

Yingying Fan

Associate Dean of the PhD Program

Centennial Chair in Business Administration

Professor of Data Sciences and Operations

Department of Data Sciences and Operations, University of Southern California

"LLM-Powered Prediction Inference with Online Text Time Series"

Abstract: Time series prediction inference is an important yet challenging task in economics and business, where existing approaches often rely on low-frequency, survey-based data. With the recent advances of large language models (LLMs), there is growing potential to leverage high-frequency online text data for improved time series prediction, an area still largely unexplored. This paper proposes LLM-TS, an LLM-based approach for time series prediction inference incorporating online text data. The LLM-TS is based on a joint time series framework that combines survey-based low-frequency data with LLM-generated high-frequency surrogates. The framework relies only on an error correlation assumption, combining a text-embedding-augmented ARX model for the observed gold-standard measurements with a VARX model for the LLM-generated surrogates. LLM-TS employs LLMs such as ChatGPT and the trained BERT models to construct LLM surrogates. Online text embeddings are extracted via LDA and BERT. We establish the asymptotic properties of the method and provide two forms of constructed prediction intervals. To demonstrate the practical power of LLM-TS, we apply it to a critical real-world example: inflation forecast. We collect a large set of high-frequency online texts from a widely used Chinese social media platform and employ LLMs to construct inflation labels for posts that are related to inflation. The finite-sample performance and practical advantages of LLM-TS are illustrated through simulations and this noisy real data example, highlighting its potential to improve time series prediction in economic applications. This is a joint work with Jinchi Lv, Ao Sun and Yurou Wang.