Poster: Boosting Interpretability of Non-Readable Deep Learning Forecasts: the Case of Buildings' Energy Consumptions Prediction

Reza Samimi
samimi01@ads.uni-passau.de
University of Passau
Passau, Germany

Ammar Alyousef
Daniele Baranzini
baranzini.daniele@hsr.it
IRCCS Ospedale San Raffaele
Milan, Italy

Hermann De Meer
hermann.demeer@uni-passau.de
University of Passau
Passau, Germany

Abstract

It is important in the energy management of a building that energy consumption forecasts made by neural networks (referred to as black boxes) are backed up by consistent explanations from the model itself. Although the existing interpretable methods provide helpful information, it is not practical enough for energy managers. Expressly, the managers are not provided with an explanation for a certain period in the forecasted time series of energy consumption. We cover this lack of explanation by proposing a novel interpretability use case: explaining the shapelet of a period's forecast based on similar patterns in the past energy consumption profile, which our forecasting model can verify. Another interpretability use case is presented to explain better the electricity consumption forecast: determining the importance of each exogenous variable in the prediction problem. Temporal Fusion Transformers (TFT), a state-of-the-art, interpretable, and accurate forecasting model is employed to address the interpretability use cases via analyzing the distribution of attention weights. The results of applying the use cases on our dataset are demonstrated.

Keywords: AI interpretability, AI explainability, energy consumption forecasting, time series forecasting

ACM Reference Format:
Reza Samimi, Ammar Alyousef, Daniele Baranzini, and Hermann De Meer. 2022. Poster: Boosting Interpretability of Non-Readable Deep Learning Forecasts: the Case of Buildings' Energy Consumptions Prediction. In *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22), June 28–July 1, 2022, Virtual Event, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3538637.3538754

e-Energy '22, June 28–July 1, 2022, Virtual Event, USA

ACM ISBN 978-1-4503-9397-3/22/06.
https://doi.org/10.1145/3538637.3538754

1 Introduction

Predicting a building's energy demand (consumption) ahead of time is critical in minimizing energy supply errors. In case of over-supply (surplus production), the exceeding generated electricity can not be easily stored, re-directed, or sold to a wider grid. On the opposite, under-supply may lead to critical energy shortages. It follows that if a machine learning model is being employed to forecast energy consumption, it must deliver both interpretable forecasts (e.g., what would explain a certain trend or peak) and small prediction errors to avoid critical and costly consequences. Therefore, at the time of making energy-related decisions, building energy managers appreciate being able to justify why a forecasting model has made its specific predictions.

Most of the advanced forecasting models available today are black-boxes (i.e., it is not clear how the models process the input and make the prediction) [4]. Interpretable models, on the other hand, offer interpretability use cases that are mostly impractical to energy managers [1]. These models only demonstrate the important periods (e.g., in attention-based models) and input features or the interpretable components (e.g., in classic and hybrid models) to predict the time-series data [3][2]. Although this information is helpful, it is too generic; in practice, energy managers can benefit from knowing which specific periods in past time series had been likely responsible for (can explain) predicting the desired period in the future. We bridge this gap by using Temporal Fusion Transformer (TFT) [3], which is a deep learning model for time series forecasting based on transformer architecture. With the help of its attention mechanism, TFT can shed light on how it makes its forecasts.

1.1 Contributions

The main contributions of this work are:

- A novel interpretability use case to explain the shapelet of a period's forecast based on similar patterns in the past energy consumption profile. Moreover, the explanation is verified by the TFT.
- Another practical interpretability use case to determine the importance of each exogenous variable in the forecasts made by TFT model.

2 Interpretability Use Cases

2.1 Justifying Forecast's Shapelet

The model used for energy consumption forecasting (TFT) is data-driven, i.e., it makes predictions into the future based on the underlying patterns that it has learned from past observations. Therefore, it makes sense to justify the model's forecasts for a period based on what the model has learned from the period's past time series. If there is a specific "temporal forward window" for which the model has forecasted an energy consumption profile, the forecasted pattern is called a shapelet or simply the query. To explain why the model has forecasted the query and its shape, one can look backward to find similar patterns to the query across preceding temporal windows in the training time series segment. Every similar shapelet/pattern found in the past predicted time series is called a match. The next step is to verify that the forecasting of the query and its matches follow the same pattern that the model has learned. The TFT model is used to approve that the query's pattern is based on its matches' pattern by checking that the query and its matches have a similar self-attention weights distribution. Figure 1-(a) shows an example of this use case.

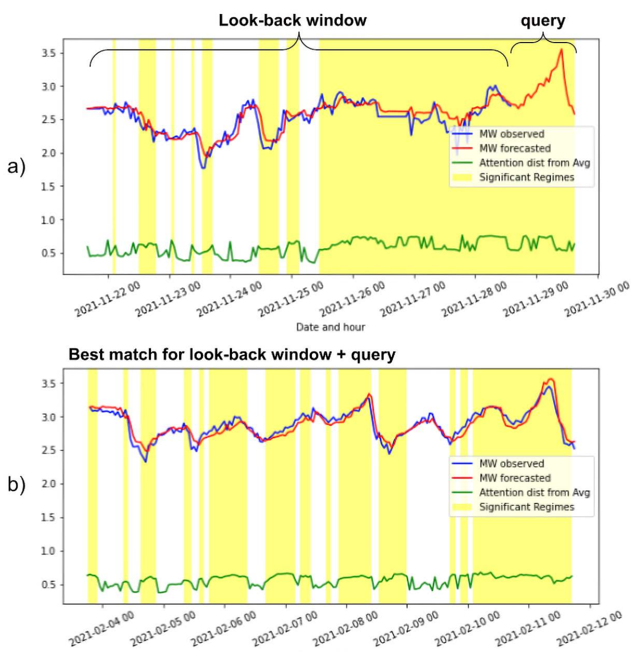

Figure 1. Interpreting forward predictions with TFT, (a) the query is a high peak which energy managers are interested in having an explanation for (b) the best-found match in the search for "similar shapelets" to query and its look-back window. Both the query and the match are identified as significant regimes (have a similar distribution of attention weights). Therefore, they have the same learned shapelet.

2.2 Determining Variable Importance

This use case determines the extent to which an exogenous variable is important to predict the energy consumption of a particular period using the TFT model. TFT utilizes a "Variable Selection Network," which provides the model with instance-wise insights into the significance of each variable for the predictions. The variable selection network weights distribution is analyzed for each of the mentioned covariate types. This way, the relative importance of each variable in the forecasting problem can be measured. The importance of variables of our dataset[1] is presented in table 1.

Water Flow	EXT Temp	Week Day	MW
0.280	0.047	0.082	0.589

Table 1. Global importance of variables in the forecasting of the test data, based on the variable selection weights.

3 Conclusion

In this work, two use cases are presented to interpret the forecasted electricity consumption of a building by employing the TFT model. The first and novel use case explains the shapelet of a period's forecast (query) based on similar patterns (matches) in its past time series. By observing the analogy between the query and its matches in terms of the distribution of TFT's attention weights, one can verify that the query and its matches are predicted based on the same learned shapelet by the model. The second use case analyzes the importance of each exogenous variable in a period's forecast.

3.1 Future Work

Future research could compare the forecast explanations of TFT and other types of intrinsically interpretable models, e.g., SARIMAX (classical), NeuralProphet (hybrid), and LightGBM (decision tree-based). Extending the dataset with other exogenous variables like occupant presence and studying their importance in the energy consumption forecasts is suggested.

References

[1] Oreshkin, B.N., Carpov, D., Chapados, N. and Bengio, Y., 2019. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv preprint arXiv:1905.10437.

[2] Taylor, S.J. and Letham, B., 2018. Forecasting at scale. The American Statistician, 72(1), pp.37-45.

[3] Lim, B., Arik, S.O., Loeff, N. and Pfister, T., 2019. Temporal fusion transformers for interpretable multi-horizon time series forecasting. arXiv preprint arXiv:1912.09363.

[4] Hochreiter, S., Schmidhuber, J"urgen, 1997. Long short-term memory. Neural computation, 9(8), pp.1735–1780.

[1] The dataset is collected under a "Horizon 2020" EU-funded project from a targeted building in a hospital in Milan, Italy.