1. Our methods

We use a combination of named entity recognition and named entity disambiguation methods to solve the Entity Linking task.

![Diagram showing named entity recognition and named entity disambiguation](image)

Named entity recognition

We extract mentions of "law names" from local assembly member's utterances.

Named entity disambiguation

We link the extracted mentions to Wikipedia title with knowledge bases i.e. Wikipedia and e-Gov.

1.1. Named entity recognition methods

We extract mentions of "law name" with BERT, and filter the extracted mentions using filter 1 and filter 2.

BERT

- We use BERT model, which is available at DeepPavlov. The model is a multilingual named entity recognition model, which was pretrained from the multilingual BERT using OntoNotes.

- We further fine-tuned the model on the training data of QA Lab-PoliInfo-2 Entity Linking task.

Filter 1

- If the sentence input into BERT does not contain the word "law", it is filtered with filter 1 and all outputs are set to "O".

Filter 2

- We extract the mentions that match following regular expressions. If the mention does not match the following phrases, the output is "O".

1.2. Named entity disambiguation methods

We disambiguate the extracted mentions and link them to Wikipedia using exact match, Wikipedia2Vec, mention-entity prior, and e-Gov.

Example of mention

- The combination methods of both the filter 1 and the filter 2 outperformed the results using only filter 1. This is probably because the wrong mention, like phrases which do not contain "law", was extracted during the mention extraction process. These results showed filter 2 is also useful to remove noise.

- Disambiguation using e-Gov alone produced lower results than using mention-entity prior. However, when e-Gov was combined with other disambiguation methods, their scores increased.

- Specifically, the combination of e-Gov and mention-entity prior showed the best results—a score of 0.6035.

- Using dictionaries such as e-Gov to process mentions that could be reliably disambiguated, the results of the combination methods were better than those obtained by other methods when they were used alone.

Example of filter 1

Input: The sentence does not contain "law", so all output is "O".

Output:

Example of Wikipedia2Vec

Input mention: "カジノ企業"

Output:

Example of mention-entity prior

Input mention: "独占禁止法(法)"

Output:

2. Our results

<table>
<thead>
<tr>
<th>New methods</th>
<th>NER methods</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BERT + filter 1 + filter 2</td>
<td>mention-entity prior + e-Gov</td>
<td>0.8035</td>
</tr>
<tr>
<td>BERT + filter 1 + filter 2</td>
<td>mention-entity prior</td>
<td>0.5863</td>
</tr>
<tr>
<td>BERT + filter 1 + filter 2</td>
<td>e-Gov</td>
<td>0.5518</td>
</tr>
<tr>
<td>BERT + filter 1 + filter 2</td>
<td>Wikipedia2Vec</td>
<td>0.5130</td>
</tr>
<tr>
<td>BERT + filter 1 + filter 2</td>
<td>Wikipedia2Vec + e-Gov</td>
<td>0.5000</td>
</tr>
<tr>
<td>BERT + filter 1</td>
<td>mention-entity prior + e-Gov</td>
<td>0.4887</td>
</tr>
<tr>
<td>BERT + filter 1</td>
<td>mention-entity prior</td>
<td>0.4747</td>
</tr>
<tr>
<td>BERT + filter 1</td>
<td>e-Gov</td>
<td>0.4666</td>
</tr>
<tr>
<td>BERT + filter 1</td>
<td>Wikipedia2Vec</td>
<td>0.3980</td>
</tr>
<tr>
<td>BERT + filter 1</td>
<td>mention-entity prior + Wikipedia2Vec</td>
<td>0.3980</td>
</tr>
<tr>
<td>BERT + filter 1</td>
<td>exact match</td>
<td>0.3247</td>
</tr>
</tbody>
</table>

Scores in the formal run | Scores in the formal run (late submissions)

3. Discussion

- The combination methods of both the filter 1 and the filter 2 outperformed the results using only filter 1. This is probably because the wrong mention, like phrases which do not contain "law", was extracted during the mention extraction process. These results showed filter 2 is also useful to remove noise.

- Disambiguation using e-Gov alone produced lower results than using mention-entity prior. However, when e-Gov was combined with other disambiguation methods, their scores increased.

- Specifically, the combination of e-Gov and mention-entity prior showed the best results—a score of 0.6035.

- Using dictionaries such as e-Gov to process mentions that could be reliably disambiguated, the results of the combination methods were better than those obtained by other methods when they were used alone.

Example of filter 2

Input: 大手カジノ企業が、カジノ推進法の提案者である…

Output:

Example of mention-entity prior

Input mention: "独占禁止法(法)"

Output:

4. Conclusion

- We achieved the best score out of all the team.

- The combination of e-Gov and mention-entity prior showed the best results—a score of 0.6035.

- Using Filter 2 and using e-Gov are useful to improve the score in this task.