DYNAMIC DIMENSIONING OF BALANCING POWER WITH FLEXIBLE FEATURE SELECTION

Anja OHSENBRÜGGE
University Oldenburg – Germany
anja.ohsenbruegge@uni-oldenburg.de

Sebastian LEHNHOFF
OFFIS - Germany
sebastian.lehnhoff.@offis.de

ABSTRACT

This paper proposes a novel dynamic design for control reserve dimensioning. In contrast to the current statistical analytic design we present a data driven approach with methods of computational intelligence. The chosen k-nearest neighbor algorithm is one of the most successfully used methods in machine learning. The model is able to predict complex nonlinear behavior by assuming that similar observations have similar outcomes. A condition for the success of this method is to determine the salient features. Therefore the core of this paper is to compare different methods of feature selection for the prediction task of control reserve. Numerical experiments for the year 2012 show that a machine learning approach has specific advantages over traditional approaches.

INTRODUCTION

To ensure a constant power frequency and thus a stable quality of supply, the permanent balance of power demand and supply is the most crucial constraint in an electrical power system. Therefore time series modeling and prediction of the power demand and supply is an important task. In recent years machine learning algorithms have drawn attention and have established themselves in the forecasting community. Especially for wind power and electricity demand the machine learning algorithms improved the prognosis accuracy. But due to the increasing share of generation from renewable resources the need for reserve and balancing power to cover these prognosis faults is still increasing.

In contrast with dynamic machine learning approaches, the current design for the dimensioning of these necessary reserves and its reliable provision is still an analytic statically method [1]. It is based on the former hierarchical and centralized structure of the European electricity sector, where the need for reserve was primarily caused by unpredictable power plant outtakes or load and generation noise, which were random and stochastically independent. Today both the reserve dimensioning and its activation critically depend on the actual state of generation and supply, the current network characteristics and also generation and load forecasts. In this paper we demonstrate the usage of machine learning algorithms for balancing power. Contrary to established approaches for wind generation and demand prediction, the prediction of the balancing power implies some challenging differences. First there is the effect of periodicity. Whereas the time series for electricity demand are highly time-sensitive with daily, weekly and annual patterns, the time dependency of the activation of balancing power is ambiguous, so that established filter mechanisms cannot easily be extended. Second there is the task of feature selection. The objective is to find the optimal subset of features which minimizes the prediction error. For predicting wind and PV generation the main influencing parameters are mostly known and primarily state specific [2], so that the accuracy of the predicted value (power output) can be optimized time-independently. In contrast, the influencing features for demand of balancing power are time- and state-dependent ─ heavily depending on generation structure, demographic effects etc. ─ and therefore bound to the system state. So the challenge is not to find the global optimum model to predict the balancing power demand for every time and state, but to find the local optimal setting. To tackle this challenge the k- Nearest Neighbor model (KNN) is extended with embedded tree methods, which generate the feature importance within the model. The third discrepancy is the accuracy metric. Whereas for generation and demand the predicted curve must coincide with the real curve, for balancing power the provided (maximum and minimum) power over a given time period is critical, thus over- and underestimation, respectively must be avoided at all costs. This paper focuses on the second aspect of the ambiguous time and space dependency of its influencing parameters. Therefore a detailed feature selection analysis is made on the basis of decision trees both for different times and states. The remainder of this paper is organized as follows: first we shortly introduce the different machine learning algorithms and the different approaches for feature selection used in this paper. Second we show first simulated numerical results of the implemented models. Last but not least we discuss the results with respect to the time and state dependency and the statistic inference of the features.

MODELING

Design of Balancing Power

Depending on its activation time, three different types of balancing power or control reserve are distinguished, primary, secondary and tertiary. Whereas the primary reserve (frequency-response reserve) is fixed to 3000 MW (outtake of two power plants) and is activated by autonomous f/P-droop controllers, the design and activation for secondary and tertiary reserve (minute reserve) falls to the transmission system operators (TSO). To prevent a contradicting activation of operating power in different areas, the German TSOs coordinate their operation reserve in a network called *Netzregelverbund* (NRV). Within this NRV the actual demand of reserve control is determined and tendered on the common internet platform [www.regelleistung.net].

Figure 1 German TSO areas[1]

The dimensioning of the control reserve to be provided is based on a probabilistic approach, which convolutes the individual probability density functions of the influencing parameters into one density function and then determining the amount depending on the given loss of load probability (LOLP). The current deficit probability is 0.1%, which means that in 0.1% of cases (9 hours a year) a lack of reserve is accepted. But due to the transformation of the energy system the actual applied method is not adequate anymore. On the one hand the influencing parameters are no longer statistically uncorrelated and on the other hand the given LOLP is originated from large capacity power plant blackouts, implying that the provided reserve power is oversized in the majority of cases, which has to be challenged when operating ever larger shares of renewable. Recently, besides the probabilistic method also simulative approaches are used. According to the described method of Graf-Haubrich[1], a Monte-Carlo-Simulation-based approach for probabilistically dimensioning of the demand of operating reserve by distinguishing the probability density function for each form of control power has been introduced by Brückl [3]. Thereupon power deficiencies due to power plant outtakes, load- or prediction-deviations were analyzed. A major result is that the wind prognosis faults have the biggest influence on the reserve amount, if the current highly over dimensioned LOLP-value is kept. As a result in [3] the question has been discussed, whether the actual reliability strategy is economically reasonable and feasible respectively.
Following studies [4] examines an adaptive control power market with capacity reserve. The objective of this research was to extend the approach of a flexible tendering for tertiary reserve on the principals of a capacity market. Thereupon the flexible dimensioning of control reserve for future energy scenarios was examined. Therefore a model was developed which is capable of calculating the control reserve within future energy scenarios by the hour.
In contrast of the aforementioned studies, this paper will focus on computational intelligence methods to dynamically predict the demand for operating reserve and is not based on probabilistic convolution-based calculation methods. In addition the aim of prior examinations was to predict the expected demand in future energy scenarios, whereas this approach concentrates on the prediction of control reserve today.

Machine Learning for Prediction

Prediction problems can be considered as a problem of supervised learning, where we have to infer from historical data the possibly nonlinear dependence between the input (feature vector) and the target output (future value)[5]. In figure 2 some of the main models are shown according to their complexity, which both describe the implementation of the model and its running time.

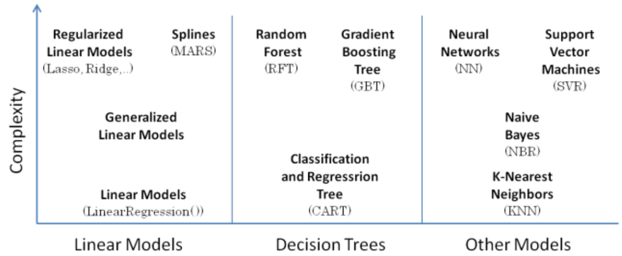

Figure 2 Machine Learning Models for Prediction[2]

Machine Learning Process

The objective of a machine learning prediction task can be described as follows. Given a dataset $(\boldsymbol{x}, y)_{i=1}^{N}$ of N samples, where $x = (x_1, \dots, x_d)$ is the d-dimensional feature vector and y is the corresponding label, the goal is to reconstruct the unknown functional dependence with an estimation $f^*(x)$, such that some specific loss function is minimized [5]. The expected prediction error (empirical risk) can be than be calculated as follows [5]:

$$E_{emp}(f) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i)) \qquad (1)$$

Common used loss functions for regression are the proportional loss function $L(y_i, f(x_i)) = |y_i - f(x_i)|$ or the quadratic loss function $L(y_i, f(x_i)) = (y_i - f(x_i))^2$. Then the expected prediction error is given by the MAE (mean absolute error) or by the RMSE (root mean squared error). To contain the search we have to restrict the function search space by choosing an appropriate prediction model from figure 2. Than the goal is to find the function $f^* = argmin E_{emp}(f|\theta)$ within this space

[1] [www.netzregelverbund.de]

[2] [http://caret.et.r-forge.r-project.org/modelList.html]

whose parameter θ minimizes the expected prediction error on the whole dataset [6]. Therefore the training set is divided into a training set and a validation set. The parameters are than optimized within a cross validation to avoid an overfitting on a single training set.
In the following we concentrate on three models, the random forest trees (RFT) [7], the gradient boosting trees (GBT)[8] and the k-Nearest Neighbor method (KNN) [9]. The KNN algorithm is one of the most popular techniques in nonlinear time series analysis. This method is already used in the energy domain for prediction of wind and PV generation or energy demand [10] [11]. Given a query feature vector $x_q \in R^d$ and a set of N d-dimensional vectors $X = x_1, x_2, \ldots, x_n$, the nearest neighbor search algorithm aims to find the subset of k items $N_k(x_q)$ from X such that its distances to the query vector are minimal. Assuming that similar observations will lead to similar outcomes of the target value, the target value $\hat{y}_q$ for a query feature vector (new observation) can be calculated as follows [6]:

$$\hat{y}_q = f_{kNN}(x_q) = \frac{\sum_{i \in N_k(x_q)} w_i y_i}{\sum_{i \in N_k(x_q)} w_i} \qquad (3)$$

w_i is the weight given to each neighbor x_i and therefore the contribution of its target value y_i is also weighted with w_i. This can either be done with a uniform weight, or with a distance weighted kernel function. Besides the weight given to each neighbor also the weight given to each feature is crucial. Therefore features are weighted according to their importance before the distance is calculated. Applying the Euclidean distance the weighted distance is calculated as follows [6]

$$d_w(x, y) = |x - y| = \sqrt{\sum_{i=1}^{d} w_i (x_i - y_i)^2} \qquad (4)$$

So for the KNN model the hyper parameters to be optimized are the weighting vector w_i for the feature weights, the distance metric $dist(x_i, x_q)$, the neighborhood (number of k) and the loss function $L(y_i, f(x_i))$. Because the KNN method is very sensitive to feature weights specified by the distance metric, in this paper we concentrate on this feature weighting. As distance function we choose the Euclidean distance, as loss function the quadratic Gaussian norm is taken and as neighborhood we choose 10 neighbors. To optimize the feature weights we use trees in form of random forests and gradient boosting trees as second model [12]. As a so called *embedded* model for feature selection[3] it derives both the partial dependence between the features and the target value y and the feature importance w. This calculated feature importance is then used for weighting the features within the KNN algorithm (KNN_RFT, KNN_GBT). The results are compared with feature weights all set to one (KNN) and a KNN model with a preprocessed principal component analysis (KNN_PCA) reducing the features to the 4 most salient ones. Figure 3 gives an overview of the main steps of the machine learning process.

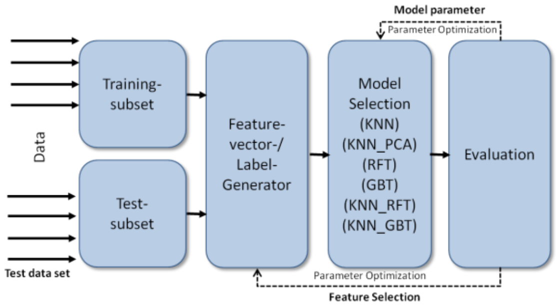

Figure 3 Machine Learning Process

The first step is to generate the data. In order to dimension the demand for control reserve the characteristic number is the joint balance of the four German TSO areas, whose geographic location can be seen in figure 1. The control area balance of each TSO is defined as the sum of all deviations of balance groups/areas within the control area (e.g. the EEG balance group) and it is equivalent to the receipt of balance power in this control area, i.e. positive balance will lead to positive balance power to counterbalance an underestimated control area and vice versa. These balances are then summed up and adjusted both among themselves and with the import and export of balance power to associated countries. The remainder has to be covered with balance power from the tendering platform.
So the target value to be predicted is the control area balance. As feature values we take different predictor variables as there are: wind and PV prediction deviations (wind error, PV error) calculated as their prediction values minus the realized generation, the vertical net load (vnetload) and the phelix base price at the EEX (PHB)[4]. All time series are given in 15 minutes intervals. Additionally some time features like the weekday and the time of the day (TOD) are examined. With these different features the next step is the generation of the feature vector ***x.***
Therefore the features are divided into three groups:

- Features-lag : Time series with values $\{x_t \mid t \leq t_0\}$, such as (balance)
- Features-future : Time series with values $\{x_t \mid t \leq t_0 + prediction\_horizon\}$[5], such as (wind_err, pv_err, vertical netload)
- Features-time : Scalar values such as (weekday[0:6], TOD [0: 23])

[3] In contrast of the filter and wrapper techniques[13]

[4] Since the introduction of a standard price (reBAP) in the year 2010 the price for balance power is related to the EEX

[5] To show the dependency of the predictor variables and the target value the errors are taken as a given

Given the initial point for the prediction t_0 the feature vector is generated as shown in figure 4 as an example for a feature window of 2 intervals (30 minutes) and a prediction horizon of 1 interval (15 minutes). In the simulation the prediction horizon is 96 intervals (one day); and the feature window is set to zero (one value each feature).

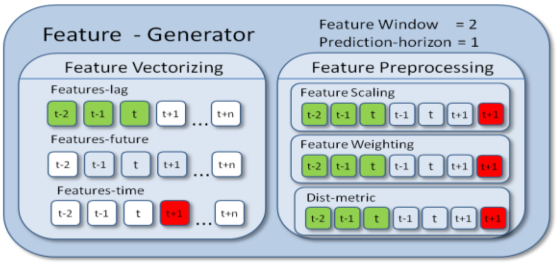

Figure 4 Vectorizing and Preprocessing of features

After vectorizing the features, each feature is standardized by removing the mean and scaling to unit variance to compensate the effect of different dimensions. This centering and scaling is done independently on each feature by computing the relevant statistics on the whole year 2012. The last step of the preprocessing is the feature weighting. As shown in formula (4) each feature is weighted independently. For setting the weights three different approaches are simulated.

- KNN : All weights set to one
- KNN_GBT : Setting the weights calculated by the Gradient Boosting Tree
- KNN_RFT : Setting the weights calculated by the Random Forest Tree

For evaluating the results two more models are simulated. The dimension reduction method “principal component analysis” (KNN_PCA), which takes the 4 most salient features and the persistence method as the simplest forecast model, which simply takes the target value of the initial point t_0.

SIMULATION

To show the time and state specific dependency of the different features and the target, the prediction is made for the each month of the year 2012, respectively three month as training set and one month as test set to predict. So the weights are generated on the latter three month. The result of the feature weights in the different control areas for winter 2012 (trained on January till March) is shown in figure 5.

Noticeable are the relative high importance of the price (PHB) in the Tennet area and the low importance of the wind genereation (WindE) in the Amprion area. The latter is due to the fact, that Amprion is the area with the highest demand and the least generation from renewable energies in relation to the vertical net load. This state specific dependency is reinforced by the partial dependency of the wind error and the price on the control balance as plotted in figure 6.

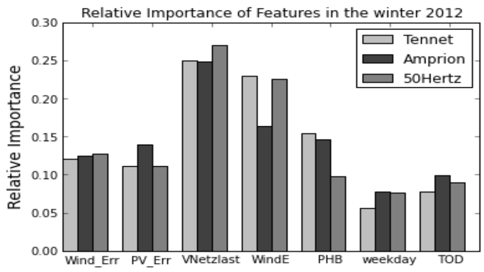

Figure 5 Feature Weights for the winter in 2012

As the variance analysis is embedded in the Tree Models they implicitly calculate the so called partial dependence between a single feature and the target value. The tick marks on the x-axis represent the deciles of the feature values in the training data. Regarding the partial dependence it can be seen, that there is approx no direct influence of the wind error on the control balance in the Amprion area. In contrast, the influence in the 50Hertz area is approx. linear with a positive gradient of 300 MW per 500 MW wind error. But even more worth noting is the linear partial dependence of the phelix base price at the EEX and the control balance in the Tennet area, where the influence has a negative gradient of approx. -300 MW per 15 EUR. These results show that the feature weights and their partial dependence vary both between the different control areas and different seasons.

Figure 7 exemplary shows the predicted curves for one day in summer 2012. The KNN_GBT curve realizes a very good conformity with the balance curve.

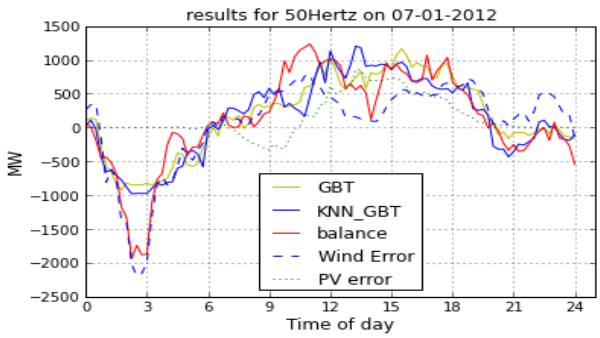

Figure 7: Prediction results for 50 Hertz one day ahead

Recognizable is the deviation from the balance at 3:00 am, where the balance coincides with the very high negative wind error ($p(x_{WindError} < -600\ \text{MW}) \sim 0.1$). So the importance of a single feature and its partial dependence is also dependent on the relative value of that feature. When there are very high edges in the wind error, it tends to have more influence on the balance. Therefore further development will concentrate on the adaptive feature weighting, both incorporating the actual correlation between the feature and the target values and the relative level of the feature value.

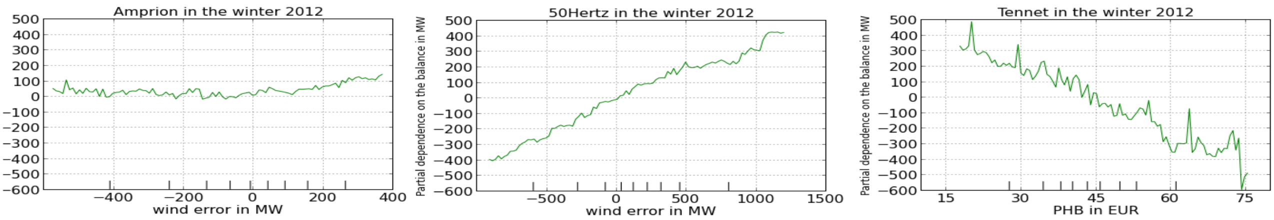

Figure 6 Partial Dependence between the wind errors respectively the price and the control balance in different areas

An overview of the numerical results for the Amprion and 50 Hertz control area is given in table 1 and 2. Displayed is the RSME for the different prediction models. In most of the cases the Gradient Boosting Tree (GBT) outperforms the other models. Especially in the second half of the year the results for the 50Hertz area are very good, so that the prediction error is 40% less then persistence. So the best result for the 50Hertz area is in November, when the GBT achieves an RSME of 280 MW in comparison with the 458 MW of the persistence.

Amprion	KNN	KNN_PCA	GBT	RFT	KNN_GBT	KNN_RFT	Persistence
Jan-12	549.02	560.15	542.94	557.75	527.65	529.70	594.24
Feb-12	575.13	614.47	531.29	555.99	581.11	553.12	535.91
Mar-12	483.19	496.34	433.27	478.47	453.48	453.84	471.74
Apr-12	514.26	529.93	478.78	509.91	530.45	491.09	536.45
May-12	540.25	563.05	520.70	549.21	560.60	549.48	627.94
Jun-12	505.36	495.96	443.19	475.00	487.02	477.93	560.86
Jul-12	426.26	447.99	403.76	433.52	430.91	417.58	460.61
Aug-12	470.64	525.79	449.32	452.84	461.63	454.49	491.22
Sep-12	403.52	448.29	365.44	393.33	393.53	383.63	434.00
Oct-12	461.25	483.17	506.14	484.23	470.92	466.13	573.02
Nov-12	439.37	452.07	406.55	425.44	431.11	427.70	495.50
Dec-12	516.18	524.25	504.48	497.83	511.14	514.05	622.41

Table 1: RSME for Amprion in 2012 with 1 day horizon

50Hertz	KNN	KNN_PCA	GBT	RFT	KNN_GBT	KNN_RFT	Persistence
Jan-12	952.34	1027.15	923.84	953.72	972.70	936.47	657.54
Feb-12	587.41	609.65	498.85	524.32	560.46	518.10	562.12
Mar-12	502.57	492.63	473.26	486.60	527.85	504.66	570.67
Apr-12	395.01	460.41	348.78	363.01	389.16	439.51	563.78
May-12	507.14	576.33	498.42	485.63	509.06	472.55	504.86
Jun-12	452.81	534.32	420.43	454.61	441.62	454.83	575.69
Jul-12	341.05	407.60	328.44	329.99	370.57	365.55	518.74
Aug-12	332.10	384.04	291.01	304.58	330.96	325.63	446.45
Sep-12	340.25	397.66	284.20	321.54	333.19	320.46	437.10
Oct-12	361.00	403.29	332.93	345.73	351.25	347.84	521.13
Nov-12	313.35	360.25	280.22	305.88	313.14	312.13	458.36
Dec-12	437.76	434.49	445.14	449.75	449.02	444.31	574.12

Table 2: RSME for 50Hertz in 2012 with 1 day horizon

Noticeable are the bad results of the KNN_PCA, so that the dimension reduction mostly produces even worse results than the simple weighting with all features to 1 (KNN). That may lead to the conclusion that at least more than 4 features are important to predict the control balance. Another thing is the bad result in the January especially in the 50Hertz area, which demonstrates the high volatility of the balance in areas with high amount of renewable energies in combination with low netload and therefore small opportunity to overcome those imbalances in the own control area.

CONCLUSION

In this paper we introduced a new methodology for predicting the amount of balancing power with methods of machine learning, namely with a two step k-Nearest Neighbor regression. The machine learning approach has the advantage to the current convolution-based method that it is more suitable to adapt the dimensioning of the balance power to the current system state. The importance of the time and space specific influence on the balance was illustrated. The numerical results showed that approaches like the KNN algorithm or Gradient Boosting Trees (GBT) are superior to statistical methods. Furthermore it reinforced the hypothesis that the weights of the influencing parameters vary as well among the control areas as between different hours of the day or years. Therefore the feature work is not only to dynamically optimize the feature-weights but to adaptively optimize them with respect to the actual system state.

REFERENCES

[1] CONSENTEC: *Gutachten zur Dimensionierung des Regelleistungsbedarfs* unter dem NRV. 17.12.2010

[2] E. Mangalova, 2012, "Wind power forecasting using the k-nearest neighbors algorithm ", *International Journal of Forecasting*, vol.30, 402-406

[3] O. Brückl, 2008, „*Wahrscheinlichkeitstheoretische Bestimmung des Regel- und Reserveleistungs-bedarfs in der Elektrizitätswirtschaft*, Muenchen, Germany

[4] S. Kippelt, 2013, " Flexible Dimensioning of Control Reserve for Future Energy Scenarios ", *Proceedings IEE Power Tech*

[5] R.Berk, 2008, *Statistical Learning from a Regression Perspective*, Springer, NY, USA

[6] Kramer, O.: *Computational Intelligence – Eine Einführung,* Springer 2009

[7] Breiman L., *Random forests*. In Machine Learning, Seiten 5-32, 2001

[8] Friedman, J.H.: "*Stochastic Gradient Boosting,* 1999

[9] Duda, R.O., Hart, P.E.: *Pattern Classification and Scene Analysis,* John Wiley & Sons, 1973

[10] Treiber, N., et.al. : *Aggregation of Features for Wind Energy Prediction with Support Vector Regression and Nearest Neighbors.* In: ECML 2013

[11] Wolff, B.; Lorenz, E.; Kramer,O.: *Statistical Learning for Short Term Photovoltaik Power Prediction.* In ECML 2013

[12] L. Breiman, J. Friedman, R. Olshen, and C. Stone, 1984, *Classification and Regression Trees,* Wadsworth Belmont, California, USA

[13] I. Guyon, A.Elisseeff, 2003, "An Introduction to Variable and Feature Selection", *Journal of Machine Learning Research*, vol.3, 1157-1182