

SOTA : DiscoEval

Archaeologist’s Findings

Karan Praharaj

February 14th

Paper Summary

I had the archaeologist role. I am reporting on a paper from Stanford titled “Pretraining with Contrastive Sentence Objectives Improves Discourse Performance of Language Models” ([Iter et al., 2020](#)), which uses the DiscoEval suite for evaluation (and thus also cites the DiscoEval paper ([Chen et al., 2019](#))).

This work proposes an inter-sentence objective called CONPONO, and it is an objective for pretraining language models that models discourse coherence and the distance between sentences. The intuition of this work is that if the model is able to predict the surrounding target sentences accurately given some anchor text, then the vector representations for these sentences should also be useful for downstream tasks.

The input to the model is a paragraph that is split into sentences. During training, a text segment is selected at random as the anchor. The anchor as well as all the targets, plus random samples (negative examples) are encoded with the transformer masked language model. The encoded representation of the anchor is used to predict each target at its target distance. So, in summary, first we input a paragraph. Next, an anchor sentence is selected, and then the model is trained to predict the text k sentences away using a sampled-softmax objective where the candidates consist of neighboring sentences and some more negative examples randomly sampled from elsewhere in the corpus. “Elsewhere” here entails the immediate neighbourhood of the input paragraph and the sentences within it. So on one hand the model is learning to tell the difference between the target sentence and negative examples in the case of sentences that are in the *same* context. But on the other hand, you are trying to tell the target from negative examples when comparing with sentences that are *not* in the same context. The missing context is intended to create a challenging objective for the model where your model may not be able to rely on trivial signals that often appear in contiguous sentences.

The proposed approach works better on discourse order classification rather than discourse relation classification tasks. But overall, the model was able to surpass the erstwhile state-of-the-art on four DiscoEval tasks, namely Sentence position, binary sentence ordering, discourse coherence and...sentence section prediction. And because they set the new SOTA on multiple DiscoEval tasks at the time, not too long ago, in 2020, is the reason I picked this paper for my summary.

References

- Mingda Chen, Zewei Chu, and Kevin Gimpel. Evaluation benchmarks and learning criteria for discourse-aware sentence representations. In *Proc. of EMNLP*, 2019.
- Dan Iter, Kelvin Guu, Larry Lansing, and Dan Jurafsky. Pretraining with contrastive sentence objectives improves discourse performance of language models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4859–4870, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.439. URL <https://aclanthology.org/2020.acl-main.439>.