Applied Energy 193 (2017) 369–380

Contents lists available at ScienceDirect

Applied Energy

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

Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks

Nastaran Bassamzadeh [a,*], Roger Ghanem [b]

[a] Department of Civil Engineering, University of Southern California, Los Angeles, CA 90089, USA
[b] Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089, USA

HIGHLIGHTS

- A probabilistic load forecasting model using Bayesian networks is proposed.
- Model learns dependencies between variables without making prior assumptions.
- The impact of real time pricing on consumption behavior of customers is studied.
- We investigate model performance at varying spatio-temporal levels of aggregation.

ARTICLE INFO

Article history:
Received 29 August 2016
Received in revised form 3 January 2017
Accepted 11 January 2017

Keywords:
Bayesian networks
Forecasting
Machine learning
Prediction
Probabilistic modeling
Smart grid

ABSTRACT

Demand management in residential buildings is a key component toward sustainability and efficiency in urban environments. The recent advancements in sensor based technologies hold the promise of novel energy consumption models that can better characterize the underlying patterns.

In this paper, we propose a probabilistic data-driven predictive model for consumption forecasting in residential buildings. The model is based on Bayesian network (BN) framework which is able to discover dependency relations between contributing variables. Thus, we can relax the assumptions that are often made in traditional forecasting models. Moreover, we are able to efficiently capture the uncertainties in input variables and quantify their effect on the system output. We test our proposed approach to the data provided by Pacific Northwest National Lab (PNNL) which has been collected through a pilot Smart Grid project.

We examine the performance of our model in a multiscale setting by considering various temporal (i.e., 15 min, hourly intervals) and spatial (i.e., all households in a region, each household) resolutions for analyzing data. Demand forecasting at the individual households' levels is a first step toward designing personalized and targeted policies for each customer. While this is a widely studied topic in digital marketing, few researches have been done in the energy sector. The results indicate that Bayesian networks can be efficiently used for probabilistic energy modeling in residential buildings by discovering the dependencies between variables.

1. Introduction

Smart Grid which represents the future generation of power systems, is composed of multiple interacting systems such as renewable supply network, distribution and storage systems and communication systems. The underlying idea which has entailed this transition is the growing need for optimized energy consumption and management. This goal is sought with the aid of digitized systems and sensors which are capable to monitor the system and collect relevant data at various scales. The information network, ensures the viability of data-driven and data-aware perspectives, both on the side of utilities and consumers. This enables both consumers and utilities to share in the responsibility and benefits of access to advanced technology. Being constituted of multiple stochastic, dynamic and distributed components, the need for an automated and intelligent framework to predict the overall system behavior in Smart Grids is accentuated. In this paper, we aim to enhance our stochastic modeling capability in complex systems by adapting a data-driven approach.

* Corresponding author.
E-mail addresses: Bassamza@usc.edu (N. Bassamzadeh), Ghanem@usc.edu (R. Ghanem).

http://dx.doi.org/10.1016/j.apenergy.2017.01.017
0306-2619/

Predictive modeling such as load forecasting is one of the core problems that is studied in power grids. However, the advent of smart meters in the context of Smart Grids introduces some new aspects to this traditional problem. Smart meters facilitate a real time interaction between power supplier and households in terms of collecting high frequency consumption data from households as well as sending incentive prices to the customers. The utility companies seek to manage demand at the customers' side by communicating incentives to them such as real time electricity prices. The pricing policy is aimed at shifting the demand from peak hours to off-peak hours. Accordingly, we encounter a huge amount of high resolution data collected at various time scales at the households' level. The rich entity of data enables decision makers to design personalized policies for each specific household. Thus, the availability of these high resolution data introduces potential benefits as well as some complexities to predictive problems in Smart Grid.

The particular characteristics of Smart Grids call for a new predictive framework that is able to efficiently capture these features. First, we need a predictive model that well suits high dimensional settings where the number of predictors is large. Historical consumption, weather variables, time parameters and real time price signals are among the contributing variables in predicting demand. Moreover, the availability of data at varying spatial and temporal granularities introduces new challenges in terms of multiscale models and the effect of model resolution on the results. Furthermore, given the detailed customers' consumption data, utility companies are interested in having models at the customers' scale which are able to recommend personalized policies. Finally, quantifying the uncertainty associated with the involved components and formalizing their effect on the prediction result in not a trivial task in this context.

In this paper, we propose Bayesian network framework as a probabilistic tool to model the dependencies between various contributing factors in demand forecasting in the Smart Grid context. Bayesian networks provide a single framework for density estimation, probability propagation and inference in complex systems. It is argued that Bayesian networks are an appropriate tool for probabilistic modeling in complex systems when the number of variables is high. In this regard, we seek to probabilistically predict demand subject to real time prices (as well as other factors) at multiple spatial and temporal scales. The load forecasting problem is studied at 15 min and hourly time resolutions. Here, the goal is to find the optimal time interval for sending, receiving and analyzing the data to/from customers. Besides, we develop a model to predict the demand at the households' scale which can be used as an aid for personalized policy making. We demonstrate the performance of our model using a real data set provided by Pacific Northwest National Lab (PNNL).

Therefore, the main contribution of this paper is to investigate whether Bayesian network model can be applied to sensor based probabilistic load forecasting problem. This modeling framework has significant implications. First, the structure of the model is purely learned from the data and no prior assumptions are inserted. Second, the high number of influencing variables as expected in sensor based models will not pose any modeling challenges, since BNs well adopt to high dimensional settings. Moreover, introducing new variables to the model can be well digested by the algorithm. Third, the full probability distribution of the demand variable can be inferred even if all predictor variables are not observed. Fourth, prediction at high granularity levels, specifically at the households' scale can be conducted which has major benefits for utility companies in designing targeted and personalized policies.

Another novelty of our work owes to the data set that we use. The scarcity of high resolution sensor data at household level makes any such data set worthy. Our data set is consisted of consumption values at 5 min intervals for 25 customers over a year. Also, the real time price that has been communicated to customers is available at 15 min intervals. This provides a unique opportunity to investigate whether pricing policy has been effective in reducing or changing the consumption behavior of customers.

2. Related work

Load forecasting problem has been extensively addressed in the literature. Generally speaking there have been two main lines of research. The first direction of research is based on "parametric" statistical models such as time-series and linear regression (e.g., [1,2],3). The second trend encompasses "nonparametric" models such as support vector regression, neural networks and fuzzy logic models [4,5]. A comprehensive literature survey of load forecasting techniques is presented in [6]. Despite being intuitive, parametric methods often suffer from the limitations posed by the underlying assumptions in their construction. The form of the parametric model is set in advance and thus it can't incorporate unspecified relationships between the target variable and the predictors. Moreover, these models are not generally suited for high dimensional problems (when the number of predictors grow) since they will need a prohibitively large number of parameters to be calibrated. Additionally, these models fail in treating the fine resolution data provided by the smart sensors at households' levels.

Nonparametric models which are heavily based on machine learning algorithms, have gained much popularity for sensor based energy forecasting [7]. In this data-driven approach, data from various sources such as energy smart meters, weather stations and resident occupancies are fed into a machine learning algorithm aiming to find the model that gives the best match between the model output and the observed data. In this category, the two most common algorithms used for sensor based energy forecasting are *artificial neural networks* [8,9] and *support vector machines* [10,11]. A comprehensive study of the performance of various machine learning algorithms for load forecasting is presented in [7].

Scarcity of sensor based data for residential buildings when contrasted to commercial buildings has caused the majority of sensor based energy forecasting literature to be focused on commercial buildings [12]. Moreover, most sensor based models for residential buildings are conducted based on monthly data gathered from monthly utility statements. Therefore, there is a need to explore additional techniques for modeling residential energy consumption with higher granularity data sets and provide more insight regarding appropriate modeling approaches for this problem. Residential load forecasting with higher granularity data sets is studied in [13,14].

While deterministic load forecasting is quite valuable, possible sensor errors, intrinsic weather uncertainties and renewable integration requirements mean that probabilistic load forecasting needs to be adopted more in energy planning and operations. The literature on probabilistic load forecasting (PLF) is quite limited especially when comoared to probabilistic wind power forecasting (PWPF) [15]. However, [16] claims that PLF should be regarded just as important as PWPF in the utility industry. The authors argue that typical 15% error in day ahead wind power forecasting, where wind penetration is around 20%, gives a similar absolute error as 3% error in day ahead load forecasting for a medium sized US utility with an annual peak of $1GW$-$10GW$. It is claimed that reducing the forecasting error by only 1%, in terms of mean absolute percentage error (MAPE), can save hundred thousand dollars per GW peak for a utility company [16]. Thus, it is quite evident that uncertainty quantification and probabilistic