## Approaches for Neural-Network Language Model Adaptation

Language Models (LMs) for Automatic Speech Recognition (ASR) are typically trained on large text corpora from news articles, books and web documents. These types of corpora, however, are unlikely to match the test distribution of ASR systems, which expect spoken utterances. Therefore, the LM is typically adapted to a smaller held-out in-domain dataset that is drawn from the test distribution. We present three LM adaptation approaches for Deep NN and Long Short-Term Memory (LSTM): (1) Adapting the softmax layer in the NN; (2) Adding a non-linear adaptation layer before the softmax layer that is trained only in the adaptation phase; (3) Training the extra non-linear adaptation layer in pre-training and adaptation phases. Aiming to improve upon a hierarchical Maximum Entropy (MaxEnt) second-pass LM baseline, which factors the model into wordcluster and word models, we build an NN LM that predicts only word clusters. Adapting the LSTM LM by training the adaptation layer in both training and adaptation phases (Approach 3), we reduce the cluster perplexity by 30% compared to an unadapted LSTM model. Initial experiments using a state-of-the-art ASR system show a 2.3% relative reduction in WER on top of an adapted MaxEnt LM.

Author: Mr FADI, Biadsy (Google Inc.)

Presenter: Mr FADI, Biadsy (Google Inc.)