Spoken Language Identification Using a Combined Language Model

Sidhi Adkoli[†] and Homayoon Beigi[†,‡]

[†]Columbia University and [‡]Recognition Technologies, Inc. New York, USA

Language identification (LID) systems have improved through use of neural networks as the acoustic model. This paper surpasses the accuracies of existing systems by augmenting a Time Delay Neural Network (TDNN) acoustic model with a combined language model. Most proposed systems that use language models require manual phonetic labeling or have individual models per language thus increasing the complexity of the LID system. In this paper, the language model is constructed by concatenating the data of all languages. The advantage of the combined model is the ability to detect language switches within a single utterance. Moreover, the same LID system can be used for speech recognition with minimal alterations. Three well documented languages - English, Mandarin, and Arabic from public data sets are chosen and a combined 4-gram language model is built. The entire architecture of LID using a TDNN acoustic model with a language model is based on the TED-LIUM (Switchboard) recipe from the ASR framework provided by Kaldi. The paper achieves a final accuracy of 99.7%. This confirms the performance boost provided by the language model unhampered by its combination across all languages.