ARTICLE IN PRESS

International Journal of Forecasting ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

International Journal of Forecasting

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

A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014

Romain Juban [a], Henrik Ohlsson [a,b], Mehdi Maasoumy [a], Louis Poirier [a], J. Zico Kolter [a,c,*]

[a] C3 IoT, Redwood City, CA, USA

[b] University of California at Berkeley, Berkeley, USA

[c] School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

ARTICLE INFO

Keywords:
Probabilistic forecasting
Quantile regression
Alternating direction method of multipliers
Solar forecasting
Wind forecasting
Load forecasting
Price forecasting
GEFCom2014

ABSTRACT

This paper proposes a generic framework for probabilistic energy forecasting, and discusses the application of the method to several tracks in the 2014 Global Energy Forecasting Competition (GEFCom2014). The proposed method uses a multiple quantile regression approach to predict a full distribution over possible future energy outcomes, uses the alternating direction method of multipliers to solve the optimization problems resulting from this quantile regression formulation efficiently, and uses a radial basis function network to capture the non-linear dependencies on the input data. For the GEFCom2014 competition, the approach proved general enough to obtain one of the top five ranks in three tracks, solar, wind, and price forecasting, and it was also ranked seventh in the final load forecasting track.

1. Introduction

Many energy forecasting methods provide point forecasts, and therefore are not appropriate for making decisions under uncertainty. As uncertainties are intrinsic to all realistic energy forecasting problems, probabilistic forecasting – predicting a distribution of possible future values rather than a point estimate – is essential for accurate energy management. There are a number of methods of producing probabilistic forecasts, which, broadly speaking, fall into two categories: methods that attempt to fit a specific parametric probability distribution to the outputs, and methods that attempt to predict some function (e.g., quantiles) of the target probability distribution directly. See e.g. Gneiting and Katzfuss (2014) for a general review of these methods.

In this paper, we focus on the second class of algorithms, and specifically develop a probabilistic forecasting method based upon (multiple) quantile regression. Although multiple quantile regression is a well-established technique in the regression literature (see Koenker & Hallock, 2001), thus far it has been applied less in the area of probabilistic energy forecasting. Our approach includes three main components: (1) we generate a complete distribution, represented by a series of quantile predictions, using an approach based upon (multiple) quantile regression for 99 quantiles[1]: $\alpha = 0.01, \ldots, 0.99$; (2) we incorporate non-linear functions of the input data using radial basis

* Corresponding author at: School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA.
E-mail address: zkolter@cs.cmu.edu (J.Z. Kolter).

[1] The α-quantile, $\alpha = 0.01, \ldots, 0.99$ for a random variable X, is defined as y_α, such that $P(X \leq y_\alpha) = \alpha$.

http://dx.doi.org/10.1016/j.ijforecast.2015.12.002
0169-2070/

Please cite this article in press as: Juban, R., et al., A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2015.12.002

function (RBF) features (see e.g. Bors, 2001) that are generated from the input data automatically using the k-means algorithm (see e.g. Arthur & Vassilvitskii, 2007; Coates & Ng, 2012; Munnoli & Bapat, 2013); and (3) we solve the resulting optimization problems efficiently using a proposed alternating direction method of multipliers (ADMM) algorithm.

We developed this framework for the GEFCom2014 energy forecasting competition (see Hong et al., 2016), where the method obtained a ranking in the top five for three of the four tasks (solar, wind, and price), and was ranked seventh in the load forecasting task. The chief merit of our approach is the fact that the exact same methodology was applied successfully to all of these tasks, with no task-specific engineering, except for minimal feature engineering. Thus, the value of the algorithm comes from its ability to obtain a high level of accuracy efficiently on a wide variety of tasks.

Section 2 defines the probabilistic forecasting problem and introduces the notation. Section 3 describes the common machine learning framework that is used to generate probabilistic forecasts for each energy quantity. Sections 4–6 present the applications of this framework to the energy forecasting tasks of GEFCom2014. We describe the specific model adjustments for each track and the corresponding results. The paper is concluded in Section 7.

2. Problem formulation

We assume a setting where the quantity of interest for forecasting, $y \in \mathbb{R}$, is a real-valued scalar quantity, such as the wind power or electricity price. We denote the relevant covariates used for generating this forecast by $x \in \mathbb{R}^p$. The task of probabilistic forecasting in which we are interested involves producing (some representation of) the conditional distribution $p(y|x)$. In our case, we are concerned with producing a distribution that is represented by its quantiles. Specifically, we let $\hat{y}_\alpha$ be a prediction of the α-quantile of the output y, and we consider methods that produce such predictions for α in some set $\mathcal{A}$ (in the case of GEFCom2014, the set is given by $\mathcal{A} = \{0.01, 0.02, \ldots, 0.99\}$). The quality of the prediction is then evaluated using the quantile or pinball loss function, defined as:

$$\ell(\hat{y}_\mathcal{A}, y) = \sum_{\alpha\in\mathcal{A}} \psi_\alpha(\hat{y}_\alpha - y),$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is the so-called "tilted absolute loss" function

$$\psi_\alpha(z) = \max\{\alpha z, (\alpha - 1)z\}.$$

The basic idea behind such a loss is that, given a set of outputs $y^{(i)}$ for $i = 1, \ldots, m$, the function $\sum_i \psi_\alpha(y^{(i)} - z)$ is minimized by letting z be the α-quantile of the $y^{(i)}$ terms.

In our quantile regression setting, we presume that we have access to a training set of m inputs and outputs, denoted by

$$\{x^{(i)}, y^{(i)}\}, \quad i = 1, \ldots, m.$$

Our goal will be to develop a model that minimizes the sum of these quantile loss functions over this data set:

$$L_\mathcal{A} = \sum_{i=1}^{m} \ell_\mathcal{A}\left(\hat{y}^{(i)}_\mathcal{A}(x^{(i)}), y^{(i)}\right)$$

where the notation $\hat{y}_\mathcal{A}(x)$ highlights the fact that the predictions depend on the input covariates x.

3. Machine learning framework for probabilistic forecasting

In this section, we present our machine learning approach for probabilistic forecasting based upon multiple quantile regression, input selection, non-linear feature generation, and ADMM optimization. We discuss each of these elements in detail below.

3.1. Multiple quantile regression

At its core, our method solves a multiple quantile regression problem using a linear model, referring to the fact that the predictions are a linear function of the parameters; the model's predictions are non-linear in the inputs, via the use of non-linear feature transformations. In particular, we assume a mapping from input vectors to feature vectors, denoted $\phi : \mathbb{R}^p \to \mathbb{R}^n$. We then predict the α-quantile of the output using the prediction function

$$\hat{y}_\alpha(x; \theta_\alpha) = \theta_\alpha^T \phi(x),$$

where $\theta_\alpha \in \mathbb{R}^n$ is a parameter vector and where we use the notation $\hat{y}_\alpha(x; \theta_\alpha)$ to emphasize that the prediction depends on both the input covariates and the parameters θ. As is apparent from the notation, and as is standard practice in multiple quantile regression, the model allows for a *separate* set of parameters θ_α being used to predict each quantile. We fit these parameters by minimizing the quantile loss function described above, plus some additional ℓ_2 regularization, which here takes the form

$$\underset{\theta_\mathcal{A}}{\text{minimize}} \sum_{i=1}^{m} \sum_{\alpha\in\mathcal{A}} \psi_\alpha\left(\theta_\alpha^T \phi(x^{(i)}) - y^{(i)}\right) + \frac{\lambda}{2} \sum_{\alpha\in\mathcal{A}} \|\theta_\alpha\|_2^2. \tag{1}$$

A crucial point in this model is that the optimization problem decomposes over each θ_α; that is, we are effectively fitting $|\mathcal{A}|$ *independent* models, one for each quantile of the output. Although it would be possible to include different features for each quantile prediction as well, we do not pursue this route in our work. As we will discuss shortly, using the same feature vector $\phi(x)$ across all models leads to significant reductions in the computational cost of the algorithm.

One potential (and well-known) issue with multiple quantile regression is that, because the quantile prediction functions are trained independently, it is possible to have predictions that violate the nature of a probability distribution: e.g., the prediction of the 0.01 quantile may be higher than the prediction of the 0.02 quantile. This is clearly inconsistent with the monotonicity constraints of a consistent distribution. Much of the work with multiple quantile regressions has been dedicated to coupling the multiple tasks in a way that enforces monotonicity constraints, either on the input data points themselves (Takeuchi, Le, Sears, & Smola, 2006) or on the input points

Please cite this article in press as: Juban, R., et al., A multiple quantile regression approach to the wind, solar, and price tracks of GEFCom2014. International Journal of Forecasting (2016), http://dx.doi.org/10.1016/j.ijforecast.2015.12.002