Data Science Seminar

Hosted by the Department of Mathematics and Statistics

- Date: Tuesday, March 18, 2025
- Time: 12:00pm – 1:00pm
- Room: Whitney Hall 100E
- Speaker: Dr. Kun Chen (University of Connecticut)
- Title: Hybrid and Integrative Learning for Rare Event Modeling with EHR Data.

Abstract

The vast repositories of Electronic Health Records (EHR) and medical claims data hold untapped potential for studying rare but critical events, such as suicide attempt. Conventional setups often model suicide attempt as a univariate outcome and also exclude any “single-record” patients with a single documented encounter due to a lack of historical information. However, patients who were diagnosed with suicide attempts at the only encounter could, to some surprise, represent a substantial proportion of all attempt cases in the data, as high as 70-80%. We innovate a hybrid & integrative learning framework to leverage concurrent outcomes as surrogates and harness the forbidden yet precious information from single-record data. Our framework employs a supervised learning component to learn the latent subspace that connect primary (e.g., suicide) and surrogate outcomes (e.g., mental disorders) to historical information. It simultaneously employs an unsupervised learning component to utilize the single-record data, through the shared latent subspace. Our general formulation covers various outcome types and model specifications, including reduced-rank regression and autoencoders. Theoretically, we show that utilizing single records leads to a faster convergence rate of recovering the shared subspace. With hospital inpatient data from Connecticut, we demonstrate that single-record data and concurrent diagnoses indeed carry valuable information and utilizing them can substantially improve suicide risk modeling.

Biography of the speaker: Kun Chen is a Professor in the Department of Statistics at the University of Connecticut (UConn) and a Research Fellow at the Center for Population Health, UConn Health Center. He is a Fellow of the American Statistical Association (ASA) and an Elected Member of the International Statistical Institute (ISI). His research focuses on large-scale multivariate statistical learning, statistical machine learning, and healthcare analytics. Dr. Chen received his B.Econ. in Finance and Dual B.S. in Computer Science from the University of Science & Technology of China in 2003 and his Ph.D. in Statistics from the University of Iowa in 2011.