

When: Friday 13:50 – 14:50

Where: ETB 1020

Speaker: Ehsan Hajiramezanali

Department of Electrical & Computer Engineering
Texas A&M University

Title: Bayesian Multi-Domain Learning for Cancer Subtype Discovery

Date: 9-14-2018

Abstract: Precision medicine aims for personalized prognosis and therapeutics by utilizing recent genome-scale high-throughput profiling techniques, including next-generation sequencing (NGS). However, translating NGS data faces several challenges. First, NGS count data are often over-dispersed, requiring appropriate modeling. Second, compared to the number of involved molecules and system complexity, the number of available samples for studying complex disease, such as cancer, is often limited, especially considering disease heterogeneity. The key question is whether we may integrate available data from all different sources or domains to achieve reproducible disease prognosis based on NGS count data. In this work, we developed Bayesian Multi-Domain Learning (BMDL) model that derives domain-dependent latent representations of over-dispersed count data based on hierarchical negative binomial factorization for accurate cancer subtyping even if the number of samples for a specific cancer type is small. Experimental results from both our simulated and NGS datasets from The Cancer Genome Atlas (TCGA) demonstrate the promising potential of BMDL for effective multi-domain learning without “negative transfer” effects often seen in existing multi-task learning and transfer learning methods.

Bio: Ehsan Hajiramezanali is a Ph.D. student with the Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX, USA. He works on computational biology and Bayesian learning.