

TOWARDS DEEP LEARNING ON SPEECH RECOGNITION FOR KHMER LANGUAGE

by

Chanmann Lim

Dr. Yunxin Zhao, Thesis Supervisor

Abstract

In order to perform speech recognition well, a huge amount of transcribed speech and textual data in the target language must be available for system training. The high demand for language resources constrains the development of speech recognizers for new languages. In this thesis the development of a low-resourced isolated-word recognition system for “Khmer” language is investigated. Data pre-processing based on Voice Activity Detection (VAD) is discussed. As by-products of this work, phoneme based pronunciation lexicon and state tying questions set for Khmer speech recognizer are built from scratch. In addition to the conventional statistical acoustic modeling using Gaussian Mixture Model (GMM-HMM), a hybrid acoustic model based on Deep Neural Network and Hidden Markov Model (DNN-HMM) trained to predict context-dependent triphone states is evaluated. Dropout is used to improve the robustness of the DNN and cross-lingual transfer learning that can make use of auxiliary training data in English is also investigated. As the first effort in using DNN-HMM for low-resourced isolated-word recognition for Khmer language, the system trained on speech data recorded via mobile phones, currently performs at 93.31% word accuracy in speaker-independent mode at the vocabulary size of 194 words on the test set.