University of Chicago and Toyota Technological Institute at Chicago Machine Learning Seminar Series

PRESENTS:

Bryan Pardon
Northwestern University

Title: Audio source separation models that learn without ground truth and are open to user correction

Abstract: Separating an audio scene into isolated sources is a fundamental problem in computer audition, analogous to image segmentation in visual scene analysis. It is an enabling technology for many tasks, such as automatic speech recognition, labeling sound objects in an acoustic scene, music transcription, and remixing of existing recordings. Source separation systems based on deep learning are currently the most successful approaches for solving the underdetermined separation problem, where there are more sound sources (e.g. instruments in a band) than channels (a stereo recording has two channels). Currently, deep learning systems that perform source separation are trained on many mixtures (e.g., tens of thousands) for which the ground truth decompositions are already known. Since most real-world recordings have no such decomposition available, developers train systems on artificial mixtures created from isolated individual recordings. Although there are large databases of isolated speech, it is impractical to find or build large databases of isolated recordings for every arbitrary sound. This fundamentally limits the range of sounds that deep models can learn to separate. Once learned, a deep model's output is take-it-or-leave it and it can be difficult for the end user to affect either the current output or to give corrective feedback for the future. In this talk Prof. Pardo discusses recent work in two areas. The first is bootstrapping learning of a scene segmentation model using an acoustic cue known to be used in human audition. This allows learning a model without access to ground-truth decompositions of acoustic scenes. The second is ongoing work to provide an interface for an end user to interact with a deep model, to affect the current separation and improve future separation by allowing for retraining of the model from corrective feedback.

Bio: Bryan Pardo is an associate professor in the Northwestern University Department of Electrical Engineering and Computer Science. Prof. Pardo received a M. Mus. in Jazz Studies in 2001 and a Ph.D. in Computer Science in 2005, both from the University of Michigan. He has authored over 100 peer-reviewed publications. He has developed speech analysis software for the Speech and Hearing department of the Ohio State University, statistical software for SPSS and worked as a machine learning researcher for General Dynamics. While finishing his doctorate, he taught in the Music Department of Madonna University.

Host: Rebecca Willett

Wednesday, March 6, 1 – 2 pm
Harper Center (Booth) Room 219
Pizza provided by UChicago CS Department