RUTGERS
School of Arts and Sciences

Department of Statistics

DEPARTMENT OF STATISTICS

Dr. Shan Yu
Department of Statistics
University of Virginia

Big Spatial Data Learning: A Parallel Solution

Wednesday, September 15th, 2021
11:45 AM EST

Zoom Meeting: Meeting ID: 924 8283 7055
Password: 378160

https://rutgers.zoom.us/j/92482837055?pwd=TktKVVU2Z1JzZnZQbm1TNUs3K0pmZz09

Virtual Coffee session before the seminar at 11:30 AM EST

Abstract:
Nowadays, we are living in the era of ``Big Data." A significant portion of big data is big spatial data captured through advanced technologies or large-scale simulations. Explosive growth in spatial and spatiotemporal data emphasizes the need for developing new and computationally efficient methods and credible theoretical support tailored for analyzing such large-scale data. Parallel statistical computing has proved to be a handy tool when dealing with big data. In general, it uses multiple processing elements simultaneously to solve a problem. However, it is hard to execute the conventional spline regressions in parallel. In this work, we develop a novel parallel smoothing technique for generalized partially linear spatially varying coefficient models, which can be used under different hardware parallelism levels. Moreover, conflated with concurrent computing, the proposed method can be easily extended to the distributed system. Regarding the theoretical support of estimators from the proposed parallel algorithm, we first establish the asymptotical normality of linear estimators. Secondly, we show that the spline estimators reach the same convergence rate as the global spline estimators. The newly developed method is evaluated through several simulation studies and an analysis of the US Loan Application Data.

Bio:
Dr. Shan Yu joined the Department of Statistics at the University of Virginia as an Assistant Professor last August 2020 after receiving her Ph.D. from Iowa State University. Her research interests focus on advanced statistical methods for complex-structured data, statistical machine learning, and "big data" analytics. Specifically, she has been engaged in projects utilizing non-/semi-parametric regression methods, spatial/spatiotemporal data analysis, biomedical imaging analysis, statistical epidemiology, and trajectory data analysis.