Public Abstract
First Name:Xin
Middle Name:
Last Name:Chen
Adviser's First Name:Yunxin
Adviser's Last Name:Zhao
Co-Adviser's First Name:
Co-Adviser's Last Name:
Graduation Term:SP 2008
Department:Computer Science
Degree:MS
Title:Ensemble Methods in Large Vocabulary Continuous Speech Recognition

Combining a group of classifiers and therefore improving the overall classification performance is a young and promising direction in Large Vocabulary Continuous Speech Recognition (LVCSR). Previous works on acoustic modeling of speech signals such as Random Forests (RFs) of Phonetic Decision Trees (PDTs) has produced significant improvements in word recognition accuracy. In this thesis, several new ensemble approaches are proposed for LVCSR and experimental evaluations have shown absolute accuracy gains up to 2.3% over the conventional PDT-based acoustic models in our telehealth conversational speech recognition task.

Unlike the implicit PDT based states tying that has been used in most ASR systems as well as in the recent RFs based PDTs, this author considers that explicit PDT (EPDT) tying that allows Phoneme data Sharing (PS) may be superior in capturing pronunciation variations. The author adopted the idea of combining multiple acoustic models and applied this idea to the EPDT models. A combination of EPDT and the implicit PDT models has been investigated to reduce phone confusions that may be introduced by the EPDT model. A 1.3% absolute gain on word accuracy is observed in this experiment on the telehealth task.

Data sampling is one of the primary ways to generate different classifiers for an ensemble classifier. In this thesis, Cross Validation (CV) based data sampling is proposed, and random sampling without replacement is used as a reference for comparison. With different datasets generated by data sampling, different PDTs and therefore different Gaussian mixture models are generated, and the diversity of the multiple models helps improve recognition accuracy. When a 10-fold-CV is used, a 2.3% absolute gain in word recognition accuracy is obtained. Several experimental parameter settings and combining methods have been investigated in the experiments and the findings are discussed in this thesis.

The word accuracy performance improvement achieved in this thesis work is significant and the techniques have been integrated in the telemedicine automatic captioning system developed by the SLIPL group of the University of Missouri - Columbia.