Learning and Inference for Low-Rank Models

Dong Xia[1]

Department of Mathematics, Hong Kong University of Science and Technology

ABSTRACT

My talk will focus on the learning and inference for low-rank matrices and tensors in the presence of missing values, heterogeneity, and adaptively collected data. These problems pose significant challenges because the estimators are typically derived using iterative algorithms and involve multiple stages of spectral decomposition. Over the past several years, we have made several contributions in this field, including specially crafted estimation procedures, a powerful spectral representation formula, the double-sample debiasing approach, false-discovery control in multiple hypothesis testing, and debiasing using inverse propensity weighting.

Keywords: Low-rank, tensor, inference, spectral methods

Federated Community Detection in Bipartite Networks from Various Platforms

Wanjie Wang

Department of Statistics and Data Science, National University of Singapore

ABSTRACT

In modern applications such as recommendation systems and e-commerce platforms, user-item interactions are naturally modelled by bipartite networks. When data are collected across multiple servers, privacy concerns and communication constraints render centralized community detection impractical. It motivated the development of federated learning. In this talk, I will present a federated framework for community detection in bipartite networks, where user data reside on disjoint servers, but all servers share a common item set.

We introduce a gradient-based power iteration method that approximates leading singular vectors using only local computations and limited inter-server communication. Building on this, we propose the BiFLICKER algorithm, which combines spectral methods with federated learning to recover global community structures while preserving data locality. Theoretical guarantees are established for the accuracy of spectral estimation and community recovery. We validate our method through simulations and real-world analysis of federated movie rating data from Douban, revealing meaningful and interpretable user and movie communities.

Keywords: Federated Learning; Spectral Analysis; Bipartite Network; Community Detection

Online Tensor Inference

Xin Wen, Will Wei Sun, **Yichen Zhang**

Department of Quantitative Methods, Daniels School of Business, Purdue University

ABSTRACT

Contemporary applications, such as recommendation systems and mobile health monitoring, require real-time processing and analysis of sequentially arriving high-dimensional tensor data. Traditional offline learning, involving the storage and utilization of all data in each computational iteration, becomes impractical for these tasks. Furthermore, existing low-rank tensor methods lack the capability for online statistical inference, which is essential for real-time predictions and informed decision-making. This paper addresses these challenges by introducing a novel online inference framework for low-rank tensors. Our approach employs Stochastic Gradient Descent (SGD) to enable efficient real-time data processing without extensive memory requirements. We establish a non-asymptotic convergence result for the online low-rank SGD estimator, nearly matches the minimax optimal estimation error rate of offline models. Furthermore, we propose a simple yet powerful online debiasing approach for sequential statistical inference. The entire online procedure, covering both estimation and inference, eliminates the need for data splitting or storing historical data, making it suitable for on-the-fly hypothesis testing. In our analysis, we control the sum of constructed super-martingales to ensure estimates along the entire solution path remain within the benign region. Additionally, a novel spectral representation tool is employed to address statistical dependencies among iterative estimates, establishing the desired asymptotic normality.

Keywords: Low-rank tensors; online learning; statistical inference; stochastic gradient descent

Back to Sessions List

Generalized Tensor Completion with Non-Random Missingness

Maoyu Zhang[1*], Biao Cai[2*], Will Wei Sun[3], **Jingfei Zhang[1]**

[1]Goizueta Business School, Emory University, Atlanta, GA, USA.

[2]Department of Decision Analytics and Operations, City University of Hong Kong, Hong Kong.

[3]Daniels School of Business, Purdue University, West Lafayette, IN, USA.

Abstract

Tensor completion plays a crucial role in applications such as recommender systems and medical imaging, where data are often highly incomplete. While extensive prior work has addressed tensor completion with data missingness, most assume that each entry of the tensor is available independently with probability p. However, real-world tensor data often exhibit missing-notat-random (MNAR) patterns, where the probability of missingness depends on the underlying tensor values. This paper introduces a generalized tensor completion framework for noisy data with MNAR, where the observation probability is modeled as a function of underlying tensor values. Our flexible framework accommodates various tensor data types, such as continuous, binary and count data. For model estimation, we develop an alternating maximization algorithm and derive non-asymptotic error bounds for the estimator at each iteration, under considerably relaxed conditions on the observation probabilities. Additionally, we propose a statistical inference procedure to test whether observation probabilities depend on underlying tensor values, offering a formal assessment of the missingness assumption within our modeling framework. The utility and efficacy of our approach are demonstrated through comparative simulation studies and analyses of two real-world datasets.

Keywords: generalized tensor completion; low-rank tensor model; missing not at random; non-convex optimization; hypothesis testing.