Master Thesis/Internship/Hiwi:

Learning Discrete Temporal Patterns for Time Series Forecasting

Modern time series forecasting methods, such as LSTMs and Transformers, typically operate directly on raw, continuous signals. While effective in many domains, these models often struggle with noisy, high-dimensional, or partially observed data—common challenges in real-world scenarios such as energy monitoring, finance, and scientific sensor networks. This project aims to investigate a novel two-stage approach for time series forecasting that first compresses continuous data into a discrete, symbolic representation before applying sequential learning.

Inspired by advances in sequence modeling and representation learning, this thesis explores methods such as Vector Quantized Variational Autoencoders (VQ-VAE) or similar discretization techniques to learn high-level temporal abstractions. These symbolic sequences are then modeled using transformer-based architectures or other sequence learners. This modular design not only improves forecasting performance but also enhances interpretability and robustness in the presence of noise or missing data.

Depending on the student's background and interests, the project may focus on one or more of the following directions:

- **Unsupervised Sequence Compression**: Learning symbolic representations of time series data using VQ-VAEs or similar methods.
- **Forecasting on Discrete Representations**: Applying transformer or autoregressive models to predict symbolic sequences and reconstruct future trajectories.
- **Benchmarking and Evaluation**: Comparing the approach against standard baselines (e.g., LSTM, N-BEATS, TCN) on publicly available forecasting datasets.
- **Robustness and Interpretability**: Analyzing the model's ability to generalize under noise, sparsity, or domain shifts.
- **(Optional)** Domain-specific Application: Applying the approach to challenging real-world datasets from domains such as energy systems, healthcare, or scientific instrumentation (e.g., KATRIN).

Required Skills

- Experience with Python and deep learning frameworks such as PyTorch or TensorFlow.
- Familiarity with time series analysis and sequence modeling.
- Understanding of autoencoders, attention mechanisms, or representation learning is a plus.

Duration & Collaboration

The project is expected to last six months and may involve collaboration with other students working on related forecasting or ML topics.

Contacts

📩 Dr. Nicholas Tan Jerome (Karlsruhe Institute of Technology)
✉ nicholas.tanjerome@kit.edu
📩 Dr. Suren Chilingaryan (Karlsruhe Institute of Technology)
✉ suren.chilingaryan@kit.edu