7th International Building Physics Conference

IBPC2018

Proceedings

SYRACUSE, NY, USA

September 23 - 26, 2018

Healthy, Intelligent and Resilient
Buildings and Urban Environments

ibpc2018.org | #ibpc2018

Comparison of data-driven building energy use models for retrofit impact evaluation

Yujie Xu[1,*], Azizan Aziz[1], Bertrand Lasternas[1], Vivian Loftness[1]

[1]Carnegie Mellon University

[*]*Corresponding email: yujiex@andrew.cmu.edu*

ABSTRACT

A change-point (piecewise linear regression) model fitted to the pre-retrofit data as the counterfactual for the savings calculation, is considered to be the best approach to evaluating the energy savings of building retrofits (ASHRAE Guideline 14). However, when applied to a large portfolio savings analysis with substantial multi-year data, the change-point model does not fit the data well in some cases. The study thus aims to improve the accuracy of the change-point model by: 1) using more advanced non-linear models, 2) incorporating additional input features, and 3) increasing the time resolution of input variables. We found that random forest regression (RF) models with an array of climate (humidity, wind, solar radiation, etc.), time (day of the week, season, holiday), and energy consumption of the immediate past 1-4 hours (energy lag terms) outperformed the change-point model, shallow neural networks, and support vector machine regression (SVR). Our result implies that high resolution smart meter data should be used in place of monthly utility bills to more accurately evaluate retrofit savings. We further explored the relative contribution of the input variables to the random forest regression model using Shapley Value, a game theoretic variable importance metric. We found that the most important input feature is the energy consumption of the immediate past (or energy lag terms). We also found that solar radiation and weekend day indicators are more important than outdoor temperature. The improved model could provide better insights to portfolio managers in planning future energy retrofits. Policy makers could also use such models to evaluate the average energy saving potential for energy policy changes, such as the requirement of minimum insulation level, and lighting equipment efficiency.

KEYWORDS

Energy Performance Prediction, Random Forest Regression, Inverse Modelling, Comparison

INTRODUCTION

In the U.S., buildings account for 40% of the nation's energy consumption (EIA 2018). Reducing the energy consumption of the building sector is critical to mitigating the global climate change. Energy efficiency upgrades could save 5% to 50% of energy consumption in buildings (DOE 2018). However, in order to accurately evaluate the impact of these upgrades, differences in weather, operating schedules and modes, occupant behavior, and other un-observed factors between pre and post retrofit needs to be accounted for. This will require a counterfactual energy model that accurately describes the potential energy consumption had the upgrade not have happened, or the counterfactual.

In ASHRAE guideline 14, the counterfactual is estimated with a change-point regression model (ASHRAE 2014). The saving is calculated as the difference between the model estimated consumption and the measured post-retrofit consumption. We, and many studies aimed to compare and improve the prediction accuracy of such counterfactual energy models (Zhang et al. 2015; Brown et al. 2012). Unlike most existing studies with only a handful of case studies, this study evaluates models on a larger sample size of 65 buildings. We tested hypothesis about whether the following three strategies improve on model accuracy: 1) advanced models using only the temperature input, 2) adding more input features, and 3) increasing the time resolution from monthly to hourly.

METHODOLOGY

The input energy data consists of 2014 to 2016 15-min interval electricity consumption data of 65 commercial buildings (24 office, 14 courthouses, 11 courthouse/office, 11 unknown) in the General Service Administration (GSA) portfolio. The climate data is retrieved from the national solar radiation database (NRSDB) Physical Solar Model (PSM).

Model evaluation procedure and metric

For each building, the input features are aligned by their timestamps. The aligned data set is then randomly split into a training set and a test set using year-month index, the first 70% of the year-months is the training data, and the rest is the validation data. The error is evaluated in Coefficient of Root-Mean-Square Error (CV-RMSE). CV-RMSE is root-mean-square error divided by mean electricity consumption (Equation 1, $\bar{y}$ is the sample mean of the energy data, n is the number of data points, y_i is the ith observed electricity consumption, $\hat{y_i}$ is the corresponding predicted electricity consumption, $(n-p)$ is the degree of freedom). However, the degree of freedom may not be well defined for models other than linear regression, for example, the effective degree of freedom for Neural Network models is changing through the training process (Anzai 2012). We believe the degree of freedom could be left out as we evaluated the model on a held-out test set. Thus, we used the adjusted CV-RMSE (Equation 2) in the evaluation.

$$\mathrm{CV}-\mathrm{RMSE} = \sqrt{\sum_i \frac{(y_i - \hat{y_i})^2}{n-p}} / \bar{y} \tag{1}$$

$$\mathrm{CV}-\mathrm{RMSE}_{\mathrm{adj}} \sqrt{\sum_i \frac{(y_i - \hat{y_i})^2}{n}} / \bar{y} \tag{2}$$

Models

We compared the prediction accuracy of the following five models: 3-parameter change-point model (baseline algorithm), neural network with one hidden layer, neural network with two hidden layers, kernelized support vector machine regression (SVR), and random forest regression. The Change-point model is implemented using Python scipy optimize package, and the other models are implemented using Python scikit-learn package (Pedregosa et al. 2011).

Change-point model splits the range of temperature into 2 to 3 different segments and fits a separate linear regression model in each segment, with continuity requirement at the boundaries of segments. The method is adopted by ASHRAE Guideline 14 (ASHRAE 2014) and IPMVP (Efficiency Valuation Organisation 2012). In this study the change-point model is implemented with a grid search over the temperature range from 40F to 80F according to (Kissock et al. 2003).

Neural network is a popular predictive method applied in a variety of fields. A neural network consists of inter-connected neurons and non-linear activation functions. Most commonly, neurons form a layered directed acyclic graph (DAG) structure with one input layer, some hidden layers, and one output layer. Directed links are formed between layers, and assigned weights using the backpropagation algorithm. Two shallow network structures are tested in this study: one with one hidden layer, and one with two hidden layers. We used the tanh activation, following the choice of (MacKay 1996; Dodier and Henze 2004). A grid-search method is applied to find the optimal number of hidden units per layer.

The support vector machine algorithm is developed by Vapnik and Chervonenkis in 1963, as a "nonlinear generalization of the Generalized Portrait algorithm" (Sewell n.d.). It can be applied to both classification and regression tasks. Support vector machine regression (SVR) tries to find a function as flat as possible that is at most ϵ from training data (Smola and Schölkopf 2004). C and ϵ are two hyper-parameters to be optimized. A grid search of the optimal C is performed in the range of 1 to 1000 with step size 100. ϵ is left as default.

Random Forest Regression is an ensemble method which constructs many regression trees at the training time, and forms predictions by taking a weighted average of the output of individual trees. In the study, the tuning parameter, maximum tree depth, is set as default 30.

RESULTS

Hypothesis 1: advanced models have better accuracy using only temperature input.

We tested whether applying more advanced nonlinear models could directly reduce the prediction error by using just temperature as the input. Our result suggests that advanced models (NN, SVM, RF) perform similarly to simple change-point model using only temperature input.

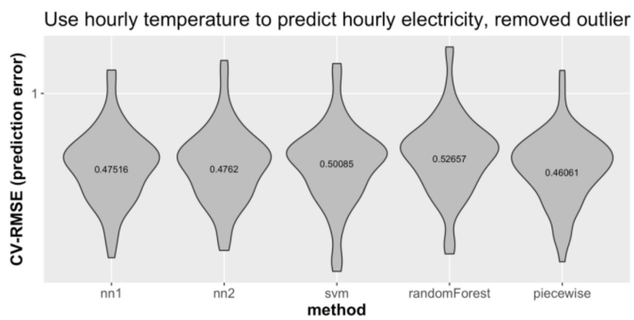

Figure 1. Violin plot, showing distribution of CV-RMSE (the prediction error) across buildings, for each method. All methods appear to have similar error distributions.

Hypothesis 2: Additional input features improves model performance.

By adding the following input variables (Table 1), in addition to temperature, we saw substantial reduction in CV-RMSE (prediction error) using random forest regression (Figure 2).

Table 1. Additional input variables.

Climate	Humidity	Dew Point
		Relative Humidity
	Solar Radiation	Diffuse Horizontal Irradiance (DHI)
		Direct Normal Irradiance (DNI)
		Global Horizontal Radiation (GHI)
		Solar Zenith Angle (Z)
	Pressure	
	Precipitable Water	
	Wind	Wind Direction
		Wind Speed
Time	Season	

	Day of the Week
	If a Day Is Federal Holiday
Energy	Energy lag term (Energy Consumption in Previous 4 Steps)

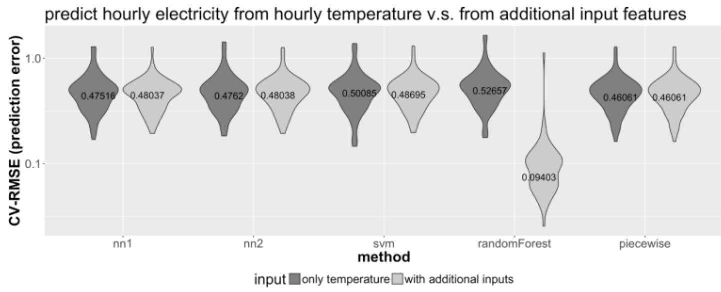

Figure 2. Violin plot, showing distribution of CV-RMSE (prediction error) across buildings, for each method: light grey uses only temperature input, dark grey is using all features in Table 1.

To compare the relative importance of these variables for prediction accuracy in random forest regression, we applied a game-theory metric, the Shapley Value imputation (Lipovetsky and Conklin 2001). The Shapley Value for the ith variable is shown in Equation 3. n is the number of input features. v is a function from feature subsets to real numbers. $v(S)$ represents the "expected sum of payoffs" ("Shapley Value" 2018) using the feature set S. In our case, $v(S)$ is CV-RMSE. R is a permutation, P_i^R is the list of features before the ith one. Computing the exact Shapley Value is #p complete (Fatima et al. 2008), i.e. intractable, so we computed an estimated Shapley Value using a random sample of permutations. The estimated Shapley Value is shown in Figure 4. The variables with lower Shapley Value are more important. The energy lag terms are the most important. A series of solar radiation related factors (GHI, DHI, etc.) ranked second, and the indicators for weekends ranked third.

$$\phi_i(v) = \frac{1}{n!} \sum_R v(P_i^R \cup \{i\}) - v(P_i^R) \tag{3}$$

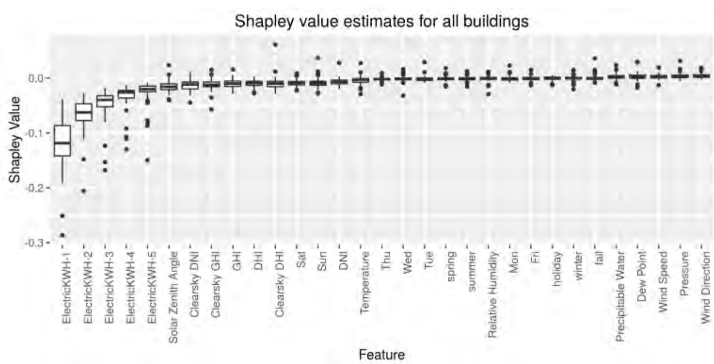

Figure 4. Shapley Value estimates for all input features

Hypothesis 3: Higher resolution input improves model performance

For each modelling method, we compared the following two approaches: 1) use monthly average of the input features in hypothesis 2 to predict monthly electricity consumption; 2) use hourly input to predict hourly electricity and aggregate the result to monthly. With higher resolution hourly input, Random Forest Regression achieved significantly better performance than using monthly average inputs (Figure 3).

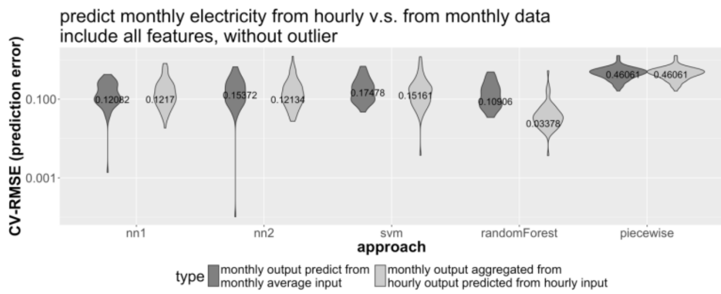

Figure 3. Violin plot, showing distribution of CV-RMSE across buildings, for each method: dark grey is using monthly average input to predict monthly electricity, light grey is using hourly input to predict hourly output then aggregate to monthly.

DISCUSSIONS

Findings

All models perform similarly when only using temperature as the input variable. Random forest regression model with an array of climate and time variables outperformed our implementation of shallow neural networks, SVR, and change-point models in terms of lowest CV-RMSE. For the random forest regression model, the most effective predictors for the hourly energy consumption are the energy lag terms, followed by solar radiation, then indicator variables for weekends.

Limitations and next stage work

We observed outliers in the 15-min electricity data that are: negative records, a large positive immediately followed a large negative record, and extremely large values followed by a data gap. A simple outlier removal rule is applied for the analysis: 1) remove negative entries 2) remove data points greater than 1.5 times the 97th percentile of the positive records. Further research into the reasons for the outliers, and more advanced outlier detection routine might help improve the model accuracy. In the Shapley Value computation, we used a sample of 100 random permutations to construct an initial estimate. The number of permutations here is chosen arbitrarily. Analyzing the convergence rate of the Shapley Value estimate, i.e. how many samples are needed to achieve certain approximation accuracy, would be a next step of the study. All the counterfactual models in this paper and many other papers under the ASHRAE Guideline 14 framework are based on the assumption of no un-measured confounding factors. In this sense, adding more covariates to the models, other than just temperature, could make the model prediction more accurate, and make it more viable to draw causal conclusions.

CONCLUSIONS

The paper compared the prediction accuracy of five data-driven energy counterfactual models under three conditions: 1) using only temperature input 2) using an array of climate and energy inputs, and 3) using higher resolution hourly inputs to predict monthly energy. We found that random forest regression model ensures the best performance when supplied with additional inputs with higher temporal resolution. The most important input features for Random Forest Regression are energy lag terms and solar radiation.

The predictive power of the energy counterfactual model determines the accuracy of the ECM savings evaluation. The result could assist ESCO companies to better predict the potential savings for implementing certain retrofit. In addition, our improvement of such energy counterfactual models could also be useful in system fault detection, and development of new control strategies (Zhang et al. 2015).

ACKNOWLEDGEMENT
This research is made possible through a collaboration with the General Service Administration to evaluate their portfolio-wide energy retrofit impact (contract PR201508040012). The author also wishes to thank Professor Matt Gormley for recommending the machine learning methods.

REFERENCES
Anzai, Yuichiro. 2012. *Pattern Recognition and Machine Learning*. Elsevier.

ASHRAE. 2014. *Ashrae Guideline 14-2014: Measurement of Energy, Demand and Water Savings*. Ashrae. https://books.google.com/books?id=zlJkAQAACAAJ.

Brown, Matthew, Chris Barrington-Leigh, and Zosia Brown. 2012. "Kernel Regression for Real-Time Building Energy Analysis." *Journal of Building Performance Simulation* 5 (4): 263–76. https://doi.org/10.1080/19401493.2011.577539.

Dodier, Robert H., and Gregor P. Henze. 2004. "Statistical Analysis of Neural Networks as Applied to Building Energy Prediction." *Journal of Solar Energy Engineering* 126 (1): 592–600.

DOE. 2018. "Why Energy Efficiency Upgrades." June 5, 2018. https://www.energy.gov/eere/why-energy-efficiency-upgrades.

Efficiency Valuation Organisation. 2012. "International Performance Measurement and Verification Protocol. Concepts and Options for Determining Energy and Water Savings." *Efficiency Valuation Organisation, Toronto*.

EIA. 2018. "How Much Energy Is Consumed in U.S. Residential and Commercial Buildings?" May 3, 2018. https://www.eia.gov/tools/faqs/faq.php?id=86&t=1.

Fatima, Shaheen S., Michael Wooldridge, and Nicholas R. Jennings. 2008. "A Linear Approximation Method for the Shapley Value." *Artificial Intelligence* 172 (14): 1673–99. https://doi.org/10.1016/j.artint.2008.05.003.

Kissock, J. K., J. S. Haberl, and D. E. Claridge. 2003. "Inverse Modeling Toolkit: Numerical Algorithms for Best-Fit Variable-Base Degree Day and Change Point Models," July. http://oaktrust.library.tamu.edu/handle/1969.1/153708.

Lipovetsky, Stan, and Michael Conklin. 2001. "Analysis of Regression in Game Theory Approach." *Applied Stochastic Models in Business and Industry* 17 (4): 319–30. https://doi.org/10.1002/asmb.446.

MacKay, David J. C. 1996. "Bayesian Non-Linear Modeling for the Prediction Competition." In *Maximum Entropy and Bayesian Methods*, edited by Glenn R. Heidbreder, 221–34. Fundamental Theories of Physics 62. Springer Netherlands. https://doi.org/10.1007/978-94-015-8729-7_18.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (October): 2825−2830.

Sewell, Martin. n.d. "History of Support Vector Machines." Support Vector Machines (SVMs). n.d. http://www.svms.org/history.html.

"Shapley Value." 2018. *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Shapley_value&oldid=840058989.

Smola, Alex J., and Bernhard Schölkopf. 2004. "A Tutorial on Support Vector Regression." *Statistics and Computing* 14 (3): 199–222.

Zhang, Yuna, Zheng O'Neill, Bing Dong, and Godfried Augenbroe. 2015. "Comparisons of Inverse Modeling Approaches for Predicting Building Energy Performance." *Building and Environment* 86 (April): 177–90. https://doi.org/10.1016/j.buildenv.2014.12.023.