Simultaneous Clustering and Estimation of Additive Shape Invariant Models for Recurrent Event Data

Zitong Zhang[1], **Shizhe Chen[1]**

Department of Statistics, University of California, Davis

ABSTRACT

Technological advancements have enabled the recording of spiking activities from large neuron ensembles, presenting an exciting yet challenging opportunity for statistical analysis. This project considers the challenges from a common type of neuroscience experiments, where randomized interventions are applied over the course of each trial. The objective is to identify groups of neurons with unique stimulation responses and estimate these responses. The observed data, however, comprise superpositions of neural responses to all stimuli, which is further complicated by varying response latencies across neurons. We introduce a novel additive shape invariant model that is capable of simultaneously accommodating multiple clusters, additive components, and unknown time-shifts. We establish conditions for the identifiability of model parameters, offering guidance for the design of future experiments. We examine the properties of the proposed algorithm through simulation studies, and apply the proposed method on neural data collected in mice.

Keywords: Point processes, shape invariant models, additive models, clustering

Multiview Manifold Learning for High-Dimensional and Noisy Data Analysis

Xiucai Ding[1], Chao Shen and Hau-Tieng Wu[2,*]

Department of Statistics, University of California, Davis

ABSTRACT

A longstanding challenge in data science is to effectively quantify systems of interest by integrating information from heterogeneous datasets, a problem known as multiview learning. In this talk, I will present recent advancements in this direction, focusing on novel algorithms based on convolutions of diffusion maps or kernel embeddings. Within the common manifold framework, the proposed algorithm can be interpreted through its spectral connection to limiting Laplacian or integral operators. Additionally, we demonstrate that the method is robust against high-dimensional noise via the analysis of the underlying kennel random matrices.

Keywords: Multiview Learning; Manifold Learning; Representation Learning; High-dimensional data analysis

Back to Sessions List

Quantile Small Area Estimation via Singh-Maddala Mixed Model Prediction

Thuan Nguyen[1], Yuzi Liu[2], Haiqiang Ma[2], Xiaohui Liu[2] and Jiming Jiang[3,*]

[1]*OHSU/PSU School of Public Health, Oregon Health and Science University*

[2]*School of Statistics and Data Science, Jiangxi University of Finance and Economics*

[3]*Department of Statistics, University of California, Davis*

ABSTRACT

We develop methods of mixed model prediction (MMP) of quantiles of interest under the Singh-Maddala (SM) distribution for the outcome variable, which is widely used for financial and economic data. Such outcome data are often non-Gaussian, having skewed distributions. The traditional quantile mixed effects models have largely focused on estimating the fixed effects and variance components in the model. However, prediction of quantiles at subject-level, such as those associated with the small areas, are also of practical interest. We develop methods of MMP for subject-level quantiles of interest under the SM outcome distribution. Specifically, we develop the optimal prediction theory under a mixed effects SM model. The best predictor (BP) under the expected pinball loss (EPL) is the conditional quantile of the target interest. We then obtain the empirical BP (EBP) of the subject-level quantiles by replacing the unknown parameters involved in the BP by their maximum likelihood estimators. We establish asymptotic optimality of the EBP under the EPL measure. As a measure of uncertainty, we develop a second-order unbiased estimator of the EPL of the EBP. Empirical performances of the EBP and its EPL estimator are evaluated via simulation studies. An application to income inequality in China is discussed.

Keywords: Best prediction, Financial and economic data, Mixed-effect SM model, Subject-level quantiles

Interpretable Transformer Regression for Functional and Longitudinal Covariates

Yuan-Jung Cynthia Juang, Jane-Ling Wang*
Department of Statistics, University of California, Davis

ABSTRACT

Predicting scalar outcomes from functional data is challenging when measurements are sparse, irregular, and noisy, as in many scientific and clinical longitudinal studies. We propose IDAT, a dual-attention Transformer that operates directly on masked sampling schedules and avoids ad-hoc imputation. IDAT couples (i) time-point attention, which captures local and long-range temporal dynamics together with the response relationship nonparametrically, with (ii) inter-sample attention, which adaptively shares information across subjects with similar trajectories to stabilize estimation under sparsity. These pathways complement one another: time-point attention captures subject-specific dynamics, whereas inter-sample attention leverages population structure to ``borrow information" from other subjects, echoing principles from random-effects model in longitudinal analysis. Under a random-effects framework that accounts for irregular sampling and measurement noise, we prove prediction-error bounds and show that IDAT consistently approaches the oracle solution. Across both simulations and real-world applications, IDAT achieves the best overall performance among 19 baselines. Only in the extremely dense case (> 80% observations) TabPFN (a recent method published in Nature) achieve a slight advantage, while IDAT still significantly outperforms all other baselines in this scenario. The learned attention weights are interpretable, revealing predictive time domains and potential clusters. In conclusion, IDAT, an end-to-end sparsity-aware Transformer architecture, offers improvements both in predictive accuracy and interpretability for scalar-on-function prediction.

Keywords: Dual-attention Transformer, interpretable attention, functional/longitudinal data, irregular sampling plan