CHALLENGES IN SLOT FILLING

Glen Pink
James Curran
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

SLOT FILLING

Slot Filling (SF) is the task of extracting values (fills) for specific named attributes (slots) for a given entity from unstructured text (see "Examples per:employee_or_member_of").

EXAMPLES per:employee_or_member_of

1) Glen Pink, University of Sydney
   Glen Pink is an employee of the University of Sydney. A straightforward case.

2) John Negroponte, McGraw-Hill
   Negroponte was a career diplomat from 1960 to 1997 before leaving public service to become a senior executive with McGraw-Hill, a financial services and publishing conglomerate. Longer distance, and need to know that ‘senior executive’ entails ‘employee of’.

3) John Negroponte, U.S State Department
   The Director of the State Council’s Taiwan Affairs Office, Chen Yunlin, discussed the Taiwan situation in a meeting with U.S. Deputy Secretary of State John Negroponte on Thursday. Having a specific title indicates being part of a specific organisation.

4) Sadia, Adriano Ferreira
   Sadia said it fired chief financial officer Adriano Ferreira. Need to know that being fired means that they were once an employee (event causality).

5) John Kerry, U.S. Senate
   Kerry said on the Senate floor that it was not the time for him to launch a presidential campaign, but he would continue to try to end the Iraq war. Taking a specific action indicates being part of an organisation.

6) Mohamed ElBaradei, New York University School of Law
   After a stint in the Egyptian Ministry of Foreign Affairs, he received a doctorate in International Law at the New York University School of Law in 1974, and later became an adjunct professor there. Multiple coreference decisions and events (where did he become an adjunct professor and what does that mean?).

SLOTS

Sample slots (from TAC Knowledge Base Population [Ji et al., 2011] tasks) include:
   per:employee_of
   per:schools_attended
   per:charges
   org:subsidiaries
   org:shareholders
   org:political/religious_affiliation

CHALLENGES IN SLOT FILLING

• SF relies on many Natural Language Processing (NLP) techniques, errors in any one of these affect results.
• Training data is very limited, and there are many relations between entities (some systems identify more than a million relations, each potentially requiring training data).
• Complex inference from document information, cross-slot relations, event causality, and background knowledge is required to identify many fills.
• Recognition of candidate fills for slots can be difficult (e.g. per:cause_of_death) [Ji and Grishman, 2011].

EVALUATION

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>All entities</td>
<td>0.1</td>
<td>88.8</td>
<td>0.2</td>
</tr>
<tr>
<td>+ path constraints</td>
<td>1.8</td>
<td>60.1</td>
<td>3.5</td>
</tr>
<tr>
<td>Phrase clustering</td>
<td>2.4</td>
<td>6.2</td>
<td>3.4</td>
</tr>
<tr>
<td>Bootstrapping</td>
<td>10.2</td>
<td>19.3</td>
<td>13.3</td>
</tr>
<tr>
<td>Weak supervision*</td>
<td>37.4</td>
<td>27.2</td>
<td>31.5</td>
</tr>
</tbody>
</table>

Table 1: TAC11 evaluation (* = state-of-the-art)

CONCLUSION

Results are still low for all approaches for SF, however there are promising aspects of all approaches. The next task is to add aspects of distant supervised approaches to a semi-supervised label propagation approach, to leverage the current feature space used for distant supervision.

REFERENCES