83 Predicting Energy Consumption of Office Buildings: A Hybrid Machine Learning-Based Approach

Kadir Amasyali and Nora El-Gohary

Abstract

Improving building energy efficiency requires an understanding of the affecting factors and an assessment of different design and operation alternatives. In this context, accurate prediction of building energy consumption gained a lot of research attention. In recent years, a significant number of building energy consumption prediction models, with various intended uses, have been proposed. However, existing data-driven models are mostly based on outdoor weather conditions, but do not take occupant behavior into account. Towards addressing this research gap, this paper presents a hybrid machine-learning and data-mining approach to develop prediction models that learn from both real data and simulation-generated data. Real data were collected from an office building, including data about building energy consumption, outdoor weather conditions, and occupant behavior. Simulation-generated data were created through simulating an office building in EnergyPlus. A feature selection algorithm was used to determine the critical features in predicting energy consumption for office buildings. A set of regression models were then trained for predicting the hourly values of an outdoor weather-related factor and an occupant behavior-related factor based on these features. Then, an ensembler model—which takes the outputs of the outdoor weather-related factor and occupant behavior-related factor models—was trained to predict cooling energy consumption. In training the models, several machine learning algorithms—such as Gaussian Process Regression (GPR), Support Vector Regression (SVR), Artificial Neural Networks (ANN), and Linear Regression (LR)—were tested. The predicted energy consumption levels showed agreement with the actual levels. This indicates that the proposed regression models can help support decision making related to office buildings.

Keywords

Building energy efficiency • Energy consumption prediction • Machine learning

83.1 Introduction

Energy consumption is on the increase. Numerous research programs and initiatives for reducing energy consumption and improving energy efficiency have been proposed to cope with the climate change and resource depletion issues caused by the increase in energy consumption and associated CO_2 emissions. Building energy consumption represents a significant portion of the primary energy consumption [1]. In recent years, the concern about the rapid increase in energy consumption of buildings due to improved living standards and economic development has triggered building energy efficiency-related researches. Energy efficiency in existing buildings can be achieved in several ways, such as building retrofitting, appliance and equipment upgrade, and improving occupant behavior. Occupant behavior is the actions and decisions taken by building occupants that affect building energy consumption [2]. In this regard, improving occupant behavior comes forward as one of the best ways for reducing building energy consumption, because it aims to improve the efficiency of energy use without sacrifice in people's demands.

K. Amasyali · N. El-Gohary (✉)
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
e-mail: gohary@illinois.edu

K. Amasyali
e-mail: amasyal2@illinois.edu

I. Mutis and T. Hartmann (eds.), *Advances in Informatics and Computing in Civil and Construction Engineering*,
https://doi.org/10.1007/978-3-030-00220-6_83

Previous studies [e.g., 3–7] investigated the impact of energy use behavior through empirical and theoretical methods and showed that behavior has a significant impact on building energy consumption. The prediction of building energy consumption that takes occupant behavior into account is thus fundamental for discovering more efficient building operation and occupant behavior-related strategies. However, existing data-driven building energy prediction efforts did not sufficiently take occupant behavior into account, and therefore cannot be utilized for improving occupant behavior.

There are two main approaches in developing data-driven prediction models: learning from simulation-generated data (e.g., using building energy simulations such as EnergyPlus and eQuest to generate data to train models) and learning from real data (using meters, sensors, and building management system data to train models). However, each approach has its own limitations and strengths. On one hand, simulation-generated data, in many cases, are limited in representing the complexity and stochastic nature of occupant behavior but can significantly reduce the data sensing efforts to collect energy consumption-related data. On the other hand, developing data-driven models that learn from real data requires a significant amount of training data to develop an accurate model, but can better capture the occupant behavior and therefore can better incorporate the impact of occupant behavior on building energy consumption.

Towards addressing the limitations of each approach, the authors are proposing an occupant behavior-sensitive energy consumption prediction model. To leverage the strengths of the both approaches in developing data-driven prediction models, a hybrid model learns from both simulation-generated data and real data was developed. The hybrid model consists of three base models: (1) a machine learning model that learns the impact of outdoor weather conditions from simulation-generated data, (2) a machine learning model that learns the impact of occupant behavior from real data, and (3) an ensembler model that predicts cooling energy consumption based on the outputs of both models.

83.2 Background

A body of research efforts has been already undertaken towards developing data-driven building energy consumption prediction models. On one hand, some research efforts utilized simulation generated data to train the models. For example, Li and Huang [8] developed short-term load prediction models using the data generated by TRNSYS simulations. On the other hand, some other research efforts utilized real data. For example, Wang et al. [9] developed hourly electricity consumption prediction models using the data collected from two educational buildings. Despite the importance of all these efforts, two primary knowledge gaps are identified. First, there is a lack of studies on data-driven and occupant behavior-sensitive building energy consumption prediction. The majority of existing efforts in the area of data-driven building energy consumption prediction either did not take occupant behavior into account at all or they did some limited efforts (e.g., taking schedules into account). Occupant behavior, as a result, still remains as one of the greatest uncertainties in building energy consumption prediction [10]. An accurate prediction model that can incorporate the impact of occupant behavior can be utilized for both discovering more efficient building operation and occupant behavior-related strategies. Second, the existing efforts either utilized simulation-generated data or real data to train the prediction models. Given that both approaches have their own advantages and disadvantages, a hybrid model that is trained using both simulation-generated data and real data can leverage the advantages of both approaches.

83.3 Methodology

A hybrid machine learning model, which learns both form simulation-generated data and real data, was developed. The hybrid model consists of three base models: (1) a machine learning model that predicts the hourly values of the weather-related factor, (2) a machine learning model that predicts the hourly values occupant behavior-related factor, and (3) an ensembler model that predicts cooling energy consumption based on the predicted values for both factors. The weather-related factor represents the impact of the outdoor weather conditions on cooling energy consumption, at a specific hour. The occupant behavior-related factor represents the impact of the occupant behavior on cooling energy consumption, at a specific hour. The simulation-generated data was utilized to train the weather-related factor prediction model, because a simulation environment allows for generating data in which the consumption is impacted by outdoor weather conditions only. On the other hand, the real data was utilized to train the occupant behavior-related factor prediction model, because the stochastic and complex nature of occupant behavior can be better captured in a real-world setting. Finally, an ensembler model that takes the weather-related and occupant behavior-related factors as features was developed to predict cooling energy consumption, and was evaluated in terms of prediction accuracy.

83.3.1 Weather-Related Factor Prediction Model Development

The development of the weather-related factor prediction model included three primary steps: energy simulations, data preprocessing and feature selection, and factor prediction model development.

EnergySimulations. A simple, one-story square office building with five thermal zones was modelled to generate the data needed for training the weather-related factor prediction model. The building model was simulated using EnergyPlus [11], a widely-used whole building energy simulation program. To understand the impact of outdoor weather conditions on building energy consumption, all the other energy consumption-related parameters (e.g., occupant behavior, operation schedule) were kept constant throughout the simulation period—the outdoor weather conditions were the only variables. The simulations were conducted from June 1 to August 31 using the typical meteorological year 3 (TMY3) weather data of Philadelphia, PA. In order to have an undisturbed consumption pattern throughout the simulation period, the holiday schedules in EnergyPlus were removed. For this pilot study, Philadelphia was chosen to match the location of the building where the real data were collected (see Sect. 83.3.2).

Data Preprocessing and Feature Selection. Four data preprocessing and feature selection steps were conducted: data cleaning and aggregation, feature selection, data normalization, and data splitting. First, the weekend and non-working hours on weekdays, where the building is unoccupied and cooling energy consumption is zero, were removed from the dataset. Also, the outdoor weather condition variables (e.g., solar radiation) that are not available for real data were removed from the simulation-generated data as well. Second, using the remaining variables, feature selection was carried out, using a Neighborhood Component Analysis (NCA), to remove the redundant and non-discriminating features from the dataset. As a result, the final, remaining features were: temperature, dewpoint temperature, wind speed, wind direction, and atmospheric pressure. Third, the features were normalized using their means and standard deviations to avoid overflowing of an individual feature. Then, in order to obtain a factor representing only the impact of weather conditions on building energy consumption, the hourly cooling energy consumption levels generated by the EnergyPlus simulations were normalized from 0 to 1, where 0 represents the minimum weather condition impact and 1 represents the maximum weather condition impact on cooling energy consumption. Fourth, the real data were split into training (90%) and validation (10%) datasets.

Factor Prediction Model Development. A weather-related factor prediction model was developed using the training dataset. In developing the model, the following four algorithms were tested: gaussian process regression (GPR), support vector regression (SVR), artificial neural networks (ANN), and linear regression (LR). The parameters of all these algorithms were tuned through parameter grid search using the validation dataset to maximize the prediction performance.

83.3.2 Occupant Behavior-Related Factor Prediction Model Development

The development of the occupant behavior-related factor prediction model included four primary steps: building instrumentation, data collection, data preprocessing, and factor prediction model development.

BuildingInstrumentation. The real data were collected from the Philadelphia Business and Technology Center (PBTC) building between October 5, 2015 and March 31, 2018. The PBTC is a 6-story masonry office building with an estimated total floor area of 272,000 ft^2. The west wing of the 4th floor was already instrumented for empirical data collection as part of an earlier research project by the Consortium for Building Energy Innovation (CBEI) [12]. The instrumented area is 10,000 ft^2, which consists of 12 offices and two thermal zones. The instrumented area is occupied on weekdays from 8 am to 5 pm, mostly. The building uses electricity for cooling.

Data Collection. Cooling energy consumption was metered in 15-min intervals using the power meters installed on the air handling units (AHUs) and monitored using the PI CoreSight web-based application. There are two AHUs. Each thermal zone has an AHU for cooling. Outdoor weather condition data were gathered from a weather station at the Philadelphia International Airport in hour intervals [13]. Occupant behavior data were captured through a preference monitoring application (PMA) [14]. The PMA was developed using an online survey tool to capture the actions taken by the occupants. The actions included turning on/off a portable heater, opening/closing a door, opening/closing a shading device, and turning on/off a light. The occupants were asked to provide feedback whenever they have taken an action.

Data Preprocessing. Five data preprocessing steps were conducted: data cleaning, data aggregation, data integration, data transformation, and data splitting. First, the data instances that have missing and/or outlier values were removed from the dataset. Non-summer months were also removed from the dataset. Second, the hourly cooling energy consumptions of the instrumented area were calculated by summing the hourly energy consumptions of the AHUs; and the 15-min cooling energy consumption data intervals were aggregated into hourly consumption levels. Third, data from multiple sources—including energy consumption data, outdoor weather conditions data, and occupant behavior data—were integrated using their date and time. Fourth, a principal component analysis was conducted to estimate the weather-normalized cooling energy consumption. The weather normalization aimed to remove the impact of weather conditions on energy consumption, and therefore all the energy consumption levels were transformed into the estimated energy consumption at the outdoor temperature level in the center of the first principal component (PC1). For the transformed cooling energy consumption levels, the outdoor weather conditions are expected to have minimum impact on cooling energy consumption due to the weather normalization. Since weather conditions have minimum impact on the transformed energy consumption data, the variance in the transformed energy consumption data can mostly be explained by other energy consumption-related parameters (e.g., occupant behavior). Finally, the real data were split into training (70%), validation (10%), and testing (20%) datasets.

Factor Prediction Model Development. An occupant behavior-related factor prediction model was developed using the training dataset. In developing the model, the same aforementioned four algorithms (see Sect. 83.3.1) were tested, and their parameters were tuned through parameter grid search.

83.3.3 Ensembler Model Development

An ensembler model was developed to predict cooling energy consumption. The weather-related factor and occupant behavior-related factor predictions were preprocessed and an ensembler model—which takes these predictions as features—was developed. Prior to the machine learningprocess, both types of features were normalized. In developing the ensembler model, the same algorithms (see Sect. 83.3.1) were tested, and their parameters were tuned through parameter grid search.

83.3.4 Performance Evaluation

The performances of the two factor prediction models were evaluated using the validation datasets. The hybrid model was evaluated using the testing dataset. The prediction performance was evaluated using coefficient of variation (CV). According to the ASHRAE Guideline 14, an hourly prediction model is considered as calibrated if hourly CV values fall below 30%. CV was calculated using Eq. (83.1).

$$\mathrm{CV}\ (\%\ ) = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(y_{predict,i}-y_{data,i}\right)^2}{n}}}{\bar{y}_{data}} \times 100 \tag{83.1}$$

where $y_{predict,i}$ is the predicted target at hour i, $y_{data,i}$ is the actual target at hour i, n is the number of hours in the validation/testing dataset, and $\bar{y}_{data}$ is the average target data.

83.4 Preliminary Results and Discussion

83.4.1 Weather-Related Factor Prediction

The simulation of the building model generated 2208 h of cooling energy consumption data. Table 83.1 presents the summary of the simulation-generated data, after feature selection but prior to data normalization and splitting. The fine-tuned GPR, SVR, ANN, and LR models achieved 18.78, 20.19, 20.42, and 31.01% CV, respectively, in predicting the values of the weather-related factor. Although the GPR model was the most accurate model, and was therefore selected, there was no clear outperformer across the GPR, SVR, and ANN. The predicted weather-related factor was used as one of the features of the ensembler model.

83.4.2 Occupant Behavior-Related Factor Prediction

As shown in Fig. 83.1, the data is scattered along PC1; there is more variation on PC1 than PC2. This can be explained by the very dominating impact of temperature on cooling energy consumption. As expected, the cooling energy consumption increased as the temperature increased. As illustrated in Fig. 83.1, all the data were transformed to the dashed vertical line which crosses the temperature in the center of PC1. The transformed data represent the occupant behavior-related factor. The upper levels of the vertical line indicate the higher occupant behavior-related factors, and vice versa. The fine-tuned GPR, SVR,

Table 83.1 Summary of the simulation generated data

Feature	Min	Mean	Median	Max
Temperature (F)	53.96	74.84	75.02	98.06
Dewpoint temperature (F)	42.08	63.31	64.94	78.08
Wind speed (mph)	27.00	70.14	72.00	100.00
Wind direction (deg)	0.00	206.72	220.00	360.00
Atmospheric pressure (mb)	0.00	8.35	8.05	21.92

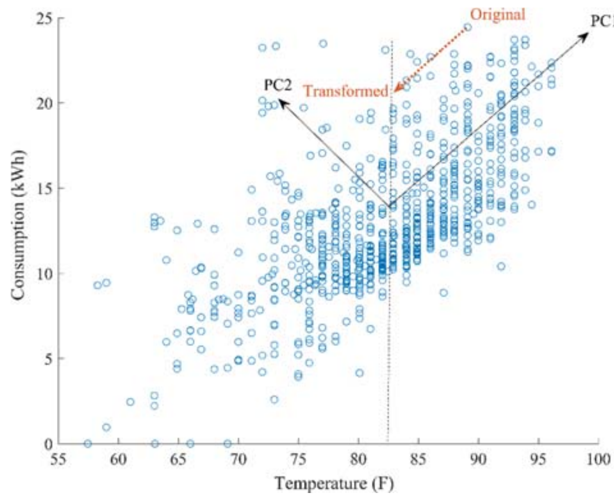

Fig. 83.1 Illustration of principal components and weather normalization

ANN, and LR models achieved 18.55, 18.19, 18.24, and 34.58% CV, respectively, in predicting the values of the occupant behavior-related factor. Similar to the weather-related factor prediction, there was no clear outperformer across GPR, SVR, and ANN. The predicted occupant behavior-related factor was used as one of the features of the ensembler model.

83.4.3 Ensembler Model Prediction

As shown in Fig. 83.2, the predicted cooling energy consumption (predicted by the proposed hybrid model) showed good fitness with the actual consumption (collected from the office building). The fine-tuned GPR, SVR, ANN, and LR models achieved 18.89, 18.78, 19.01, and 19.55% CV, which is considered calibrated according to ASHRAE Guideline 14.

For comparative purposes, a typical prediction model, which does not take occupant behavior into account and learns only from real data, was also developed (using the same dataset). The model achieved 30.21% CV, which is significantly lower than the performance of the proposed hybrid model. These results, thus, indicate that the proposed hybrid model is promising and can be utilized for better understanding and improvement of occupant behavior. The proposed model can be used to discover more efficient building operation and occupant behavior-related strategies under a set of given weather conditions.

83.5 Conclusion

In this paper, the authors proposed a hybrid model, which learns both from simulation-generated data and real data, for predicting cooling energy consumption of an office building. The hybrid model consists of three base models: (1) a machine learning model that learns the impact of outdoor weather conditions from simulation-generated data, (2) a machine learning model that learns the impact of occupant behavior from real data, and (3) an ensembler model that predicts cooling energy consumption based on weather-related and occupant behavior-related factors predicted by the first two models. The model was validated using a testing dataset collected from an office building in PA. The prediction results showed that the proposed model has the potential to be successfully used for better understanding and improvement of occupant behavior.

In future work, the authors will model and simulate a set of new buildings with different geometries and different building properties and retrain the weather-related factor prediction model to learn from this extended set of data. Also, the occupant

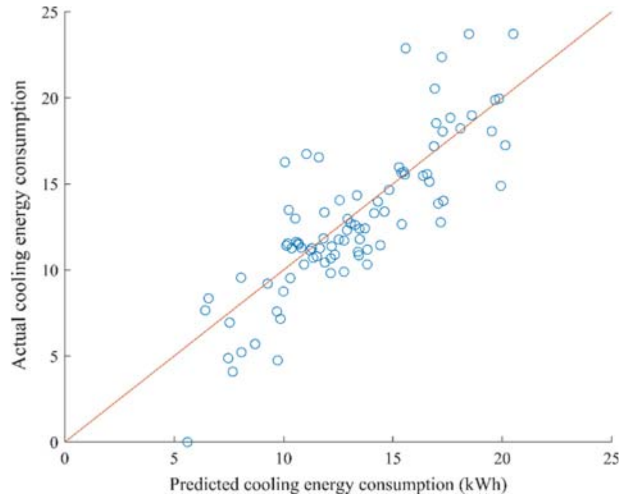

Fig. 83.2 Prediction results of the hybrid model

behavior-related factor prediction model will further be improved through including additional types of occupant behaviors. Currently the authors are conducting a set of empirical energy studies in residential and office buildings to collect such real data, which include energy consumption, indoor environmental conditions, outdoor weather conditions, and occupant behavior data.

Acknowledgements This publication was made possible by NPRP Grant #6-1370-2-552 from the Qatar National Research Fund (a member of Qatar Foundation). The findings achieved herein are solely the responsibility of the authors. The authors would like to thank the Philadelphia Business and Technology Center (PBTC) and the Penn State Consortium for Building Energy Innovation (CBEI) for providing access to building energy data, and Prof. Chimay Anumba and Yewande Abraham for helping the authors with the data collection.

References

1. Zhao, Z., Amasyali, K., Chamoun, R., El-Gohary, N.: Occupants' Perceptions about Indoor Environment Comfort and Energy Related Values in Commercial and Residential Buildings. Procedia Environ. Sci. **34**, 631–640 (2016). https://doi.org/10.1016/J.PROENV.2016.04.055
2. Klein, L., Kwak, J., Kavulya, G., Jazizadeh, F., Becerik-Gerber, B., Varakantham, P., Tambe, M.: Coordinating occupant behavior for building energy and comfort management using multi-agent systems. Autom. Constr. **22**, 525–536 (2012). https://doi.org/10.1016/j.autcon.2011.11.012
3. Clevenger, C.M., Haymaker, J.R., Jalili, M.: Demonstrating the impact of the occupant on building performance. J. Comput. Civ. Eng. **28**, 99–102 (2014). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000323
4. Azar, E., Menassa, C.C.: A comprehensive analysis of the impact of occupancy parameters in energy simulation of office buildings. Energy Build. **55**, 841–853 (2012). https://doi.org/10.1016/j.enbuild.2012.10.002
5. Petersen, J.E., Shunturov, V., Janda, K., Platt, G., Weinberger, K.: Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. Int. J. Sustain. High. Educ. **8**, 16–33 (2007). https://doi.org/10.1108/14676370710717562
6. Delmas, M.A., Fischlein, M., Asensio, O.I.: Information strategies and energy conservation behavior: a meta-analysis of experimental studies from 1975 to 2012. Energy Policy. **61**, 729–739 (2013). https://doi.org/10.1016/j.enpol.2013.05.109
7. Jain, R.K., Taylor, J.E., Culligan, P.J.: Investigating the impact eco-feedback information representation has on building occupant energy consumption behavior and savings. Energy Build. **64**, 408–414 (2013). https://doi.org/10.1016/j.enbuild.2013.05.011
8. Li, Z., Huang, G.: Re-evaluation of building cooling load prediction models for use in humid subtropical area. Energy Build. **62**, 442–449 (2013). https://doi.org/10.1016/j.enbuild.2013.03.035
9. Wang, Z., Wang, Y., Zeng, R., Srinivasan, R.S., Ahrentzen, S.: Random forest based hourly building energy prediction. Energy Build. (2018). https://doi.org/10.1016/j.enbuild.2018.04.008
10. O'Brien, W., Gunay, H.B.: Mitigating office performance uncertainty of occupant use of window blinds and lighting using robust design. Build. Simul. **8**, 621–636 (2015). https://doi.org/10.1007/s12273-015-0239-2
11. Yousefi, F., Gholipour, Y., Yan, W.: A study of the impact of occupant behaviors on energy performance of building envelopes using occupants' data. Energy Build. **148**, 182–198 (2017). https://doi.org/10.1016/j.enbuild.2017.04.085
12. CBEI: Consortium for Building Energy Innovation, http://cbei.psu.edu/. Accessed 26 Apr 2018
13. Pennsylvania State Climatologist: Pennsylvania State Climatologist, http://climate.psu.edu/. Accessed 26 Apr 2018
14. Abraham, Y., Zhao, Z., Anumba, C., Asadi, S.: Implementation of a preference monitoring application for office building occupants. In: Lean and Computing in Construction Congress—Volume 1: Proceedings of the Joint Conference on Computing in Construction, pp. 793–801. Heriot-Watt University, Heraklion (2017)