Udacity Machine Learning Nanodegree - Capstone Project Proposal

Fritz Arnold

October 4, 2017

1 Domain Background

On the liberalized German energy market electricity can be traded on specialized energy exchanges. The primary organized electricity trading scheme is the so called *Day-Ahead market*. Trades on the Day-Ahead market contract the obligation to deliver or accept a certain quantity of energy during an one hour block of the next day. Contracts are determined in an anonymous auction: Market participants place bidding orders until 12:00 noon the day before delivery. Orders are defined by the amount of energy to be bought (sold) together with the maximum (minimum) accepted buying (selling) price and the hour of delivery. When the bidding period ends, a *Market Clearing Price* (MCP), also denoted as *uniform price*, is determined by the so called *Merit Order Model*. The market clearing price is applied to all orders that are accepted by the merit order, independent of the offered quantity or original bidding price. Due to the anonymous nature of the bidding process, Day-Ahead prices are unknown prior to market closing; orders of other market players are hidden until the auction process is finished and all accepted orders become binding. In order to optimize their bidding strategy, market participants are interested in estimating the market clearing price for all 24 hourly blocks of the next day. Precise knowledge of Day-Ahead prices is crucial for the economic success of a market actor.

For an in-depth review about common electricity market mechanisms, the motivation to forecast electricity prices and recent approaches and algorithms, the reader is referred to [6].

2 Problem Statement

The problem to be solved in this project is to forecast the hourly Day-Ahead market clearing price of every hour of the next day for the German bidding zone on the European Power Exchange (EPEXSPOT). The predictions for the day d have to be conducted before market closing time, in this case 12:00 noon of day $d-1$. The temporal dependencies are visualized in Figure 1. The actual prediction has to rely solely on information that is available at that point of time to ensure applicability in a real world deployment.

3 Dataset and Inputs

Electricity prices are determined by (expected) supply and demand. Electricity demand generally has a strong time dependency, with peaks on evenings and weekends and lows during the night and weekly working hours. The supply side is characterized by the electricity generation mix, i.e. coal and nuclear thermal power plants, hydropower plants, and renewable energy generation (mainly wind and solar energy). Additionally, external factors like the weather might influence supply and demand. In order include all possible input factors, the following primary data will be used as an input for the forecasting model:

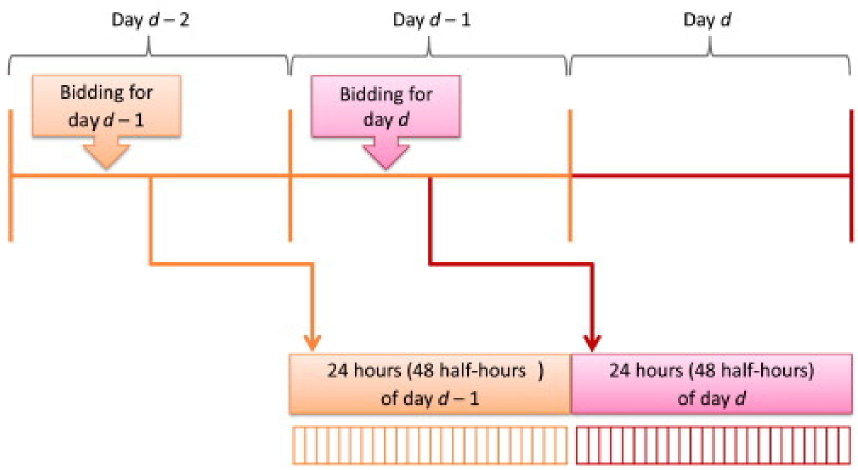

Figure 1: The bidding schedule on the Day-Ahead electricity market. Taken from [6].

- Historical hourly Day-Ahead market clearing prices for the German bidding zone. These prices can be used to model short- and long-term temporal patterns in electricity prices. Historical prices can be obtained from the EPEX SPOT website [3].
- Generation forecasts for the next day. Different electricity generation technologies have different generation costs. The respective share of generation technologies should have a major impact on the price level. Historical Day-Ahead generation forecasts can be obtained from the transparency platform of the European Network of Transmission System Operators (ENTSO-E) [2].
- Load forecasts for the next day. The load is equivalent to the electricity demand and therefore influences the electricity price. Historical load forecast data can be obtained from the ENTSO-E website [2].
- Historical load and generation data. Short- or long-term patterns in the energy generation mix might have an influence on the characteristics of supply and demand and thus the price. Historical load and generation data can be obtained from the ENTSO-E website [2].

Although all of the above information could potentially be useful for price forecasting, not all of it is available before market closing. According to European jurisdiction [4], load forecast data has to be published two hours before the Day-Ahead market gate closure and is therefore available for the prediction. Generation forecasts have to be published until 18:00 Brussels time and therefore might not be available for the prediction.

However, especially wind and solar power generation are strongly dependent on weather conditions. A model might incorporate weather variables (solar radiation, temperature, wind speed etc.) to give an indication about potential renewable generation levels. The following weather data will be used:

- Historical geo-spatial grid weather data of air temperature, precipitation, direct and diffuse solar radiation, wind speed and wind direction in Germany. The data is available via the website of the German National Meteorological Service (DWD) in hourly resolution for a 5 by 5 km grid [1].

4 Solution Statement

The solution of the problem is an hourly forecast of the Day-Ahead market clearing price for the German bidding zone in EUR/MWh at the EPEX SPOT for the upcoming day.

5 Benchmark Model

The simplest benchmark model would be to use the Day-Ahead prices of day $d-1$ as a forecast for the prices of day d. Due to the strong temporal patterns electricity prices this benchmark should yield a relatively good estimation. A more advanced option would be to take average historical prices for every individual hour of the day as a forecast, for example the average hourly prices of the last 30 days prior to day d.

6 Evaluation Metrics

The evaluation of the forecasts will be based on a mean absolute error between predicted and actual prices. Let $\hat{P}_{d,t}$ be the predicted market clearing price for hour t of day d and $P_{d,t}$ the actual market clearing price. The Mean Absolute Error of the prediction can be defined as:

$$MAE = \frac{1}{24}\sum_{t=1}^{24}|\hat{P}_{d,t} - P_{d,t}| \tag{1}$$

7 Project Design

The project will be subdivided into three major and one optional step, each of which is evaluated for its potential to improve the overall solution.

The first part will rely on pure time series prediction. The developed model should be able to predict the long- and short term temporal patterns present in Day-Ahead electricity prices. In recent literature, a combination of Convolutional and Recurrent Neural Networks has been proposed for short- and long-term pattern detection in time series prediction [5]. This idea will be used as a basis for the time series approach. The data inputs in this step will be historical electricity prices. This data most likely has to be preprocessed (scaling, outlier removal) in order to be a valid input for the model. A time series model relies on the input data to be a rolling window of historical values that are used to predict (a window of) future values. The input data has to be reshaped in order to fit that structure. The goal of that step is to beat the benchmark model which is a kind of 'naive' time series model on its own.

In a second step, the time series data will be enriched with historical load and generation data. This data might include patterns of shifting demand and supply (especially trends in the energy generation technology mix). A first idea on how to incorporate this data is to treat it as additional time series variables, which will result in a multivariate times series prediction model.

So far, only historical data is considered. However, additional information about the prediction horizon is available in the form of hourly consumption and (renewable) generation forecasts. This data might contain information about a possible deviation of the electricity price from its 'usual' temporal pattern. One way to incorporate this data is to treat the output of the time series model, together with load and generation predictions, as input features for a secondary model. A simple linear regression or Multi Layer Perceptron model should be a first feasible option for this step. The main challenge in this step is to find out how to best merge the two models together. Python libraries like Keras offer the option to merge multiple models ('branches') into one bigger model. However, it is questionable whether a merged model will still be able to train properly. Another idea is to use the outputs of the pre-trained time series model as bottleneck features for the secondary model, together with the additional data. This decision will be made based on the empirical success of both approaches.

As described above, generation forecasts are usually not available until market closing time. Wind and solar power generation are dependent on weather conditions. Weather data can be used to predict the levels of wind and solar power generation, which in turn can be used as an input for the model described above. The available historical weather data is organized in a 2D grid. This data could be fed into a Convolutional Neural Network, comparable to raw pixel data in image recognition. Since renewable generation power plants are not distributed homogeneously over Germany, the geo-spatial component of the weather data should be critical to determine the actual generation levels. In a real world deployment the historical weather data would have to be substituted by weather forecasts. Please note that the chance of success for this last step is somewhat speculative. If the generation forecasting model does not yield sufficient results, the available generation forecasts can still be used for prediction, leaving exact generation predictions as possible future steps.

References

[1] DWD. Regional reanalysis COSMO-REA6 FTP Server. `ftp://ftp-cdc.dwd.de/pub/REA/`.

[2] ENTSO-E. ENTSO-E Transparency Platform. `https://transparency.entsoe.eu/dashboard/show`.

[3] EPEXSPOT. EPEX SPOT SE: Day-Ahead Auction. `http://www.epexspot.com/en/market-data/dayaheadauction`.

[4] European Commission. Commission Regulation (EU) No 543/2013 of 14 June 2013 on submission and publication of data in electricity markets and amending Annex I to Regulation (EC) No 714/2009 of the European Parliament and of the Council Text with EEA relevance. `http://eur-lex.europa.eu/eli/reg/2013/543/oj`, 2013.

[5] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. `http://arxiv.org/abs/1703.07015`, 2017.

[6] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014.