Person Name Linking

Named Entity Disambiguation and Linking to a knowledge base

Goal: uniquely assign a real-world entity to a person mention in text document

Problem: person mentions can be ambiguous

Social Network Named Entity Linking:
- only uses neighboring person mentions, information from Wikidata and the Wikipedia Social Network
- language independent, reliable, simple

Disambiguation Model

Disambiguation by comparison of entity candidates to uniquely identified persons

1. Person Name Extraction
   Stanford NER

2. Entity List Look-Up
   matching the mention to Wikidata labels and alternative names:
   - (a) no match (grey)
   - (b) seed sp: one match (green)
   - (c) ambiguous mention m: multiple matches (blue), each match for m is called an entity candidate c ∈ C_m

3. WSN based Disambiguation
   - selecting the candidate c with the strongest relation to the neighborhood N consisting of seed persons sp.
   - weight function ω: (c, N) → R for the neighborhood relation between a candidate c ∈ C_m for a person mention m and a neighborhood N of seed persons sp in the document d:
     ω(c, N) = \sum_{sp \in N} r(c, sp)

Experiments

Dataset: 1.5 M person wikilinks
- base 1: lowest Wikidata id
- base 2: random selection

<table>
<thead>
<tr>
<th>SocNNEL</th>
<th>base 1</th>
<th>base 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>A</td>
<td>P</td>
</tr>
<tr>
<td>all mentions</td>
<td>96.4</td>
<td>94.1</td>
</tr>
<tr>
<td>seed persons</td>
<td>98.7</td>
<td></td>
</tr>
<tr>
<td>ambiguous</td>
<td>89.8</td>
<td>82.4</td>
</tr>
</tbody>
</table>

Table 1: Accuracy (A) and Precision (P) in %

Number of candidates
- number of candidates influences the precision
- maximum number of candidates should be 10 - 40

Conclusion
- disambiguation model that employs a mention’s neighborhood and the Wikipedia Social Network.
- measure for the relationship between a person and its neighborhood
- comparable to state-of-the-art approaches, simple, reliable and easily adoptable to other document types and languages
- comparable to state-of-the-art approaches

Ongoing Work
- different named entity types (e.g., locations, organizations)
- refining the model to cover more mentions and become independent of seeds persons
- limiting the number of candidates per mention

References

Evaluation

Dataset: AIDA CoNLL-YAGO testset b (984 person mentions)

<table>
<thead>
<tr>
<th>SocNNEL</th>
<th>base 1</th>
<th>base 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>A</td>
<td>P</td>
</tr>
<tr>
<td>all m</td>
<td>84.6</td>
<td>94.1</td>
</tr>
<tr>
<td>truth possible</td>
<td>91.5</td>
<td></td>
</tr>
<tr>
<td>ambiguous m</td>
<td>86.9</td>
<td>87.6</td>
</tr>
<tr>
<td>true possible</td>
<td>86.8</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Accuracy (A) and Precision (P) in % on the AIDA CoNLL-YAGO testset-b.