Solving Hard Coreference Problems
Cognitive Computation Group, University of Illinois

Problem Description
- **Standard Coreference Instnaces**
  - [John] adores [Mary] because [she] is pretty. [Prof. Roth] is satisfied with his [students] because [they] work hard.
- **Problems with Existing Coref Systems**
  - Rely heavily on gender/plurality information.
- **Hard Coreference Instnaces**
  - [A bird] perched on the [limb] and [it] sang.
  - [Jim] was afraid of [Robert] because [he] gets scared around [pollen].
- **Goal**
  - A better overall coreference system
  - Improve on solving hard coreference problems
  - Maintain the state-of-art performance on standard coreference problems

Predicate Schemas
- **Type 1 Schema**
  - \( \text{pred}_{1}(m, a) \) : Sub a : Obj
  - [The bee] landed on [the flower] because [it] had pollen. \( S(\text{have}(\text{the flower}), \text{be}(\text{pollen})) \rightarrow S(\text{have}(\text{the bee}), \text{be}(\text{pollen})) \)
- **Type 2 Schema**
  - \( \text{pred}_{2}(m, a) \mid \text{pred}_{2}(m, a), cn \)
  - [Jim] was afraid of [Robert] because [he] gets scared around new people. \( S(\text{be afraid of}(\text{Robert}), \text{get scared around}(\text{new people})) \rightarrow S(\text{be afraid of}(\text{Jim}), \text{get scared around}(\text{new people})) \)

Schema Variations
- **Type 1**
  - \( S(\text{pred}_{1}(m, a)) \)
  - \( S(\text{pred}_{1}(m, a), m) \)
  - \( S(\text{pred}_{1}(m, a), m, a) \)
  - \( S(\text{pred}_{1}(m, a), m, a, c) \)
- **Type 2**
  - \( S(\text{pred}_{2}(m, a)) \)
  - \( S(\text{pred}_{2}(m, a), m) \)
  - \( S(\text{pred}_{2}(m, a), m, a) \)
  - \( S(\text{pred}_{2}(m, a), m, a, c) \)

Example Beyond Above Schemas
- [Lakshman] asked [Vivan] to get him some ice cream because [he] was hot.

Utilizing Knowledge
- **Knowledge as Features**
  - \( w_{u,v} = w_{u} \phi(u, v) + w_{v} s(u, v) \)

Knowledge Acquisition
- **Gigaword co-occurrences**
  - Extract triplets from Gigaword
  - \( S_{\text{giga}} = \{S_{\text{giga}}^{1}(u, v), S_{\text{giga}}^{2}(u, v) \}
  - \( S_{\text{giga}}^{1}(u, v) = S(\text{pred}_{1}(m, a, a = a_{c})) \)
  - \( S_{\text{giga}}^{2}(u, v) = S(\text{pred}_{2}(m, a, a = a_{c}), \text{pred}_{2}(m, a, a = a_{c})), cn \)

- **Wikipedia Disambiguated Co-occurrences**
  - Extract disambiguated noun, verbs and entities, etc., in Wikipedia (Illinois-Wikifier)
  - Collect co-occurrence statistics: \( 1 \) immediately after \( 2 \) immediately before \( 3 \) before \( 4 \) after

Web Search Statistics
- Generate queries to get the score
  - \( 1) “u a_{c}” 2) “u pred_{c}” 3) “u pred_{c} a_{c}” 4) “a_{c} u” \)

- **Polarity Information**
  - Initialize \( P_{\text{pred}_{1}}(m, a) \) and \( P_{\text{pred}_{2}}(m, a) \) (Wilson et al. 2005)
  - Negate polarity when mention role is object
  - If there is a polarity reversing connective (such as “but”) preceding the predicate, reverse the polarity.

Learning and Inference Framework
- **Learning**
  - **Mention Detection**
  - **Learning**
  - **Inference**
  - **Coreference Resolver**

Results
- **Datasets:**
  - Winograd (Rahman & Ng, 2012)
  - Winocoref: Winograd with more mentions
  - Coref: ACE-2004, Ontonotes
- **Metrics:**
  - Accuracy: for “Winograd” (binary classification)
  - MUC, BCUB, CEAF (general coref)
- **AntePre:**
  - \( k \) pronouns in sentence
  - Each pronoun has \( n_{k} \) antecedents
  - \( m \) correct binary decisions

Evaluations:
- **Hard Coreference Problems**

<table>
<thead>
<tr>
<th>System</th>
<th>MUC</th>
<th>BCUB</th>
<th>CEAF</th>
<th>AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>79.42</td>
<td>78.97</td>
<td>78.74</td>
<td>78.97</td>
</tr>
<tr>
<td>IIIlCons</td>
<td>78.17</td>
<td>81.64</td>
<td>78.45</td>
<td>79.42</td>
</tr>
<tr>
<td>KnowComb</td>
<td>77.51</td>
<td>81.97</td>
<td>77.44</td>
<td>78.97</td>
</tr>
<tr>
<td>OntoNotes</td>
<td>84.10</td>
<td>78.30</td>
<td>68.74</td>
<td>77.05</td>
</tr>
<tr>
<td>KnowComb</td>
<td>84.33</td>
<td>78.02</td>
<td>67.95</td>
<td>76.76</td>
</tr>
</tbody>
</table>

- **Ablation Study**
- We categorized instances in Winogrd data

<table>
<thead>
<tr>
<th>Category</th>
<th>Cat1</th>
<th>Cat2</th>
<th>Cat3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>317</td>
<td>1060</td>
<td>509</td>
</tr>
<tr>
<td>Portion</td>
<td>16.8%</td>
<td>56.2%</td>
<td>27.0%</td>
</tr>
</tbody>
</table>

Evaluated on each category:

<table>
<thead>
<tr>
<th>Schema</th>
<th>AntePre(Tst)</th>
<th>AntePre(Train)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1</td>
<td>76.67</td>
<td>86.79</td>
</tr>
<tr>
<td>Type 2</td>
<td>79.55</td>
<td>88.86</td>
</tr>
<tr>
<td>Type 1 (Cat1)</td>
<td>90.26</td>
<td>93.04</td>
</tr>
<tr>
<td>Type 2 (Cat2)</td>
<td>83.38</td>
<td>92.49</td>
</tr>
</tbody>
</table>

Haoruo Peng, Daniel Khashabi and Dan Roth