

A Systematic Comparison of Alignment and Attention Models

Anahita Mansouri Bigvand

November 14, 2016

Objective: The attention model has been successfully used to improve neural machine translation (NMT) and achieved state-of-the-art performances by jointly learning to align and translate. Due to the correspondence between the notions of attention and alignment in traditional machine translation, we focus on the attention model (Bahdanau et al., 2014) and plan to incorporate the Markov conditioning in the attention model. The HMM-based word alignment model (Vogel et al., 1996) is motivated by the localization effect in aligning the words in parallel texts. In this model, alignment at each position depends on the previous alignment. The attention model does not take advantage of the previous context vector at each time step. We hypothesize that the context vector at each time step should depend on the previous context vector and we plan to investigate the effect of this Markov conditioning.

Methodology: Tu et al. (2016) make use of the past alignment information by introducing the coverage vector into the attention mechanism to avoid under-translation and over-translation. As a primary step to our goal, we want to look at this study that raises opportunities for further improvement. We want to perform an empirical study in which we compare the alignment quality of the NMT coverage model (Tu et al., 2016) with the alignment quality of two baselines: Simple attention model (Bahdanau et al., 2014) and the HMM-based word alignment model (Vogel et al., 1996).

Experiment: For the experiments, we compare the three models on German-English data with a test set of 150 sentences that have been hand-aligned and marked with *sure* and *possible* alignments. We evaluate the alignment quality using the alignment error rate (AER) measure.

References

- Bahdanau, D., K. Cho, and Y. Bengio (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Tu, Z., Z. Lu, Y. Liu, X. Liu, and H. Li (2016). Modeling coverage for neural machine translation. *ArXiv eprints, January*.
- Vogel, S., H. Ney, and C. Tillmann (1996). Hmm-based word alignment in statistical translation. In *Proceedings of the 16th conference on Computational linguistics-Volume 2*, pp. 836–841. Association for Computational Linguistics.