Joint Posterior Revision of NLP Annotations via Ontological Knowledge

Marco Rospocher (rospocher@fbk.eu), Francesco Corcoglioniti (corcoglio@fbk.eu)

1. Problem: Incoherent Mention-level NLP Annotations

Eric Clapton is one of the greatest guitar players.

Mr. Washington was runner-up at Wimbledon in 1996.

The GW Bridge is a suspension bridge over the Hudson.

2. Solution: Coherence via Ontology

ontological background knowledge

\[ \text{Estimating NERC} \]
\[ P(C | \delta_{\text{NERC}}; K) = \alpha \cdot P(C | K) + (1 - \alpha) \cdot P(C | \delta_{\text{NERC}}; G) \]

\[ \text{Estimating EL} \]
\[ P(C | \delta_{\text{EL}}; K) = 1 \cdot \{C_K(a_0)\} \{C\} \]

3. General Probabilistic Model

Variables

- \( m \) entity mention
- \( a \) \((a_1, \ldots, a_k)\) NLP annotations
- \( B \) NLP Background Knowledge
- \( K \) “The” Ontological Knowledge
- \( C \) Entity’s ontological class set (from K)

Conditional Independence Assumptions

1. \( P(a|m, B, K, C) = \prod_i P(a_i|m, B, K, C) \)
2. \( P(a_i|m, B, K) = P(a_i|m, B) \)
3. \( P(C | a_i, m, B, K) = P(C | a_i, K) \)

\[ P(a|m, B, K) = \sum_C P(a, C|m, B, K) \]

\[ P(a, C|m, B, K) = \frac{P(C|m, B, K) \cdot P(a|m, B, K, C)}{P(C|m, B, K) \cdot \prod_i P(a_i|m, B, K, C)} \]

\[ P(C|m, B, K) = \left( \prod_i P(a_i|m, B, K) \right)^\frac{1}{n} \]

\[ P(C | a_i, m, B, K) = \frac{P(a_i|m, B, K) \cdot P(C | a_i, m, B, K)}{P(a_i|m, B)} \cdot P(C | a_i, K) \]

4. Model Instantiated on NERC + EL

Ontological Background Knowledge

- 6,016,695 entities
- Taxonomy of 568,255 classes

Estimating NERC

\[ \frac{n_K(C)}{\sum_{C'} n_K(C')} \text{ Prior (popularity based on entity ingoing links)} \]

\[ \frac{n_G(C, a_{\text{NERC}})}{\sum_{C'} n_G(C', a_{\text{NERC}})} \text{ Computed from gold corpus (NERC + class annotations)} \]

Consider only class sets restricted to popular classes (seen at least \( n \) times in the gold corpus)

Estimating EL

Deterministically computable from the classes of the entity (possibly leveraging alignments between EL Knowledge Base and YAGO)

5. Evaluation

Tools: Stanford CoreNLP, DBpedia Spotlight

Gold Corpus (NERC): AIDA CoNLL-YAGO (train)

Datasets:

1. AIDA CoNLL-YAGO (test-b)
2. MEANTIME
3. TAC-KBP

Research question: Does the JPARK posterior joint revision of the annotations from Stanford CoreNLP (NERC) and DBpedia Spotlight (EL), via YAGO, improve their performances?

Measures: NERC / EL / NERC+EL

<table>
<thead>
<tr>
<th>Tools</th>
<th>NERC P</th>
<th>R</th>
<th>F1</th>
<th>EL P</th>
<th>R</th>
<th>F1</th>
<th>NERC+EL P</th>
<th>R</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDA standard</td>
<td>94.30%</td>
<td>90.80%</td>
<td>90.80%</td>
<td>87.50%</td>
<td>84.50%</td>
<td>84.50%</td>
<td>63.00%</td>
<td>62.50%</td>
<td>63.00%</td>
</tr>
<tr>
<td>with JPARK</td>
<td>95.00%</td>
<td>88.10%</td>
<td>91.40%</td>
<td>78.10%</td>
<td>65.40%</td>
<td>66.20%</td>
<td>65.50%</td>
<td>63.70%</td>
<td>63.40%</td>
</tr>
<tr>
<td>MEANTIME</td>
<td>90.40%</td>
<td>72.00%</td>
<td>80.50%</td>
<td>78.50%</td>
<td>55.70%</td>
<td>62.20%</td>
<td>67.00%</td>
<td>51.10%</td>
<td>59.20%</td>
</tr>
<tr>
<td>TAC-KBP</td>
<td>3.20%</td>
<td>2.00%</td>
<td>2.00%</td>
<td>0.20%</td>
<td>0.10%</td>
<td>0.10%</td>
<td>3.20%</td>
<td>2.80%</td>
<td>3.10%</td>
</tr>
</tbody>
</table>

Restricted to gold mentions; similar improvements also considering all mentions, and macro-averaging by document or by NERC type.