Semantic Role Labelling for Dutch Law Texts

Motivation

- Legal texts are difficult to interpret
- Formalising them will make an interpretation explicit
- Creating formalisations is very labour-intensive
- We propose a method for automated extraction of knowledge representations using two different NLP approaches

Related work

- Rule-based approaches with syntactic parsing [1]
- Flint frames as a framework for representing norms in acts and facts [2]
- State-of-the-art performance with BERT for semantic role labelling [3]

Architecture

Law texts → Preprocessing → Syntactic/Semantic Tagging → Frame building → Flint Frames

Method

- Law text cleaning and preprocessing
- Law text annotation
- Finetuning BERTje (Dutch BERT)
- Syntactic tagging + rules VS Semantic role labeling
- Evaluation: performance and comparison

Data

- 1854 unique sentences from law texts
- Annotated with four semantic roles: action, actor, object, recipient
- 409 sentences with actions, 324 actors, 337 objects, 85 recipients
- Inter-annotator agreement: Fleiss $\kappa = .785$
- "The processor [actor] collects [action] personal data [object] of the data subject [recipient] (GDPR art. 5, simplified)"

Results: Mean accuracy per method and test set

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Rule</th>
<th>Transformer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotated dataset</td>
<td>0.42</td>
<td>-</td>
</tr>
<tr>
<td>Test set</td>
<td>0.52</td>
<td>0.81</td>
</tr>
<tr>
<td>Aliens Act (positives)</td>
<td>0.42</td>
<td>0.80</td>
</tr>
<tr>
<td>Aliens Act (all)</td>
<td>0.74</td>
<td>0.69</td>
</tr>
</tbody>
</table>

Figure: Rule-based matrix and Transformer-based matrix on test set

Conclusion and Future Work

- The transformer-based method outperforms the rule-based method and is promising with an accuracy of 81%
- Future work: extraction of other elements such as recipients, improve the rules and combine both methods, test methodology on English law texts

References

