Crowd-sensing with Polarized Sources

Tanvir Amin, Tarek Abdelzaher, Dong Wang
University of Illinois at Urbana Champaign
{maamin2, zaher, dwang24}@illinois.edu

Bolek Szymanski
Rensselaer Polytechnic Institute
szymab@rpi.edu

Presented by
Tanvir Amin

IEEE DCOSS 2014
Humans as Sensors

- Egypt Unrest
- Fukushima Disaster
- Sandy Gas Outage
- Crimea Annexation
- Syria Chemical Attack

Twitter
Facebook
Google+
Instagram
Flickr

Credibility Estimation
Anomaly Detection
Timeline Reconstruction
...
Binary Sensor Model

• Assume that each observation is either **True** or **False**
  • **True** means independently observable events.

• Some observations are neither True nor False, representing Non-Factual claims.
  • May be slogans or emotions or opinions.
Source Claim Network

Sources

Claims

Attribute: Credibility

Attribute: True / False
State of the Art

- Independent Sources, Independent Claims  IPSN 2012
  - On Truth Discovery in Social Sensing
- Confidence Bounds  SECON 2012
  - On scalability and robustness limitations of real and asymptotic confidence bounds in social sensing
- Admission Control  INSS 2012
  - On Diversifying Source Selection in Social Sensing
- Conflicting Claims  RTSS 2013
  - Exploitation of Physical Constraints for Reliable Social Sensing
- Non-independent Sources  IPSN 2014
  - Using Humans as Sensors: An Estimation Theoretic Perspective
- Polarized Sources  DCOSS 2014
  - This paper
Case Study: Egypt 2013

- Event: 2013 Uprising regarding former Egyptian President Mohamed Morsi
- Crawler starting July 2013, and continued for more than four months.
- 17 GB of tweets collected. 600K were “English” containing the word “Morsi”, which belonged to 173K cascades of different claims / observations.
- The largest 1000 cascades were manually annotated as being Pro-Morsi or Anti-Morsi or Neither
  - Accounted for 44K sources and 95K tweets
Pro Claims

Pro Claims

#EGYPT
#Morsi supporters denied right amid reports of arrests and beating
#Military_Coup
facebook.com/photo.php?fbid…

2:44 AM - 19 Jul 2013

Egyptian police fire tear gas to disperse supporters of ousted President Mohammed Morsi, reports say bbc.in/17mafk3

8:10 AM - 13 Aug 2013
Anti Claims

#RT " @DrBassemYoussef Helicopter footage of Anti Morsi protests, very impressive youtube.com/watch?v=Vux_v... ⚪ ⚪ ... "

5:48 AM - 2 Jul 2013

Amnesty International | Egypt: Evidence points to torture carried out by Morsi supporters amnesty.org/en/for-media/p...

2:09 PM - 18 Aug 2013
Neither
• Some users are highly polarized, and mostly forwards tweets favoring camp they belong to (Pro or Anti)
• Some users are neutral
Effect of Polarization

• When sources are biased towards a topic, their observation errors on that topic are more correlated.
  • When they do not share a bias, errors are independent.
  • Corroboration among correlated sources carry less statistical weight than when they are independent.

• Polarity unaware algorithm improperly computes the correlation between sources.
Polarity-aware Fact Finder

- Computes the latent networks from pro, anti, and neutral claims (SIGMETRICS 2012)
- Uses each network to estimate correlated errors in a manner that depends on content type.
- Accounts for correlation in credibility analysis.
Polarity-aware Fact Finder

Social Propagation Network

Original Problem

Pro Network
Attribute: Pro Cred
Attribute: True / False

Anti Network
Attribute: Anti Cred
Attribute: True / False

Neutral Network
Attribute: Neu Cred
Attribute: True / False

C1
C2
C3
C4
C5
## Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Combined</th>
<th>Polarized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro Claims</td>
<td>199</td>
<td>147</td>
<td>128</td>
</tr>
<tr>
<td>Anti Claims</td>
<td>109</td>
<td>88</td>
<td>76</td>
</tr>
<tr>
<td>Neutral Claims</td>
<td>692</td>
<td>543</td>
<td>496</td>
</tr>
</tbody>
</table>
### Evaluation

<table>
<thead>
<tr>
<th>Definition</th>
<th>Set A</th>
<th>Set B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarized Exclusive</td>
<td>38</td>
<td>116</td>
</tr>
<tr>
<td>Combined Exclusive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Factual</td>
<td>26</td>
<td>82</td>
</tr>
<tr>
<td>Non-factual (0)</td>
<td>12</td>
<td>34</td>
</tr>
<tr>
<td>True (1)</td>
<td>25</td>
<td>72</td>
</tr>
<tr>
<td>False (-1)</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>False Claims</td>
<td><strong>2.6 %</strong></td>
<td><strong>8.6 %</strong></td>
</tr>
<tr>
<td>Factual True</td>
<td><strong>96 %</strong></td>
<td><strong>88 %</strong></td>
</tr>
</tbody>
</table>

- **662** output claims were common to both algorithms.

- **2.6 %** of False Claims in Set A vs **8.6 %** in Set B.
Conclusion

• Separation of claims by polarity prevents estimation of false dependencies between neutral sources.
• Probability of error reduced by factor of three for the factual claims.
• More than 18% improvement in overall Quality of Information.
• Easily extensible to incorporate ML or NLP analysis which may improve the fact-finding performance.
• Idea of polarities can be extended to “topics” with arbitrary relations and hierarchy.
• Did not consider adversarial sources.
Questions
Why Largest Cascades?
Polarity-aware Fact Finder

- Computes the latent networks from pro, anti, and neutral claims (SIGMETRICS 2012)
- Uses each network to estimate correlated errors in a manner that depends on content type.
- Accounts for correlation in credibility analysis.
Effect of Polarization

The non polarized algorithm confuses the dependency of the neutral sources
Effect of Polarization

Dependency of the strongly polarized sources are correctly determined by both polarized and non-polarized algorithms.
### Evaluation

<table>
<thead>
<tr>
<th></th>
<th>Input</th>
<th>Combined</th>
<th>Polarized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro Claims</td>
<td>199</td>
<td>147</td>
<td>128</td>
</tr>
<tr>
<td>Anti Claims</td>
<td>109</td>
<td>88</td>
<td>76</td>
</tr>
<tr>
<td>Neutral Claims</td>
<td>692</td>
<td>543</td>
<td>496</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>778</td>
<td>700</td>
</tr>
</tbody>
</table>

- 662 output claims were common to both algorithms
Epidemic Propagation

• Nodes get infected, infect other nodes, and the process continues resulting in a propagation graph.

• Inverse Problem*: Given the observation of infections, find the structure.

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012
Standard Independent Cascade Model*

- Proceed in discrete step
- Initial seeds random with probability $p_{\text{init}}$
- $i$ infect $j$ with $p_{ij}$
- Active node inactive after one step

* Proposed by Goldenberg, Libai, and Muller; "Talk of the Network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters 2001. Also appears in Kempe, Kleinberg, and Tardos "Maximizing the spread of influence through a social network, KDD 03

Illustrative example taken from author's presentation
Standard Independent Cascade Model

- Proceed in discrete step
- Initial seeds random with probability $p_{\text{init}}$
- $i$ infects $j$ with $p_{ij}$
- Active node inactive after one step

---

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012
Standard Independent Cascade Model

- Proceed in discrete step
- Initial seeds random with probability $p_{\text{init}}$
- $i$ infect $j$ with $p_{ij}$
- Active node inactive after one step

*Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012*
Standard Independent Cascade Model

- Proceed in discrete step
- Initial seeds random with probability $p_{init}$
- $i$ infect $j$ with $p_{ij}$
- Active node inactive after one step

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012
Standard Independent Cascade Model

- Proceed in discrete step
- Initial seeds random with probability $p_{\text{init}}$
- $i$ infect $j$ with $p_{ij}$
- Active node inactive after one step

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012
Structure Learning Problem

Given node activation times

Find network structure

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012
The example cascade had no evidence of the black edges. Hence, one cascade is not sufficient to learn the structure.

Q: How many cascades necessary?

Q: Given the cascades, how to find the structure?

* Praneeth Netrapalli, Sujay Sanghavi, Learning the Graph of Epidemic Cascades. SIGMETRICS 2012