# IMPROVEMENT OF DECODING ENGINE & PHONETIC DECISION TREE IN ACOUSTIC MODELING FOR ONLINE LARGE VOCABULARY CONVERSATIONAL SPEECH RECOGNITION

Jian Xue

Dr. Yunxin Zhao, Dissertation Supervisor

## ABSTRACT

In this work, new approaches are proposed for online large vocabulary conversational speech recognition, including a fast confusion network algorithm, novel features and a Random Forests based classifier for word confidence annotation, new improvements in speech decoding speed and latency, novel lookahead phonetic decision tree state tying and Random Forests of phonetic decision tree state tying for acoustic modeling of speech sound units.

The fast confusion network algorithm significantly improves the time complexity from $O(T^3)$ to $O(T)$, with $T$ equaling the number of links in a word lattice. Several novel features, as well as Random Forests based classification technique are proposed to improve word annotation accuracy for automatic captioning. In order to improve the speed of speech decoding engine, we propose to use complementary word confidence scores to prune uncompetitive search paths, and use subspace distribution clustering hidden Markov modeling to speed up computation of acoustic scores and local confidence scores. We further integrate pre-backtrace in decoding search to significantly reduce captioning latency.

In this work we also investigate novel approaches to improve the performance of phonetic decision tree state tying, including two lookahead methods and a Random Forests method. Constrained lookahead method finds an optimal question among $n$ pre-selected questions for each split node to decrease effects of outliers, and it also discounts the contributions of likelihood gains by deeper decedents. Stochastic full lookahead method uses sub-tree size instead of likelihood gain as a measure for phonetic question selection, in order to produce small trees with better generalization capability and consistent with training data. The Random Forests method uses an ensemble of phonetic decision trees to derive a single strong model for each speech unit. We investigate several methods of combining the acoustic scores from multiple models obtained from multiple phonetic decision trees in decoding search. We further propose clustering methods to compact the Random Forests generated acoustic models to speed up decoding search.