Energy and Buildings 133 (2016) 478–488

Contents lists available at ScienceDirect

Energy and Buildings

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

Applied machine learning: Forecasting heat load in district heating system

Samuel Idowu*, Saguna Saguna, Christer Åhlund, Olov Schelén

Luleå University of Technology, Luleå, Sweden

ARTICLE INFO

Article history:
Received 23 January 2015
Received in revised form
28 September 2016
Accepted 29 September 2016
Available online 8 October 2016

Keywords:
Data driven modeling
District heating
Energy efficiency
Machine learning
Smart cities

ABSTRACT

Forecasting energy consumption in buildings is a key step towards the realization of optimized energy production, distribution and consumption. This paper presents a data driven approach for analysis and forecast of aggregate space and water thermal load in buildings. The analysis and the forecast models are built using district heating data unobtrusively collected from 10 residential and commercial buildings located in Skellefteå, Sweden. The load forecast models are generated using supervised machine learning techniques, namely, support vector machine, regression tree, feed forward neural network, and multiple linear regression. The model takes the outdoor temperature, historical values of heat load, time factor variables and physical parameters of district heating substations as its input. A performance comparison among the machine learning methods and identification of the importance of models input variables is carried out. The models are evaluated with varying forecast horizons of every hour from 1 up to 48 h. Our results show that support vector machine, feed forward neural network and multiple linear regression are more suitable machine learning methods with lower performance errors than the regression tree. Support vector machine has the least normalized root mean square error of 0.07 for a forecast horizon of 24 h.

1. Introduction

The current increase in global demand for energy is a major challenge for energy production and its distribution [1]. To improve energy sustainability, there is a need to approach the consumer end rather than focus at the production side alone. This is an important consideration since a fully optimized energy grid can only be realized with optimization at both the production and the consumer ends [2]. Forecasting total energy consumption is essential for the economical and technical planning of power generation. Also, forecasting energy consumption in buildings is crucial for optimizing its energy use. Energy consumption optimization can either be carried out on a grid scale (e.g., large central accumulators in district heating systems (DHS) and the conversion of electricity energy to thermal energy storage) or on a smaller scale such as in buildings.

A survey of recent methods employed in the estimation of energy-demand, for instance, thermal and electric energy in buildings show that there are two broad categories, the *forward (classical)* and the *data-driven (inverse)* methods [3]. The former uses equations that describe the physical behavior of a system to predict an output, while the latter approach relates to supervised machine learning (ML) methods where measurements of input and output variables of a system are collected, and then used to mathematically describe the system [3]. Recently, several ML methods are used for predicting energy-demand, such as, support vector machine (SVM) [4], multiple regression [5,6] and neural networks based methods [7,8]. In specific to thermal load forecast in DHS, some of the advantages of a data-driven approach over a classical approach include the ability to discover models from large volume of data and the ability to adapt and update models based on new data [9].

A more comprehensive classification of methods used in building load forecast can be broadly grouped into physical models, black-box models and Grey-box models [10,11]. In relation to the two categories mentioned in [3], physical models relate to the classical methods, while black-box models relate to data-driven methods. In comparison to black-box models, physical models are costly and time consuming due to their requirement of very large number of parameters and building or system information as input. Black-box models on the other hand are easy to build and usually require data over a long period for its training purpose. Grey-box models are generally used to address the drawbacks in physical models and the difficulty in determining optimal parameters associated with black-box models. While physical models suffer from poor generalization capabilities, grey-box models provide a balance between high accuracy and good generalization capabilities, by extracting the mathematical model/structure from system's

* Corresponding author.
E-mail addresses: sam.idowu@outlook.com, samup4web@gmail.com (S. Idowu).

http://dx.doi.org/10.1016/j.enbuild.2016.09.068
0378-7788/

physics, and estimation of model parameters from measured data. As a consequence, grey-box models require more effort to develop, have good generalization capabilities in reference to black-box models, and demonstrate higher accuracy compared to white-box models [11].

Several DHS related work are limited to predicting net consumption energy at the production end [12–14,8]. These studies fall under the category of the *top-down* approach [3]. Forecasting thermal load in buildings can be prone to high error rates due to the highly stochastic nature of the consumer load pattern [12]. Grosswindhager et al. [12] stated that it is required to build several individual models for the consumer ends of a DHS due to the stochastic nature of its data. Gadd and Werner [14] presented a novel assessment method which describes daily variations of heat load in DHS. It describes variation in heat load patterns and could be used for the design and planning of storage in a DHS network.

In related work, a number of studies have looked more closely at thermal load forecasting methods for DHS [12,8,7,15,13]. Grosswindhager et al. [12] presented an approach for on-line short term load forecast using *seasonal autoregressive integrated moving average* models in state space representation. Kato et al. [8] proposed a thermal load prediction method which uses a recurrent neural network to deal with the dynamic variations of heat load and its characteristics. The approach shows decent prediction accuracy for non-stationary heat load. Sakawa and Ushiro [7] proposed a load prediction method which is robust enough to handle cases of outliers and missing data. The method uses a simplified robust filter and a three-layered neural networks. Nielsen et al. [15] used a grey-box approach to identify the model that links the heat consumption in a large geographical area to its climate and the calendar information. The process involved a theoretical based identification of an overall building model structure followed by a data based modeling.

Zhou et al. [10] integrated air temperature/relative humidity and solar radiation prediction modules with a grey-box model for hourly building thermal load prediction. The author reported performance of the building load prediction model is satisfactory, while the weather prediction is occasionally affected by temporary weather changes. Afram et al. [11] developed and compared the performance of black-box models and grey-box models of a residential heating, ventilation and air conditioning (HVAC) system. All models performed well and were able to predict the outputs correctly, while black-box model, artificial neural network (ANN) performed overall best in most prediction tasks. Wang and Tian [13] applied wavelet analysis in combination with neural network, and its evaluation shows that the approach is suitable for short-term heat load forecast.

Furthermore, while the top-down approach emphasizes on the total energy consumption in a grid at the production side, the bottom-up approach on the other hand focuses on individual consumption end [3]. A comparative review of recently developed models for the prediction of building energy consumption is provided in Zhao and Magoulès [16]. The identified methods in the review are classified into engineering, statistical, neural networks, support vector machines and grey models. A recent study [17], applied probabilistic methods and heuristics for fault detection and ranking of anomalies. The methods are applied on hourly data from district heating substations. Bacher et al. [18] recently employed a *bottom-up* approach with focus on the consumer end and specifically consider single family buildings. The work used computationally effective recursive least squares scheme with meteorological variables as model input. The model presented provides forecast up to 42 h horizon [18]. Serban and Popescu [19] presented a methodology for prognosis of domestic hot water consumption in DHS using time-series analysis. The work modeled hot water heating load in a block of flats with 60 apartments. The study concluded that time-series analysis is powerful and appropriate for predicting thermal load in DHS. Nikovski et al. [20] proposed a general method for controlling building zone air temperature by setting temperature set-points. The method uses building thermal model based on thermal circuit identified from collected sensor data. The building thermal dynamics was reduced to a markov decision process (MDP) whose decision variables are the sequence of temperature set-points over a suitable horizon. This method saved cost significantly, sometimes exceeding 50% with respect to control strategies in buildings such as the night set-up and demand limiting strategy.

Generally, previous studies consider meteorological variables as major influencing factors to forecasting load in buildings. Likewise, Wang and Tian [13] classifies the thermal load influencing factors in a DHS as *external* and *internal* factors. The former refers to meteorological variables and building occupant's behavioral factors. The latter refers to factors, which relates to the operational characteristics of a district heating substation such as the supply temperature.

The research in this paper is a step further from the preliminary research project described in Idowu et al. [21]. The project aims to achieve a more efficient co-operation of a combined heat and power (CHP) plant and a DHS, where the key focus is to reduce (a) energy consumption, (b) emissions such as CO_2, (c) fossil fuel consumption in CHP plant and (d) peak demands. It will employ energy saving strategy (ESS) in buildings. The ESS will maintain the operation of substations within set values or ranges, e.g., maintaining a set value of return temperature for a specific condition. The forecast of heat demand estimate is key for the optimal control of the ESS.

This paper presents a bottom-up approach to the analysis and forecasting of heat load (building space and domestic hot water) in buildings using ML techniques. The ML methods used are SVM, feed forward neural network (FFNN), multiple linear regression (MLR) and regression tree. The forecasting models are built using district heating data collected in a non intrusive manner from five multi-family and five commercial buildings located in Skellefteå, Sweden, combined with outdoor temperature measurements and forecast. The data collection period is between mid February and early April 2014. The basic model inputs used include outdoor temperature, historical values of heat load, time factor variables and parameters of district heating substations. As a contribution, the application and comparison among the ML methods and the identification of the importance of the internal factors as model input variables is carried out. The forecast model for each building is evaluated with varying forecast horizon up to 48 h. This paper shows a comparative analysis among the energy consumption patterns in building using methods such as partial least square method. The outcome of this paper is intended to be used for the optimal control systems in a building energy management system, which can also be applied in other related work.

The content of this paper is organized as follows. Section 2 presents the background theory of DHS, the description of the target system and theory of ML methods used. Section 3 describes the analysis and modeling processes in this work. Section 4 presents the results while Section 5 presents the discussion. Section 6 presents the conclusion and the future work.

2. Theory

This section presents a technical overview of district heating systems. A high-level description of the target system corresponding to the research project under which this work is done is presented. In this paper, our focus lies mainly on the consumption side. Finally, this section also presents four ML methods used in this paper.