Statistical Thinking and AI Transformers: A Two-Way Exchange Between Time Series and Attention Mechanisms

Jiecheng Lu (student), **Shihao Yang**

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology

ABSTRACT

Despite Transformers' success across AI domains, their application to time series forecasting remains lukewarm. For example, in infectious disease forecasting it showed no clear advantage over classical statistical baselines, motivating us to bridge attention mechanisms and classical time series principles.

We first develop a statistical account of when and why attention should work: a single linear attention layer behaves like a low-rank Vector Autoregression (VAR), while stacking layers induces higher-rank lag interactions, yielding an attention design aligned with VAR structure. We then observe that typical Transformers are autoregressive only, missing the moving-average component in statistical time series models. By introducing an attention pathway over residuals inspired by ARMA models, we improve forecasting accuracy. For efficiency and long contexts, we reinterpret linear attention as a truncated softmax and add dedicated pathways (beyond Q, K, V) capturing higher-order Taylor series to recover expressivity at linear time. We show how "prompt engineering" for time series enables zero-shot and few-shot forecasting through in-context learning, how auxiliary "scratch-paper" channels act as a chain-of-thought analogue for multivariate series, and how targeted training techniques stabilize and accelerate convergence.

Together, these contributions demonstrate how statistical thinking enables better Transformer designs for time series and reciprocally, how time series insights yield improved attention mechanisms that transfer to vision and text domains. This two-way exchange between classical statistical principles and modern deep learning architectures opens new directions for both fields.

Keywords: Time Series; Attention Mechanisms; Transformers; AI Foundation Models

HeteroJIVE: Joint Subspace Estimation for Heterogeneous Multi-View Data

Jingyang Li[1], Zhongyuan Lyu[2]
[1]*University of Michigan, Ann Arbor*
[2]*The University of Sydney*

ABSTRACT

Many modern datasets consist of multiple related matrices measured on a common set of units, where the goal is to recover the shared low-dimensional subspace. While the Angle-based Joint and Individual Variation Explained (AJIVE) framework provides a solution, it relies on equal-weight aggregation, which can be strictly suboptimal when views exhibit significant statistical heterogeneity (arising from varying SNR and dimensions) and structural heterogeneity (arising from individual components). In this paper, we propose HeteroJIVE, a weighted two-stage spectral algorithm tailored to such heterogeneity. Theoretically, we first revisit the ``non-diminishing" error barrier with respect to the number of views K identified in recent literature for the equal-weight case. We demonstrate that this barrier is not universal: under generic geometric conditions, the bias term vanishes and our estimator achieves the $O(K^{-1/2})$ rate without the need for iterative refinement. Extending this to the general-weight case, we establish error bounds that explicitly disentangle the two layers of heterogeneity. Based on this, we derive an oracle-optimal weighting scheme implemented via a data-driven procedure. Extensive simulations corroborate our theoretical findings, and an application to TCGA-BRCA multi-omics data validates the superiority of HeteroJIVE in practice.

Keywords: JIVE, heterogeneity, Multi-view data analysis

Back to Sessions List

A Riemannian Factor Model for Manifold-valued Time Series

Shuo-Chieh Huang[1], Rong Chen[1], Yaqing Chen[1]

[1]*Department of Statistics, Rutgers University*

ABSTRACT

In this paper, we propose the Riemannian factor model, a novel framework for analyzing time series taking values in Riemannian manifolds in potentially high dimensions. Such time series is encountered in many applications, including economics, finance, medical imaging, and genomics and microbiome research. The proposed model is geometry-aware and accounts for the inherent nonlinearity in the data. Under a high-dimensional asymptotic regime, where the manifold dimension is allowed to diverge with n, the sample size, we establish convergence rates for the estimated loading space. In particular, under short-memory and strong factor conditions, we obtain a dimension-free n^{-1/2} rate, which matches the convergence rates of the fixed-dimensional Riemannian principal component analysis and the high-dimensional linear factor models with strong factors. Applied to the covariance matrices of selected U.S. stock returns, viewed as time series in the Bures-Wasserstein manifold, the proposed method yields interpretable factors and competitive predictions.

Keywords: dimension reduction, non-Euclidean time series; compositional time series; Bures-Wasserstein metric; covariance prediction.

Enhancing Generalizability and Fairness of HIV Risk Predictions: A Machine Learning Approach Using EHR Data

Hulin Wu

The Betty Wheless Trotter Professor, Department of Biostatistics & Data Science,

School of Public Health, University of Texas Health Science Center at Houston

ABSTRACT

Despite significant progress in HIV prevention and treatment, underdiagnosis remains a major driver of new infections. We developed and validated a machine learning–based HIV-1 risk prediction model using the Cerner Health Facts® nationwide electronic health record (EHR) database, encompassing more than 69 million patients across 85 U.S. health systems (2000–2018). An automated HIV phenotyping algorithm identified 38,310 HIV-1 cases and 118,090 controls, from which 4,442 candidate predictors were screened. A refined LASSO logistic regression model retained 281 predictors and achieved high discrimination (AUC = 0.927; sensitivity = 79.2%; specificity = 92.7%). To address fairness, we evaluated performance across demographic subgroups and implemented a threshold-based post-processing method to mitigate disparities in false-positive and true-positive rates, particularly across race groups. This study is the first to systematically integrate algorithmic fairness into HIV-1 prediction, demonstrating that EHR-based models can effectively and equitably identify underdiagnosed individuals. Our approach offers a scalable framework for improving HIV case finding and supporting equitable clinical decision-making in diverse healthcare settings. This is the work with my students, Tianheng Zhang and Yuxuan Gu.

Keywords: AIDS; EMR; Machine Learning Fairness

Back to Sessions List