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

Brendan O’Connor  http://brenocon.com
Machine Learning Department
Carnegie Mellon University
Presentation at AAPOR, May 2012

Joint work with: Ramnath Balasubramanyan,
Bryan R. Routledge, Noah A. Smith; CMU

Originally presented at ICWSM-2010: the AAAI International Conference on Weblogs and Social Media. Published in its proceedings; paper available on website
Measuring public opinion through social media?

People in U.S.

-I like Obama
-I do not

Can we derive a similar measurement?

Aggregate Text Sentiment Measure
Contributions

• Correlations between
  1. Very simple text sentiment analysis
  2. Telephone public opinion polls
     • Consumer confidence and Presidential job approval
• Negative results as well!
• Also
  – Time-series smoothing is a critical issue
  – Topic selection, topic volumes, text leads polls, stemming, election polling
Text Data: Twitter

• A large and public social media service
  – Good: Has people writing their thoughts
  – Bad: News, celebrities, “media” stuff (?)

• Sources
  1. Archiving the Twitter Streaming API
     “Gardenhose”/”Sample”: ~10-15% of public tweets
  2. Scrape of earlier messages
     thanks to Brendan Meeder, CMU

• ~2 billion messages before topic selection, 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 #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,
   }
}
Message data we use

```
{
    "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"
}
```

1. Text
2. Timestamp

- Message data we do not use:
  - Locations from GPS
  - Locations from IP addresses – not public
  - User information (name, description, self-described location)
  - Conversation structure: retweets, replies
  - Social structure: follower network
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?
Text sentiment analysis method

- Two-stage process
  1. **Topic Selection**: select topical tweets, via hand-selected keywords
  2. **Sentiment Analysis**: For one topic’s messages, count “sentiment polarity” words, from pre-existing lexicon

(More sophisticated methods are possible, for both stages)
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% of tweets contain “obama”
Sentiment analysis: word counting

• Subjectivity Clues lexicon from OpinionFinder (Univ. of Pittsburgh)
  – 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 a particular topic word, compute score

\[
\text{SentimentRatio}( \text{topic\_word, } t ) = \frac{\text{MessageCount}_t(\text{pos. word AND topic word})}{\text{MessageCount}_t(\text{neg. word AND topic word})}
\]

\[
= \frac{p(\text{pos. word | topic word, } t)}{p(\text{neg. word | topic word, } t)}
\]
Sentiment Ratio Moving Average

- High day-to-day volatility.
- Average last $k$ days.
- $\text{SentRatio(“jobs”), } k = 1$
- (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.
• $\text{SentRatio(“jobs”), } k = 1, 7$
• (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
SentimentRatio("jobs")

window = 1, r = 0.064
Smoothed comparisons

SentimentRatio("jobs")

\[ \text{window} = 2, \ r = 0.380 \]
Smoothed comparisons
SentimentRatio(”jobs”)
Smoothed comparisons
SentimentRatio("jobs")

window = 4, r = 0.591
Smoothed comparisons

SentimentRatio("jobs")

window = 5, r = 0.677
Smoothed comparisons

SentimentRatio("jobs")

window = 6, r = 0.766
Smoothed comparisons

SentimentRatio("jobs")

window = 7, r = 0.766
Smoothed comparisons

SentimentRatio("jobs")

window = 8, r = 0.735
Smoothed comparisons
SentimentRatio(”jobs”)

window = 9, r = 0.756
Smoothed comparisons
SentimentRatio("jobs")

window = 10, r = 0.770
Smoothed comparisons
SentimentRatio("jobs")

window = 11, r = 0.781
Smoothed comparisons

SentimentRatio("jobs")

window = 12, r = 0.798
Smoothed comparisons

SentimentRatio("jobs")

window = 13, r = 0.823
Smoothed comparisons
SentimentRatio("jobs")

window = 14, r = 0.819
Smoothed comparisons
SentimentRatio("jobs")

window = 15, r = 0.804
Smoothed comparisons

SentimentRatio("jobs")

- Sept. 15, 2008: Lehman collapse, AIG bailout
- Feb 2009: Stock market bottoms out, begins recovery

$r = 0.804$
Which leads, poll or text?

• Cross-correlation analysis: between
  – SentimentRatio(“jobs”) on day $t$
  – Poll for day $t+L$

• “jobs” leading indicator for the poll
Keyword message selection

• How much do sentiment scores for different keywords correlate to Gallup?

• 15-day windows, no lag
  – SR("jobs")   \( r = +0.80 \)
  – SR("job")   \( r = +0.07 \)
  – SR("economy")   \( r = -0.10 \)

• Look out for stemming
  – SR("jobs" OR "job")   \( r = +0.40 \)
Presidential elections and job approval

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

• 2008 elections
  – SR(“obama”) and SR(“mccain”) sentiment do not correlate
  – But, “obama” and “mccain” volume: $r = .79, .74$ (!)
  – Simple indicator of election news?

• 2009 job approval
  – SR(“obama”): $r = .72$
  – Looks easier: simple decline
Challenges

• Preliminary results that sentiment analysis on Twitter data can give information similar to opinion polls
  – **But**, still not well-understood!
  – Who is using Twitter?
    • Massive changes over time (2009 Twitter ≠ 2012 Twitter)
  – News vs. opinion?
    • Other data sources might better distinguish?
  – Better text analysis
    – Very wide linguistic variation on Twitter
    – Word sense ambiguity: “steve jobs”
  – Better data sources
  – Suggestion for future work: analyze correlations to pre-existing surveys and other attitude measurements
    – If someone tries to sell you a sentiment analysis system, make sure they at least correlate to a gold standard you trust
    – There is lots of mediocre work on Twitter now
• Related work: demographic and geographic-specific words
• Not a replacement for polls, but seems potentially useful.
Thanks!

Paper at: http://brenocon.com
General population 

Twitter users 

Tweet texts 

Natural language processing
Attempting to estimate \( P(\text{opinion} | \text{text}) \)

Topic selection 

Sentiment estimation 

Measuring endpoint correlations

Surveys 

\( P(\text{twitter user}) \) 

\( P(\text{text} | \text{twitter user, opinion}) \) 

Ignoring effects of social group! e.g. Teenagers, young adults, African-Americans overrepresented
A partial taxonomy of Twitter messages

Official announcements
- BritishMonarchy TheBritishMonarchy
  On 6 Jan: Changing the Guard at Buckingham Palace - Starts at approx 11am http://www.royal.gov.uk/G
  17 hours ago

Business advertising
- bigdogcoffee bigdogcoffee
  Back to normal hours beginning tomorrow.........Monday-Friday 6am-10pm Sat/Sun 7:30am-10pm
  2 Jan

Links to blog and web content
- crampell Catherine Rampell
  Casey B. Mulligan: Assessing the Housing Sector - http://nyti.ms/hcUKK9
  10 hours ago

Celebrity self-promotion
- THE_REAL_SHAQ THE_REAL_SHAQ
  fill in da blank, my new years shaqulation is __________
  4 Jan

Status messages
- emax electronic max
  1.1.11 - britons and americans can agree on the date for once. happy binary day!
  1 Jan

Group conversation
- _siddx3 Evelyn Santana
  RT @ _LusciousVee: #EveryoneShouldKnow Ima Finally Be 18 This Year ^^
  3 minutes ago

Personal conversation
- xoxoJuicyCee CeeCee^*
  @fxxknCelly aha kayy goodnightt (: 4 Jan

Slide from Jacob Eisenstein, collaborator for papers: “A Latent Variable Model for Geographic Lexical Variation” and “A Mixture Model of Demographic Lexical Variation.”
Geographic Variation, Slang
<table>
<thead>
<tr>
<th>Location</th>
<th>“basketball”</th>
<th>“chit chat”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston</td>
<td>CELTICS victory</td>
<td>ese exam suttin sippin</td>
</tr>
<tr>
<td>N. California</td>
<td>THUNDER KINGS GIANTS</td>
<td>hella flirt hut iono OAKLAND</td>
</tr>
<tr>
<td>New York</td>
<td>NETS KNICKS</td>
<td>wassssup nm</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>#KOBE #LAKERS AUSTIN</td>
<td>wyd coo af nada tacos messin fasho bomb</td>
</tr>
</tbody>
</table>

PISTONS KOBE LAKERS game DUKE NBA CAVS STUCKEY JETS KNICKS

lol smh jk yea wyd coo ima wassup somethin JP
**Related work as of 2010: aggregate sentiment**

<table>
<thead>
<tr>
<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>Keywords related to poll</td>
<td>Word counting (OpinionFinder)</td>
<td>Opinion polls</td>
</tr>
<tr>
<td>Mishne and de Rijke 2006</td>
<td>N/A</td>
<td>Linear model (words, time)</td>
<td>Mood labels</td>
</tr>
<tr>
<td>Dodds and Danforth 2009</td>
<td>N/A</td>
<td>Word counting (LIWC)</td>
<td>Exploratory (mostly)</td>
</tr>
<tr>
<td>Gilbert and Karahalios ICWSM-2010</td>
<td>N/A</td>
<td>Decision tree + NB (words)</td>
<td>Stocks</td>
</tr>
<tr>
<td>Asur and Huberman 2010</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>N/A</td>
<td>Word counting (POMS)</td>
<td>Stocks, politics</td>
</tr>
<tr>
<td>Tumasjan et al ICWSM-2010</td>
<td>Party name</td>
<td>Word counting (POMS)</td>
<td>Elections</td>
</tr>
<tr>
<td>Kramer 2010</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>
</tr>
</tbody>
</table>
• See http://brenocon.com for the paper, published 2010
  – [And related work: statistical discovery of geographic dialects and slang on Twitter; Censorship of political content on Chinese social media]
General population

Twitter users

Tweet texts

Text sentiment measure

Who uses twitter?

How do people express themselves on Twitter?

How do natural language processing algorithms assess expressed opinions?

Every stage needs probability estimates

These vary by demographics, social groups, sentiment, topics, time...