Impact of data availability and probabilistic feature engineering on predicting load peaks

Sektorkopplung und Flexibilität

Mario Ernst[[1]](#footnote-1)(1), Luisa Brüser(1), Hans Auer(2)

(1)Verbund AG, (2)TU Wien

Motivation

Due to particular incentive systems in some countries (e.g. Germany), special forms of peak shaving on distribution grid level can be used as additional use-cases for large-scale batteries. To combine this form of peak shaving with intraday trading and Frequency Containment Reserve (FCR), a forecast-based operation strategy with quarter-hourly sensitivity is required. However, this is often made more difficult due to poor short-term grid load data availability and quality.

This paper deals with the problem of peak sensitive forecasting under the circumstance of not usable short-term grid load data by combining probabilistic and machine learning (ML) methods. Therefore, the following question is addressed:

*“How can probabilistic feature engineering and extraction improve ML forecasts for peak shaving?”*

Methodology

State of the art short-term grid load forecasting methods are machine learning based [1]. However, the use of forecasts for these special use-cases, more specific metrics than mean absolute percentage error (MAPE) for evaluating forecasting quality are required. The particular investigated battery use-case requires feeding in at the highest peak per year.

Five years of historical grid load data from three different German DSOs, Day Ahead and Intraday Auction prices as well as local solar radiation and temperature data are used to compare the ability of different univariate forecasting methods to hit daily peak loads.

The baseline scenario assumes that grid load data are available at all times and in sufficient quality.

In three other scenarios the effect of a grid load data availability delayed by one day, by one week and by one year is investigated on different forecasting algorithms. Manually extracted, probabilistic-based features are added to the dataset for developing forecasting strategies with limited data availability [2].

To measure the forecast ability for hitting high-load daily peaks, test peaks above a predefined threshold are identified on a monthly base. Assuming a battery operation with a one-hour-capacity[[2]](#footnote-2) battery, the hit-rate is calculated for each grid area on a yearly basis.

Results and Discussion

As battery scheduling requires the availability of load level data, a regression forecast followed by peak classification is done instead of a singular peak classification forecast due to highly imbalanced data. First results show that the addition of probabilistic features in the process of widely used ML models like XGBoost outperforms the prediction results without feature engineering. The best results can be obtained by using statistical peak times combined with a load forecast.

Literatur

[1] Andriopoulos, N.; Magklaras, A.; Birbas, A.; Papalexopoulos, A.; Valouxis, C.; Daskalaki, S.; Birbas, M.; Housos, E.; Papaioannou, G.P. Short Term Electric Load Forecasting Based on Data Transformation and Statistical Machine Learning. Appl. Sci. 2021, 11, 158.

[2] Maryasin, O. Y.; Lukashov, A. I.; Analyzing and Forecasting Peak Load Hours. 2021 Internal Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). IEEE 2021

1. Lustgasse 16/26 1030 Wien, 0043 664 8285481, mario.ernst@verbund.com [↑](#footnote-ref-1)
2. Battery capacity allows discharge with maximum power for one hour [↑](#footnote-ref-2)