Coreference Resolution

Working Title

Patrick Kahardipraja, Olena Vyshnevska

1 Introduction

- Some nice long paragraphs about place of coreference in NLP
- Paragraphs about problems with coref
- Narrow down to loca/non-local coreference
- Very broadly what we intended to to
- Broadly what we have found

2 Related Work

- some long paragraphs about history of coref. mention mention-ranking, and others
- end-to-end and higher order
- bert, bert for coref

Tenney et al. (2019)

Liu et al. (2019)

Joshi et al. (2019)

Lee, He, and Zettlemoyer (2018)

Lee, He, Lewis, et al. (2017)

3 Approach

3.1 Span Representation and Long-Range Coreference

We attempt to investigate to what extent the span representations proposed using BERT embeddings in Joshi et al. (2019) can encode coreference information, and whether it is able to encode non-local coreference phenomena or is it just simply modeling local coreference.

4 Experiment

In order to analyse this, we consider 2 kinds of span representations: 1. BERT-based span representations finetuned on OntoNotes in Joshi et al. (2019) with first and last word-pieces (concatenated with the attention version of all word pieces in the span). 2. pre-trained BERT embeddings (not finetuned on OntoNotes) for all tokens within the mention span, which is then passed through a convolutional layer (with kernel width of 3 and 5) to incorporate the local context and followed by self-attention pooling operator to produce a fixed-length span representations. This is to model head words, inspired by approach from Tenney et al. (2019).

Both span representations will be then used as inputs for coreference arc prediction task Liu et al. (2019), where a probing model (in this case a simple FFNN) is used to predict coreference relations. The probing model is designed with limited capacity to focus on what information that can be extracted from the span representations. The probing model itself has a sigmoid output layer, which is trained to minimize binary cross entropy. Each negative samples (*w_entity*, *wb*) will be generated for every positive samples (*wa*, *wb*) where *wb* occurs after *wa* and *w_entity* is a token that occurs before *wb* and belong to a difference coreference cluster, to ensure a balanced data. By comparing the performance of the probing model using these two span representations, we can hypothesize to what extent that the proposed span representation in Joshi et al. (2019) can capture coreference information. We will also experiment with mention span separation distance to see how the probing model performs and whether if there is a degradation of accuracy and F1 score of the probing model with distant spans.

5 Results

- some nice plots here
- some tables here

6 Discussion

- long long analysis of what we've seen
- generalisations, parallels
- what do this results tell us about coref and nlp in general
- discussion of why they are the way the are

6.1 Future Work

- everything we don't have the time for
- mention other corpora that could be used for finetuning (Winogrande, GAP...)

7 Conclusion

- so what have we learned about coref in general and local dependencies in particular

References

Joshi, Mandar et al. (2019). "BERT for Coreference Resolution: Baselines and Analysis". In: *Empirical Methods in Natural Language Processing (EMNLP)*.

Lee, Kenton, Luheng He, Mike Lewis, et al. (2017). “End-to-end Neural Coreference Resolution”. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, pp. 188–197. DOI: 10.18653/v1/D17-1018. URL: https://www.aclweb.org/anthology/D17-1018.

Lee, Kenton, Luheng He, and Luke Zettlemoyer (2018). “Higher-Order Coreference Resolution with Coarse-to-Fine Inference”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. New Orleans, Louisiana: Association for Computational Linguistics, pp. 687–692. DOI: 10.18653/v1/N18-2108. URL: https://www.aclweb.org/anthology/N18-2108.

Liu, Nelson F. et al. (2019). “Linguistic Knowledge and Transferability of Contextual Representations”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, pp. 1073–1094. DOI: 10.18653/v1/N19-1112. URL: https://www.aclweb.org/anthology/N19-1112.

Tenney, Ian et al. (2019). “What do you learn from context? Probing for sentence structure in contextualized word representations”. In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=SJzSgnRcKX.