Auto-Encoding Variational Neural Machine Translation
B. Eikema

Abstract

Translation data is often a byproduct of mixing different sources of data. This could be intentional such as by mixing data of different domains or including back-translated monolingual data, but often also is the result of how the bilingual dataset was constructed: a combination of different documents independently translated in different translation directions, by different translators, agencies, etc. Most neural machine translation models do not explicitly account for such variation in their probabilistic model. We attempt to model this by proposing a deep generative model that generates source and target sentences jointly from a shared sentence-level latent representation. The latent representation is designed to capture variations in the data distribution and allows the model to adjust its language and translation model accordingly. We show that such a model leads to superior performance over a strong conditional neural machine translation baseline in three settings: in-domain training where the training and test data are of the same domain, mixed-domain training where we train on a mix of domains and test on each domain separately, and in-domain training where we also include synthetic (noisy) back-translated data. We furthermore extend the model to be used in a semi-supervised setting in order to incorporate target monolingual data during training. Doing this we derive the commonly employed back-translation heuristic in the form of a variational approximation to the posterior over the missing source sentence. This allows for training the back-translation network jointly with the rest of the model on a shared objective designed for source-to-target translation with minimal need of pre-processing. We find that the performance of this approach is not on par with the back-translation heuristic, but does lead to improvement over a model trained on bilingual data alone.