On Critical Event Observability using Social Networks: A Disaster Monitoring Perspective

Dong-Anh Nguyen, Tarek Abdelzaher, Steven Borbash, Xuan-Hong Dang, Raghu Ganti, Ambuj Singh, Mudhakar Srivatsa
Agenda

• Problem Definition
• Data
• Models
• Experiments
• Conclusion
Problem Definition

• Extreme events
  – Hurricane Sandy

• Event observability
  – Broadcast comm. media
  – Online social networks

• Event predictability?
Data

• Gas Shortage Damage
  – All Hazard Consortium
  – Daily fuel availability with address, longitude, latitude

• Twitter Social Sentiment
  – Sentiment Analysis
  – 711 tweets
Data

- **Gas Shortage Damage**
  - All Hazard Consortium
  - Daily fuel availability with address, longitude, latitude
- **Twitter Social Sentiment**
  - Sentiment Analysis
  - 711 tweets
Models

• Social Sentiment Modeling
  – Regression
  – ARMA model

\[
S(t) = \sum_{i=d_D}^{d_D+n_D-1} a_i \cdot D(t - i) + \sum_{j=1}^{n_S} b_j \cdot S(t - j)
\]

where:

d_D: Delay time on Gas shortage (D) time series
n_D: No. of days looking back on D starting from d_D
n_S: No. of days looking back on Social Sentiment time series
Models

• Social Sentiment Modeling
  – Regression
  – ARMA model

\[ S(t) = \sum_{i=d_p}^{d_p+n_p-1} a_i \cdot D(t - i) + \sum_{j=1}^{n_s} b_j \cdot S(t - j) \]

eample:

\[ S(t) = 0.168 \cdot D(t - 1) + 0.428 \cdot D(t - 2) + 0.217 \cdot D(t - 3) - 0.414 \cdot S(t - 1) \]
## Experiments – Social Sentiment Modeling

<table>
<thead>
<tr>
<th>Params</th>
<th>$n_D$</th>
<th>$d_D$</th>
<th>$n_S$</th>
<th>$d_S$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0.066</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0.065</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0.065</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.079</td>
</tr>
</tbody>
</table>

**TABLE I.** Comparison of different linear models

<table>
<thead>
<tr>
<th>Delay</th>
<th>$n_D$</th>
<th>$n_S$</th>
<th>$d_S$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_D = 1$</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td><strong>0.065</strong></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0.070</td>
</tr>
<tr>
<td>$d_D = 2$</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0.075</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>0.090</td>
</tr>
</tbody>
</table>

**TABLE II.** Diminishing return
Experiments – Social Sentiment Modeling

• Influenced by
  – Damage 2 days ago
  – Sentiment 1 day ago
• Auto-regressive
  – Function of its pasts
• Non-linear model
  – Panic state
Models

• Gas Shortage Modeling
  – Regression
  – ARMA model

\[
D(t) = \sum_{i=d_D}^{d_D+n_D-1} a_i \cdot D(t-i) + \sum_{j=d_S}^{d_S+n_S-1} b_j \cdot S(t-j)
\]

where:

- \(d_D\): Delay time on Gas shortage (D) time series
- \(n_D\): No. of days looking back on D starting from \(d_D\)
- \(d_S\): Delay time on Social Sentiment (S) time series
- \(n_S\): No. of days looking back on S starting from \(d_S\)
Models

• Gas Shortage Modeling
  – Regression
  – ARMA model

\[
D(t) = \sum_{i=d_D}^{d_D+n_D-1} a_i \cdot D(t-i) + \sum_{j=d_S}^{d_S+n_S-1} b_j \cdot S(t-j)
\]

example:

\[
D(t) = 0.368 \cdot D(t-1) + 0.487 \cdot S(t-1) - 0.172 \cdot S(t-2) - 0.148 \cdot S(t-3)
\]
Experiments – Gas Shortage Modeling

<table>
<thead>
<tr>
<th>Params</th>
<th>$n_D$</th>
<th>$d_D$</th>
<th>$n_S$</th>
<th>$d_S$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.067</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0.074</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.076</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.085</td>
</tr>
</tbody>
</table>

**TABLE III.** Comparison of models
Experiments – Gas Shortage Modeling

• Influenced by
  – Damage 1 day ago
  – Sentiment 1 day ago

• Effect of Damage delay
  – Important for physical damage reconstruction
Experiments – Gas Shortage Modeling

• Influenced by
  – Damage 1 day ago
  – Sentiment 1 day ago

• Effect of Damage delay
  – Important for physical damage reconstruction
Experiments – Gas Shortage Regional Modeling

• Motivation
  – Can model be applied to specific regions?
  – Any relationship between damage in different regions?

• Regions
  – New York, New Jersey

• Regional Damage Models
  – Function of regional damage, regional sentiment
  – Function of regional damage, global sentiment
  – Function of global damage, regional sentiment
  – Function of global damage, global sentiment
Experiments – Gas Shortage Regional Modeling

• Model fitness
  – Relates more to actual Damage
  – New Jersey: Relates more to regional Damage
  – New York: Relates more to global Damage

<table>
<thead>
<tr>
<th>Region</th>
<th>Model</th>
<th>Params=2</th>
<th>Params=3</th>
<th>Params=4</th>
<th>Params=5</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Jersey</td>
<td>Model a)</td>
<td>0.0349</td>
<td>0.0307</td>
<td>0.0301</td>
<td>0.0278</td>
</tr>
<tr>
<td></td>
<td>Model b)</td>
<td>0.0351</td>
<td>0.0309</td>
<td>0.0303</td>
<td>0.0277</td>
</tr>
<tr>
<td></td>
<td>Model c)</td>
<td>0.0345</td>
<td>0.0313</td>
<td>0.0306</td>
<td>0.0279</td>
</tr>
<tr>
<td></td>
<td>Model d)</td>
<td>0.0344</td>
<td>0.0312</td>
<td>0.0306</td>
<td>0.0279</td>
</tr>
<tr>
<td>New York</td>
<td>Model a)</td>
<td>0.0323</td>
<td>0.0296</td>
<td>0.0284</td>
<td>0.0271</td>
</tr>
<tr>
<td></td>
<td>Model b)</td>
<td>0.0326</td>
<td>0.0297</td>
<td>0.0285</td>
<td>0.0273</td>
</tr>
<tr>
<td></td>
<td>Model c)</td>
<td>0.0275</td>
<td>0.0269</td>
<td>0.0267</td>
<td>0.0223</td>
</tr>
<tr>
<td></td>
<td>Model d)</td>
<td>0.0280</td>
<td>0.0272</td>
<td>0.0270</td>
<td>0.0220</td>
</tr>
</tbody>
</table>

TABLE IV. REGIONAL MODELS COMPARISON
Experiments – Gas Shortage Regional Modeling

• Hypothesis
  – People in New York (NY) respond more to New Jersey (NJ) damage

• Test Models
  – Damage NY as function of
    • NJ Damage, regional Sentiment
    • NJ Damage, global Sentiment
  – Damage NJ as function of
    • NY Damage, regional Sentiment
    • NY Damage, global Sentiment
Experiments – Gas Shortage Regional Modeling

• New York

• New Jersey

– Respond strongly to damage of New Jersey
  (New Jersey was more heavily affected)
Experiments – Cross Validation

• Model generalizability
  – Learn model in one region and apply to other region
  – Reasonable result in regions suffered from same event
Conclusions

• **Social sentiment model**
  – Follows non-linear model
  – Sentiment follows damage initially, but gradually follows its past values

• **Damage model**
  – Best approximated using balanced past damage and social sentiment values

• **Regional Damage model**
  – Highly correlates with regional damage than sentiment
  – Responds strongly to the more severely affected region
Question & Answer