The Chinese University of Hong Kong
Department of Statistics

Seminar

Double Feature Allocation for Phenotyping with Electronic Health Records Data

By

Dr. Yang NI
Department of Statistics
Texas A&M University

Abstract

In this talk, we will introduce a categorical matrix factorization method to infer latent diseases from electronic health records data in an unsupervised manner. A latent disease is defined as an unknown biological aberration that causes a set of common symptoms for a group of patients. The proposed approach is based on a novel double feature allocation model which simultaneously allocates features to the rows and the columns of a categorical matrix. Using a Bayesian approach, available prior information on known diseases greatly improves identifiability of latent diseases. This includes known diagnoses for patients and known association of diseases with symptoms. We validate the proposed approach by simulation studies including mis-specified models and comparison with sparse latent factor models. In the application to Chinese electronic health records (EHR) data, we find interesting results, some of which agree with related clinical and medical knowledge.

Date: May 30, 2019 (Thursday)
Time: 2:30 p.m. - 3:30 p.m.
Venue: Lady Shaw Building, Room LT2
The Chinese University of Hong Kong

ALL INTERESTED ARE WELCOME !!