From Tweets to Polls:
Linking Text Sentiment to Public Opinion Time Series

Brendan O’Connor
Ramnath Balasubramanyan
Bryan R. Routledge
Noah A. Smith

Carnegie Mellon University
Measuring public opinion through social media?

People in U.S.

Can we derive a similar measurement?

I do not like Obama

I do not

Aggregate Text Sentiment Measure
Contributions

• Correlations between
  1. Very simple text sentiment analysis
  2. Telephone public opinion polls
     • Consumer confidence and Presidential job approval

• Time-series smoothing is a critical issue

• Also
  – Topic selection, topic volumes, text leads polls, stemming, election polling
Rest of talk

• Data Overview
• Analysis
• Discussion and Related Work
• New Results!
Text Data: Twitter

• Twitter is large, public, and all in one place

• Sources
  1. Archiving Twitter Streaming API
     “Gardenhose”/”Sample”: ~15% of public tweets
  2. Scrape of earlier messages via API
     thanks to Brendan Meeder

• Sizes
  – 0.7 billion messages, Jan 2008 – Oct 2009
  – 1.5 billion messages, Jan 2008 – May 2010
Message data

{
  "text": "Time for the States to fight back!!! Tenth Amendment Movement: Taking On the Feds http://bit.ly/14t1RV #tcot #teaparty",
  "created_at": "Tue Nov 17 21:08:39 +0000 2009",
  "geo": null,
  "id": 5806348114,
  "in_reply_to_screen_name": null,
  "in_reply_to_status_id": null,

  "user": {
    "screen_name": "TPO_News",
    "created_at": "Fri May 15 04:16:38 +0000 2009",
    "description": "Child of God - Married - Gun carrying NRA Conservative - Right Winger hard Core Anti Obama (Pro America), Parrothead - www.ABoldStepBack.com #tcot #nra #BlogHer2010",
    "followers_count": 10470,
    "friends_count": 11328,
    "name": "Tom O'Halloran",
    "profile_background_color": "f2f5f5",
    "profile_image_url": "http://a3.twimg.com/profile_images/295981637/TPO_Balcony_normal.jpg",
    "protected": false,
    "statuses_count": 21147,
    "location": "Las Vegas, Baby!!",
    "time_zone": "Pacific Time (US & Canada)",
    "url": "http://www.tpo.net/1dollar",
    "utc_offset": -28800,
  }
}
Message data we use

{
  "created_at": "Tue Nov 17 21:08:39 +0000 2009",
  "geo": null,
  "id": 5806348114,
  "in_reply_to_screen_name": null,
  "in_reply_to_status_id": null,

  "user": {
    "screen_name": "TPO_News",
    "created_at": "Fri May 15 04:16:38 +0000 2009",
    "description": "Child of God - Married - Gun carrying NRA Conservative - Right Winger hard Core Anti Obama (Pro America), Parrothead - www.ABoldStepBack.com tcot nra iPhone",
    "followers_count": 10470,
    "friends_count": 11328,
    "name": "Tom O'Halloran",
    "profile_background_color": "f2f5f5",
    "profile_image_url": "http://a3.twimg.com/profile_images/295981637/TPO_Balcony_normal.jpg",
    "protected": false,
    "statuses_count": 21147,
    "location": "Las Vegas, Baby!",
    "time_zone": "Pacific Time (US & Canada)",
    "url": "http://www.tpo.net/1dollar",
    "utc_offset": -28800,
  }
}
Poll Data

• Consumer confidence, 2008-2009
  – Index of Consumer Sentiment (Reuters/Michigan)
  – Gallup Daily (free version from gallup.com)

• 2008 Presidential Elections
  – Aggregation, Pollster.com

• 2009 Presidential Job Approval
  – Gallup Daily

• Which tweets correspond to these polls?
Message selection via topic keywords

- Analyzed subsets of messages that contained manually selected topic keyword
  - “economy”, “jobs”, “job”
  - “obama”
  - “obama”, “mccain”

- High day-to-day volatility
  - Fraction of messages containing keyword
  - Nov 5 2008: 15% contain “obama”
Sentiment analysis: word counting

• Subjectivity Clues lexicon from OpinionFinder / U Pitt
  – Wilson et al 2005
  – 2000 positive, 3600 negative words

• Procedure
  1. Within topical messages,
  2. Count messages containing these positive and negative words
A note on the sentiment list

• This list is not well suited for social media English.
  – “sucks”, “ :) ”, “ :( ”

• Examples for one day.

(Top examples)

<table>
<thead>
<tr>
<th>word</th>
<th>valence</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>will</td>
<td>positive</td>
<td>3934</td>
</tr>
<tr>
<td>bad</td>
<td>negative</td>
<td>3402</td>
</tr>
<tr>
<td>good</td>
<td>positive</td>
<td>2655</td>
</tr>
<tr>
<td>help</td>
<td>positive</td>
<td>1971</td>
</tr>
</tbody>
</table>

(Random examples)

<table>
<thead>
<tr>
<th>word</th>
<th>valence</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>funny</td>
<td>positive</td>
<td>114</td>
</tr>
<tr>
<td>fantastic</td>
<td>positive</td>
<td>37</td>
</tr>
<tr>
<td>cornerstone</td>
<td>positive</td>
<td>2</td>
</tr>
<tr>
<td>slump</td>
<td>negative</td>
<td>85</td>
</tr>
<tr>
<td>bearish</td>
<td>negative</td>
<td>17</td>
</tr>
<tr>
<td>crackdown</td>
<td>negative</td>
<td>5</td>
</tr>
</tbody>
</table>
Sentiment Ratio over Messages

For one day $t$ and topic word, compute score

$$\frac{\text{MessageCount}_t(\text{pos. word AND topic word})}{\text{MessageCount}_t(\text{neg. word AND topic word})} = \frac{p(\text{pos. word} \mid \text{topic word}, t)}{p(\text{neg. word} \mid \text{topic word}, t)}$$
Sentiment Ratio Moving Average

- High day-to-day volatility.
- Average last $k$ days.
- Keyword “jobs”, $k = 1, 7, 30$
- (Gallup tracking polls: 3 or 7-day smoothing)

$$MA_t = \frac{1}{k} \left( x_{t-k+1} + x_{t-k+2} + \ldots + x_t \right)$$
Sentiment Ratio Moving Average

- High day-to-day volatility.
- Average last $k$ days.
- Keyword “jobs”, $k = 1, 7, 30$
- (Gallup tracking polls: 3 or 7-day smoothing)

$$MA_t = \frac{1}{k} \left( x_{t-k+1} + x_{t-k+2} + \ldots + x_t \right)$$
Sentiment Ratio Moving Average

- High day-to-day volatility.
- Average last $k$ days.
- Keyword "jobs", $k = 1, 7, 30$
- (Gallup tracking polls: 3 or 7-day smoothing)

$$MA_t = \frac{1}{k} \left( x_{t-k+1} + x_{t-k+2} + \ldots + x_t \right)$$
Smoothed comparisons

“jobs” sentiment

window = 1, r = 0.064
Smoothed comparisons
“jobs” sentiment

window = 2, r = 0.380
Smoothed comparisons
“jobs” sentiment

window = 3, r = 0.513
Smoothed comparisons
“jobs” sentiment

window = 4, r = 0.591
Smoothed comparisons
“jobs” sentiment

window = 5, r = 0.677
Smoothed comparisons
“jobs” sentiment

window = 6, r = 0.766
Smoothed comparisons
“jobs” sentiment

window = 7, r = 0.766

Gallup Poll
Twitter Sentiment
Smoothed comparisons

“jobs” sentiment

window = 8, r = 0.735
Smoothed comparisons
“jobs” sentiment

window = 9, r = 0.756
Smoothed comparisons
“jobs” sentiment

window = 10, r = 0.770
Smoothed comparisons
“jobs” sentiment

\[ \text{window} = 11, \ r = 0.781 \]
Smoothed comparisons

“jobs” sentiment

window = 12, r = 0.798

Gallup Poll
Twitter Sentiment
Smoothed comparisons
“jobs” sentiment

window = 13, r = 0.823

Gallup Poll
Twitter Sentiment
Smoothed comparisons
“jobs” sentiment

window = 14, r = 0.819

Gallup Poll
Twitter Sentiment
Smoothed comparisons

“jobs” sentiment

window = 15, r = 0.804
Smoothed comparisons
“jobs” sentiment

Sept. 15, 2008:
Lehman collapse, AIG bailout

Feb 2009:
Stock market bottoms out, begins recovery

r = 0.804

Gallup Poll
Twitter Sentiment
Which leads, poll or text?

- Cross-correlation analysis: between
  - Sentiment score for day $t$
  - Poll for day $t+L$
- “jobs” leading indicator for the poll
- (Can turn into forecasting model: see paper)
Keyword message selection

• 15-day windows, no lag
  – “jobs” \( r = 80\% \)
  – “job” \( r = 7\% \)
  – “economy” \( r = -10\% \)

• Look out for stemming
  – (“jobs” OR “job”) \( r = 40\% \)
Presidential elections and job approval

• 2008 elections
  – “obama” and “mccain” sentiment do not correlate
  – But, “obama” and “mccain” volume => 79%, 74% (!)
  – Simple indicator of election news?
Presidential elections and job approval

• 2008 elections
  – “obama” and “mccain” sentiment do not correlate
  – But, “obama” and “mccain” volume => 79%, 74% (!)
  – Simple indicator of election news?
• 2009 job approval
  – “obama” => r = 72%
  – Looks easy: simple decline
## Related work: aggregate sentiment

<table>
<thead>
<tr>
<th></th>
<th>Text</th>
<th>Message Selection</th>
<th>Opinion Estimation</th>
<th>External Correlate</th>
</tr>
</thead>
<tbody>
<tr>
<td>This work – O’Connor et al ICWSM-2010</td>
<td>Microblogs (Twitter)</td>
<td>Keywords related to poll</td>
<td>Word counting (OpinionFinder)</td>
<td>Opinion polls</td>
</tr>
<tr>
<td>Mishne and de Rijke 2006</td>
<td>Blogs (Livejournal)</td>
<td>N/A</td>
<td>Linear model (words, time)</td>
<td>Mood labels</td>
</tr>
<tr>
<td>Dodds and Danforth 2009</td>
<td>Blogs, Speeches, Songs</td>
<td>N/A</td>
<td>Word counting (LIWC)</td>
<td>Exploratory (mostly)</td>
</tr>
<tr>
<td>Gilbert and Karahalios ICWSM-2010</td>
<td>Blogs (Livejournal)</td>
<td>N/A</td>
<td>Decision tree + NB (words)</td>
<td>Stocks</td>
</tr>
<tr>
<td>Asur and Huberman 2010</td>
<td>Microblogs (Twitter)</td>
<td>Movie name</td>
<td>NB-like model (char. n-grams)</td>
<td>Movie sales</td>
</tr>
<tr>
<td>Bollen et al 2010</td>
<td>Microblogs (Twitter)</td>
<td>N/A</td>
<td>Word counting (POMS)</td>
<td>Stocks, politics</td>
</tr>
<tr>
<td>Tumasjan et al ICWSM-2010</td>
<td>Microblogs (Twitter)</td>
<td>Party name</td>
<td>Word counting (POMS)</td>
<td>Elections</td>
</tr>
<tr>
<td>Kramer 2010</td>
<td>Microblogs (Facebook Wall)</td>
<td>N/A</td>
<td>Word counting (LIWC)</td>
<td>Life satisfaction answers</td>
</tr>
<tr>
<td>... many more!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

• Preliminary results that sentiment analysis on Twitter data can give information similar to traditional opinion polls
  – But, still not well-understood
  – Twitter bias?
  – News vs. opinion?

• Issues
  – Relevant message selection
  – Time series smoothing

• Replacement for polls? Promising but not quite yet