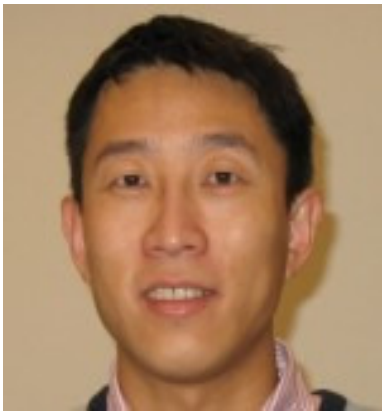
RUTGERS
School of Arts and Sciences

Department of
Statistics and Biostatistics

DEPARTMENT OF STATISTICS AND BIOSTATISTICS



Jun Yan

Department of Statistics
University of Connecticut

Online Updating Method with New Variables for Big Data Streams

October 12, 2016
3:20 – 4:20pm
Light refreshments will be served
110 Frelinghuysen Road
Hill Center, Room 552

Abstract: For big data arriving in streams, online updating is an important statistical method that breaks the storage barrier and the computational barrier under certain circumstances. Assuming no changes in the number of variables, however, online updating algorithms cannot handle newly available variables. A naive approach would be to discard all saved information and start the updating with new variables from scratch. We propose a method that utilizes the information from earlier data in the online updating algorithm with bias corrections to improve efficiency.
The method is developed for linear models first, and then extended to estimating equations for generalized linear models. The performance in comparison with the naive approach is assessed in simulation studies with data generated from a normal linear model and a Logistic regression model. The method is applied to a study on airline delay, where reasons for delays were only available in more recent years of data.

Bio: Jun Yan earned his PhD at the University of Wisconsin. He held a position at the University of Iowa before moving to the University of Connecticut. An elected member of the International Statistical Institute, he is known for his work in multivariate dependence, estimating functions, and longitudinal data analysis. He is co-editor of the book Extreme Value Modeling and Risk Analysis.