Student: &
Supervisor/s: &
Client: & Fraunhofer Institute for Integrated Circuits (IIS),
&
Thesis: & M.Sc.

Pitfalls of the Recurrent Neural Network Family

Keywords: Time Series, (S)ARIMA(X), (I)RNN, GRU, LSTM, Capacity, Consistency, Error Carusel, HCNN, ECNN, Forecasting, Stacked-, Dense-, Skipped- Architecture.

Motivation. The analysis of the applicability and the strategic utilization has shown that Recurrent Neural Networks (RNN), especially Historical-Consistent Neural Networks (HCNN), offer great potential for industrial, macroeconomic, sports, healthcare, and localization applications, as they provide higher forecast quality compared to state-of-the-art methods, such as Gradient Boosting [9], VAR, and ARIMA [4]. However, in the scientific discourse and in practical tests, three major challenges have emerged that have hindered the widespread and successful use of HCNN so far: (1) the optimal feature extraction, (2) the robustness w.r.t. uncertainty, capacity, and consistency of recurrent models, and (3) the comparison with state-of-the-art methods. Thus, this thesis will address these challenges.

(1) Architecture. The student will investigate feature extraction mechanisms that allow both, a flexible (variable length of the input sequence) input embedding and an optimal feature pre-processing to enable long-term dependencies and free memory capacity. Hence, the student will employ synthetic datasets to evaluate effects of different sequence lengths and dimensions on the model capacity of the context and hidden state vectors of RNNs. The student will examine the research question *"Which potentially conceivable influencing variables actually provide predictive added value?"*. Thus, to address this question, the student will investigate either feature selection procedures or manual selection, paired with domain knowledge. Furthermore, the student will investigate to what extent this idea can be generalized to LSTMs, ECNNs, and CRCNNs.

(2) Robustness. To investigate the robustness w.r.t. uncertainty and consitency, the student will compare different mechanics that s/he will adapt to various RNN methods: (2.1) creating a large ensemble of HCNNs and determining the uncertainty about the spread of the results, (2.2) repeated sampling of the forecast values from a Gaussian distribution or (2.3) additional Monte Carlo dropouts. It is also unclear to what extent the errors in the models are also related to the errors in the forecasts. Hence, there is a need to analyze whether the spread of RNN, especially HCNN, ensembles is suitable to clarify the uncertainty of the prognosis [6]. In addition, the student will investigate how the quantification of the uncertainty contributes to the explainability of the models [3].

(3) Evaluation. To show the potential of RNNs, especially HCNNs, in (macroeconomic) applications, it is necessary to achieve significantly better solutions than current state-of-the-art methods (LSTMs, XGBoost). Thus, the student will perform a comprehensive benchmark with scientific rigor on several relevant datasets and state-of-the-art methods, such as VAR, (S)ARIMA(X), Prophet, etc. The student will share these findings with the research community in a scientific publication.

Overall goals. In this qualification thesis, the limits of modern data-driven regression methods for analysis and prediction on time series data are to be examined. Specifically:

- (1) (I)RNN, ECN, Elman, GRU, LSTM, and HCNN cells will be implemented;
- (2) Suitable public datasets for evaluating different criteria, e.g., capacity, sequence length, sampling rate, data consistency, frequency, information complexity, short- and long-time dependencies, and input invariance will be generated;
- (3) The methods will be evaluated against the state-of-the-art, e.g., VAR, (S)ARIMA(X), n-Beats, Prophet, Transformer, TCN, ResNet, etc. ..;
- (4) An optimal a priori feature extraction mechnism will be investigated that is robust against dynamic sequence length with gaps, delays, and varying length;

- (5) State-of-the-art uncertainty estimation mechnisms will be adapted to the models and evaluated thereof;
- (6) An academic paper will be written.

Timetable (6 months, in person weeks [PW]).

4 PW Literature and patent research; Familiarization with relevant work on the subject areas;

10 PW Methodological work: adaptation of the individual components to the state-of-the-art methods and advances to the state-of-the-art based on recent deep learning methods;

4 PW Evaluation and real-world demonstration;

6 PW Transcript.

Expected results and scientific contributions.

- RNN model with specific cells (e.g., HCNN) and uncertainty estimation mechanism;
- Benchmark results of state-of-the-art and the novel models on prominent datasets;
- Scientific publication and GitHub publication.

References

[1] Zimmermann H. G., Tietz C., Grothmann R. (2012). "Forecasting with Recurrent Neural Networks: 12 Tricks". In: Neural Networks: Tricks of the Trade: Second Edition, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 687–707, 10.1007/978-3-642-35289-83_7

[2] Ioannides R.T., Pany T., Gibbons G. (2016). Known Vulnerabilities of Global Navigation Satellite Systems, Status, and Potential Mitigation Techniques. Proc. IEEE. 104:1174–1194.

[3] Dominik Seuß. (2021). Bridging the Gap Between Explainable AI and Uncertainty Quantification to Enhance Trustability. arXiv:2105.11828 [cs.AI]

[4] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1), 54-74.

[5] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780. Chicago

[6] Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. arXiv preprint arXiv:1612.01474.

[7] Gal, Y., & Ghahramani, Z. (2015). On modern deep learning and variational inference. In Advances in Approximate Bayesian Inference workshop, NIPS (Vol. 2).

[8] Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, B., Maddix, D., Turkmen, C., ... & Januschowski, T. (2020). Neural forecasting: Introduction and literature overview. arXiv preprint arXiv:2004.10240.

[9] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.