Statistics and Its Interface Volume 17 (2024) 143–144

Editorial: Special issue on statistical learning of tensor data

Tensor, or multidimensional array, is arising in various scientific and business applications. Research on learning of tensor data has been rapidly expanding during the last few decades, extending to modern datasets such as medical images, social networks, and personalized recommendation systems, and widely used in many fields, including medicine, biology, public health, engineering, finance, economics, sports analytics, and environmental sciences. The rapid developments also lead to many challenges in estimation, inference, prediction, and computation in learning tensor data. SII promotes an interface between statistical theory, methodology, and applications. All submissions went through a regular review process. Four co-guest editors handled the peer review of all invited and contributed submissions. In total, we have collected eleven articles that include cutting-edge research on statistical and computational methods for tensor data analysis and their applications.

Zhang, Gao, Pan, and Wang consider community detection in temporal citation networks via a tensor-based approach. The authors specifically analyze the temporal citation network from publications in 44 statistical journals between 2001 and 2018 by the Tensor-based Directed Spectral Clustering On Ratios of Eigenvectors (TD-SCORE) method to effectively handle degree heterogeneity and detect community structures. The study begins by examining the temporal network's characteristics, including in-degree distribution and the changing nature of community structures and key nodes over time. By applying TD-SCORE to the network's core, the authors successfully identify seven distinct communities, including variable selection, Bayesian analysis, and functional data analysis. The authors track the evolution of these communities, which offers valuable insights into their dynamics and transformation over the studied period.

Qi, Zhu, and Wang develop a random projection method for large-scale community detection. By introducing a random Gaussian matrix that generates several projections on the column space of the network adjacency matrix, the k-means algorithm is then applied with the low-dimensional projected matrix. Under appropriate model assumptions (e.g., the SBM model), a strong consistency result of the proposed algorithm is provided. Two real-world network datasets illustrate the effectiveness of the proposed methods.

Lan focuses on correlated Wishart matrices classification via an expectation-maximization composite likelihood-based algorithm. The composite likelihood is an inference function derived by multiplying a collection of component likelihood, primarily to resolve the issue that the full likelihood is unavailable. The proposed method enjoys accurate parameter estimation and fast computational speed and has significant contributions in both the methodological development of matrix-variate analysis and the application of computer vision.

Lai and Yin discuss the construction of a conditional dependence graph for unstructured data, where variables do not follow the i.i.d. assumption. By assuming that variables follow a matrix normal distribution with sparse precision matrices and using GloVe embeddings and word co-occurrence data, the authors propose an empirical test for this assumption and derive its asymptotic distribution. This approach captures essential semantic relationships without listing all related variable pairs. The authors introduce Matrix-GloVe, a penalized matrix normal graphical model (MNGM) using the MDMC optimization method, which is faster and can accommodate new concepts. They also propose a sentence granularity bootstrap to enhance the MNGM algorithm.

Wang and Xu study a Bayesian tensor-on-tensor regression to predict a multidimensional tensor of arbitrary dimensions from another tensor of arbitrary dimensions, building upon the Tucker decomposition of the regression coefficient tensor. The proposed model can simultaneously estimate the model dimension (the dimension of the core tensor) and other model parameters. An efficient Markov chain Monte Carlo (MCMC) sampling algorithm and an optimization-based computing algorithm are designed for the proposed method. The proposed Bayesian framework provides a natural way for both parameter estimation and uncertainty quantification.

Wang, Zhou, He, and Ni consider a Bayesian multiway overlapping clustering approach to cluster genes, tissues, and individuals simultaneously, motivated by the Genotype-Tissue Expression (GTEx) RNA-seq data. A latent categorical variable combined with a mixture of one Gaussian and two uniform distributions is introduced to model the normalized gene expression data. To reduce the dimensionality and to perform overlapping clustering, the latent variable is linked with three lower-dimensional matrices by a multi-class logistic model. Conditions are provided to guarantee the identifiability of the model.

Wang, Zhou, Yang, and Mai introduce a new classifier for tensor data, Density-Convoluted Tensor Support Vector Machines (DCT-SVM), to address the limitations of traditional vector-based classification methods in modern applications like image processing and digital marketing. DCT-SVM is based on applying kernel density convolution to the SVM loss, thereby generating a novel family of classification loss functions. The authors thoroughly explore the theory of DCT-SVM, particularly focusing on the probabilistic order of magnitude for its excess risk. To compute DCT-SVM efficiently, the authors develop and validate the convergence of a fast monotone accelerated proximal gradient descent algorithm. The superior efficacy of DCT-SVM is validated through simulation studies, showcasing its advantages over prevalent classification methods. Moreover, the practical utility of DCT-SVM is illustrated in a real-world application within the context of online advertising.

Shi and Shen provide a comprehensive overview of methodologies for tensor completion and regression. The review covers model formulation, prior assignment, posterior computation, and theoretical properties. The authors also shed light on potential future research in this domain.

Hsu, Huang, Tsay, and Kao propose rank-r matrix autoregressive models for modeling spatio-temporal data. The authors incorporate a banded neighborhood structure for AR coefficient matrices and utilize a flexible nonstationary low-rank covariance model for the spatial innovation process, leading to a parsimonious model without sacrificing its exibility. A computationally efficient alternating direction method of multipliers algorithm is used for maximizing likelihood.

Wang and Zhang examine robust and covariance-assisted tensor response regression based on a recently proposed tensor t-distribution. The proposed model assumes that the tensor regression coefficient has a low-rank structure that can be learned more effectively by using the additional covariance information. This approach is insensitive to heavy-tailed data and potential outliers. Theoretical and numerical studies confirm the superior performance of the proposed method.

Mu, Hu, and Wu study model-based statistical depth for matrix data, focusing on the under-explored ranking problem motivated by applications ranging from medicine and finance to sports analytics. The authors propose a new approach using a model-based depth framework that involves estimating the eigendecomposition of a 4th-order covariance tensor. The methodology is enhanced by employing a Kronecker product form on the covariance, improving robustness and reducing the complexity of the model. The authors present an efficient algorithm to estimate this model-based statistical depth, demonstrate the method's effectiveness through simulations and real-world applications, and analyze real datasets on NBA players' field goal attempts and global temperature anomalies.

We hope that this special issue can help stimulate more research interests on statistical learning of tensor data and shed light on future research directions. We also hope that this special issue makes SII a friendly home to many more exciting developments of theory and methods and novel applications in statistical learning of tensor data.

Guanyu Hu (Co-Guest Editor)
University of Texas Health Science Center at Houston

HaiYing Wang (Co-Guest Editor)
University of Connecticut

Jing Wu (Co-Guest Editor)
University of Rhode Island

Anru Zhang (Co-Guest Editor)
Duke University

Ming-Hui Chen (Co-Editor-in-Chief)
University of Connecticut

Yuedong Wang (Co-Editor-in-Chief)
University of California, Santa Barbara