THE UNIVERSITY OF CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

PUYUAN PENG

Department of Statistics
The University of Chicago

A Correspondence Variational Autoencoder for Unsupervised Acoustic Word Embeddings

WEDNESDAY, February 17, 2021, at 4:00 PM
ZOOM Meeting

ABSTRACT

We propose a new unsupervised model for mapping a variable-duration speech segment to a fixed-dimensional representation. The resulting acoustic word embeddings can form the basis of search, discovery, and indexing systems for low- and zero-resource languages. Our model, which we refer to as a maximal-sampling correspondence variational autoencoder (MCVAE), is a recurrent neural network (RNN) trained with a novel self-supervised correspondence loss that encourages consistency between embeddings of different instances of the same word. Our training scheme improves on previous correspondence training approaches through the use and comparison of multiple samples from the approximate posterior distribution. In the zero-resource setting, the MCVAE can be trained in an unsupervised way, without any ground-truth word pairs, by using the word-like segments discovered via an unsupervised term discovery system. Experiments show that MCVAE outperforms previous state-of-the-art models, such as Siamese-, CAE- and VAE-based RNNs. In addition, we also studied the recently proposed large scale speech pre-training model wav2vec2.0 and show the discriminative properties of the learned representation.

For information about building access for persons with disabilities, please contact Keisha Prowoznik at 773.702-0541 or send an email to kprowoznik@statistics.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: https://lists.uchicago.edu/web/subscribe/statseminars.