

Controlling Politeness in Neural Machine Translation via Side Constraints

1 Idea

The authors intend to control politeness while translating one language to another by using honorifics. The politeness is not grammatically implied in the source text, but should be predicted in the target text.

The term side constraints in this paper is just another way of saying attributes that the generated text is conditioned on.

2 Background

The method relies on the attention model [1] for neural translation. Each word y_i is predicted based on 3 variables: the recurrent state s_i , the previously predict word y_{i-1} and the context vector c_i .

The model relies on a GRU-backed bi-directional sequence autoencoder and the hidden state from both directions are combined to form the annotation vector. The weight of each of these annotation vectors is computed via a single layer feed-forward neural network. The weights are computed through an alignment model α_{ij} , which models the probability that x_j is aligned with y_i .

A beam search decoder is used for the generation phase.

3 Method

- The authors use a classifier first to automatically annotate the target language sentences as either formal or informal.

- The data used was movie dialogues from OpenSubtitles
- To avoid over-dependence on side-constraints and to learn to ignore side-constraints when they're irrelevant, the authors use a controlling co-efficient for how many neutral sentences are marked with a politeness feature. They also control how many labelled training instances are marked. Both co-efficients are 0.5 in their experiments.
- The entire system is parameterized by an attentional encoder-decoder NMT model.

4 Observations

Side constraints can be applied to other phenomena too, given that we already possess a classifier that can be used to label the training set without manual annotations.

References

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.