Kazuaki Hanawa, Akira Sasaki, Naoaki Okazaki*, Kentaro Inui
(Tohoku University, *Tokyo Institute of Technology)

Abstract

Goal
- Annotate causal relation instances in Wikipedia

Approach
- Integrate a crowdsourcing service and brat

Contributions
- Collected 95,008 causal relation instances in 1,494 Wikipedia articles (http://www.cl.ecei.tohoku.ac.jp/wikipedia_pro_sup/)
- The corpus can be used as supervision data for automatic recognition of causal relation instances
- Revealed valuable facts for improving the annotation process of this task

Annotation policy

- $X$ promotes $Y$
- $Y$ is activated when $X$ is activated
- $X$ suppresses $Y$
- $Y$ is inactivated when $X$ is activated

Wikipedia article “Nyctalopia”

Nyctalopia, also called night-blindness, is a condition making it difficult to see in relatively low light. Nyctalopia may exist from birth, or be caused by injury or severe malnutrition.

(PRO, nyctalopia, night-blindness)
(SUP, nyctalopia, see in relatively low light)
(PRO_BY, nyctalopia, injury) = (PRO, injury, nyctalopia)
(PRO_BY, nyctalopia, severe malnutrition)

Using brat in crowdsourcing

Complete the task
Enter the password 0.3 or more
Correct password F9pw4JkD0lk3
Incorrect password Yd2UwmHr51p

Annotation interface of brat
One out of ten is a test question

Desertification is a type of land degradation where the land becomes increasingly arid, well as vegetation and wildlife.

Collective ten annotations per an article

Example

... and result in high numbers of abnormal white blood cells.
Symptoms may include bleeding and bruising problems,...

Micro-F1 between gold standard

- $m$: Number of annotators
- $n$: Adopt only spans with $n$ or more exactly matched annotations

Percentage of POS of head words

<table>
<thead>
<tr>
<th>POS</th>
<th>Noun</th>
<th>90.17</th>
<th>Mark</th>
<th>2.27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
<td>5.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auxiliary verb</td>
<td>1.09</td>
<td></td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>Adjective</td>
<td>0.41</td>
<td></td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

- It may be sufficient to limit annotation spans to noun phrases

Numbers of words and bunsetsu chunks

<table>
<thead>
<tr>
<th>bunsetsu chunks</th>
<th>word tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO</td>
<td>1</td>
</tr>
<tr>
<td>SUP</td>
<td>2</td>
</tr>
<tr>
<td>SUP_BY</td>
<td>3</td>
</tr>
</tbody>
</table>

- Allowing crowd workers to choose their segment boundaries may be necessary

Automatic recognition

- Use $n = 2$ data as training and test data
- IOB2 notation was applied to the causal relations (e.g., B-PRO, I-PRO, B-SUP, I-SUP)
- Use one-layer bi-directional LSTM

<table>
<thead>
<tr>
<th>Label</th>
<th>precision</th>
<th>recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO</td>
<td>0.507</td>
<td>0.364</td>
<td>0.424</td>
</tr>
<tr>
<td>SUP</td>
<td>0.354</td>
<td>0.275</td>
<td>0.310</td>
</tr>
<tr>
<td>PRO_BY</td>
<td>0.470</td>
<td>0.344</td>
<td>0.397</td>
</tr>
<tr>
<td>SUP_BY</td>
<td>0.259</td>
<td>0.178</td>
<td>0.211</td>
</tr>
</tbody>
</table>