 ± .19</td>
</tr>
<tr>
<td>6</td>
<td>.98 ± .45</td>
<td>.75 ± .23</td>
</tr>
<tr>
<td>7</td>
<td>1.05 ± .50</td>
<td>.83 ± .28</td>
</tr>
<tr>
<td>8</td>
<td>1.12 ± .54</td>
<td>.89 ± .32</td>
</tr>
<tr>
<td>9</td>
<td>1.18 ± .57</td>
<td>.93 ± .34</td>
</tr>
<tr>
<td>10</td>
<td>1.19 ± .55</td>
<td>.91 ± .36</td>
</tr>
</tbody>
</table>
INFLUENZA DATA REVISIONS

An important caveat about historical data...

• Weekly ILINet values are subject to future revisions

• We were careful to train the models on the data that would have been available at the time of the prediction
  • But we evaluated on the gold standard value from the final report for the season
INFLUENZA DATA REVISIONS

An important caveat about historical data...

- The value initially reported has an average absolute difference from the final value of 0.18

- The value reported after 3 weeks still has an average difference of 0.10
## INFLUENZA DATA REVISIONS

<table>
<thead>
<tr>
<th>Data</th>
<th>Average error per season</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2011-12</td>
</tr>
<tr>
<td>ILINet (current)</td>
<td>.19</td>
</tr>
<tr>
<td>ILINet (final)</td>
<td>.11</td>
</tr>
</tbody>
</table>

- Error is greatly underestimated when using the final gold values instead of values available at time of forecast.
Forecasting milestones of the 2013-2014 season
We also compared to Google Flu Trends

<table>
<thead>
<tr>
<th>Data</th>
<th>Average error per season</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2011-12</td>
</tr>
<tr>
<td>ILINet</td>
<td>.19</td>
</tr>
<tr>
<td>ILINet+Twitter</td>
<td>.15</td>
</tr>
<tr>
<td>ILINet+Google</td>
<td>.20</td>
</tr>
</tbody>
</table>

Twitter improved nowcasting and forecasting more than Google
CONCLUSION #1

• Twitter improves influenza forecasting
  • For a given level of accuracy, including Twitter can give you 2-4 weeks of additional forecasting ability

• Twitter outperforms Google
  • At least in these three seasons
  • Google recently updated their model so comparison is difficult
CONCLUSION #2

• When using historical data, be careful to use data that actually would have been available at the time of model training

• Others have assumed these were the same
  • Our results showed that this has a substantial effect on performance
CONCLUSION #3

• Always compare to a simple time series baseline

• Our results showed that Twitter by itself is worse than using lagged ILINet data
  • No one had compared to this (using Twitter)
  • But we then showed that you can do even better by combining both!
THANK YOU