Description of Grid Loss

A power grid transports the electricity from power producers to the consumers. But all that is produced is not delivered to the customers. Some parts of it are lost in either transmission or distribution. In Norway, the grid companies are responsible for reporting this grid loss to the institutes responsible for national transmission networks. They have to nominate the expected loss day ahead to the market so that the electricity price can be decided.

The physics of grid losses are well understood and can be calculated quite accurately given the grid configuration. Still, as these are not known or change all the time, calculating grid losses is not straightforward.

Content

Grid loss is directly correlated with the total amount of power in the grid, which is also known as the grid load.

We provide data for three different grids from Norway that are owned by Tensio (Previously Trønderenergi Nett).

Features:
In this dataset, we provide the hourly values of all the features we found relevant for predicting the grid loss.

For each of the grids, we have:

1. Grid loss: historical measurements of grid loss in MWh
2. Grid load: historical measurements of grid load in MWh
3. Temperature forecast in Kelvin
4. Predictions using the Prophet model in MWh
5. Trend, daily, weekly, and yearly components of the grid loss, also from the Prophet Model.

Other than these grid-specific features, we provide:

1. Calendar features: year, season, month, week, weekday, hour, in the cyclic form (see Notes 1.) and whether it is a holiday or not.
2. Incorrect data: whether the data was marked incorrect by the experts, in retrospect. We recommend removing this data before training your model.
3. Estimated demand in Trondheim: predicted demand for electricity in Trondheim, a big city in the middle of Norway, in MWh (see Note 2.)

We have split the dataset into two parts: training and testing set.

Training set:
This file (train.csv) contains two years of data (December 2017 to November 2019). All the features mentioned above are provided for this duration.

Test set:
This file (test.csv) contains six months of data (December 2019 to May 2020). All the features from training data are provided for the test set as well. Occasionally, some of the features could be missing.

Additionally, we provide a copy of the test dataset (test_backfilled_missing_features.csv) where the missing features are backfilled.

Note:

1. Calendar features are cyclic. If we encode the weekdays (Monday to Sunday) as 0 to 6, we find that while Sunday and Monday are next to each other, the distance between their embeddings does not reflect it. To reflect this cyclic nature of the calendar features, we created cyclic calendar features based on cosine and sine which together place the highest and lowest value of the features close to each other in the feature space.

2. We don't have an estimate of demand for all the grids. We used the demand predictions for Trondheim, the biggest city closest to the three grids.

3. The grid load is directly proportional to the grid loss. While we don't have predictions for grid load, since we have historical measurements for them, it makes sense to predict it and use it as a feature for predicting the grid loss.

4. While the Prophet Model did not perform nicely as a prediction tool for our dataset, we found it useful to include its prediction and other components as features in our model.

5. Grid 3 has less training data than Grid 1 and Grid 2.

6. We published our solution. For more details, please refer to:

Dalal, N., Mølnå, M., Herrem, M., Røen, M., & Gundersen, O. E. (2020). Day-Ahead Forecasting of Losses in the Distribution Network. In AAAI (pp. 13148-13155).

Bibtex format for citation:

@incollection{dalal2020a,
author = {Dalal, N. and Mølnå, M. and Herrem, M. and Røen, M. and Gundersen, O.E.},
date = {2020},
title = {Day-Ahead Forecasting of Losses in the Distribution Network},
pages = {13148–13155},
language = {en},
booktitle = {AAAI}
}

Challenges

Working with clean and processed data often hides the complexity of running the model in deployment. Some of the challenges we had while predicting grid loss in deployment are:

1. Day-ahead predictions: We need to predict the grid loss for the next day before 10 am the current day at an hourly resolution i.e on 10:00 May 26, 2020, we need to predict the grid loss on May 27, 2020, from midnight to 23:00 May 28, 2020, at an hourly resolution (24 values) for each grid.

2. Delayed measurements: We don't receive the measured values of load and loss immediately. We receive them 5 days after. Sometimes, there can be additional delays for a few more days. While grid loss and load are provided for the test data set as well, DO NOT USE them as features, unless they are 6 days old i.e. while predicting grid loss for 27th January 2020, you can use the grid loss values will 20th January 2020. Using grid loss or grid load data after that date is unfair and will be discarded.

3. Missing data: Sometimes, we don't receive some of the features. For example, whether the client might be out of service. You should make sure that your model should work even when some features are unavailable/missing.

4. Incorrect data: There have been times when the measurements we received were incorrect, by a big margin. They have been marked in the dataset in the incorrect_data column. It is recommended to remove those data points before you start analyzing the data.

5. Less training data: For one of the grids, grid 3, we only have a few months of data.

6. Changes in the grids: Grid structures can keep changing. Sometimes new big consumers are added, or small grids can be merged into big ones.