Interpreting, Training, and Distilling Seq2Seq Models

ALEXANDER RUSH

is an Assistant Professor at Harvard University studying NLP, and formerly a Post-doc at Facebook Artificial Intelligence Research (FAIR).

He is interested in machine learning and deep learning methods for large-scale natural language processing and understanding.

His past work has introduced novel methods for structured prediction with applications to syntactic parsing and machine translation.

ABSTRACT

Deep Sequence-to-sequence models have rapidly become an indispensable general-purpose tool for many applications in natural language processing, such as machine translation, summarization, and dialogue.

Many problems that once required careful domain-specific engineering can now be tackled using off-the-shelf systems by interested tinkerers. However, even with the evident early success of these models, the seq2seq framework itself is still relatively unexplored.

In this talk, I will discuss three questions we have been studying in the area of sequence-to-sequence NLP: (1) Can we interpret seq2seq’s learned representations? [Strobelt et al, 2016], (2) How should a seq2seq model be trained? [Wiseman and Rush, 2016], (3) How many parameters are necessary for the models to work? [Kim and Rush, 2016]. Along the way, I will present applications in summarization, grammar correction, image-to-text, and machine translation (on your phone).