Poster No. 11

Title: Addressing Human Subjectivity via Transfer Learning: An Application to Predicting Disease Outcome in Multiple Sclerosis Patients

Presented by: Yijun Zhao

Authors: Yijun Zhao and Carla E. Brodley, Department of Computer Science, School of Engineering, Tufts University; Tanuja Chitnis and Brian C. Healy, Department of Neurology, Harvard Medical School, Partners MS Center, Brigham and Womens Hospital

Abstract:

Predicting disease course is critical in chronic progressive diseases such as multiple sclerosis (MS). In our work we are applying machine learning methods to longitudinal records of MS patients to build a classifier that predicts whether a patient will have a significant increase in disability at the five-year mark using information from the first two years of clinical visits. This prediction is key for choosing among the available treatments as some have more troubling side-effect profiles. Two challenges arise while learning with this data. First, patient data involves the physician's (possibly subjective) evaluation. Because a patient's data may come from one doctor on the first visit and different doctors on subsequent visits, it can be difficult to form an accurate predictor of disease outcome. In particular, some physicians may be biased in one direction, scoring each patient as more severe than would other physicians, while others may be biased in the opposite direction. Another challenge is that it is much easier to classify the cases with low future disability compared to cases with high future disability at early stages of the disease. This asymmetric property is due to the nature of the disease rather than class imbalance. In this paper we introduce a new transfer learning approach to handle these challenges. The algorithm builds a single SVM classifier for each doctor by dividing the entire dataset into primary (instances from the doctor) and auxiliary (instances from other doctors) sets. When applied to our dataset of MS patients our new approach is able to realize a significant increase in prediction performance over approaches that form a single SVM classifier for the entire dataset or for the physician's dataset alone.