Master thesis

Topic: “Global models: Comparing the performance of global models on real-life datasets”

Prof. Dr. Frauke Schleer-van Gellecom
Honorary Professor – Decision Intelligence, Justus-Liebig University Gießen
Partner –PricewaterhouseCoopers GmbH WPG

Background: Traditionally, a model for time series forecasting is trained using the historical observations of a single time series and external variables to predict future values. For a so-called Local Model, an individual predictive function is fitted for each time series.
In contrast to Local Models, Global Models pool the historical observations of multiple time series to train only one predictive function that can be used to forecast all available time series. Global Models have gained popularity over the past years, winning competitions such as the prestigious M4 and M5 Forecasting Competitions. Global Models are scalable and - although needing to fit many different series - may achieve astoundingly good generalization abilities. In particular, they can leverage the information from other time series and therefore perform cross-learning. Models that can train on multiple related time series include Neural Networks (RNNs, DeepAR, 1D-CNNs) or Gradient Boosting Algorithms (LGBM, XGBoost).

While at the beginning the consensus was that the time series within the dataset for training Global Models should be similar to achieve the necessary generalization, nowaday practitioners and some researchers claim that even with heterogeneous training and evaluation datasets high accuracy can be achieved.
This leaves the question if a global model built on an overarching dataset performs better than global models and more homogeneous datasets which might also have more explaining variables available.

Research tasks/questions:

- You will be provided with a dataset of time series from real-life businesses.
- Optionally enhance the dataset with publicly available time series datasets.
- Build several global models using Python and different underlying models for both the full dataset as well as several smaller subsets.
- Analyze the effect of adding explanatory variables to the smaller subsets.
- Compare accuracy and run-time between the global models and benchmark local models (local model forecasts will be provided).

Supervisor: Prof. Dr. Frauke Schleer-van Gellecom. The master thesis will be in cooperation with PricewaterhouseCoopers GmbH WPG. You will be supported by a forecasting expert from PwC.

Start: Anytime

Prior knowledge / Requirements:

- Excellent base knowledge in statistics and machine learning models (e.g. Zeitreihenökonometrie und computergestützte Verfahre, Prof. Peter Winker, “Advanced Econometrics”, Prof. Dr. Peter Winker oder Data Science for Management, Prof. Dr. Nicolas Pröllochs)
- Coding experience in Python or similar language
- Basic knowledge about time series forecasting.

Literature:

- Pablo Montero-Manso, Rob J. Hyndman: Principles and algorithms for forecasting groups of time series: Locality and globality, International Journal of Forecasting, Volume 37, Issue 4, 2021 ,Pages 1632-1653, ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2021.03.004.
- Local vs Global Forecasting: What You Need to Know (https://towardsdatascience.com/local-vs-global-forecasting-what-you-need-to-know-1cc29e66cae0)