Truth Finding Group

Yanglei Song, Shi Zhi, Ahmend El-Kishky

October, 7, 2013
Data set

- News data set
  - Extract entity attributes from news by entity name

- 18 systems, 10 news agencies, 100 queries
- Query: 50 Person (25 attributes), 50 Organization (16 attributes)
Progress

- Generate data in link form (query-system-agency-answer)
- Hardness of attributes—performance of easy attributes have some indications of hard ones
- Decomposition of quality of systems and news agencies
- Latent topic behind each entity
- Incorporate constraints on claims in model
Next Step

- Lack of ground truth
  - Manually label some, but not enough

- Careful data cleaning and constraint definition
  - Different kinds of attributes (names, places), define hard/soft constraints on claims

<table>
<thead>
<tr>
<th>Format constraint</th>
<th>numerical, date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location constraint</td>
<td>Hierarchical locations (country, state, city)</td>
</tr>
<tr>
<td>Semantic-similarity</td>
<td>Simple similarity (attack, heart attack)</td>
</tr>
<tr>
<td>preserving representation</td>
<td>Word2vec, etc. (other?) -- Clustering</td>
</tr>
<tr>
<td>Constraints on names</td>
<td>Names of person, names of org, religion</td>
</tr>
</tbody>
</table>

- Single value/List value attributes (TruthFinder vs. LTM) to balance precision/recall

- Decompose quality of systems and news agencies