Language Model Contextualization for Automatic Speech Recognition by Dynamic Adjustment

Anna Currey
Université de Lorraine and Universität des Saarlandes

M.Sc. Thesis, 2016

Abstract

Out-of-vocabulary (OOV) words can pose a particular problem for automatic speech recognition (ASR) of broadcast news. The language models (LMs) of ASR systems are typically trained on static corpora, whereas new words (particularly new proper names) are continually introduced in the media. Additionally, such OOVs are often content-rich proper nouns that are vital to understanding the topic. In this work, we explore methods for dynamically adding OOVs to language models by directly estimating their language model parameters. We concentrate on adaptation of the bigram language model used in the first pass of our ASR system.

We propose two strategies for estimation of \(n\)-gram parameters. The first relies on finding in-vocabulary (IV) words similar to the OOVs; OOV behavior is modeled after the behavior of these similar IVs. We use word embeddings to define similarity and examine various word vector training architectures, finding skip-gram with a context size of two words to work best for our application.

Our second strategy leverages a small contemporary corpus to estimate OOV unigram probabilities and to find bigrams containing the OOVs. We find that it is best to limit the number of bigrams added to the language model; this allows us to increase the coverage of the model without creating a significant amount of noise.

We use two experimental setups to evaluate our proposed methods. In experiment 1, we create only one adapted LM for each algorithm; these LMs are then incorporated into our existing ASR system and recognition error rates (word and proper noun) are calculated. In experiment 2, we create a separate language model for each article using each of the algorithms and find the average perplexity over all of the articles. In both experiments, our adapted models improve over the baseline in perplexity; improvements are greatest in experiment 2. Only the corpus-based adaptation method improves significantly over the baseline in recognition error rate.