Pre-trained Forecasting Models: Strong Zero-Shot Feature Extractors for Time Series Classification

Andreas Auer, PhD student @ JKU Linz, Researcher @ NXAI

Joint work with
Daniel Klotz, Sebastian Böck, Sepp Hochreiter

JOHANNES KEPLER
UNIVERSITY LINZ

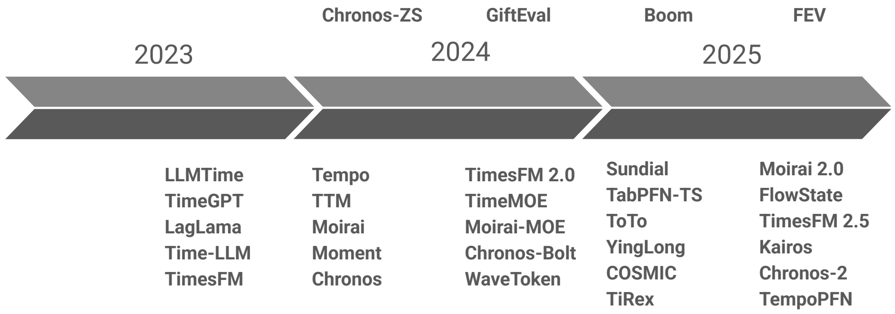
NXAI

The advent of Zero-Shot Time Series Forecasting Models

Pre-Trained Models for Forecasting

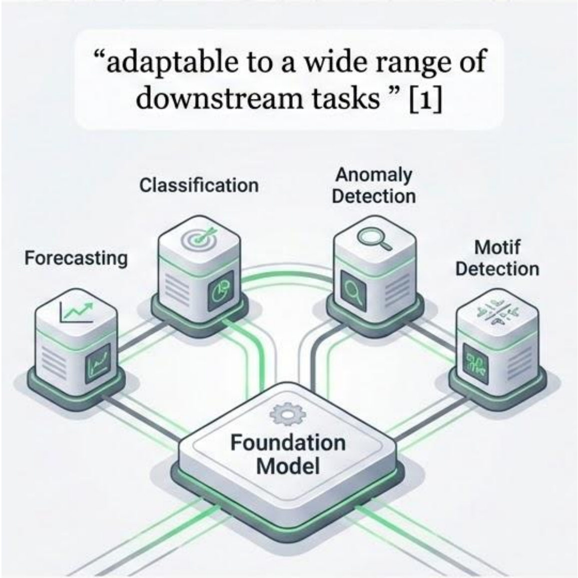
✅ Competitive forecasting in <u>zero-shot</u> mode

✅ Accessible with limited expertise and low operational overhead (think of agents)



Foundation Models ?

- Recent focus on forecasting for pre-trained time series models *
- Important tasks beyond forecasting in the time series domain
- Generalizable representations or downstream task specific pre-training ?



* Important work on multi-purpose models and pre-training for different downstream task does exist! [2,3,4,5,..]

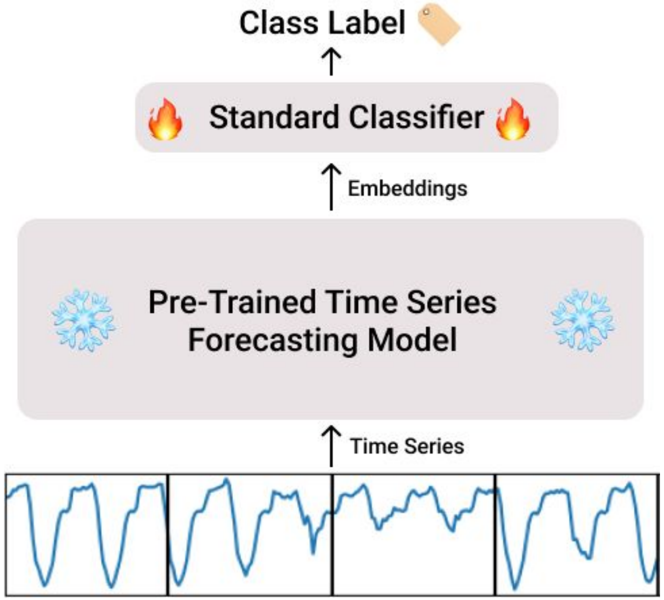
How well do representations from pretrained forecasting models transfer to classification tasks?

Method/Analysis Setup

1. Extract Embedding from pre-trained forecasting model
   - Analyse sequence/layer aggregation

2. Apply/Concatenate embedding augmentation
   - Absolute sample statistics
   - Time Series Differencing

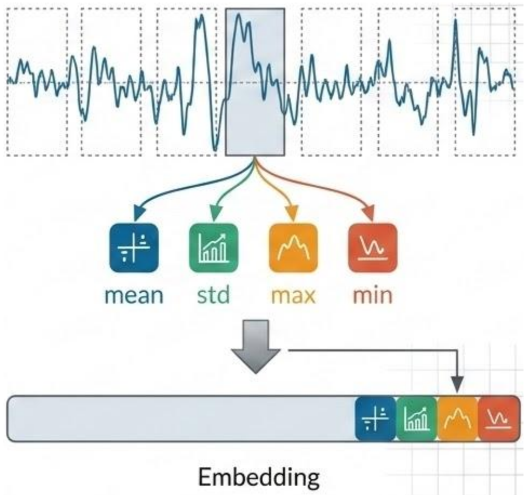
3. Fit classifier based on representations



Augmentation 1: Absolute Sample Statistics

- Why:
  Models use instance normalization, hence loose absolute value information

- How:
  - Divide into *k* patches
  - Signal features for each patch (*mean*, *std*, *max*, *min*)
  - Append features to embedding

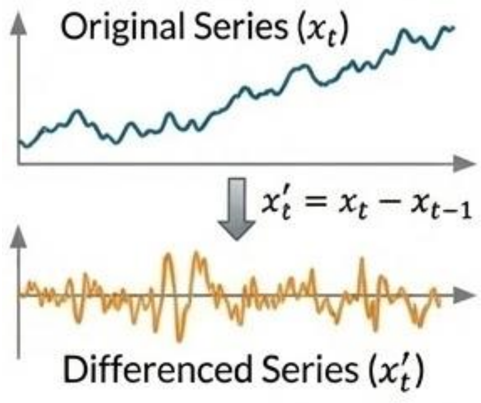


JMU | NXAI

Augmentation 2: Time Series Differencing

- Why:
  Trends might dominate the signal and overlap more subtle patterns

- How:
  - Difference time series: $x'_t = x_t - x_{t-1}$
  - Embed original and differenced series
  - Concatenate both embeddings



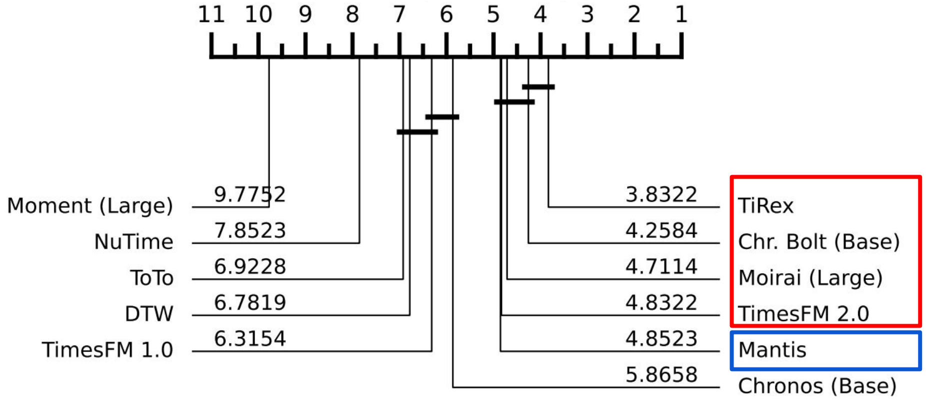
Experiment Setup

- Datasets: UCR[6] and UAE[7] classification benchmark
  ➪ zero-shot for forecasting models
- No training/fine-tuning of pre-trained model
- Selection of models covering different generations and SOTA:
  TiRex[8], Toto[9], Chronos[10], TimesFM[11], Moirai[12]
- Standard Classifier: *kNN, Linear, RandomForest*
- Compare to SOTA pre-trained classification models and
  multi-purpose models (*Mantis*[4], *NuTime*[3], *Moment*[2])

Results: Strong Performance of Forecasting Models

- Best forecasting models on-par/outperform best pre-trained classification models

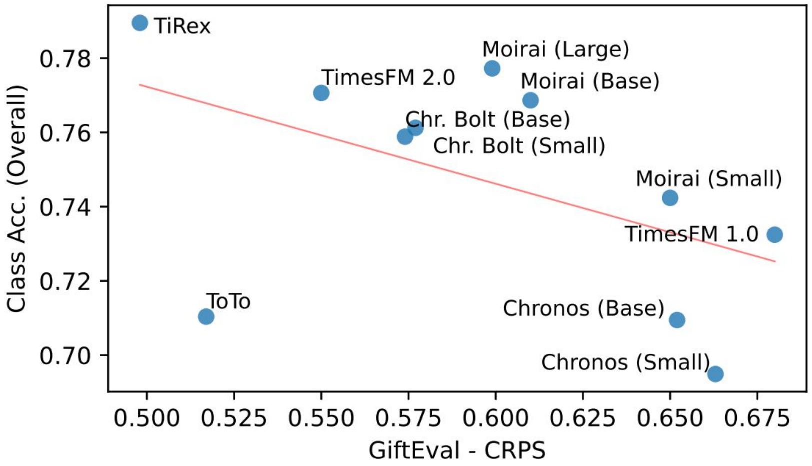
- Results robust across classifier, metrics and benchmark configurations



JKU | NXAI

Results: Forecast & Classification Performance Correlate

- Positive correlation between forecasting performance classification performance

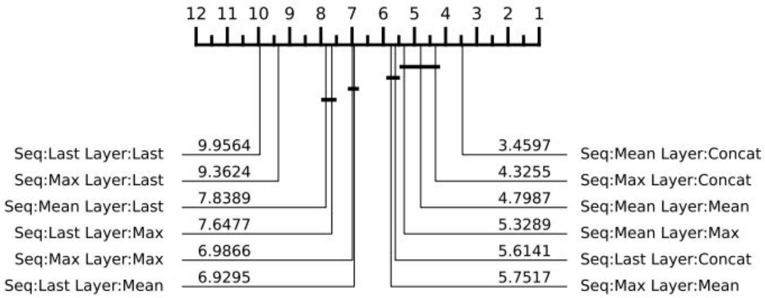


JKU | NXAI

Results: Architecture and Embedding Extraction

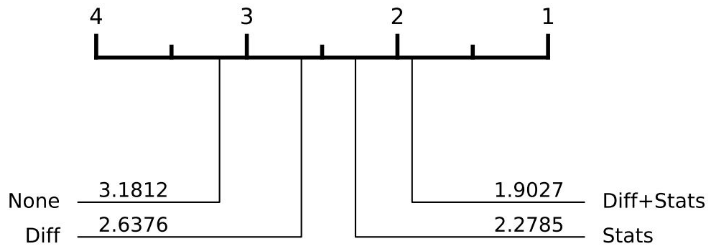
- No superior architecture paradigm
  (enc/dec-only, enc-dec)

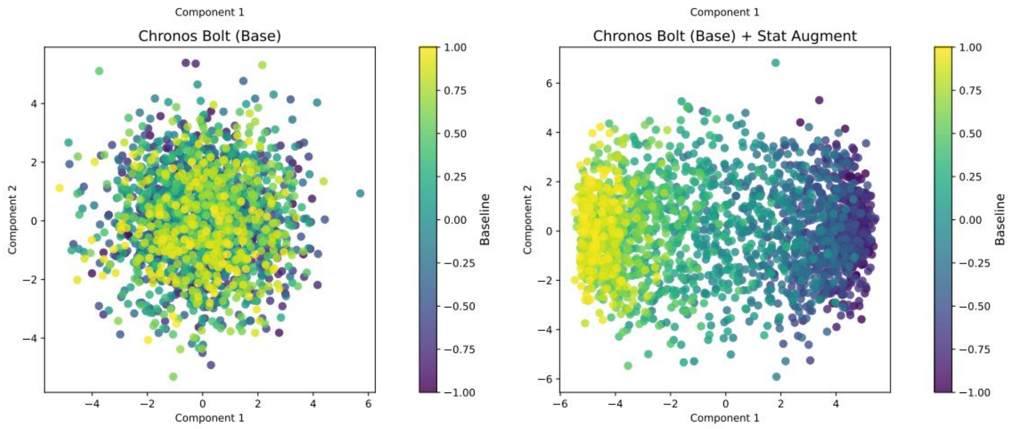
- Embedding aggregation along sequence and layers is important
  - Last layer likely to close to “forecast representation”
  - Final token does not retain everything important from sequence



Results: Augmentations are important

- Augmentations significantly improve the results across most models





JKU | NXAI

Conclusion

Pre-trained forecasting models are effective zero-shot feature extractors for classification

+

Classification and Forecasting performance are correlated positively

- Representation of forecasting models seem to generalize
- Forecasting as pre-training task might be a viable path for generating foundational time series models

- Augmentation and embedding aggregation important

Limitations & Future Work

- Limited amount of data points (= models)
  - □ Continuously analyse new models
- Beyond Zero-Shot ?
  - □ Analyse fine-tuning approaches
- Beyond Classification ?
  - □ Anomaly detection, Motif discovery

Paper & Socials

References

[1] Bommasani et al. "On the opportunities and risks of foundation models." arXiv:2108.07258 (2021)

[2] Goswami et al. “Moment: A family of open time-series foundation models”. ICML 2024

[3] Lin et al. “Nutime: Numerically multi-scaled embedding for large-scale time-series pretraining”. TMLR 2024

[4] Feofanov et al. “Mantis: Lightweight calibrated foundation model for user-friendly time series classification.” FMSD @ ICML 2025

[5] Gao et al. “UniTS: A unified multi-task time series model.” NeurIPS 2024

[6] Dau et al. “The ucr time series classification archive”. arXiv, 1810.07758 (2018)

[7] Bagnall et al. “The UEA multivariate time series classification archive”, arXiv, 1811.00075 (2018)

[8] Auer et al. “Tirex: Zero-shot forecasting across long and short horizons with with Enhanced In-Context Learning” NeurIPS 2025

[9]Cohen et al. “This time is different: An observability perspective on time series foundation models.” NeurIPS 2025

[10] Ansari et al. “Chronos: Learning the Language of Time Series.” TMLR 2024

[11] Das et al. “A decoder-only foundation model for time-series forecasting.” ICML 2024

[12] Woo et al. “Unified Training of Universal Time Series Forecasting Transformers.” ICML 2024