Energy and Buildings 159 (2018) 109–122

Contents lists available at ScienceDirect

Energy and Buildings

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

A novel ensemble learning approach to support building energy use prediction

Zeyu Wang [a], Yueren Wang [a,*], Ravi S. Srinivasan [b]

[a] School of Management, Guangzhou University, Guangzhou, Guangdong, China
[b] M.E. Rinker, Sr. School of Construction Management, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Article history:
Received 27 October 2016
Received in revised form 17 October 2017
Accepted 25 October 2017

Keywords:
Ensemble learning
Bagging trees
Building energy use prediction
Parameter selection

ABSTRACT

Broadly speaking, building energy use prediction can be classified into two categories based on modeling approaches namely engineering and Artificial Intelligence (AI). While engineering approach requires solving physical equations representing the thermal performance of systems and components that constitute the buildings, the AI-based approach uses historical data to predict future performance. Although engineering approach estimates energy use with greater accuracy, it falls short in the overall complexity of model building and simulation in which detailed data that represent the building geometry, systems, configurations, and occupant schedule is needed. Whereas, the AI-based approach offers a rapid prediction of building energy use and, if appropriately trained and tested, may be used for quick and efficient decision-making of energy use reduction. Nevertheless, for robust integration with and to improve automated building systems management and intelligence, the need for consistent, stable, and higher prediction accuracy cannot be understated. To alleviate the instability issue, and to improve prediction accuracy, we have exploited and tested an ensemble learning technique, 'Ensemble Bagging Trees' (EBT), using data obtained from meteorological systems and building-level occupancy and meters.Results showed that the proposed EBT model predicted hourly electricity demand of the test building with improved accuracy of Mean Absolute Prediction Error that ranged from 2.97% to 4.63%. Additionally, results showed that proposed variable selection method could reduce the computation time of EBT by 38–41% without sacrificing the prediction accuracy. The proposed ensemble learning model that exemplifies improved prediction accuracy over other AI techniques can be used for real-time applications such as system fault detection and diagnosis.

1. Introduction

Improving building energy efficiency has become a major concern for energy conservation during the past few decades because buildings, as the main energy consumer, account for more than 41% of U.S. primary energy usage [1]. Particularly, in electricity consumption, the operation of both commercial and residential buildings consumes 74% of all the electricity produced in the U.S. in 2015 [2]. The prediction of building energy usage, as a tool, plays an important role in implementing different building energy efficiency measures, e.g., demand response based Heating, Ventilating, and Air-Conditioning (HVAC) system design and control strategies for effective system fault detection and diagnosis, etc. Accurate building energy prediction can provide benefit to understand the energy behavior of buildings [3]. An effective building energy prediction tool can help owners estimate the potential energy savings and implement adequate building energy management system [4]. Building owners can plan ahead the energy usage over time and potentially shift operation of non-critical equipment to off-peak periods, and make more effective energy purchase plans based on the prediction result [5]. Researchers estimated that building energy system with accurate energy prediction is expected to save 10% to 30% of total building energy usage [6]. Therefore, efforts spent toward improved and effective building energy prediction are meaningful and crucial for building energy efficiency.

1.1. Literature review

A considerable number of research studies have been conducted during the past three decades to develop efficient building energy prediction tools. Based on the applied algorithms and models, these tools can be further classified into three main categories: engineering, machine learning, and hybrid methods [7]. Engineering

* Corresponding author.
E-mail addresses: wangzeyu@gzhu.edu.cn (Z. Wang), wangyueren1@Hotmail.com (Y. Wang), sravi@ufl.edu (R.S. Srinivasan).

https://doi.org/10.1016/j.enbuild.2017.10.085
0378-7788/

method uses physical principles to calculate the thermodynamics and energy behaviors of individual building energy components which can then be aggregated at the whole building level. Since the inference is made based on fundamental thermodynamics, the procedure of computation is strictly proved and analyzed. Hence, engineering method is also known as "white-box" method. On the contrary, machine learning approach is considered as a "black-box" method since the learning process relies on establishing an empirical model to generalize mapping relationship between input and output. While the hybrid method is also known as "gray-box method" since it integrates both engineering and machine learning methods for the purpose of eliminating the limitation of each method.

Despite the success of engineering and hybrid methods in the design phase, they are susceptible to several limitations. The greatest shortcomings are the difficulty in data collection and model development. Notably, both engineering and hybrid methods may require detailed building information, e.g., building geometry, material and component specifications, and HVAC system, lighting and equipment specifications, as their inputs to simulate the inner relation across building components and build the energy prediction model. However, it is difficult and sometimes impossible to collect all data related. Needless to say, this issue is exacerbated in existing buildings. What's more, to construct a theoretical model requires careful engineering work and considerable domain expertise, making it hard to be widely applied. In contrast, machine learning method develops fast energy predictors which predict building energy usage according to its correlated variables, e.g., environmental conditions and occupancy status, which are easy to be obtained. Because of the prominent prediction performance and the light-weight implementation, machine learning method has been widely applied by many researchers to predict building energy usage.

1.1.1. Single prediction model

Traditionally, machine learning based building energy prediction is conducted by using one learning algorithm and training a monolithic model throughout the model development process. This conventional approach is also known as 'single' prediction approach. Various machine learning algorithms such as Multiple Linear Regression (MLR) [8], Artificial Neural Networks (ANN) [9], decision tree [10], and Support Vector Regression (SVR) [11], have been introduced to building energy prediction and provided promising prediction results during the past two decades. However, one of the major disadvantages of single prediction approach is the instability issue within each learning algorithm. Learning algorithms such as ANN and decision tree are unstable learners which may introduce significant variation in the output value due to some small changes made in the input data [12]. This instability issue could impede these algorithms from implementation in real-time, on-the-field applications as some energy efficiency measurements rely on the reliability of the prediction, for example, an unstable learner may lead to high false alarm ratio for building system fault detection. To overcome the limitation of instability as well as to improve the prediction accuracy, the concept of ensemble learning has been recently introduced by researchers to solve both classification and regression problems [13].

1.1.2. Ensemble prediction model

An ensemble prediction model consists of a set of individually trained base models, e.g., neural networks and decision trees, whose outputs are combined to make a prediction [13]. By taking advantage of model complementarity, ensemble prediction model can provide more stable and accurate predictions than the conventional single prediction model [14]. Ensemble prediction method can be further classified into two categories according to the base model generation strategy namely, the heterogeneous ensemble and the homogeneous ensemble. Heterogeneous ensemble model generates its base models by applying the same training data on different learning algorithms or the same algorithms with different parameter settings. Homogeneous ensemble model generates its base models by applying different training data which are resampled from the original data to the same learning algorithm with same parameter settings. In other words, heterogeneous ensemble model improves the prediction by simply taking advantages of the complementarity among different types of learning algorithms, while homogeneous ensemble model is similar to an optimization process that improves the performance of a certain learning algorithm by training it multiple times with varied datasets and combining their predictions. Because of its appealing prediction performance, ensemble prediction approach has received greater attention and become a sought-after research topic in many fields, particularly in data classification [15], disease diagnosis [16], and power system load predictions [17]. However, this novel method is not prevalent in the area of building energy prediction, and a few related studies were not started until 2014 which are discussed below.

Fan et al. [14] developed a data mining based heterogeneous ensemble model to predict next-day energy usage and peak load of a commercial building. The proposed ensemble model comprised of eight base models which were trained independently using different prediction algorithms. A Genetic Algorithm (GA) was used to combine the prediction of each base model and output the final results as the prediction of the ensemble model. Preliminary results showed that the proposed ensemble model provided higher prediction accuracy than the typical single model. Similarly, [18] applied ensemble learning application by integrating multiple AI-based models to predict residential building's cooling and heating loads in building design stage. And, for their research, twelve building types were simulated using an energy simulation software. Besides, six AI models were used as base models for prediction. Eight building features were used as input to predict cooling and heating loads. Research results indicated that the ensemble model was suitable for predict cooling and heating loads. Their research supported the feasibility of using ensemble model to facilitate early designs of energy conservative buildings. More recently, [19] used a neural network-based ensemble model to predict daily heating energy consumption. Three artificial neural networks were used as base models to build the ensemble model. Three different integration methods namely, the simple average, weighted average, and median-based averaging were used to combine the base model and compute the final output for the ensemble model. The results showed that combining different neural networks provided better prediction results than only using the individual neural network.

1.2. Motivation and objectives

It is worth noting that the research on the application of homogeneous ensemble model in building energy prediction is still underexplored. Furthermore, it is evident that heterogeneous ensemble model requires more time and efforts to develop the model due to its complex training, adjustment, and integration strategy. Needless to say, for robust integration with and to improve automated building systems management and intelligence, the need for consistent, stable, and higher prediction accuracy cannot be understated. To alleviate the instability issue, and to improve prediction accuracy, this paper discusses the development and testing of an ensemble learning technique namely 'Ensemble Bagging Trees' (EBT), using data obtained from meteorological systems and building-level occupancy and meters. EBT improves the prediction performance of conventional decision tree method, i.e., Classification and Regression Tree (CART) by introducing bagging technique