Data Secure Transfer Learning from Heterogeneous Low Rank and Sparse Panel VAR Models

George Michailidis

University of California, Los Angeles

ABSTRACT

Transfer learning (TL) enhances high-dimensional inference by borrowing strength from auxiliary data sets. We extend TL to multivariate time series settings through a panel Vector Autoregression (VAR) model, whose coefficients decompose into heterogeneous low-rank and sparse components. The proposed algorithm treats the two components independently and automatically detects the transferable auxiliary sources for each. Importantly, it operates solely on *pre-trained estimators* from the *auxiliary* panel, thereby preserving data privacy. We establish that the source detection mechanism achieves high accuracy under mild technical conditions. The proposed TL approach delivers significant performance improvements in data-scarce regimes---precisely the scenarios that motivate transfer learning. Extensive numerical experiments based on synthetic data, conducted for both correctly specified and misspecified time series models, along with applications to macroeconomic time series, confirm the practical effectiveness of our method. The results highlight the potential of TL to enhance predictive tasks in complex multivariate time series models when primary data are limited.

Fast Segmentation of Watermarked Texts from Large Language Models through Epidemic Change-Points Framework

Soham Bonnerjee, Subhrajyoty Roy, **Sayar Karmakar**

University of Florida

ABSTRACT

With the increasing popularity of large language models, concerns over content authenticity have led to the development of myriad watermarking schemes. These schemes can be used to detect a machine-generated text via an appropriate key, while being imperceptible to readers with no such keys. The corresponding detection mechanisms usually take the form of statistical hypothesis testing for the existence of watermarks, spurring extensive research in this direction. However, the finer-grained problem of identifying which segments of a mixed-source text are actually watermarked, is much less explored; the existing approaches either lack scalability or theoretical guarantees robust to paraphrase and post-editing. In this work, we introduce a unique perspective to such watermark segmentation problems through the lens of epidemic change-points. By highlighting the similarities as well as differences of these two problems, we motivate and propose WISER: a novel, computationally efficient, watermark segmentation algorithm. We theoretically validate our algorithm by deriving finite sample error-bounds, and establishing its consistency in detecting multiple watermarked segments in a single text. Complementing these theoretical results, our extensive numerical experiments show that WISER outperforms state-of-the-art baseline methods, both in terms of computational speed as well as accuracy, on various benchmark datasets embedded with diverse watermarking schemes. Our theoretical and empirical findings establish WISER as an effective tool for watermark localization in most settings. It also shows how insights from a classical statistical problem can lead to a theoretically valid and computationally efficient solution of a modern and pertinent problem.

Keywords: Epidemic change, Large language model, Anomaly detection

Monitoring and Early Detection of Instability in Manufacturing Process Using Vector Autoregression Models

Yi-Ting Wang[1], Ying-Chao Hung[2]

Institute of Industrial Engineering, National Taiwan University

ABSTRACT

Instability in manufacturing processes (e.g., semiconductor wafer fabrication) can cause severe economic losses if not detected early. This talk presents a monitoring framework based on Vector Autoregression (VAR) models to detect and predict whether a process is drifting out of control. We first show how linear predictors can characterize stability and serve as monitoring statistics for online data streams. We then design monitoring strategies with rigorous control of the family-wise error rate (FWER), ensuring reliable inference in multivariate settings. An optimal strategy is introduced to balance early detection with false alarms by minimizing a weighted combination of FWER and Type II error. The framework also extends to cases with partial variable information and incorporates fault propagation and causal analysis under a Bayesian framework. Together, these methods form a practical toolkit for early detection and diagnosis of process instability, supporting more resilient and efficient manufacturing operations.

Keywords: Vector Autoregression; Process Monitoring; Fault Detection; Family-wise ErrorRate; Bayesian Causal Analysis