International Journal of Forecasting 32 (2016) 1038–1050

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting

Pierre Gaillard [a,b], Yannig Goude [a,*], Raphaël Nedellec [a]

[a] EDF R&D, Clamart, France

[b] GREGHEC: HEC Paris – CNRS, Jouy-en-Josas, France

A R T I C L E   I N F O

Keywords:
Electric load
Electricity price
Forecast
Semi-parametric models
Combining forecasts
Sparse regression

A B S T R A C T

We summarize the methodology of the team TOLOLO, which ranked first in the load forecasting and price forecasting tracks of the Global Energy Forecasting Competition 2014. During the competition, we used and tested many different statistical and machine learning methods, such as random forests, gradient boosting machines and generalized additive models. In this paper, we only present the methods that showed the best results. For electric load forecasting, our strategy consists of producing temperature scenarios that we then plug into a probabilistic forecasting load model. Both steps are performed by fitting a quantile generalized additive model (quantGAM). Concerning the electricity price forecasting, we investigate three methods that we used during the competition. The first method follows the spirit of that used for the electric load. The second one is based on combining a set of individual predictors. The last one fits a sparse linear regression to a large set of covariates. We chose to present these three methods in this paper because they perform well and show the potential for improvements in future research.

1. Introduction

This paper presents the methodology employed for the probabilistic electric load and electricity price forecasting tracks of the Global Energy Forecasting Competition 2014 (GEFCom2014). We participated in both tracks, but with different levels of intensity and motivation. We were familiar with the field of load forecasting before the competition, but were inexperienced with price forecasting. As a consequence, we converged rapidly to a unique solution for load forecasting, but changed our method for electricity price forecasting constantly as our knowledge and understanding increased.

Quantile regression based on pinball loss minimization (see Koenker & Bassett, 1978) and generalized additive models (see Hastie & Tibshirani, 1990; Wood, 2006) are the main tools of our work. To the best of our knowledge, no off-the-shelf program for quantile generalized additive models was available, and we implemented our own solution for that. We originally designed it for load forecasting, but it turned out to be the most efficient method for both tasks. It is presented in Section 2. We also tested a wide range of other approaches for the price forecasting task, of which we describe those that we consider to deserve sharing because they show the potential for improvement and can be applied to other forecasting problems.

The aggregation of experts is considered in Section 4.3. We were inspired by the work of Nowotarski and Weron (2015), and extend it to the case where the weights of the combination can vary over time. More precisely, we adapt the setting of the robust online aggregation of experts

* Corresponding author.
E-mail addresses: pierre@gaillard.me (P. Gaillard), yannig.goude@edf.fr (Y. Goude), raphael.nedellec@edf.fr (R. Nedellec).

http://dx.doi.org/10.1016/j.ijforecast.2015.12.001

(see Cesa-Bianchi & Lugosi, 2006), which has already been applied successfully to point-wise load forecasting by Devaine, Gaillard, Goude, and Stoltz (2013) and Gaillard and Goude (2015), to quantile regression. Our set of 13 experts consists of forecasters from the price forecasting literature, and includes AR (autoregressive) models, TAR (threshold autoregressive) models, ARX (autoregressive exogenous) models, TARX (threshold autoregressive exogenous) models, PAR (spike preprocessed autoregressive) models (as presented by Weron & Misiorek, 2008), GAMs (generalized additive models), random forests (see Breiman, 2001), and gradient boosting machines (see Friedman, 1999).

The third approach, presented in Section 4.4, is based on covariate selection using the ℓ_1 selection procedure, commonly known as Lasso regression and introduced by Tibshirani (1996). This was motivated by the fact that we generated a lot of extra covariates (192) on top of those used for price forecasting. Thus, at the end of the competition, we were curious to see how an automatic procedure would perform in selecting an optimal subset from among them. ℓ_1 selection and quantile regression were studied by Belloni and Chernozhukov (2011), but have never been applied to either price or load forecasting. As far as we could determine, there is no open source code that satisfies our needs for the competition. Section 4.4 presents the kernel-based approach that we developed for this purpose. For the price forecasting task, the results obtained during the competition differ slightly (sometimes better, sometimes worse) from those obtained using the three methods (quantile GAM, quantile mixture and quantile GLM (generalized linear model)). This is due largely to the other approaches that we used over the course of the competition, and to hybrid variants of those presented here. For the sake of conciseness, this paper deliberately focuses on these three methods.

2. Quantile regression with generalized additive models

We consider a supervised regression setting where we are asked to forecast a univariate response variable $Y_t \in \mathbb{R}$ (such as the load) according to several covariates $\boldsymbol{X}_t = (X_{t,1}, \dots, X_{t,d}) \in \mathbb{R}^d$ (such as the temperature). A training sample $\{(\boldsymbol{X}_t, Y_t)\}_{t=1}^n$ is available.

2.1. Generalized additive models

GAMs were introduced by Hastie and Tibshirani (1990), and explain the output Y_t as a sum of smooth functions of the different covariates $X_{t,j}$. More formally, we assume that for all time $t = 1, \dots, n$, $Y_t = \mu(\boldsymbol{X}_t) + \varepsilon_t$, where μ is the unknown function to be estimated and ε_t are zero mean random variables from some exponential family distribution that are independent and identically distributed (i.i.d.).[1] GAMs assume that there exists a link function g such that

$$g\big(\mu(\boldsymbol{X}_t)\big) = f_1(X_{t,1}) + f_2(X_{t,2}) + f_3(X_{t,3}, X_{t,4}) + \cdots, \quad (1)$$

where the f_j are smooth functions of the covariates $X_{t,k} \in \mathbb{R}$. In what follows, the link function g is the identity and the smooth functions f_j are cubic splines (unless specified otherwise). Basically, cubic splines are polynomials of degree 3 that are joined at points known as "knots" by satisfying some continuity constraints (see Wood, 2006 for details). We call $\mathcal{S}(K_i)$ the class of cubic splines for some fixed set of knots, K_i.

We fit the smooth effects f_i using penalized regression methods. To do this, we first choose the knots K_i for each effect f_i, then use the ridge regression that minimizes the following criterion over all effects $f_1 \in \mathcal{S}(K_1), f_2 \in \mathcal{S}(K_2), \dots$:

$$\sum_{t=1}^{n}\left(Y_t - \sum_{i=1}^{p} f_i\big(X_t^i\big)\right)^2 + \sum_{i=1}^{p} \lambda_i \int \left\|f_i''(x)\right\|_2^2 dx, \quad (2)$$

where X_t^i are one or two covariates of $\boldsymbol{X}_t$ that correspond to each effect f_i. Here, $\lambda_1, \dots, \lambda_p > 0$ are regularization parameters that control the degree of smoothness of each effect (the higher λ_i is, the smoother f_i is), and have to be optimized. The knots K_i are distributed uniformly over the range of the covariate(s) X_t^i corresponding to effect f_i. The number of knots (i.e., the cardinal of K_i) is another way to control the smoothness of the effect f_i, and should be optimized as well. These problems are solved by using the methodology presented by Wood (2006), which consists of minimizing the generalized cross validation criterion (GCV). The method is implemented in the R package `mgcv` (see Wood, 2006).

2.2. Quantile regression

Quantile regression was introduced by Koenker and Bassett (1978). Let Y be a real value random variable and $\boldsymbol{X}$ be a set of explanatory variables. If $F_{Y|\boldsymbol{X}}$ denotes the conditional cumulative distribution of Y given $\boldsymbol{X}$, then the conditional quantile q_τ of order $\tau \in [0, 1]$ of Y knowing X is defined as the generalized inverse of $F_{Y|\boldsymbol{X}}$:

$$q_\tau(Y|\boldsymbol{X}) = F_{Y|\boldsymbol{X}}^{-1}(\tau) = \inf\left\{y \in \mathbb{R}, F_{Y|\boldsymbol{X}}(y) \geq \tau\right\}. \quad (3)$$

Now, the idea of quantile estimation arises from the observation that the median (i.e., $q_{0.5}(Y|\boldsymbol{X})$) minimizes the expected absolute error. More generally, it can be shown that the conditional quantile $q_\tau(Y|\boldsymbol{X})$ is the solution of the minimization problem:

$$q_\tau(Y|\boldsymbol{X}) \in \arg\min_g \mathbb{E}[\rho_\tau(Y - g(\boldsymbol{X}))|\boldsymbol{X}], \quad (4)$$

where ρ_τ is the pinball loss defined as $\rho_\tau(u) = u(\tau - \mathbf{1}_{\{u<0\}})$ for all $u \in \mathbb{R}$.

Linear quantile regression is implemented in the R-package `quantreg` (see Koenker, 2013), and assumes that $\{(\mathbf{X}_t, Y_t)\}_{t=1,\dots,n}$ are i.i.d. such that $Y_t = \mathbf{X}_t^\top \boldsymbol{\beta} + \varepsilon_t$, where $\boldsymbol{\beta} \in \mathbb{R}^d$ is a vector of unknown parameters. Linear quantile regression solves the convex minimization problem

$$\widehat{\boldsymbol{\beta}}_\tau \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^d} \sum_{t=1}^{n} \rho_\tau\big(Y_t - \mathbf{X}_t^\top \boldsymbol{\beta}\big), \quad (5)$$

and estimates q_τ using $\widehat{q}_\tau : \boldsymbol{x} \mapsto \boldsymbol{x}^\top \widehat{\boldsymbol{\beta}}_\tau$.

[1] ε_t are i.i.d. error terms throughout the paper, but their distributions may change between displays.