No. 20 | December 2016

CEEME MEMOS

Machine learning applied to intraday optimization and trading in Germany

Machine learning is a highly promising field of study in artificial intelligence, specifically when applied to trading and optimization. This issue of the Memo describes a concrete application developed by CEEME in collaboration with BU GEM and the outlook for these new approaches. Enjoy reading!

Olivier Lecointe, CEEME Director

Driven by the growth of renewable sources of energy, intraday electricity markets are becoming increasingly important on the trading floors of the major energy groups. Intraday enables generators to meet their electricity supply commitments when production and/or consumption deviates from their day-ahead forecast (since generation plans must be submitted to the power exchanges no later than noon the day before).

German intraday market: a fertile ground for strategies based on machine learning

The German intraday market exhibits several characteristics that make it particularly well suited to the development of automated trading and optimization strategies.

Firstly, it is a highly physical market, with very little speculation. Price movements are often the result of measurable fundamentals, the volumes involved are relatively low and there are many players.

On the other hand, there are numerous factors that can trigger large intraday price movements, including: changing weather forecasts, unscheduled unavailability of generation units, the status of the transmission grid, the flexibility of generating facilities, and so on. It is sometimes difficult to assess the impact a specific event will have because there are so many factors at play and because those factors are interdependent.

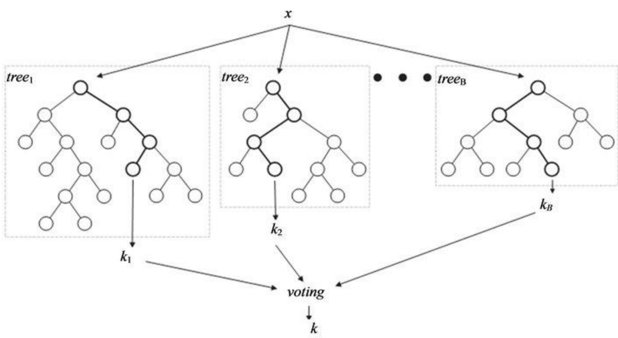
Given this context, machine learning is useful for detecting and exploiting market situations in which intraday prices differ from their *expected* behaviour, and which would not be possible to spot using conventional statistical approaches because there are too many variables simultaneously influencing price formation.

Detecting opportunities for gains via a machine learning algorithm

Machine learning makes it possible to detect *combinations of factors* in which a sought-after phenomenon is observed statistically. In our case, the aim is to detect market situations in which prices deviate significantly from what a model predicts. This model is built by *learning* the relationships between observations of recent months or years.

When that happens, we can then take a directional position in the market, anticipating that the market price will evolve towards the price predicted by the model. By systematising the approach, we can generate a substantial profit when the system correctly predicts the price trend with an error rate low enough to cover transaction fees and the price differential between purchase and sale price (bid-ask spread).

We can use decision trees to classify market conditions into buy or sell opportunities. In the figure below, we illustrate how to combine several trees to increase the predictive capabilities of the model. While *Boosted Trees* try to find optimal linear combination of trees, *Random Forests* use for each tree a random subset of the training data.



Using a decision 'forest' to achieve a 57% correct classification rate

A wide range of algorithms is available, so it is necessary to choose the approach best suited to the problem in question. We opted for methods based on the juxtaposition of a large number of decision trees (boosted trees, random forests, etc.). These methods are highly recommended for data mining applications that aim to extract knowledge from huge datasets. The principle consists in creating a large group of decision trees, each of which gives an indication about the decision to be taken. By combining these indications, we are able to create a 'committee' with high predictive power.

After a long iterative process of data mining and modelling, we were able to correctly predict the direction of price trends in around 57% of the cases. This may seem low at first glance, but it is a quite remarkable score in this context of market analysis.

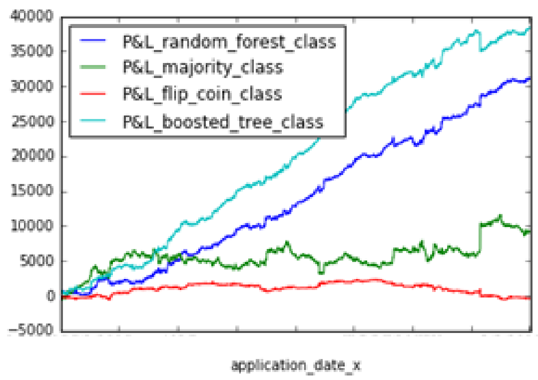
Developing a strategy and estimating its performance

The two curves in blue in the chart below show the results the strategies *could achieve* over a period of 12 months when one megawatt hour (MWh) is systematically bought or sold every quarter-hour via the most liquid products.

The results presented were obtained via the *k-fold cross-validation* method (where k=12). In other words, we deploy an iterative process in which 11 months of data are used to train a model, which will then be applied to the remaining 12th month (which was not itself part of the model training process). In each iteration, by taking a different month to apply the model and then aggregating the results, we can develop an estimate of what the strategy could yield over 12 months of unknown data.

Presenting the results

The dark blue curve represents the results of *random forest* learning, while the light blue curve represents the results of the *boosted trees* method. These two methods, while similar, differ in how they aggregate the many decision trees constructed to produce a final result.



By way of comparison, it is interesting to confront the results with those obtained from a *majority rule* (green curve), which consists in systematically buying or selling by choosing *a posteriori* the most lucrative of these two options. In red, the *flip-coin* approach decides, at each quarter-hour, to purchase or sell (heads or tails).

The best model (based on boosted trees) achieves an average profit of around €0.80/MWh, with average transaction fees estimated at around €0.20/MWh (not taken into account in this simulation). This means that the results presented here should be considered as an upper limit of what it is possible to achieve if the market were perfectly liquid and there were no transaction fees.

What do these models teach us?

The simulation presented here showed us that it would appear possible to use machine learning to enhance trading and optimization strategies in the German intraday market. These machine learning algorithms are currently being tested in production to determine if they deliver satisfactory results in a real-world setting.

Another lesson we can draw from this kind of approach is of *exploratory nature:* which variables have the most influence on intraday prices and which variables have little impact?

By analysing the frequencies of occurrence of the different variables in the trees (as well as their respective position in those trees), they can be classified in decreasing order of relative influence.

In our example, we see that the hourly day-ahead price is the most important variable and the variable that will be most often used *in conjunction with others* to determine the expected price direction.

In the influence ranking, we see in decreasing order of influence:

- the residual load
- the quantity of wind generation anticipated on a day-ahead basis
- the trend in intraday prices observed in the quarter-hour preceding the decision
- the total deviation (compared to the day-ahead) of renewable generation forecasts (wind + solar) four hours before the product expiration
- the quantity of solar generation forecasted in day-ahead
- the hour of the product
- deviations in renewable generation announced in the hour preceding the decision
- the time remaining until expiration of the product
- the spread between the intraday price and the day-ahead price, three hours before the decision
- ….

The study also revealed one of the most interesting aspects of machine learning: the ability to use data to truly understand the intraday price dynamics.

<u>CEEME contacts</u>: *Olivier Martin and Marcelo Espinoza*