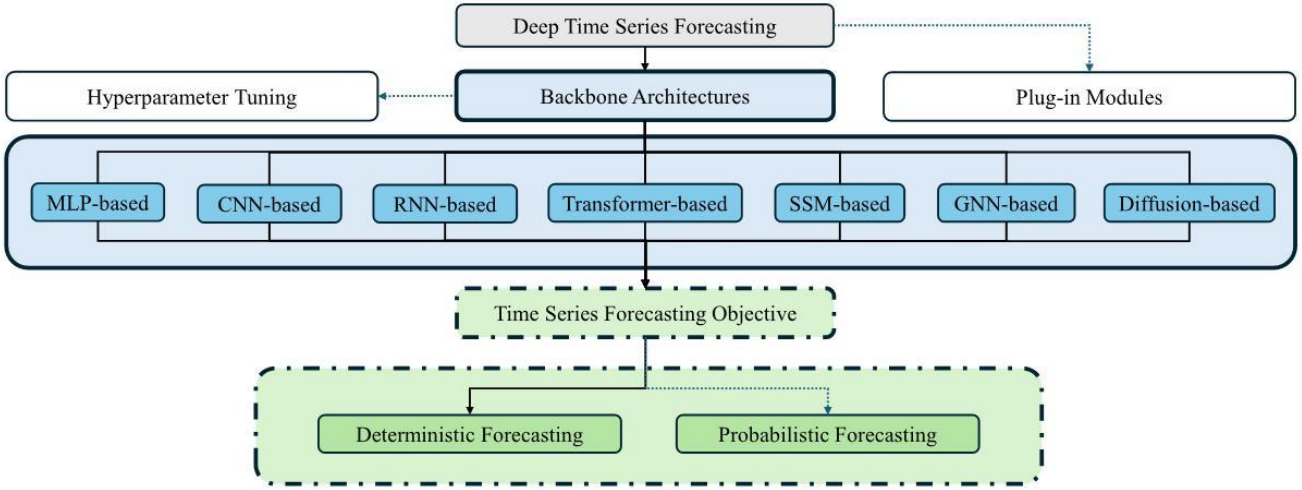
A Survey of Deep Time Series Forecasting Backbone Architectures: Progress, Pitfalls, and a Systematic Comparison

Xiang LI, Yanping ZHENG, Zhewei WEI

Frontiers of Computer Science, DOI: 10.1007/s11704-026-50462-z

Problems & Ideas

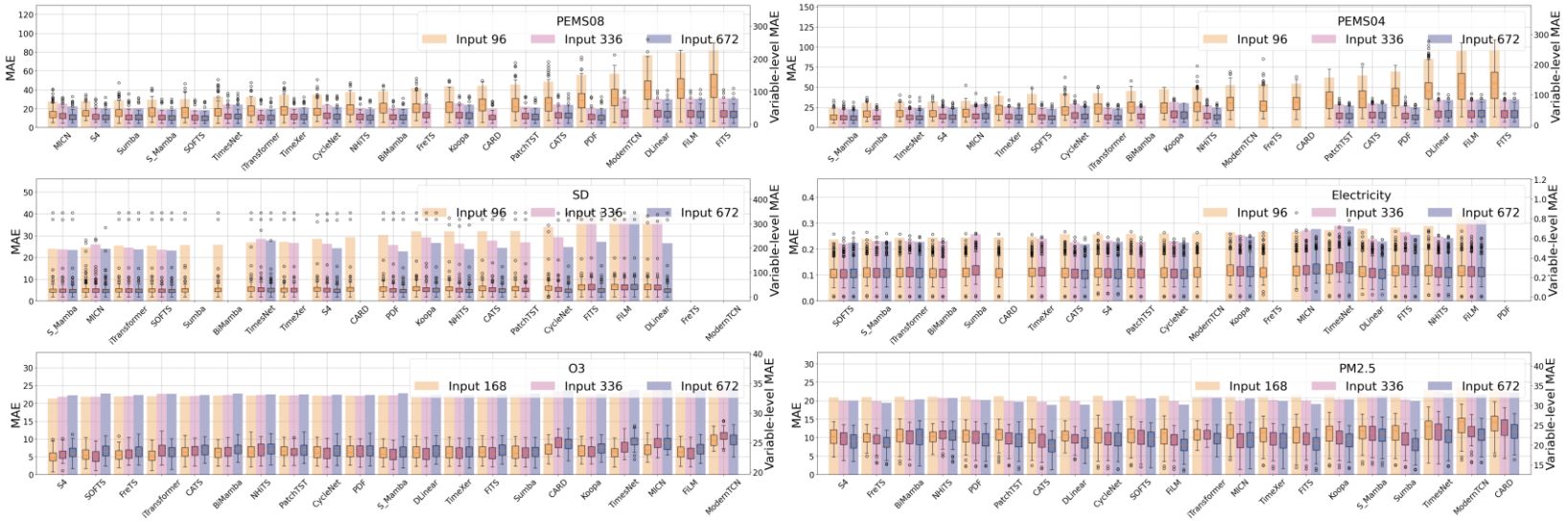
- Problems of deep time series forecasting tasks:
  - Various architectures have been proposed with similar modules, yet their performance remains sensitive to factors such as training hyperparameters and experimental settings.
  - Mainstream evaluation protocols rely on metrics averaged over the entire forecasting horizon, obscuring heterogeneity across variables and across different prediction ranges.
- Ideas: Conducting a survey supported by experimental results enables a deeper understanding of current developments and future research trends.



Although forecasting performance can also be affected by hyperparameter tuning, plug-in modules, and the choice of forecasting objective, this survey isolates the role of the backbone by evaluating all models under a unified deterministic objective. This controlled setting enables us to uncover behavioral differences that are obscured by standard averaging protocols.

Main Contributions

- Contributions:
  - A historical and evolutionary review of major backbone architectures.
  - A unified controlled benchmark that isolates architectural effects.
  - Fine-grained variable- and horizon-level analysis showing that
    - larger input windows often outperform architectural innovations
    - averaged metrics obscure heterogeneity
    - architectures differ substantially in short- and long-term forecasting performance.



Conventional overall MAE (bar) and variable-level MAE distributions (box). Overall MAE reflects the widely used mean-based evaluation protocol, whereas the variable-level distributions expose substantial heterogeneity.