Maximal Mutual Information Predictive Coding for Natural Language Processing

Karl Stratos

Toyota Technological Institute at Chicago

3/25/2019 at 10:30 am
CoRE A 301

Abstract

Neural predictive coding is an enormously successful approach to unsupervised representation learning in natural language processing. In this approach, a large-scale neural language model is trained to predict the missing signal (e.g., next word, next sentence) and the trained model is used in downstream tasks to produce useful text representations. While effective, it is computationally difficult to work with and yields uninterpretable representations.

In this talk, I will present a novel approach to neural predictive coding based on maximal mutual information (MMI). Instead of predicting the raw missing signal, we define a set of interpretable latent "codes" and directly predict the underlying code of the missing signal. The model is trained by maximizing the mutual information between the predicted codes. I will first present a simple and effective MMI predictive coding neural model that pushes the state-of-the-art performance in unsupervised part-of-speech tagging. In the general case in which exact calculation of entropy is intractable, a popular workaround is to maximize a sample-estimated lower bound on mutual information. I will next show that this approach suffers from fundamental statistical limitations and present an alternative approach free of these limitations.

Bio

Karl Stratos is a research assistant professor at Toyota Technological Institute at Chicago (TTIC). His research centers on statistical approaches to unsupervised learning in natural language processing. He obtained his PhD from Columbia University where he worked with Michael Collins and Daniel Hsu and spent summers at Microsoft Research New England and Google New York. He was a senior research scientist at Bloomberg L.P. and an adjunct assistant professor at Columbia University before coming to TTIC in 2017.

Faculty Host: Sungjin Ahn