Energy and Buildings 109 (2015) 23–34

Contents lists available at ScienceDirect

Energy and Buildings

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

Gated ensemble learning method for demand-side electricity load forecasting

Eric M. Burger*, Scott J. Moura

Energy, Control, and Applications Lab, University of California, Berkeley, USA

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form 9 October 2015
Accepted 9 October 2015
Available online 22 October 2015

Keywords:
Smart grid
Buildings
Electricity forecasting

ABSTRACT

The forecasting of building electricity demand is certain to play a vital role in the future power grid. Given the deployment of intermittent renewable energy sources and the ever increasing consumption of electricity, the generation of accurate building-level electricity demand forecasts will be valuable to both grid operators and building energy management systems. The literature is rich with forecasting models for individual buildings. However, an ongoing challenge is the development of a broadly applicable method for demand forecasting across geographic locations, seasons, and use-types. This paper addresses the need for a generalizable approach to electricity demand forecasting through the formulation of an ensemble learning method that performs model validation and selection in real time using a gating function. By learning from electricity demand data streams, the method requires little knowledge of energy end-use, making it well suited for real deployments. While the ensemble method is capable of incorporating complex forecasters, such as Artificial Neural Networks or Seasonal Autoregressive Integrated Moving Average models, this work will focus on employing simpler models, such as Ordinary Least Squares and *k*-Nearest Neighbors. By applying our method to 32 building electricity demand data sets (8 commercial and 24 residential), we generate electricity demand forecasts with a mean absolute percent error of 7.5% and 55.8% for commercial and residential buildings, respectively.

1. Introduction

Commercial and residential buildings account for 74.1% of U.S. electricity consumption, more than either the transportation sector or the industrial sector (0.2% and 25.7%, respectively) [1]. Maintaining a continuous and instantaneous balance between generation and load is a fundamental requirement of the electric power system [2]. To reliably match supply with demand, the forecasting of grid-level electricity loads has long been a central part of the planning and management of electrical utilities [3]. The accuracy of these forecasts has a strong impact on the reliability and cost of power system operations. Trends, such as vehicle electrification and distributed renewable generation, are expected to pose new challenges for grid operators and may undermine the accuracy of load forecasts.

To improve the accuracy of electricity demand forecasts and aid in the management of power systems, recent attention has been placed on short-term building-level electricity demand forecasting using a wide range of models [4,5]. The ability to accurately and adaptively forecast demand-side loads will play a critical role in maintaining grid stability and enabling renewables integration. Additionally, many novel optimal control schemes, under research umbrellas such as demand response and microgrid management, require short-term building electricity demand forecasts to aid in decision making [6].

The supply-side and load-side time series forecasting of electricity demand has been a topic of research for many decades. The literature is filled with a variety of well-cited modelling approaches, each differing in algorithmic complexity, estimation procedure, and computational cost. Of particular note are the variants of Artificial Neural Networks (ANN) [3–5,7–10], Support Vector Regression (SVR) [11–14] and Autoregressive Integrated Moving Average (ARIMA) models [3,12,13,15–18]. Lesser but nonetheless noteworthy attention has been given to approaches such as Multiple Linear Regression [3,11,19], Fuzzy Logic [3,20], Decision Trees [4], and k-Nearest Neighbors (k-NN).

These studies provide a broad catalog of use-cases and demonstrate the performance of certain forecasting algorithms when applied to specific building types. In particular, [3,4,10,17] provide a survey of electricity forecasting methods and a high-level

* Corresponding author at: Energy, Control, and Applications Lab (eCAL), 611 Davis Hall, Department of Civil and Environmental Engineering, University of California, Berkeley, Berkeley, CA 94720, USA.
E-mail address: ericburger@berkeley.edu (E.M. Burger).

http://dx.doi.org/10.1016/j.enbuild.2015.10.019
0378-7788/

comparison of techniques. Hippert et al. [8] provides a detailed description of ANNs and their application to load forecasting, including data pre-processing and ANN architectures. Jetcheva et al. [5] details the development of a seasonal ANN approach and the advantage over a Seasonal ARIMA (SARIMA) model when applied to 6 building datasets. Newsham and Birt [18] focuses on the introduction of motion sensor data to improve the accuracy of an ARIMA model. In [9,11,15,18,20], the authors perform an in-depth analysis of the power demand patterns of a particular building in order to customize a forecasting model.

In papers with experimental results, the authors have generally applied their electricity demand forecasting technique to only a small number of datasets. Consequently, the literature is rich with forecasting algorithms customized for individual buildings. This leads us to the following question: Is it possible to design a single minimally-customized forecasting algorithm that is widely applicable across a diversity of building types, enabling scalability? We pursue this question by proposing a novel ensemble learning method for electricity demand forecasting.

Specifically, due to unique building characteristics, occupancy patterns, and individual energy use behaviors, we argue that no single model structure is capable of accurately forecasting electricity demand across all commercial and residential buildings.

For example, some forecasting models may produce accurate predictions under certain observable or unobservable conditions, such as a seasonal trend, a morning routine, or an extended absence. Other models may be ideal for buildings with energy use behaviors that are stable over long periods of time. For buildings with frequent changes in occupancy patterns, models that are trained over a moving horizon may yield the highest accuracy. In short, this work will develop an ensemble learning method that trains and validates multiple forecasting models before applying a gating method to select a single model to perform electricity demand forecasting.

In this way, the ensemble method is able to learn from real-time data and to produce short-term electricity demand forecasts that are automatically tailored to a particular building and instance in time. In addition to forecast accuracy, this paper will place an emphasis on method adaptability and ease of use. While we have implemented certain forecasting models, the method is intended to allow the models to be interchangeable.

To demonstrate the use of our ensemble method to produce short-term forecasts, this paper includes 3 experimental studies: Single Model Studies, Multiple Model Study, and Residential Study. For each of these studies, we will make the following assumptions with respect to the availability of building electricity demand data:

A1. We have access to hourly historical building electricity demand at the meter.
A2. We have access to hourly historical weather data near the building location.
A3. We do not have access to submetered electricity demand data or building operations data, such as occupancy measurements or mechanical system schedules.

The limited access to input data with which to produce forecasts is representative of the challenge faced by grid operators. Accordingly, this paper will demonstrate the potential of our ensemble method to non-invasively forecast total electricity demand using data-driven methods. Additionally, unlike in [9,11,15,18,20], where the authors perform an in-depth analysis of the power demand patterns in order to customize a model to a particular building, this paper will focus on developing a forecasting approach that is generally applicable to all buildings without customization.

This paper is organized into five sections: Regression Models, Single Model Studies, Ensemble Method, Multiple Model Study, and Residential Study. Section 2, Regression Models, briefly presents background theory for 5 regression models that will be employed in this paper. In Section 3, Single Model Studies, we apply the forecasting models to 8 commercial/university building electricity demand datasets using batch and moving horizon training approaches. Section 4, Ensemble Method, presents our method for training and validating multiple models and for selecting the optimal model using a gating method. Section 5, Multiple Model Study, applies our ensemble learning method to 8 commercial/university building electricity demand datasets and quantifies and qualifies the advantage over a single model approach. Finally, in Section 6, Residential Study, we apply our ensemble learning method to 24 residential building electricity demand datasets and summarize the results. Key conclusions and future research directions are summarized in Section 7.

2. Regression models

In this paper, we will consider one parametric regression model, Ordinary (Linear) Least Squares with ℓ_2 Regularization (Ridge), and four nonparametric models, Support Vector Regression with Radial Basis Function (SVR), Decision Tree Regression (DTree), *k*-Nearest Neighbors with uniform weights and binary tree data structure (k-NN), and Multilayer Perceptron (MLP), a popular type of feed-forward Artificial Neural Network (ANN). In this section, we will briefly describe the structure of each regression model.

2.1. Ordinary Least Squares with ℓ_2 regularization

Ordinary Least Squares with ℓ_2 regularization (Ridge) fits a linear model with coefficients $w \in \mathbf{R}^n$ to minimize the residual sum of squared errors between the observed and predicted responses while imposing a penalty on the size of coefficients according to their ℓ_2-norm. The linear model of a system with univariate output is given by

$$\begin{aligned}\hat{y} &= w_0x_0 + w_1x_1 + \cdots + w_nx_n \\ &= \sum_k w_kx_k = w^Tx \end{aligned} \tag{1}$$

with variables $x \in \mathbf{R}^n$, the model input, $\hat{y} \in \mathbf{R}$, the predicted response, n, the number of inputs or features in x, and $k = 1, \ldots, n$.

The linear model is trained on a set of inputs and observed responses by optimizing the function

$$\underset{w}{\text{minimize}} \sum_i \|w^Tx_i - y_i\|_2^2 + \lambda\|w\|_2^2 \tag{2}$$

with variables $x_i \in \mathbf{R}^n$, the model input for the ith data point, $y_i \in \mathbf{R}$, the ith observed response, $w \in \mathbf{R}^n$, the weighting coefficients, and $i = 1, \ldots, N$, where N is the number of data samples and n is the number of features in x_i. Lastly, λ is a weighting term for the regularization penalty.

For a system with a multivariate output $\hat{y} \in \mathbf{R}^m$, we will treat the outputs as uncorrelated and define a set of coefficients $w_j \in \mathbf{R}^n$ for each predicted response $\hat{y}_j \in \mathbf{R}$ for $j = 1, \ldots, m$. Thus, the multivariate linear model is given by

$$\hat{y}_j = w_j^Tx, \quad \forall j = 1, \ldots, m \tag{3}$$

The weights of the multivariate model are determined by optimizing the function:

$$\underset{w}{\text{minimize}} \sum_i \sum_j \|w_j^Tx_i - y_{i,j}\|_2^2 + \sum_j \lambda\|w_j\|_2^2 \tag{4}$$

with variables $x_i \in \mathbf{R}^n$, the model input, $y_i \in \mathbf{R}^m$, the observed multivariate response, $w_j \in \mathbf{R}^n$, the weighting coefficients of the jth