EPE 2017:
The Sherlock Negation Resolution Downstream Application

Emanuele Lapponi♣ and Stephan Oepen ♠♠ and Lilja Øvrelid♣♠
♣ University of Oslo, Department of Informatics
♠ Center for Advanced Study at the Norwegian Academy of Science and Letters
{emanuel oe liljao}@ifi.uio.no

September 20, 2017
Negation resolution (NR)

- For a given sentence, find negation cues and the words they affect.
- 2012 *SEM Shared Task (Morante & Blanco, 2012) is one of the most notable NR-related effort in recent years:
  - A non-biomedical, human-annotated corpus for negation
  - Empirical NR results from 8 competing teams
  - Sherlock predecessor (Lapponi et al., 2012) ranked 1st in the open and 2nd in the closed track
A collection of fiction works by Sir Arthur Conan Doyle

- Training: 3644 sentences drawn from *The Hound of the Baskervilles*
- Development: 787 sentences taken from *Wisteria Lodge*
- Held-out: 1089 sentences from *The Cardboard Box* and *The Red Circle*

- Pre-processed with sentence boundaries, tokens, lemmas, pos-tags and constituency trees
1. Since \{we have been so\} \langle\text{un}\rangle\{\text{fortunate}\} \text{as to miss him} [\ldots]

2. If \{he was\} in the hospital and yet \langle\text{not}\rangle \{\text{on the staff}\} he could only have been a house-surgeon or a house-physician: little more than a senior student.

- **Cues** (angle brackets): (multiple) tokens or sub-tokens
- **Scopes** (braces): extend to full propositions, can be discontinuous
- **Events** (underlined): in-scope events or states, if factual
we have never gone out without keeping a sharp watch, and no one could have escaped our notice.

Assumes classified cues

NR as a classical sequence labeling problem, ‘flattening‘ scopes

Fine-grained label set

Wapiti (Lavergne, Cappé, & Yvon, 2010), an open-source implementation of a Conditional Random Field (CRF) classifier
System Description

- Features include different n-gram combinations of token, lemma, pos-tags and dependency relations.
- Cue-aware features include surface and dependency distance, as well shortest dependency path from a cue.
- Adapted to be robust to a wider range of dependency graphs:
  - full set of dependency relations
  - dependency path, distance: assumes graphs with re-entrancies and unconnected nodes, records only one of several shortest paths.
EPE submissions come in different tokenization flavors

Original CD annotations are token-oriented and CoNLL-like

We developed a separate ‘projection‘ step that

1. converts the gold-standard negation annotations into character spans
2. projects them onto a dependency graph provided by a participating parser
3. serializes the enriched graph in the token-oriented back to the *SEM 2012 format

I.e. a ‘personalized’ version of the negation annotations for each individual segmentation
## Results

<table>
<thead>
<tr>
<th></th>
<th>UiO₂</th>
<th>Elming et al.</th>
<th>Stanford–Paris 6</th>
<th>Szeged 0</th>
<th>Paris–Stanford 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST</td>
<td>85.75</td>
<td>—</td>
<td><strong>88.57</strong></td>
<td>86.64</td>
<td>88.19</td>
</tr>
<tr>
<td>SM</td>
<td>80.00</td>
<td><strong>81.27</strong></td>
<td>80.43</td>
<td>78.42</td>
<td>80.14</td>
</tr>
<tr>
<td>ET</td>
<td><strong>80.55</strong></td>
<td>76.19</td>
<td>76.55</td>
<td>75.47</td>
<td>71.77</td>
</tr>
<tr>
<td>FN</td>
<td>66.41</td>
<td><strong>67.94</strong></td>
<td>65.37</td>
<td>62.15</td>
<td>60.48</td>
</tr>
</tbody>
</table>

- Token-level $F_1$ for in-scope tokens (ST) and event tokens (ET)
- Scope-level $F_1$ with (FN) and without (SM) events
Keep in mind that
  - Sherlock was designed around a specific set of linguistic annotations
  - Very possible bias

For each submission, we really should
  - Experiment with adding/discarding features for the CRF
  - Design new features!
Conclusions

- Sherlock, Negation Resolution for Extrinsic Parser Evaluation
  - *SEM 2012 annotation projected to arbitrary segmentations
  - System updated to be robust to more dependency representations
Future work

- Sift through the tea leaves: systematic qualitative and quantitative error analysis of EPE submissions
- Tune (and design) features and heuristics around a sub-set of the EPE submissions
Thank you!

https://github.com/ltgoslo/sherlock
