International Journal of Forecasting 32 (2016) 585–597

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Electric load forecasting with recency effect: A big data approach

Pu Wang [a], Bidong Liu [b], Tao Hong [b,*]

[a] SAS Institute, USA

[b] University of North Carolina at Charlotte, USA

ARTICLE INFO

Keywords:
Interaction regression
Group analysis
Lagged temperature
Linear models
Moving average temperature
Multiple linear regression
Naive models
Benchmark

ABSTRACT

Temperature plays a key role in driving the electricity demand. We adopt the "recency effect", a term drawn from psychology, to represent the fact that the electricity demand is affected by the temperatures of the preceding hours. In the load forecasting literature, the temperature variables are often constructed in the form of lagged hourly temperatures and moving average temperatures. In the past, computing power has limited the amount of temperature variables that can be used in a load forecasting model. In this paper, we present a comprehensive study to model the recency effect using a big data approach. We take advantage of modern computing power to answer a fundamental question: *how many lagged hourly temperatures and/or moving average temperatures are needed in a regression model in order to capture the recency effect fully without compromising the forecasting accuracy?* Using a case study based on data from the load forecasting track of the Global Energy Forecasting Competition 2012, we first demonstrate that a model with the recency effect outperforms its counterpart (a.k.a. Tao's Vanilla Benchmark Model) by 18% to 21% for forecasting the load series at the top (aggregated) level. We then model the recency effect in order to customize load forecasting models at the bottom level of a geographic hierarchy, again showing a superiority over the benchmark model of 12% to 15% on average. Finally, we discuss four different implementations of the recency effect modeling by hour of a day. In addition, this paper also presents two interesting findings: 1) the naive models are not useful for benchmark purposes in load forecasting at aggregated level due to their lack of accuracy; and 2) slicing the data into 24 pieces to develop one model for each hour is not necessarily better than building one interaction regression model using all 24 hours together.

1. Introduction

At the inception of electric power systems, lighting was the only end use of electricity, meaning that the electricity demand was driven primarily by calendar variables. As more and more electricity-powered appliances were invented, the end use became diversified. The increasing penetration of electrical air conditioning systems made the role of weather more and more important in driving the electricity demand. Ever since the 1940s, people have realized that the electric load depends strongly on the weather (Dryar, 1944).

In the pre-PC (personal computer) era, utility planners and operators created lookup tables and charts, based on historical data and past experience, for capturing the relationship between the load and weather variables such as temperature and humidity. They then forecasted the load using these charts and tables, together with rulers and intuitions (Hong, 2014).

* Corresponding author.
E-mail address: hongtao01@gmail.com (T. Hong).

http://dx.doi.org/10.1016/j.ijforecast.2015.09.006
0169-2070/© 2016 Published by Elsevier B.V. on behalf of International Institute of Forecasters.

When people started using computers for load forecasting in 1980s, the computing power was very limited. Quite often, the model building procedures for selecting variables and estimating parameters had to be conducted offline, meaning that the computer was performing online calculations of the load forecast based on the new data and previously calculated variables and parameters (Gross & Galiana, 1987). The offline model building scheme meant that the model(s) could not be updated in real time to reflect the most recent status of the power system. As a consequence, the forecasting accuracy was more or less compromised.

The technological advancement through the late 1990s quickly eliminated the need for offline computation for many load forecasting techniques. The benefits of the increase in computing capability also meant that people started to apply some more computationally intensive techniques such as artificial neural networks (Hippert, Pedreira, & Souza, 2001) and autoregressive integrated moving average models (Weron, 2006) to load forecasting. At the same time, people also started using large numbers of variables in load forecasting models (Hippert et al., 2001).

In psychology, the recency effect refers to the fact that human beings tend to remember the most recent items. The power grid is similar, in that its demand tends to be affected significantly by the recent temperatures. Hong (2010) was the first to adopt this term for illustrating part of a systematic load forecasting methodology, which used lagged temperatures to enhance the load forecasting accuracy of a benchmark model. Since then, this term "recency effect" has been accepted widely in the US utility industry, and has become part of a commercial software package (SAS® Energy Forecasting) that is currently being used by many power companies worldwide. Note that many papers in the load forecasting literature have reported the use of lagged temperatures. Papalexopoulos and Hesterberg (1990) used lagged temperatures to calculate lagged heating and cooling degree days for regression models. One winning team (Ben Taieb & Hyndman, 2014) of the Global Energy Forecasting Competition 2012 (GEFCom2012) used lagged hourly temperature and average daily temperature variables in the competition. Another winning team of GEFCom2012 (Nedellec, Cugliari, & Goude, 2014) used exponentially smoothed temperature variables. Nevertheless, there has never been an in-depth study of different model sizes, to investigate whether large numbers of lagged and average temperature variables can help to improve the forecast accuracy.

Despite the big improvement on the computation side, model building can still take significant amount of time if one wants to test many variables. Indeed, at times people still have to juggle the tradeoff between the frequency of model updates and the sufficiency of the variables. As was discussed by Hong (2010), for example, lagged temperature variables were limited to the past three hours of temperatures due to computational constraints. The limitation of computing power constraints was also an issue in earlier decades, as was discussed by Gross and Galiana (1987).

The methodology presented in this paper is a continuation and extension of the work of Hong (2010). We attempt to take advantage of modern computing power in order to answer a fundamental question:

How many lagged hourly temperatures and/or daily moving average temperatures are needed in a regression model so as to capture the recency effect fully without compromising the forecasting accuracy?

Note that this paper does not explore exponentially weighted temperature variables, primarily in order to avoid discussing algorithms for fine-tuning the exponential weights. For instance, one heuristic method for selecting the base for the exponential weights was discussed by Hong (2010). Other notable discussions of exponential smoothing for electric load forecasting are provided by Taylor and McSharry (2007) and Weron (2006). The lagged temperatures and moving average temperatures covered in this paper can be regarded as a typical representation of "recency", in the sense that we assign a weight of one to each observation in the moving window and a weight of zero to observations outside the moving window. Nevertheless, the proposed framework does not exclude the use of exponentially weighted temperatures.

Following Hong, Wang, and White (2015), we develop a case study based on the load forecasting data from GEFCom2012 published by Hong, Pinson, and Fan (2014a). There are two big-data aspects of this paper: (1) we customize the model for each zone of a geographic hierarchy and each hour of the day; and (2) we take advantage of modern computing power to develop large load forecasting models with thousands of variables. Section 4.2 provides further discussion about big data in load forecasting.

This paper makes the following significant contributions to the load forecasting literature: (1) this is the first comprehensive study on the modeling of the recency effect without computational constraints; (2) this is the first time that the recency effect has been applied to hierarchical load forecasting, in the context of both geographical and temporal hierarchies, where the recency effect is being modeled in a customized fashion for each zone and each hour of a day; and (3) publicly available data are used to conduct the case study, so that future researchers can reproduce our results.

2. Background

2.1. Data description

One of the objectives of GEFCom2012 was to establish a benchmarking data pool to make it easy for researchers in the energy forecasting community to compare models. In this paper, we use the data from the hierarchical load forecasting track of GEFCom2012, which includes 4.5 years of hourly load and temperature across 21 zones (Z_i, $i = 1, 2, \ldots$) of a US utility, of which Z_{21} was the sum of the first 20 zones (Hong, Pinson et al., 2014a). We use the first 4 years of load and temperature data in this paper.

As per the case study by Hong et al. (2015), we are conducting out-of-sample tests instead of selecting models based on in-sample fits. Here, we slice the data to three pieces, the first two years (2004–2005) for training (or in-sample fit, for parameter estimation), the next year (2006) for validation (or post-sample fit, for model selection), and the last full calendar year (2007) for testing (or out-of-sample testing, for a summary of error statistics). The four years of Z_{21} load data are shown in Fig. 1.