Low Resource Speech Recognition

Project Code: FP1-10

Student Name: Li HUO

Supervisor: Pascale Fung

Overview

- In this thesis, a new method to train an acoustic model of speech recognizer for Cantonese is presented. The new method is based on an unsupervised training method with additional training data in Hansards format.
- My objective is to train a Cantonese acoustic model with English recognition ability in my new method and to demonstrate its improvement compared with conventionally trained acoustic models.

ASR System

- Automatic Speech Recognition (also known as speech recognition or computer speech recognition) converts spoken words to text.

Data

- I have collected a total of 2700 hours of audio data and their corresponding Hansards from the website of Hong Kong Legislative Council.
- The audio data used for our experiments consist of 42 council meetings in 2009 between Jan and Dec. The speeches are spontaneous and fluent, major in Cantonese and code switching with some English and Mandarin.
- A total of 30.6 millions word tokens are collected from 344 documents.

Algorithm

**Initialization**

1. Preparing the reference training data for SVM Classifier by aligning the Hansards and manual transcription.
2. Extract features such as Log Likelihood Score, TFIDF and Cosine Similarity from the aligned Hansards and manual transcription.
3. Train the initial model of the SVM Classifier, legislative council meeting data in normal approach.

**Iteration Process**

4. Apply the classifier on the test data set to select high-confidence words.
5. Perform adaptation on the initial acoustic model by using the high-confidence words obtained in Step 4 as training data.
6. Evaluate the acoustic model on the development set data.
7. Forced-align the Hansards with the audio speech according to the updated acoustic model.
8. Retrain the classifier with forced-aligned Hansards obtained in Step 6.

Results

- Data decoded by the ASR system with updated acoustic models are aligned with the true transcription. Decoding accuracy can be calculated by Accuracy = (Total – Substitution – Insertion – Deletion) / Total.

Conclusions

- Most of the updated models out-perform the baseline model.
- The model updated with the data selected by SVM classifier based on acoustic likelihood score feature performs the best.
- Confidence measure with SVM classifier out-performs the one with threshold.