Chinese-English Mixlingual Automatic Speech Recognition System

Emily Hua, Kelly Chen, Wendy Wang

Department of Computer Science, Engineering, QMSS @ Columbia University in the City of New York

Introduction

Our project aims to build a Mandarin-English mixlingual ASR. The mixlingual speaking habit brings a code-switching phenomenon, where more than one language occurs within an utterance. This is very common in many multilingual countries.

For example, the following sentence involves one utterance code-switching: “Deep learning project presentation. An example code-switching sentence”.

We tackle this code-switching challenge using hybrid models. Our approach uses GMM-HMM system, which assumes Gaussian distribution of signal and associates each of GMM states. However, this fixed distribution assumption is not necessarily true. Neural Nets come in to learn features and provides posterior for decoding without assuming any particular structure of the data. Our project shows how systems improves traditional GMM- HMM systems by 10% absolute on this Mandarin-English Code-Switching in South-East Asia (SEAME) mixlingual dataset we acquired from Linguistic Data Consortium (LDC).

GMM-HMM

Hybrid System

The mixlingual speaking habit brings a code-switching phenomenon, where more than one language occurs within an utterance. This is very common in many multilingual countries. For example, the following sentence involves one utterance code-switching: “Deep learning project presentation. An example code-switching sentence”.

We tackle this code-switching challenge using hybrid models. Our approach uses GMM-HMM system, which assumes Gaussian distribution of signal and associates each of GMM states. However, this fixed distribution assumption is not necessarily true. Neural Nets come in to learn features and provides posterior for decoding without assuming any particular structure of the data. Our project shows how systems improves traditional GMM-HMM systems by 10% absolute on this Mandarin-English Code-Switching in South-East Asia (SEAME) mixlingual dataset we acquired from Linguistic Data Consortium (LDC).

GMM-HMM

Hybrid System

In the hybrid model, instead of using GMM, we use Deep Neural Nets to estimate posterior probabilities. The output of the neural network is the probability of a phone class given the feature $P(y|x)$. In order to compute the estimation probability $P(x|y)$ for HMM, we use the Bayes Rule: $P(x|y) = P(y|x)P(x)/P(y)$, which can be simplified as $P(y|x)P(x)$. This is the case, as $P(x)$ does not depend on the class $y$.

This being said, we scaled the neural net output by class priors $P(y|x)$ in order to get the estimation probability for HMM.

DNN

Network Description:

We used a Mafnet (p-grams) Network for our baseline DNN. The output is a “network-centric” re-ranking. p-grams

$$w = \{x_1, x_2, \ldots, x_n\}$$

where the vector $x$ represents a small group of input.

The network consists of 2 layers: $w$ and $\theta$.

The loss is to be minimized:

$$\sum_{w \in W} E(w)$$

where $w$ is the input.

This is applied directly after p-grams (without a layer of weight in-between) to stabilize unbounded output non-linearities.

Further Thoughts

We are going to check the following aspects to improve the performance, if time permits:

Transcript Improvement:

- change Brits English to American English as in the CMU lexicon (e.g., asleep -> asleep)
- misapplied English words (e.g., available -> available)

Vocabulary Improvement:

- improve single words in lexicon to reduce OOV rate (e.g., handphone, hat, network)

We are planning to cover the above aspects in our presentation.

References


Fig 1. University Of Edinburgh-ASR Course Slides, 2013

Fig 2. Some Kaldi Notes, Josh Meyer, 2016

Fig 3. Hybrid System Diagram. Here GMM-HMM creates the several alignments between features and phone states. These phone states serve as input to the Deep Neural Network, while the inputs to DNN are the usual features. The Neural Network provides the conditional probability, which can be used as estimation probability for the HMM.