**PhD Dissertation Proposal Presentation**

## Thursday, April 18, 2013

**9:00 ­– 10:00 AM**

**Dean’s Seminar Room**

**Engineering Building**

## Nonparametric Bayesian Approaches for Acoustic Modeling in Speech Recognition

## Amir Harati

## Department of Electrical and Computer Engineering

 **Committee:** Dr. Joseph Picone (Advisor, College of Engineering)

 Dr. Iyad Obeid (College of Engineering)

 Dr. Slobodan Vucetic (College of Science and Technology)

 Dr. Marc Sobel (Fox School of Business)

 Dr. Chang-Hee Won (College of Engineering)

### Abstract:

Recently, nonparametric Bayesian (NPB) methods have become a popular alternative to Bayesian approaches. In such approaches, we do not fix the complexity a priori and instead place a prior over the complexity (or model structure). This prior usually biases the system towards sparse or low complexity solutions. This helps to control the number of parameters in the model yet allows the structure to be learned during a data-driven training process. Therefore models can adapt to new data encountered during the training process.

In this proposal, our goal is to investigate the application of NPB modeling to acoustic modeling. Three important problems fundamental to the acoustic modeling component of a large vocabulary speaker independent continuous speech recognition system are addressed: (1) automatic discovery of sub-word acoustic units; (2) statistical modeling of sub-word acoustic units; and (3) supervised training algorithms for nonparametric acoustic models. We propose a NPB algorithm based on an ergodic Hierarchical Dirichlet Process HMM (HDP-HMM) that automatically segments and clusters the speech signal. We apply this algorithm to the problems of automatic discovery of acoustic sub‑word units and generation of a pronunciation lexicon.

A new type of HDP-HMM is presented that preserves the useful left-to-right properties of a conventional HMM, yet still supports automated learning of the structure and complexity from data. We will introduce a NPB algorithm for training these models for continuous speech recognition that allows us to infer different HDP-HMM models and segment the training data simultaneously. Moreover, a NPB approach is introduced that replaces the phonetic decision tree used in state of the art speech recognizers to tie triphone states.

Our NPB approaches improve a model’s flexibility and its ability to adapt to previously unseen events. This is critical when training speech recognition systems on imperfect data where there might be channel mismatches or noisy transcriptions. We expect our proposed solutions for these well-known acoustical modeling problems to outperform conventional approaches. This will enable a new generation of speech recognition systems capable of being trained on vast archives of data (e.g., YouTube) and to enable the rapid development of speech recognition systems in new languages.