MLJ Contribution Information Sheet for *AALF: Almost Always Linear Forecasting*

Matthias Jakobs[1,2], Thomas Liebig[1,2]

1. Lamarr Institute for Machine Learning and Artificial Intelligence, Dortmund, Germany
2. TU Dortmund University, Dortmund, Germany

What is the main claim of the paper? In this work we investigate whether or not we can improve the interpretability of online model selection methods for time-series forecasting by restricting the set of models to choose between just two models: One high performance, black-box neural network and one interpretable, simple linear model. We chose to investigate these two model families for the reason that both have been found to perform very well in the past, with linear models in the form of for example ARIMA still being a strong baseline that deep learning methods have to surpass.

We claim that it is possible to learn a model selector in such a way that a trade-off between high performance (i.e., more selection of the deep learning model) and high interpretability (i.e., more selection of the linear model) is possible. More specifically, a practitioner can control this trade-off via a intuitive hyperparameter.

What is the evidence provided to support claims? We conduct three experiments in total to explore the claim. First, we investigate the scenario where we have full knowledge about the error both models would make should they be chosen. That way we are able to show that (in the best case scenario) we can outperform the deep learning model on the used datasets by just selecting it about 10 to 15 percent of the time, which significantly increases the overall interpretability. Next, we investigate how good we are able to fit a classifier, acting as a selector, on held-out validation data that has been labeled using the previously mentioned full-knowledge paradigm. We are able to show that we can find satisfactory fits, which we test (in our last experiment) against state-of-the-art model selection methods. There, we find that we can achieve comparable performance to these state-of-the-art methods (depending on the selected hyperparameter) while also being able to provide an overall more interpretable forecast compared to the other methods.

Most closely related contributions

- **ADE** [1] tries to learn the expected error for each model on unseen data based on measured validation error. Afterwards, a weighting scheme is proposed, assigning each model a weight which can be used for ensembling.
- **DETS** [2] is an ensembling method that selects a subset of models for prediction based on recent errors. These errors are computed over a window of past predictions for each model, smoothed and aggregated into a weighting for ensembling.

By choosing the model with the highest weight these two methods can be "converted" to single model selection methods. However, none of those assume some knowledge about the individual models (such as that one might be more preferable over the other for interpretability reasons).

- Smyl [3] combined an interpretable time-series forecasting algorithm, namely Exponential Smoothing (ES), with a Deep Neural Network, more specifically a Long-Short Term Memory (LSTM) network. However, the combination was done sequentially, i.e., ES was used to predict first and the LSTM was used to "correct" the predictions by utilizing the residuals. This approach is thus strictly aiming to increase the prediction performance and not to enhance interpretability.

References

[1] Vitor Cerqueira et al. “Arbitrated Ensemble for Time Series Forecasting”. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Michelangelo Ceci et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 478–494. ISBN: 978-3-319-71246-8.

[2] Vitor Cerqueira et al. “Dynamic and Heterogeneous Ensembles for Time Series Forecasting”. In: *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. Oct. 2017, pp. 242–251. (Visited on 02/21/2024).

[3] Slawek Smyl. “A Hybrid Method of Exponential Smoothing and Recurrent Neural Networks for Time Series Forecasting”. In: *International Journal of Forecasting*. M4 Competition 36.1 (Jan. 2020), pp. 75–85. ISSN: 0169-2070. DOI: `10.1016/j.ijforecast.2019.03.017`. (Visited on 02/26/2024).