**PhD Dissertation Defense**

## Thursday, June 25, 2015

**1:00 PM ­– 2:00 PM**

**ENGR 304**

## Nonparametric Bayesian Approaches for Acoustic Modeling

## Amir Harati

## Department of Electrical and Computer Engineering

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

 Dr. Iyad Obeid (College of Engineering)

 Dr. Marc Sobel (Fox School of Business)

 Dr. Slobodan Vucetic (College of Science and Technology)

 Dr. Chang-Hee Won (College of Engineering)

 **External Reader:** Dr. Kevin Buckley (Villanova University)

### Abstract:

The goal of Bayesian analysis is to reduce the uncertainty about unobserved variables by combining prior knowledge with observations. A fundamental limitation of a parametric statistical model, including a Bayesian approach, is the inability of the model to learn new structures. The goal of the learning process is to estimate the correct values for the parameters. The accuracy of these parameters improves with more data but the model’s structure remains fixed. Therefore new observations will not affect the overall complexity (e.g. number of parameters in the model). Recently, nonparametric Bayesian methods have become a popular alternative to Bayesian approaches because the model structure is learned simultaneously with the parameter distributions in a data-driven manner.

We have developed a Doubly Hierarchical Dirichlet Process Hidden Markov Model (DHDPHMM) with a non-ergodic structure that can be applied to problems involving sequential modeling. DHDPHMM shares mixture components between states using two Hierarchical Dirichlet Processes (HDP). An inference algorithm for this model has been developed that enables DHDPHMM to outperform both its hidden Markov model (HMM) and HDP HMM (HDPHMM) counterparts. This inference algorithm is shown to also be computationally less expensive than a comparable algorithm for HDPHMM.

In addition to sharing data, the proposed model can learn non-ergodic structures and non-emitting states, something that HDPHMM does not support. This extension to the model is used to model finite length sequences. We have also developed a generative model for semi-supervised training of DHDPHMMs. Semi-supervised learning is an important practical requirement for many machine learning applications including acoustic modeling in speech recognition. The relative improvement in error rates on classification and recognition tasks is shown to be 22% and 7% respectively. Semi-supervised training results are slightly better than supervised training (29.02% vs. 29.71%). Context modeling was also investigated and results show a modest improvement of 1.5% relative over the baseline system.

We also introduce a nonparametric Bayesian transducer based on an ergodic HDPHMM/DHDPHMM that automatically segments and clusters the speech signal using an unsupervised approach. This transducer was used in several applications including speech segmentation, acoustic unit discovery, spoken term detection and automatic generation of a pronunciation lexicon. For the segmentation problem, an F-score of 76.62% was achieved which represents a 9% relative improvement over the baseline system. On the spoken term detection tasks, an average precision of 64.91% was achieved, which represents a 20% improvement over the baseline system. Lexicon generation experiments also show automatically discovered units (ADU) generalize to new datasets.

In this dissertation, we have established the foundation for applications of non-parametric Bayesian modeling to problems such as speech recognition that involve sequential modeling. These models allow a new generation of machine learning systems that adapt their overall complexity in a data-driven manner and yet preserve meaningful modalities in the data. As a result, these models improve generalization and offer higher performance at lower complexity.