

Neural Coreference Resolution for Dutch Parliamentary Meetings

Ruben van Heusden, Maarten Marx and Jaap Kamps

University of Amsterdam

Abstract

The task of coreference resolution revolves around the linking of words and phrases describing real world entities to their correct antecedents in text. This task is an integral part of Natural Language Processing as it encapsulates a variety of aspects of the field, such as Named Entity Recognition, reading comprehension and the need for external knowledge, as well as a variety of others. In this paper we evaluate the current state-of-the-art models for coreference resolution on a newly annotated dataset of Dutch Parliamentary meetings. We expand the annotation scheme of [Schoen et al., 2014] for Dutch, to also include references to treaties, documents and laws, and introduce a new entity type specific to this dataset, the ‘parliamentary speech acts’. We perform an error analysis, and demonstrate that the current state-of-the-art, the neural e2e model for Dutch [van Kuppevelt and Attema, 2021] has difficulties in capturing the characteristics of this dataset, even after fine-tuning on the parliamentary data, owing to some very interesting characteristics of this data. Furthermore we propose several slight alterations to the e2e model regarding the addition of external knowledge to further improve the performance of the model on the parliamentary dataset, in particular the resolution of pronouns and the correct clustering of parliamentary speech acts.

References

- A. Schoen, C. van Son, M. van Erp, and H. van Vliet. Newsreader document-level annotation guidelines-dutch techreport 2014-8. Technical report, Technical report, VU University, 2014.
- D. van Kuppevelt and J. Attema. e2e-dutch, 2021. URL <https://github.com/Filter-Bubble/e2e-Dutch>.