Measure Country-Level Socio-Economic Indicators with Streaming News: An Empirical Study

Bonan Min, Xiaoxi Zhao
Motivation

- **Social-economic indicator**
  - Measuring economic conditions
  - Input for policy maker (e.g. president)

- **Very difficult to obtain**
  - Expensive (census is usually conducted every 10 years)
  - Inaccurate (e.g. unemployment rate)
  - Laggy

- **Hypothesis**
  - Correlation between events in streaming news and economic events.
  - Streaming news can be used to measure these indicators
Framework

- Data: English Gigaword corpus (1997-2010)

<table>
<thead>
<tr>
<th>Event triggers</th>
<th>Location</th>
<th>Time</th>
<th>Related*</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEPT. 1, 2011 The bankruptcies of three American solar power companies in the last month</td>
<td>American (USA)</td>
<td>Last month (2011-08)</td>
<td>Y</td>
</tr>
<tr>
<td>The man accused of fatally shooting his estranged wife inside a New Jersey church last Tuesday</td>
<td>New Jersey (USA)</td>
<td>Last Tuesday (2009-07)</td>
<td>N</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Extract events

1. Label predicates using SRL (He et al 2017)
2. Extract predicates as triggers, entity, time, location mentions
3. Refine location into country-level GPE
4. Refine time entity (absolute and relative time entities)
Find relevant events

- Process 500 documents (as above)
- Annotate trigger words for target indicators
- Using wordnet to extend the word list
- Refine word list by another annotator
Measure indicator

1. Aggregate and count in 1-month bin
2. Normalize
3. Average smoothing (7 months)

\[
ECIM_{i,t} = \frac{1}{T} \sum_{t' \in [t-T/2,t+T/2]} \frac{\sum_{e \in E_i} N_{e,t'}}{M_{t'}}
\]
Results

Unemployment Rate

ECIM
MA of ECIM
Unemployment Rate
MA of Unemployment Rate

VIX

ECIM
MA of ECIM
VIX
MA of VIX

Year


ECIM

VIX

10 20 30 40 50 60
Correlation test

Pearson correlation test: To test the confidence of the correlation between two variables

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Pearson</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unemployment rate</td>
<td>0.4286</td>
<td>0.0000</td>
</tr>
<tr>
<td>$(1/M_{comp})$</td>
<td>(0.4877)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>EPU News</td>
<td>0.5136</td>
<td>0.0000</td>
</tr>
<tr>
<td>VIX</td>
<td>0.4115</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 2: Correlation coefficients between ECIMs and indicators. For unemployment rate, we also show its correlation with $1/M_{comp}$ (in parentheses).