Title: Universal Features – Information Extraction for Transfer Learning

Abstract:

Deep neural networks have been used successfully in a wide range of applications. At a conceptual level, we know that the statistical dependence between the data and the label is learned and some approximated version of the conditional distribution stored in the weights of DNN. By trying to understand the operations of DNN, our goal is to give a mathematical explanation to how the statistical quantities are represented inside the networks, so that we can integrate the learned knowledge stored in one DNN with knowledge from other sources, such as prior knowledge, structural knowledge, learning results from other neural networks, or simply just use it in a new and related problem. In this talk, we try to address this issue by developing a theoretical structure to measure the meaning of information by its relevance to specific inference problems, and from that we explain the behavior of neural networks as extracting "universal features", defined as the solution to a specific optimization problem. We show that this learning procedure is closely connected with a number of well-known concepts in statistics and information theory. Based on this theoretic framework, we demonstrate some flexible ways to use neural networks in transfer learning, especially in conjunction with some conventional signal processing techniques.