The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Cross-Modal Knowledge Transfer in Time Series AI via Large Vision Models

Jingchao Ni

Department of Computer Science, University of Houston
jni7@uh.edu

Abstract

Time series analysis has progressed from traditional autoregressive models to deep learning, Transformers, and foundation models (FMs), including large language models (LLMs) and large vision models (LVMs). These advances expand model design possibilities and enable time series problem-solving across multiple modalities. This talk will provide an overview of recent developments in large FMs for time series, highlighting frameworks for transferring knowledge from other modalities to time series, and identifying the advantages of LVMs over LLMs in cross-modal knowledge transfer. I will then delve into our recent research on LVMs for time series, discussing (1) mainstream techniques for imaging time series; (2) key strengths and limitations of LVMs in time series modeling; and (3) multimodal frameworks that integrate LVMs for time series encoding. This talk will conclude with an application of LVMs to brain time series analysis in neuroscience. The aim of the talk is to review state-of-the-art (SOTA) AI techniques for time series, highlight unique challenges, and share our recent findings in this promising area.

Introduction

This research establishes **a well-structured pipeline** for understanding LVMs for time series, ranging from a taxonomy of existing techniques to our exclusive findings of LVMs' strengths and limitations, along with their extensions to multimodal learning and real-world applications. This research **lays the groundwork** for future developments of time series FMs that can **fuse multimodal knowledge** for powering representation learning. This talk will be structured in a progressive manner as: our recent survey (Ni et al. 2025) $\rightarrow$ our empirical findings (Zhao et al. 2025) $\rightarrow$ our multimodal framework (Shen et al. 2025). The talk starts with an **overview of time series AI**, covering (1) background information; (2) recent developments of large FMs for time series including LLMs, LVMs, time series FMs, and vision-language models (VLMs); (3) a high-level framework of cross-modal knowledge transfer; and (4) the advantages of LVMs over LLMs in time series modeling.

Techniques for Imaging Time Series

By reviewing existing works on vision models for time series, we have identified 8 major techniques for imaging time series (Ni et al. 2025), including (1) line plot; (2) univariate heatmap (UVH); (3) multivariate heatmap (MVH); (4) Gramian angular field (GAF); (5) short-time Fourier transform (STFT); (6) wavelet transform; (7) filterbank; and (8) recurrence plot (RP). Different techniques encode different temporal or frequency patterns in time series. This part will introduce and compare these methods in terms of their encoded data types, strengths and limitations, preparing for an empirical comparison that is introduced in the next part.

Are LVMs Useful for Time Series?

We performed **the first benchmark** for assessing LVM's effectiveness in time series modeling (Zhao et al. 2025). This work compared 4 typical LVMs, 8 imaging techniques, and 26 non-LVM baselines on 18 datasets across both high-level (*e.g.*, classification) and low-level (*e.g.*, forecasting) tasks, with extensive ablation analysis. We summarized the current best ways to adapt LVMs to time series, assessed the adapted LVMs in terms of pre-training, imaging, decoding, fine-tuning, architecture, temporal modeling, and computational costs. This part will highlight the key findings on LVMs' strengths and challenges when used for time series.

Multimodal View Enhanced LVMs

Integrating LVMs with models in other modalities is promising in enhancing time series encoding. However, optimal integration should accommodate individual modalities' biases. Our findings indicate that the SOTA LVM-based time series forecaster has a bias toward periodicity. Our **DMMV** model (Shen et al. 2025) introduces the first **adaptive integration** of LVMs and numerical forecasters, harnessing LVMs' bias and complementing it with numerical forecasters' strengths in trend modeling. This part will use DMMV as an example to showcase the **challenges**, **caveats**, and **insights** of integrating LVMs in multimodal learning farmeworks.

Applications and Future Directions

To illustrate the impact of LVM-based time series modeling in real applications, I will introduce our ongoing research of LVM-enhanced **brain foundation models** for encoding EEG and fMRI signals in **BCI and medical tasks**. Additionally, this part will conclude with our vision of future directions on large FMs-inspired **small time series models** in resource-constrained scenarios for **sustainable AI**.

References

Ni, J.; Zhao, Z.; Shen, C.; Tong, H.; Song, D.; Cheng, W.; Luo, D.; and Chen, H. 2025. Harnessing Vision Models for Time Series Analysis: A Survey. In *IJCAI*.

Shen, C.; Yu, W.; Zhao, Z.; Song, D.; Cheng, W.; Chen, H.; and Ni, J. 2025. Multi-Modal View Enhanced Large Vision Models for Long-Term Time Series Forecasting. In *NeurIPS*.

Zhao, Z.; Shen, C.; Tong, H.; Song, D.; Deng, Z.; Wen, Q.; and Ni, J. 2025. From Images to Signals: Are Large Vision Models Useful for Time Series Analysis? *arXiv preprint arXiv:2505.24030*.