High-dimensional Tensor Learning: The Good, the Bad, and the Pragmatic

Miaoyan Wang, July 2021

Overview

High dimensional tensor methods are making enormous impacts on science and society. The empirical successes, however, are uncovering pressing new challenges that stand in the way of further progress: decisions arising from classical statistical premises are sensitive to model misspecification; greedy algorithms are yielding biased solutions; and concerns on prediction-interpretability balance continue to mount.

These challenges can be addressed only by an appeal to a genuine combination of multidisplinary approaches. **We propose to develop a suite of statistical learning theory, efficient algorithms, and data-driven solutions for high-dimensional tensor problems.** The research agenda spans the full spectrum of multilinear analysis tools in modeling, statistical-computational balance in estimation, and robust algorithms with accuracy guarantees. Unlike matrices, higher-order tensor problems are NP-hard in general. The PI's work will investigate the intrinsic low-dimensionality for a wide range of structured tensors including, but not limited to, low-rankness, non-negativity, block-structure, and smoothness. The new parametric and non-parametric framework for addressing higher-order high-dimensional tensor problems will provide solutions that were previously impossible.

Tensor data analysis is an integral part of modern domain sciences. The new directions have been particularly motivated by three applications: the classification and comparison of brain connectivity data, the integration of omic (genome, transcriptome, proteome) data, and the pattern detection in social recommendation system. **The developed tools will allow domain scientists to examine complex interactions among tensor entries and between multiple tensors, thereby providing solutions to questions that cannot be addressed by traditional analysis.**

Intellectual Merit. The proposal lays out an ambitious plan for a full spectrum of parametric, non-parametric, higher-order, high-dimensional tensors with multimodal data integration. The PI has an excellent record of pushing disciplinary boundaries by synthesizing knowledge and experience from diverse areas. The PI has a unique combination of post-graduate training in statistics (2010-2015), mathematics (2015-2017), computer science (2015-2018), and genomics (2013-2018). The proposed research builds on the PI's established high-impact tensor work ranging from theory [2, 5, 6, 9, 14] to methodology [3, 8, 10, 11, 13]) to domain science [1, 4, 7, 12, 15]. PI has established herself as an outstanding teacher-scholar in multiple disciplines, including two 2021 Best Student Paper Awards (mentored by the PI as advisor) from American Statistical Association, Charles J. Epstein Trainee Semifinalist Award from American Society of Human Genetics, and IGES Williams Finalist Award from International Genetic Epidemiology Society. The PI's work has a transformational nature involving connections across diverse disciplines, and the PI strives to push the boundary of interdisciplinary research further.

Broader Impacts. The PI will integrate the research and eduction goals by establishing a unique research-training experience, thereby broadening participation in STEM fields. Comprehensive plans for education and outreach will draw on PI's past success and leverage institutional resources through the four-university TRIPODS and Data Science Hub that the PI is affiliated with. The PI will organize and host summer research positions that engage a diverse student body at all levels of seniority. New courses on *fairness in data science* will be created in both traditional and online formats. The course focuses on critical thinking skills and development of equitable access for underrepresented groups. The PI is committed to training students the interdisciplinary skills for them to make independent, insightful contributions to the larger society. Collaborations with domain science researchers in academia and non-profit labs will continue through the established channels. The research-training agenda is well integrated into the education, which as a whole, creates an inclusive environment for every student to succeed in their own individual careers.