Creating Measures of Employment Phenomena using Social Media

Dolan Antenucci
Mike Cafarella
Margaret Levenstein
Matthew Shapiro
University of Michigan
Describing Real-World Phenomena

- Social media indicative of real world phenomena
- Deriving from online activity can be cheap and quick
- Would be useful to economists and policy makers
- Example: U.S. unemployment rate

Phone polls → Social Media

2 weeks later

Unemployment Rate

Instant
Prediction vs. Indication

Prediction:
• Using online activity to predict a future real-world event
• Example: the 2011 UK riots before they happen

Indication:
• Using online activity to indicate what is happening “now”
• Example: this month’s job growth numbers for U.S.
Decision Makers Want Timely Data

- Ten-day auto sales
- Weekly retail sales
- Housing permit data

Predicts end of mid-1970’s recession
1. Introduction/Motivation

2. Related Work

3. Our Approach

4. Future Work
Google Flu Trends

- Attempts to predict flu outbreaks based on users’ search queries
Other Indicators via Social Media

Similar uses of Google Searches

• German unemployment rates and US mortgage delinquencies
• Predict future housing sales in each of 50 U.S. states

Twitter has been used in a similar manner

• Earthquake detection
• Expanded illness-tracking
1. Introduction/Motivation
2. Related Work
3. Our Approach
4. Future Work
5. Conclusion
Case Study: Unemployment Behavior

Proposed experiment:

- Short-term: Use Twitter activity as indicator of new job loss
- Long-term: “Live” indicators of job loss, creation, and search

Why use Twitter?

- Large dataset of timestamps, text, geography and personal information
- Growing diversity of users

Why use Unemployment Insurance claims?

- Weekly dataset to benchmark our indicator’s performance
About Our Experiment

500M Tweets from 2009

6/30: “I just lost my job again.”

8/5: “I finally found a job!”

Weekly Social Media Signals

<table>
<thead>
<tr>
<th>Event</th>
<th>Week</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lost a job</td>
<td>27</td>
<td>14</td>
</tr>
<tr>
<td>Found a job</td>
<td>19</td>
<td>6</td>
</tr>
<tr>
<td>Found a job</td>
<td>32</td>
<td>3</td>
</tr>
</tbody>
</table>

1. Use search strings to build time-varying signals
2. Test signals for correlation with unemployment data
3. Using correlated signals, build linear regression model in order to predict unemployment
Results of Our Experiment

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE (# new UI Claims)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (Previous week’s UI claims)</td>
<td>89,300</td>
</tr>
<tr>
<td>Top Indicator Model</td>
<td>48,300</td>
</tr>
</tbody>
</table>
1. Introduction/Motivation
2. Related Work
3. Our Approach
4. Future Work
Challenges

- Difficult to identify relevant social media signals

- A high-quality indicator may require many social media signals

Current Work

- Takes “unemployment behavior” and outputs “found a job”, “lost a job”, etc.
More challenges

- Relevant media-derived signals rare

> Total signals found: 0.01% of all Tweets

- Correlated media-derived signals that have nothing to do with each other
From “Gee Whiz” to National Statistics

• Social media based measures allow us to capture households often missed by traditional survey instruments
  – People without landlines still tweet
  – People voluntarily tweet or post personal information they resist giving when asked directly
• Developing weighting procedures to provide adjustments for differences between workforce and Twitter population
• Use ongoing, traditional surveys to measure use of different kinds of social media by different groups
  – Current Population Survey
  – Health and Retirement Study
  – American Life Panel
From Proof-of-Concept to Production

• Create portal to provide broad user access to automated production of economic measures based on query parameters
  – More than counting posts or searches
  – Models that map from Tweets to meaningful economic series
    • Extracting meaning from Tweets
    • Weighting Tweets from different people and places
• New, high frequency University of Michigan times series of key employment measures
What’s better about these measures?

• Timely – nearly instantaneous – measures of changes in employment
  – New reports of job loss
  – New reports of hiring

• Very local measures of employment change
  – Geography built into tweets
  – Inferred from text and photographs
Improving our understanding of unemployment dynamics

• Better measures of job search strategy and the causes and consequences of long term unemployment

  – Interview reports: why didn’t I get this job?
    • Over-qualified? Under-qualified? Salary requirements too high? Too far from my under-water home?

  – New reports of self-employment
    • Informal economy

  – Loss of UI benefits
    • Timing of job losses during recession and expected end of extended unemployment compensation
Highlights

• Initial results suggest that we can reasonably replicate time series properties of most frequent extant unemployment measures with Twitter-based measure
  – Continuing to improve with better measures for capturing meaning from Tweets

• Moving to production
  – High frequency, social media based measures of employment creation and loss
  – Portal to allow user queries to generate new measures

• Insights into current research and policy questions on the changing nature of unemployment in the U.S.