91 Machine-Learning-Based Model for Supporting Energy Performance Benchmarking for Office Buildings

Lufan Wang and Nora M. El-Gohary

Abstract

Buildings are dominant contributors of global energy consumption. Enhancing building energy efficiency has long been recognized as an important way to achieve energy saving goals and sustainability targets. In response, a body of building energy performance benchmarking models and tools have been proposed during the past decades. The degree of similarity between the compared buildings is the core of the benchmarking process. However, existing benchmarking tools mainly classify buildings only based on building use types, instead of fully considering a wider range of impacting factors. To address this gap, this paper proposes a machine-learning (ML)-based model for classifying buildings—based on building characteristics, occupant behaviors, and geographical and climate features—into three energy-consumption levels: low, medium, and high. Support vector regression models are then fitted to define the predicted energy consumption for benchmarking. The proposed ML-based building energy consumption prediction model was tested on the office buildings in the commercial building energy consumption survey (CBECS) dataset. Principal component analysis (PCA) was used for data dimensionality reduction and feature extraction. Different ML algorithms were tested and compared, including Naïve Bayes (NB), support vector machines (SVM), decision trees (DT), and random forests (RF). The classification algorithms were evaluated in terms of precision and recall; the regression models were evaluated in terms of root mean square error; and the energy consumption prediction results were further compared with the prediction results by EnergyStar. The performance results indicate that, compared with EnergyStar, the proposed model can reduce the prediction error by 13%.

Keywords

Energy benchmarking • Energy consumption • Classification • Machine learning • Office buildings

91.1 Introduction

Human beings spend an average of 87% of their time inside buildings [1], which indicates a large amount of energy demand for supporting indoor activities and thermal comfort. With the rapid urbanization around the world, enhancing building energy efficiency is becoming more crucial for achieving energy saving goals and global sustainability targets. Energy benchmarking is a key first step to understand the energy efficiency of buildings. It evaluates the energy consumption of an individual building by comparing it to its peers. The benchmarking result can provide important decision support for identifying energy conservation opportunities, when the assessed building consumes more energy than other similar buildings.

L. Wang (✉) · N. M. El-Gohary
Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign,
205 North Mathews Ave., Urbana, IL 61801, USA
e-mail: lwang105@illinois.edu

N. M. El-Gohary
e-mail: gohary@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_91

A number of building energy performance benchmarking models and tools have been proposed during the past decades, including energy performance indices (e.g., energy usage intensity [2]), point-based rating systems (e.g., Leadership in Energy and Environmental Design (LEED) [3]), simulation-based systems (e.g., EnergyPlus [4]), and statistical models (e.g., ordinary least squares models [5] and EnergyStar [6]). Despite the importance of the existing efforts, the way they select similar buildings for benchmarking is limited. The degree of similarity between the compared buildings is the core of the benchmarking process, because the energy performance of the target building needs to be compared to that of other similar buildings. However, existing benchmarking models/tools only classify buildings based on a single characteristic. For example, EnergyStar classifies buildings only by building use types, instead of fully considering a wider range of impacting factors, which can provide unreliable results when assessing the building energy performance [7].

To address this gap, this paper proposes a machine-learning (ML)-based model for supporting building energy consumption benchmarking. The proposed model first classifies buildings into three energy consumption levels—low, medium, and high—based on a range of features (e.g., building characteristics, occupant behaviors, and geographical and climate features). Separate support vector regression models are then fitted to define the predicted energy consumption for each class. The proposed model considers the impacts of a large amount of features that could impact building energy performance, and discovers the underlying similarities between the buildings, which could provide more robust prediction results and can better assist the building stakeholders in improving building energy efficiency. To limit the scope of this paper, only the electricity consumption of office buildings was considered.

In the remainder of this paper, Sect. 91.2 briefly introduces the background. Section 91.3 presents the research methodology for the proposed model. Section 91.4 discusses the preliminary results. Finally, Sect. 91.5 summarizes the conclusions and future work.

91.2 Background

Numerous building energy benchmarking models/tools have been developed since the 1990s. The existing energy benchmarking methods can be divided into five categories: energy performance indices, point-based rating systems, simulation-based systems, statistical models, and ML-based models.

Energy performance indices are commonly obtained by normalizing the building energy use relative to a primary determinant of energy use [8]. For example, energy use intensity (EUI) provides normalized energy use per building floor area per year. However, such simple indices can be unreliable because they only consider a single factor (e.g., building floor area) without accounting for other impacting factors.

Point-based rating systems, such as LEED, use predefined standards or guidelines to measure the energy efficiency of a building. However, they do not allow comparisons against other buildings, and the scoring system without considering actual energy consumptions can be misleading [7].

Simulation-based systems compare the actual building energy consumption with the energy use of simulated buildings. Although this method considers a wide range of impacting factors, the simulation results might be inaccurate because of insufficient calibration to the actual building data.

Statistical models generate a regression line between the actual energy consumption and its impacting factors, like in EnergyStar [6]. However, such models are highly sensitive to outliers, which could provide wrong conclusions because of the skewed regression line.

Recently, some efforts have used ML-based models (e.g., artificial neural networks [9]) for energy benchmarking. These models aim to learn from data of similar buildings. However, they select similar buildings based on building type only (e.g., office, residential).

As such, all five categories of models have a common drawback: they only define similarity of buildings (i.e., classify buildings) based on a single characteristic (e.g., building type), which can provide unreliable results when predicting the energy consumption of a building. Therefore, there is a need for a building classification and energy consumption prediction model for supporting a more accurate energy performance benchmarking.

91.3 Research Methodology for the Proposed ML-Based Model for Supporting Energy Performance Benchmarking

The proposed ML-based building classification and energy consumption prediction model for supporting energy performance benchmarking is composed of four main steps (as per Fig. 91.1): (1) data preparation; (2) feature extraction; (3) model development; and (4) model evaluation.

91.3.1 Data Preparation

Feature Prescreening and Data Cleaning. Feature prescreening aims to remove the data attributes that are irrelevant, redundant, and/or with mostly missing values. Each data sample has a total of 516 data attributes (not including imputation flags and weights). But they contain a large amount of irrelevant and redundant attributes, as well as attributes with many missing values. For example, the attribute "natural gas used for cooling" is irrelevant, because the scope of this study is focused on electricity consumption. The attribute "computer used" is redundant in the presence of the attribute "number of computers". And, 90% of the data samples have an empty value for the attribute "central plant in building". After feature screening, the original 516 data attributes were reduced to 193 attributes. The attributes are composed of building characteristics (e.g., square footage, year of construction, and floor to ceiling height), occupant behaviors (e.g., principal building activity, number of businesses, and percent occupancy), and geographical and climate features (e.g., census division, heating degree days, and cooling degree days).

Data cleaning then aims to removes the data entries with incomplete building information or with zero electricity consumption data, as well as duplicate data entries. Only three building samples were removed as a result of data cleaning.

Data Transformation and Standardization. The dataset includes a mixture of categorical and continuous data attributes. For example, the attributes "wall construction material", "main heating equipment", and "census division" are categorical, while the attributes "total hours open per week", "square footage", and "heating degree days" are continuous. Data transformation and standardization aims to transform the mixed-type data into a standard format for the training and testing of the ML models.

For categorical data, a standard one-hot encoding method was conducted. The data samples with n categories were expanded into $n-1$ dummy variables. For example, for the feature $x_i = k(k \in 1, 2, 3, \ldots, n)$, x_i was expanded to $n-1$ dummy variables, where the kth variable is 1, and the rest is 0.

Continuous data have different magnitudes. For example, the gross floor area of the buildings ranges from 1001 to 1.5 million square feet. The high-valued variables are likely to dominate the prediction [10]. Therefore, the datasets need to be normalized, to scale the continuous data with different magnitudes down to a common scale. In this study, the continuous data were normalized by using the min-max normalization method, as shown in Eq. (91.1), where x_i' and x_i are the normalized and original data, respectively; and $\max(x)$ and $\min(x)$ are the maximum and minimum values of x, respectively.

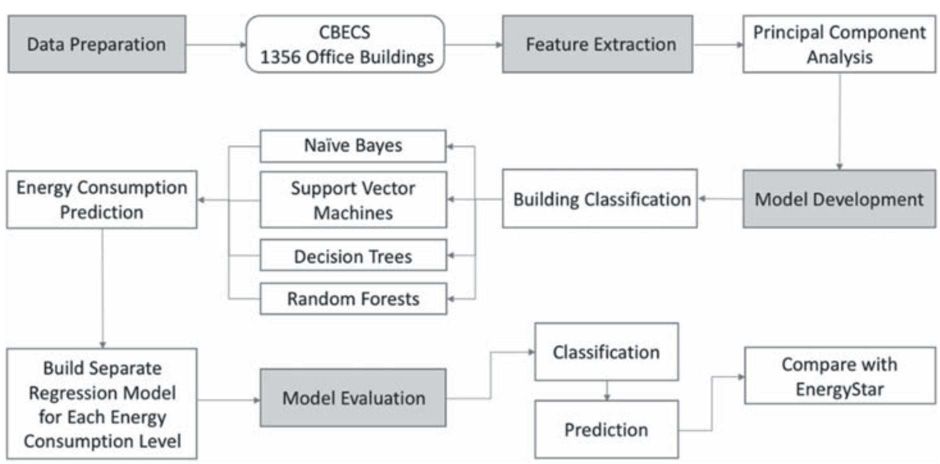

Fig. 91.1 Proposed ML-based building energy performance benchmarking framework

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{91.1}$$

Energy Consumption Discretization. The continuous target value is the total annual electricity consumption, which needs to be discretized to a few categories for building classification. In this study, three energy consumption levels were defined, i.e., low, medium, and high. Quantile-based discretization method was used for dividing the continuous target value into the three intervals. Figure 91.2 shows a boxplot of the logarithmic electricity consumption for all the three energy consumption levels.

91.3.2 Feature Extraction

All the data attributes can be used as features of the ML-algorithms. As the feature space was largely expanded by the one-hot encoding in the data transformation step, the high dimensional dataset and the large amount of correlated features could deteriorate the prediction performance of the ML-algorithms. Feature extraction aims to reduce both the curse of dimensionality and the multicollinearity deficiency of the data. This study used principal component analysis (PCA) for feature extraction and dimensionality reduction.

PCA projects the raw features to a lower dimensional space by using singular value decomposition (SVD). It thus constructs a series of linear combinations of the original features. The results of PCA are a set of independent orthogonal vectors, which are called principal components (PCs). The PCs maintain the major characteristics of the original features, and are ranked based on the magnitudes of their corresponding eigenvalues [11]. For example, the first PC is the most important PC, which has the largest eigenvalue and explains the largest possible variance of the original dataset. Each subsequent PC represents the largest remaining variance. By only using the several PCs that rank first, the number of features can thus be reduced.

91.3.3 Model Development

The ML-based model is composed of two parts: (1) building classification and (2) energy consumption prediction.

Building Classification. This step aims to classify the buildings—based on the building characteristics, occupant behaviors, and geographical and climate features—into three predefined classes: low-, medium-, and high-energy-consumption buildings. This research selected four of the most-popular ML algorithms, including Naïve Bayes (NB), support

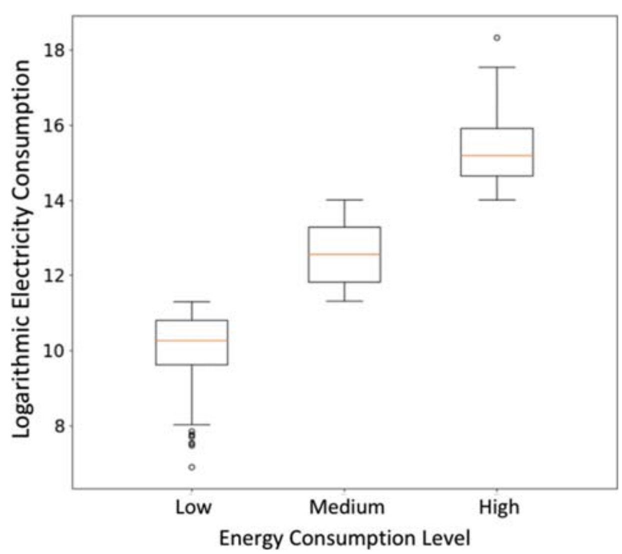

Fig. 91.2 Boxplot of logarithmic electricity consumption for the three levels

vector machines (SVM), decision trees (DT), and random forests (RF). The dataset was randomly partitioned into two—90% training and 10% testing. The aforementioned ML algorithms were implemented using the Scikit-learn module [12] written in Python programming language.

Energy Consumption Prediction. This step develops separate regression models for each of the three classes. When testing, the class label of each testing sample is first determined, then the regression model of the corresponding class is used to determine the target energy consumption value. Support vector regression (SVR) was selected for developing the regression models, because it has been found in other studies (e.g., [13–15]) to outperform other machine learning algorithms (i.e., artificial neural networks, multiple linear regression, etc.) in solving the nonlinear energy consumption prediction problem.

91.3.4 Performance Evaluation

The model was evaluated based on the testing data. The performance evaluation includes the evaluation of the classification algorithms, the regression models, and the overall energy consumption prediction results. The classification algorithms were evaluated in terms of precision and recall; the regression models were evaluated in terms of root mean square error (RMSE); and the prediction results were further compared with the prediction results generated by EnergyStar.

Precision indicates the proportion of the positive identifications that were actually correct. Recall indicates the proportion of the actual positives that were identified correctly. RMSE indicates how close the model's predicted values are to the observed data points. The definitions of these three metrics are shown in Eqs. (91.2)–(91.4), where TP, FP, FN refers to true positive, false positive, and false negative class predictions, respectively; y_i and $\hat{y}_i$ are the actual and predicted energy consumption value at the ith data point, respectively; and N is the total number of data points in the dataset.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{91.2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{91.3}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{91.4}$$

The performance of the overall energy consumption prediction results was evaluated by comparing it to that of the prediction results generated by EnergyStar. EnergyStar is the best-known and most technically-robust building energy benchmarking tool [7]. It is developed based on an ordinary least squares (OLS) regression model. The OLS model for benchmarking office buildings developed by EnergyStar only considers the following features: gross floor area, weekly operating hours, number of workers, number of computers, percentage of area heated, percentage of area cooled, and heating/cooling degree days. Therefore, for this step, the same set of features were used to generate an OLS model. And the prediction results by the proposed ML-based model and the OLS model were compared in terms of RMSE.

91.4 Preliminary Results and Discussion

Figure 91.3 shows the cumulative percentage of explained variance of the PCs. As the feature space has been expanded to 655 features by the data transformation step, PCA produced a total of 655 PCs. As shown in Fig. 91.3, the top 112 PCs can represent more than 85% of the variance. Therefore, only the top 112 PCs were selected as the new features for model development. Figure 91.4 shows the percentage of explained variance by each selected PC.

Table 91.1 presents the performance of the building classification algorithms. Comparing the algorithms, SVM performed best in terms of both precision and recall. It achieved an average precision and recall of 86%. Decision trees performed the worst, with an average precision and recall 16% lower than SVM. Naïve Bayes and random forests achieved a similar performance, but random forests performed slightly better. The high performance of the SVM algorithm indicates that SVM is suitable for handling and capturing the nonlinearity of the building classification problem.

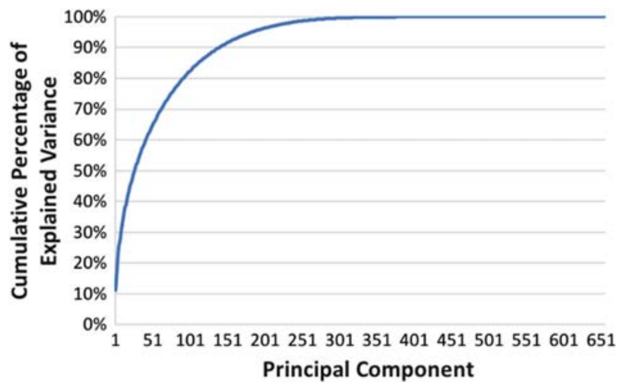

Fig. 91.3 Cumulative percentage of explained variance of principal components

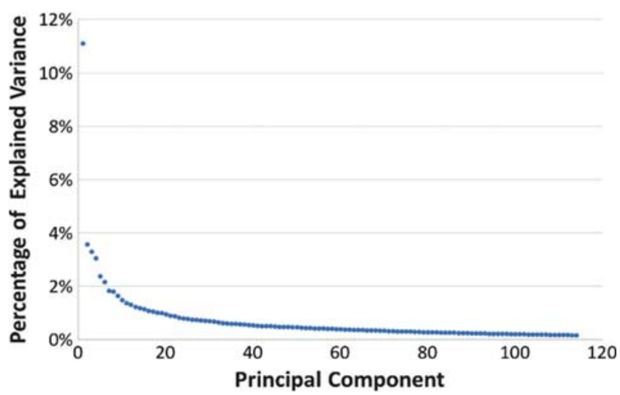

Fig. 91.4 Percentage of explained variance by each principal component

Table 91.1 Performance of the building classification algorithms

Algorithm	Performance	Energy consumption level			Average
		Low	Medium	High	
Naïve Bayes	Precision (%)	82	70	78	76.7
	Recall (%)	82	64	84	76.7
Decision trees	Precision (%)	59	62	89	70.0
	Recall (%)	74	50	87	70.3
Random forests	Precision (%)	73	79	85	79.0
	Recall (%)	85	61	96	80.7
Support vector machines	Precision (%)	83	82	93	86.0
	Recall (%)	85	80	93	86.0
Average precision (%)		74.3	73.3	86.3	77.9
Average recall (%)		81.5	64.8	90.0	78.4

For the classification of the buildings, the building class with high energy consumption had the best performance. For example, on average, the prediction precision and recall for the high-consumption buildings is 86.3 and 90.0%, respectively, compared with 74.3 and 81.5% for the low-consumption buildings and 73.3 and 64.8% for the medium-consumption buildings. This result indicates that the high-consumption buildings are easier to identify, because they have more distinctive features than the other two classes, such as much larger gross floor area, more occupants, and more energy consuming appliances. However, the classification of the low- and medium-consumption buildings might be more challenging and sensitive, which could be attributed to the different occupant behaviors in the buildings with similar weather and physical features.

The classification model generated by the SVM algorithm was further used in predicting the amount of building energy consumption. The RMSE of the proposed model is 2.7×10^6 kWh, while the RMSE of the EnergyStar model is 3.1×10^6 kWh. Overall, compared to EnergyStar, the proposed ML-based building classification and energy consumption prediction model reduced the prediction error by 13%.

91.5 Conclusions and Future Work

Building energy performance benchmarking provides important potential to increase building energy efficiency and reduce building carbon footprint. In this paper, the authors proposed a ML-based model for supporting building energy performance benchmarking. The proposed model is composed of two main components: (1) building classification, and (2) energy consumption prediction. The building classification aims to classify the buildings into three energy consumption levels: low, medium, and high, based on a wide range of building characteristics, occupant behaviors, and geographical and climate features. Four of the most popular ML algorithms, i.e., NB, SVM, DT, and RF, were tested and compared. The energy consumption prediction aims to build separate regression models by using the SVR algorithm for predicting the target energy consumption values of the buildings. The proposed ML-based model was tested on the office buildings from the CBECS dataset, and PCA was used for data dimensionality reduction and feature extraction. The preliminary experimental results indicate that: (1) for this problem, SVM outperforms all the other ML classification algorithms, which achieved 86% average precision and recall; and (2) the proposed energy consumption prediction model can reduce the prediction error by 13% compared to the most widely used model, EnergyStar.

The experimental results indicate the promise of the proposed ML-based model in supporting building energy benchmarking. But, the classification accuracy for the low- and medium-energy-consumption buildings still needs improvement. Therefore, in their future work, the authors will further improve the model by identifying better ways to classify these two groups of buildings. The authors will also extend the proposed model to the residential sector, apply it in different cities and contexts, and assess if and how it can enhance energy efficiency decision making.

Acknowledgements This research is based upon work supported by the Strategic Research Initiatives (SRI) Program by the College of Engineering at the University of Illinois at Urbana-Champaign.

References

1. Klepeis, et al.: The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants. J. Expo. Sci. Environ. Epidemiol. **11**(3), 231 (2001)
2. Filippın, C.: Benchmarking the energy efficiency and greenhouse gases emissions of school buildings in central Argentina. Build. Environ. **35** (5), 407–414 (2000)
3. U.S. Green Building Council: LEED v4 for Building Design and Construction. https://new.usgbc.org/leed-v4. Last accessed 27 Apr 2018
4. U.S. Department of Energy: EnergyPlus. https://energyplus.net. Last accessed 27 Apr 2018
5. Chung, W.: Review of building energy-use performance benchmarking methodologies. Appl. Energy **88**, 1470–1479 (2011)
6. U.S. Environmental Protection Agency (EIA): EnergyStar. https://www.energystar.gov. Last accessed 27 Apr 2018
7. Gao, X., Malkawi, A.: A new methodology for building energy performance benchmarking: an approach based on intelligent clustering algorithm. Energy Build. **84**, 607–616 (2014)
8. Sharp, T.: Energy benchmarking in commercial office buildings. Proc. ACEEE Summer Study Energy Effic. Build. **7996**, 321–329 (1996)
9. Yalcintas, M.: An energy benchmarking model based on artificial neural network method with a case example for tropical climates. Int. J. Energy Res. **30**, 1158–1174 (2006)
10. Grolinger, K., Capretz, M.A., Seewald, L.: Energy consumption prediction with big data: balancing prediction accuracy and computational resources. In: IEEE International Congress on Big Data Congress, pp. 157–164 (2016)
11. Roth, J., Rajagopal, R.: Benchmarking building energy efficiency using quantile regression. Energy **152**, 866–876 (2018)
12. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**(10), 2825–2830 (2011)
13. Wang, L., El-Gohary, N. M.: Data-driven residential building energy consumption prediction for supporting multiscale sustainability assessment. In: Computing in Civil Engineering 2017, pp. 324–332. ASCE (2017)

14. Wang, L., El-Gohary, N. M.: Data-driven approach to identify the impacts of urban neighborhood characteristics on building energy consumption. In: Construction Research Congress 2018, pp. 664–674. ASCE (2018)
15. Edwards, R.E., New, J., Parker, L.E.: Predicting future hourly residential electrical consumption: a machine learning case study. Energy and Build. Energy Build **49**, 591–603 (2012)