Multiple-Instance Learning with Heterogeneous Instances

Chun-Hao Yang
Institute of Statistics and Data Sciences, National Taiwan University

Abstract

Multiple-instance learning (MIL) is a weakly supervised learning problem, where a single class label is assigned to a bag of instances, and the instance labels are not directly observed. The standard assumption for the relationship between bags and instances is that a bag has a positive label if at least one instance is positive. This assumption allows us to infer the instance-level information through the bag-level observations. MIL has received a lot of attention in the past decade due to its wide applications in biology, chemistry, computer vision, etc. In this talk, I will briefly introduce some of my research works in MIL including (i) the multiple-instance logistic regression (MILR) model, and (ii) the heterogeneous-instance logistic regression (HILoR) model, which is a modification of MILR to accommodate the heterogeneity in predictors among different instances.

As an application of the proposed HILoR model, I will use the multiple-criterion diagnoses for mild cognitive impairment (MCI) and Alzheimer's disease (AD) as an example. MCI is a prodromal stage of AD that causes a significant burden in caregiving and medical costs. Clinically, the diagnosis of MCI is determined by the impairment statuses of five cognitive domains. If one of these cognitive domains is impaired, the patient is diagnosed with MCI, and if two out of the five domains are impaired, the patient is diagnosed with AD. The proposed model is validated in terms of its estimation accuracy, latent status prediction, and robustness via extensive simulation studies and the National Alzheimer's Coordinating Center-Uniform Data Set.