JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Department of Biostatistics

BIOSTATISTICS SEMINAR

Iterative Random Forests to Discover Predictive and Stable High-order Interactions

Sumanta Basu, Assistant Professor
Shayegani Bruno Family Faculty Fellow, Department of Statistics and Data Science, Department of Computational Biology, Cornell University

Abstract: Genomics has revolutionized biology, enabling the interrogation of whole transcriptomes, genome-wide binding sites for proteins, and many other molecular processes. However, individual genomic assays measure elements that interact in vivo as components of larger molecular machines. Understanding how these high-order interactions drive gene expression presents a substantial statistical challenge. Building on random forests (RFs) and random intersection trees (RITs) and through extensive, biologically inspired simulations, we developed the iterative random forest algorithm (iRF). iRF trains a feature-weighted ensemble of decision trees to detect stable, high-order interactions with the same order of computational cost as the RF. We demonstrate the utility of iRF for high-order interaction discovery in two prediction problems: enhancer activity in the early Drosophila embryo and alternative splicing of primary transcripts in human-derived cell lines. We also propose a method to refine the interactions identified by iRF to explicitly map responses as a function of interacting features. This method, signed iRF, describes subsets of rules that frequently occur on RF decision paths. We refer to these rule subsets as signed interactions. Signed interactions share not only the same set of interacting features but also exhibit similar thresholding behavior, and thus describe a consistent functional relationship between interacting features and responses.

Papers/preprints: https://www.pnas.org/content/115/8/1943.full, https://arxiv.org/abs/1810.07287

Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics
Monday, October 14, 2019, Room W2008 (Refreshments 12:00pm)

For disability access information or listening devices, please contact the Office of Support Services at 410-955-1197 or on the web **www.jhsph.edu/SupportServices**. **EO/AA**

Department of Biostatistics, 615 N. Wolfe Street, Suite E3527, Baltimore, MD 21205