International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 0586

PREDICTING ENERGY CONSUMPTION IN MIXED-USE BUILDINGS USING MACHINE LEARNING TECHNIQUES

Aaron Jules R. Del Rosario[1], Aristotle T. Ubando[2], Alvin B. Culaba[2*]

[1]Graduate Student, Mechanical Engineering Department, De La Salle University, 2401 Taft Avenue, Manila, Philippines 0922

[2]Mechanical Engineering Department and Center for Engineering and Sustainable Development Research, De La Salle University, 2401 Taft Avenue, Manila, Philippines 0922 (*Corresponding Author)

ABSTRACT

The rise of mixed-use buildings has contributed to the sustainable development of cities, but they still lack energy design guidelines. Energy consumption forecasting models have been crucial to the improvement of energy efficiency and sustainability of buildings, but their application to mixed-use buildings have been challenging and less tackled in literature. This study presented a novel forecasting model to predict energy consumption in mixed-use buildings using machine learning techniques. The model integrated k-means clustering algorithm and support vector regression to improve predicting performance. The model was demonstrated to a case study considering mixed-use buildings in a tropical area. Clustering results found major differences in the consumption behavior of building clusters, especially on peaking characteristics. The proposed forecasting model was able to capture these variations due to clustering, leading to an increase in predicting performance. The model also performed within building modeling standards and better than statistical approaches in the literature.

Keywords: energy consumption forecasting, mixed-use buildings, machine learning, energy conservation in buildings, energy modeling, urban energy systems

1. INTRODUCTION

The building sector accounts for a significant portion of the global energy consumption, especially in tropical areas where demand for space cooling is high [1]. The increasing energy demand and climate change concerns have raised the need for improvements in energy efficiency and sustainability of this sector. One trend is the rise of mixed-use buildings that combine areas with different purposes (e.g., residential and office) in a single built environment. They aid in sustainable development by providing benefits such as reducing costs and emissions from transport; however, they still lack energy efficiency and sustainability design guidelines [2].

A previous paper on the energy system optimization for mixed-use buildings has emphasized the need to develop energy consumption forecasting models to improve system performance [3]. However, most models in the literature focus on single-type buildings. These models do not consider metadata on the variation of consumption characteristics in different areas of a mixed-use building, which can help improve model accuracy [1]. A previous study on modeling the energy consumption of mixed-use buildings used a statistical approach, with errors ranging from 12-18% [4].

This study aims to present a novel energy consumption predictive model by applying machine learning (ML) techniques, which were found to be more accurate than statistical approaches, to mixed-use buildings, which are growing but are not yet extensively considered in the literature. This paper integrates a clustering model that can extract information on the consumption characteristics of the building to a regression model that can predict future consumption. The model adopts the k-means algorithm for clustering due to its balanced simplicity and accuracy [5], and support vector machines (SVM) for regression due to its superior performance for building applications [6].

2. THEORY

This section discusses the theoretical background of the adopted ML techniques, namely, K-means and SVM.

2.1 K-Means

The K-means algorithm partitions data based on the similarity of several properties through the following procedure. First, k points are randomly selected as initial cluster centroids, where k is an exogenously defined parameter. Other points are assigned to a cluster based on their proximity to the centroid, defined by the chosen distance function. Then, centroids are iteratively updated until the chosen stopping criterion is met [5].

2.2 Support Vector Machines

SVM is a predictive algorithm based on statistical learning and structural risk minimization. Support vector regression (SVR), the implementation of SVM for regression problems, determines a linear function to approximate the value of the output vector based on the input vectors. Since most problems are non-linear, SVR first maps the data into high-dimensional feature space using kernel functions. The corresponding weights are determined through an optimization problem that balances complexity and accuracy [1].

3. MATERIAL AND METHODS

This section details the considered case study to demonstrate the model and the applied methodology to develop the model.

3.1 Case Study

The case study considered mixed-use buildings located in a city with a tropical monsoon climate (Köppen: *Am*), which is the condition in sections of Central and South America, West and Central Africa, and South and Southeast Asia. The model was developed using the MATLAB R2019a software and a 16.0-GB Intel® Core™ i7-6500 CPU at 2.50 GHz hardware.

3.2 Data Acquisition

Two datasets were acquired: an energy consumption dataset of 30 buildings in Miami, Florida, from the Open Energy Information (OpenEI) database [7]; and a weather dataset from the National Renewable Energy Laboratory (NREL) database [8]. The datasets contained observations for 17 years, from the year 1998 to 2014, taken in 30-minute intervals.

3.3 Data Pre-Processing

The data were pre-processed in five steps to make them suitable for modeling. First is to reduce the size of data by regularizing observations into hourly intervals through averaging. Second is to remove redundant and unnecessary variables in the weather dataset, only retaining the following variables: solar irradiance (SI), dry bulb temperature (DBT), dew point temperature (DPT), and wind speed (WS). Third is to clean the dataset by filling missing data, replacing outliers, and smoothening noise through the moving mean method and polynomial interpolation. Fourth is to standardize the data through Z-normalization. Fifth is to extract relevant features from the data. The extracted features include the following types: temporal features, such as the hour of the day, the day of the week, and the month of the year; and lagged predictors, such as the consumption from the previous hour, previous week, and the previous year [1].

3.4 Clustering Model

The energy consumption data were clustered using the K-means algorithm. Euclidean Distance (ED) was used as the distance function. In order to determine the appropriate number of clusters, the algorithm was implemented iteratively, incrementing the parameter k from 2 to 6. Initialization effects were avoided by creating ten replicates per iteration. The final number of clusters was selected based on the silhouette parameter, which is a quality measure that describes how well each object lies in a cluster. The k parameter in the iteration with the highest mean silhouette value for the considered range was selected [5].

3.5 Forecasting Model Training and Validation

In developing the model, the pre-processed data were first partitioned into training and testing sets, consisting of 70% and 30% of the data, respectively. The SVR model was trained and tuned in the Regression Learner application of the MATLAB using the training set. The training model was validated by holding out 25% of the training data and using it to compute for performance metrics. Multiple SVR training models are developed using different kernel functions. The best training function was chosen based on accuracy, measured by root mean square error (RMSE) and mean absolute error MAE [9].

3.6 Model Evaluation and Comparison

The model was evaluated using the unbiased test set. The performance metrics of the model were compared to a model that does include the clustering technique. Finally, the performance of the model was compared to previous models using absolute percent error (APE); and against building modeling standards, using mean bias error (MBE) and coefficient of variation RMSE (CVRMSE) metrics [10].

4. RESULTS AND DISCUSSION

This section presents the results and discussion of the development of the model for the case study.

4.1 Data Description

The acquired datasets each have 298,032 observations. Fig. 1 shows the graph of the consumption dataset, while Table 1 presents the statistical summary of the weather dataset.

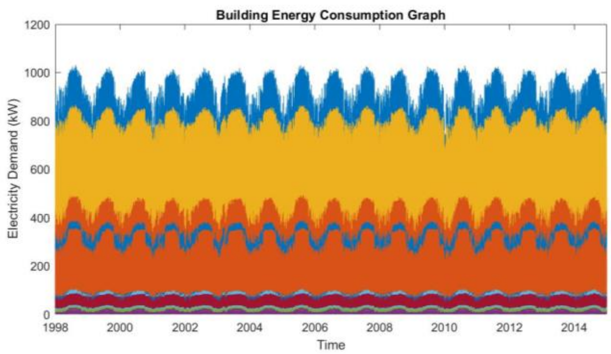

Fig. 1. Plot of the acquired building consumption dataset.

Table 1. Summary of the acquired weather dataset.

Variable	Unit	Min.	Max.	Avg.	Std. Dev.
SI	W/m^2	0	1065	222.38	304.27
DBT	°C	5.0	32.0	24.77	3.53
DPT	°C	1.0	27.0	20.50	4.06
WS	m/s	0	25.7	4.22	1.87

4.2 Clustering Results

Table 2 shows the silhouette value for each parameter *k* iteration. The *k* value with the best performance is 2. Fig. 2 then visualizes the energy behavior of the centroids of the resulting two clusters.

Table 2. Mean silhouette value for each iteration of the *k* parameter.

k	2	3	4	5	6
Mean Silhouette	0.713	0.675	0.678	0.688	0.677

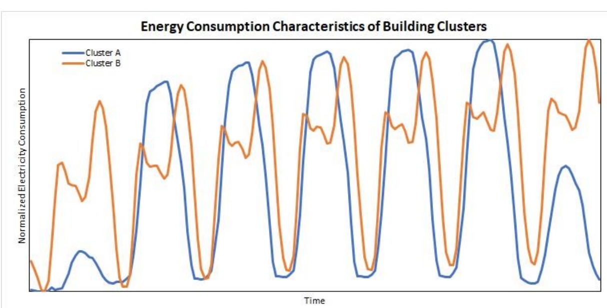

Fig. 2. Energy consumption characteristics of building cluster centroids for a sample week.

Fig. 2 shows that the consumption characteristics of the two clusters are significantly different. Cluster A has a relatively flat peak that shifts twice during a week, presumably during weekdays and weekends. On the other hand, the consumption of cluster B fluctuates more often within a day than a week, with the daily peak happening right after that of cluster A.

4.3 Forecasting Results

Table 3 shows the results from training and validating the forecasting models for datasets with and without clusters. The kernel functions compared for training were linear, quadratic, cubic, fine Gaussian, medium Gaussian, and coarse Gaussian functions. Fig. 3 then shows sample results from the evaluation of the trained models using the test set. Finally, Table 4 shows the evaluation performance of the models using the unbiased testing set.

Table 3. Training and validation results of the developed forecasting models.

Parameter	Clustered	Not Clustered
Selected Function	Medium Gaussian	Medium Gaussian
Validation RMSE	0.0181	0.0192
Validation MAE	0.0140	0.0154

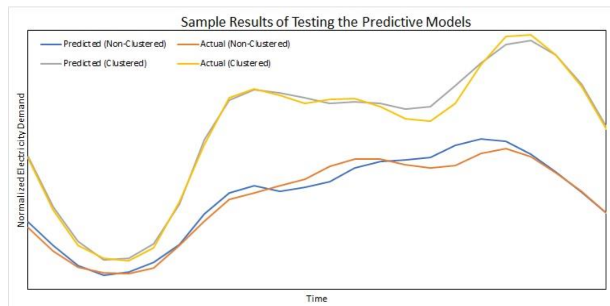

Fig. 3. Sample evaluation results of the developed model using the test set.

Table 4. Evaluation results of the developed models.

Parameter	Clustered	Not Clustered
APE	3.3020%	3.3920%
MBE	0.0264%	0.3519%
CVRMSE	4.0991%	4.2896%

As shown in Table 3, both developed models used the medium Gaussian kernel function for training, which had the best accuracy while only having an average training time compared to other functions. The accuracy of the model improved when building clusters were considered. This effect can be visualized in Fig. 3, as the proposed model was able to capture demand variations unique to each cluster as opposed to the model that did not consider clustering. Also, it was clear from Table 4 that the proposed model had better evaluation performance over the generic model without clusters. Finally, the APEs of the models are significantly better than those in literature ranging from 12-18%; and both their MBE and CVRMSE are within standards of 5-10% and 20-30%, respectively.

5. CONCLUSION

This study presented an energy consumption predictive model for mixed-use buildings using ML techniques. The study contributes to the literature through the consideration of mixed-use buildings in tropical areas and the novel integration of clustering and regression techniques to effectively characterize and forecast the energy consumption of thirty buildings from a case study. The methodology in developing the model applied data pre-processing techniques, k-means algorithm for clustering, and SVR for regression. Datasets were acquired from OpenEI and NREL databases, and the model was developed in MATLAB. The clustering model was optimized by considering the silhouette parameter, while the forecasting model was trained using different kernel function to find the best model. The model was validated and evaluated using a portion of the dataset that was not used in training to avoid bias. The performance of the model was compared to previous models and standards. The results of clustering show the variation in characteristics of the two selected clusters of buildings from the case study, where one cluster exhibits a flat peak while the other fluctuates throughout the day. This characterization helped the forecasting model capture the said variations, leading to improved accuracy. The performance of the model is also significantly better than statistical approaches done in the literature and within building modeling standards. Recommendations for future studies include the identification of other input features that can improve model accuracy, and the application of the model to optimize energy efficiency and sustainability technologies in mixed-use buildings, such as in reducing electricity peak demand and time-of-use charges and evaluating the viability of renewable energy systems.

ACKNOWLEDGMENT

The first author expresses his gratitude to the Department of Science and Technology-Science Education Institute (DOST-SEI) and the Engineering Research and Development for Technology (ERDT) consortium for the scholarship award for his Master of Science in Mechanical Engineering degree program.

REFERENCES

[1] M. Bourdeau, X. Q. Zhai, E. Nefzaoui, X. Guo and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustainable Cities and Society,* vol. 48, p. 101533, July 2019.

[2] C. Hachem, "Design of a base Case Mixed-use Community and its Energy Performance," *Energy Procedia,* vol. 78, pp. 663-668, 2015.

[3] I. Aldaouab and M. Daniels, "Renewable energy dispatch control algorithms for a mixed-use building," in *2017 IEEE Green Energy and Smart Systems Conference (IGESSC)*, Long Beach, California, USA, 2017.

[4] X. Gao, A. M. Malkawi and Y. K. Yi, "A New Method for Predicting Mixed-Use Building Energy: The Use of Simulation to Develop Statistical Models," in *Proceedings of BS2013: 13th Conference of International Building Performance Simulation Association*, Chambéry, France, 2013.

[5] G. Tardioli, R. Kerrigan, M. Oates, J. O'Donnell and D. P. Finn, "Identification of representative buildings and building groups in urban datasets using a novel pre-processing, classification, clustering and predictive modelling approach," *Building and Environment,* vol. 140, pp. 90-106, August 2018.

[6] F. Kayetz, M. C. Taplamacioglu, E. Cam and F. Hardalac, "Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems,* vol. 67, pp. 431-438, May 2015.

[7] N. Darghouth, "Simulated load profiles for DOE commercial reference buildings (17 years using NSRD data)," OpenEI, 11 September 2016. [Online]. Available: https://openei.org/datasets/dataset/simulated-load-profiles-17year-doe-commercial-reference-buildings. [Accessed 31 May 2019].

[8] NREL, "National Solar Radiation Data Base," National Renewable Energy Laboratory, 15 March 2006. [Online]. Available: https://rredc.nrel.gov/solar/old_data/nsrdb/. [Accessed 31 May 2019].

[9] MathWorks, "Regression learner App," MathWorks, 9 March 2017. [Online]. Available: https://www.mathworks.com/help/stats/regression-learner-app.html. [Accessed 31 May 2019].

[10] W. Li, Y. Zhou, K. Cetin, J. Eom, Y. Wang, G. Chen and X. Zhang, "Modeling urban building energy use: A review of modeling approaches and procedures," *Energy,* vol. 141, pp. 2445-2457, 15 December 2017.