Data-driven Identification of Occupant-thermostat Interactions in Small Commercial Buildings

Brent Huchuk*
National Research Council of Canada
Ottawa, Ontario, Canada
Brent.Huchuk@nrc-cnrc.gc.ca

Farid Bahiraei
National Research Council of Canada
Ottawa, Ontario, Canada
Farid.Bahiraei@nrc-cnrc.gc.ca

Saptak Dutta
National Research Council of Canada
Ottawa, Ontario, Canada
Saptak.Dutta@nrc-cnrc.gc.ca

ABSTRACT

Small commercial buildings owners and utility managers often look for opportunities for energy and greenhouse gas emission savings through various energy efficiency approaches. However, in Canada, the small commercial buildings are currently underserved by energy conservation tools because of their dispersion and lower payback potential. Connected thermostats provide a low-cost solution to collect massive amounts of data from a portfolio of these buildings that can be used to improve the understanding of their energy use behaviors. Similar to their residential application counterparts, these thermostats can be overridden by building occupants and potentially hinder more advanced controls or coordination across a managed building portfolio. In this work, we investigated the temperature setpoint overrides from more than 620 thermostats across 250 small commercial buildings located in Ontario, Canada. In particular, we developed a global model to estimate the fraction of buildings in the portfolio that experienced overrides each hour. We were able to correctly predict the percentage of overridden buildings within 2%. Such a predictive model may help the portfolio manager to forecast future conditions in order to create more efficient energy saving and peak reduction programs without compromising the occupants' thermal comfort and organizational productivity.

CCS CONCEPTS

• **Information systems → Data mining**; • **Applied computing → Engineering**.

KEYWORDS

Connected thermostats, Small commercial buildings, Energy efficiency

ACM Reference Format:
Brent Huchuk, Farid Bahiraei, and Saptak Dutta. 2021. Data-driven Identification of Occupant-thermostat Interactions in Small Commercial Buildings. In *The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '21), November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3486611.3491119

BuildSys '21, November 17–18, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9114-6/21/11…$15.00
https://doi.org/10.1145/3486611.3491119

1 INTRODUCTION

Small commercial buildings (<10 Ksqft) comprise 68% of all commercial buildings and 18% of commercial building energy use [7], and yet are radically under-served by energy conservation tools, because of their dispersion and lower payback potential. They are too small for a full commercial building automation systems (BAS), so there is an interest to develop BAS-lite systems from low-cost components. Recent advances in connected (i.e., internet-connected) thermostats (CTs) have begun to address some of these challenges by providing a price-efficient method of centralizing controls and monitoring of small commercial buildings (SCB) without the need of installing full building automation systems. CTs allow for energy efficiency and demand response possibilities [10]; however, given the control logic, any sort of overrides will possibly prevent these advanced applications.

It has been shown that more than 30% in energy savings can be achieved by properly adjusting temperature setpoints [6]. However, the users often do not properly operate CTs which results in little or no savings compared to traditional thermostats [9]. In the case of a portfolio of buildings, one approach to deliver savings is to apply standard default schedules to groups of buildings based on their working hours and climate zone. Nevertheless, the occupants' interactions with CTs can be highly unique, and therefore, relying on only default schedules may result in user dissatisfaction and/or minimal savings [2]. Various aspects of CT performance in residential buildings have been studied [3, 11]. However, the applications of these devices in SCB buildings (or a portfolio of SCBs) have not been addressed in the literature. Therefore, in this study, we use CT data collected from a portfolio of 250 SCBs over a period of 13 months to develop two global models based on various features. The models can be used to estimate the fraction of buildings that override the scheduled temperature setpoints over a certain forecasting horizon. Such a tool is useful for portfolio managers as it enables them to predict upcoming holds rates. Furthermore, it can be used to potentially forecast the impact of any broad changes to thermostat temperature settings (for example as part of energy conservation programs). The main contributions of this paper are:

- A longitudinal and cross-sectional study of the override patterns of a SCB portfolio,
- The exploration of key override prediction model features, and
- The development and evaluation of a machine learning model for future portfolio-level override prediction.

2 METHODOLOGY

For this analysis we first needed to format and finalize our data (Section 2.1) before selecting candidate features (Section 2.2) and finally developing/evaluating our predictive models (Section 2.3-2.5).

2.1 Data

We conducted our study on a novel dataset constructed by Energy Metrics, an IoT technology service provider. Energy Metrics collects data from installed CTs in commercial buildings and stores it in a proprietary database as event-based data. The data from the CTs includes indoor air temperature, the setpoint temperatures, heating or cooling equipment state (i.e., on or off), and if the CT setpoints were being overridden.

We limited the dataset for this analysis to only buildings in Ontario, Canada and only to buildings owned or operated by the same organization (i.e., part of one portfolio). The 250 buildings contained 621 CTs which were installed starting in 2017. The number of CTs per branch varied from one to seven and the average size of the buildings was 506 m^2 (5448 ft^2). The buildings are conditioned by multiple roof top units (RTU) with each RTU serving its own zone and controlled by a single CT. Each CT is unaware of the other devices in the building.

The event-driven data for each CT was resampled first to an hourly frequency for each building before being aggregated to hourly values for the entire portfolio. Continuous values had various statistical measurements (discussed in Section 2.2) calculated for that hour, while any binary values were mapped as a `logical-OR` for that hour. If no information was available during an hour for a building the value was left as `NULL`. Missing information could be a result of network, power, or service outages either by the CT manufacturer's API or Energy Metrics' system. It is not possible with the available data to distinguish between these various causes for missing data.

We mapped each building to a local weather station based on its latitude and longitude. The sample of 250 buildings was mapped to 68 unique weather stations. We recorded average daily temperature, relative humidity, wind speed, and precipitation from the weather stations. Finally, we calculated heating and cooling degree days for each station. For both types of degree days we selected a balance point temperature at 18°C (≈65°F); similar to other definitions [12].

2.2 Candidate Feature Selection

We selected candidate features based on a combination of observed insights, previous works, and the assumption that overrides were a result of occupant dissatisfaction with the current indoor environmental conditions. Our intention was to introduce a large number of features and eventually select the most important features (Section 3.1).

The first feature type we included related to temporal elements of the schedule override. Huchuk et al. [4] observed that in the residential context there was a relation between overrides and the time of day — a similar trend we observed in this data. Figure 1 shows the fraction of buildings with an override over the 12 month period we later used for model training (Section 2.4). From Figure 1 we see there is a clear repeating daily pattern of an increasing fraction of buildings being overridden. For our analysis, we specifically looked at including a feature for the hour of the day (i.e., {0, ..., 23}) and the day of the week (i.e., {0, ..., 6}, with the index 0 being Monday). Seasonality components were not explicitly included, instead the inclusion of the outdoor conditions (namely average outdoor air temperature and degree days) was expected to be a proxy.

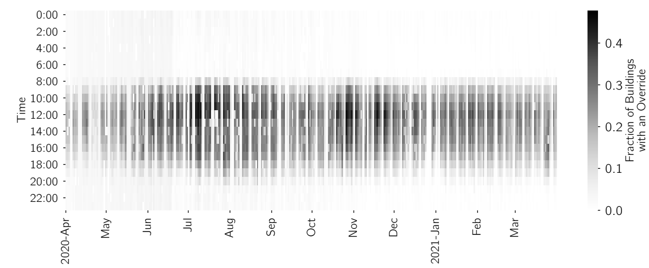

Figure 1: Heatmap of the fraction of buildings in an override at every hour over our 12 month training period.

Beyond temporal patterns, assuming that overrides were an indicator of dissatisfaction with the internal building conditions, we included features related to the internal air temperatures or the heating and cooling setpoints. For each building we resampled the air and setpoint temperatures using hourly *max*, *min*, and *mean* functions. When aggregating across the portfolio of buildings, the means, max, and mins were re-aggregated (i.e., the max of the max). In addition, the 90^{th} and 10^{th} percentiles were calculated on the min and max values of the setpoints.

The final class of features selected were autoregressive features. In particular we considered previous indoor air temperatures, heating and cooling setpoints, and the previous fraction of the building portfolio in an override. We varied the order (i.e., how far back in recent history to include) from one to four previous hours. In practice, the autoregressive features would require predictions further than one hour ahead to have the predictions chained together or a new model generated for every prediction horizon desired. In total, there were 42 features considered during the initial model training.

2.3 Model Selection

After testing various linear and non-linear machine learning methods, we ultimately selected a random forest regression (RF) model. The RF is a non-linear estimator made up of multiple tree estimators [1]. Unlike a single decision tree (which is a high variance estimator), an RF aggregates the results from multiple decision trees. In our regression task, the predictions from the various trees were averaged together.

In addition to the non-linear decision space provided by an RF model, it does not require features to be standardized or cyclically encoded (e.g., encoding hour of day in both sine and cosine components). The model was implemented using the Scikit-learn library [8]. We tuned the hyperparameters of the RF model using a random sample search with 750 samples and five-fold cross-validation. The specific hyperparameters tuned and their considered ranges are shown in Table 1.

Table 1: Tuned hyperparameters and values sampled during cross-validation.

Hyperparameter	Values	Description
n_estimators	{10, 15, 20, ..., 95 }	The number of trees in the forest.
max_features	{auto, sqrt}	The number of features to consider at every split.
max_depth	{1,2,3,...,10}	The maximum depth of the tree.
min_samples_split	{2,5,10}	The minimum number of samples required to split a node.
min_samples_leaf	{1,2,4}	The minimum number of samples required at each leaf node.
bootstrap	{True, False}	Whether bootstrap samples are used when building trees.

2.3.1 Baseline Model. To understand how well the RF model was predicting, we constructed a simple baseline prediction method to compare against. The baseline method calculated the average fraction of buildings with an override based on the hour of day and the day of week. These override fractions were stored in a lookup table. During testing, predictions were made by looking up the average override rate from the table.

2.4 Model Training

Both the baseline and RF models were trained on data logged during a 12 month period starting April 1, 2020. All testing occurred during the month of April 2021. We used a time series split for cross-validation. Figure 2 shows the cross-validation training method used for hyperparameter tuning of the RF model. With the time series split, the training data was kept sequential (i.e., not shuffled) and split into five equal segments. Following the hyperparameter tuning, the feature combination with the highest R^2 score across iterations was kept for evaluation and exploration on the test data.

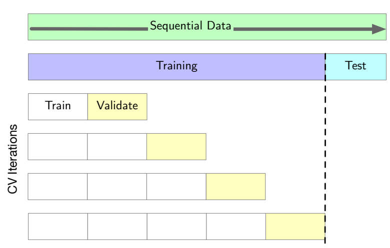

Figure 2: Demonstration of the five-fold time series cross-validation used for hyperparameter tuning.

2.5 Evaluation Metrics

The performance of the predictions by the models was evaluated using three metrics. We calculated the root mean squared error (RMSE), the mean biased error (MBE), and the mean absolute error (MAE) between the measured (y) and predicted ($\hat{y}$) building override fraction at each timestep (t) over the prediction time frame (N timesteps) as,

$$RMSE = \sqrt{\frac{\sum_{t=0}^{N-1} (y_t - \hat{y}_t)^2}{N}} \tag{1}$$

$$MBE = \frac{\sum_{t=0}^{N-1} (y_t - \hat{y}_t)}{N} \tag{2}$$

$$MAE = \frac{\sum_{t=0}^{N-1} |y_t - \hat{y}_t|}{N} \tag{3}$$

3 RESULTS AND DISCUSSION

Our analysis first looks at the feature importance for the trained RF model and discuss potential feature reduction (Section 3.1). Next, we look at the predictive capabilities of the trained RF models with both the full and reduced (based on importance) feature set (Section 3.2).

3.1 Feature Importance

We evaluated feature importance in the RF model using SHapley Additive exPlanations (SHAP) [5]. Figure 3, shows a SHAP value summary plot for the five most important features of RF model. We limited the number of features in the plot to five as beyond this the features had quickly diminished impacts.

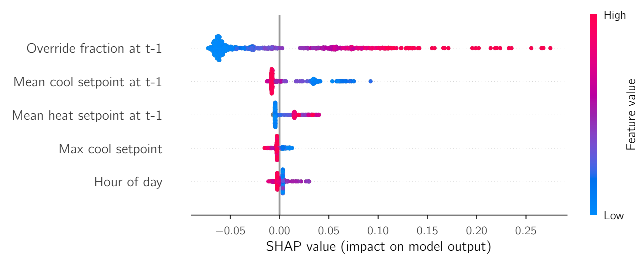

Figure 3: SHAP value summary plot for the five most important features from RF model.

Based on Figure 3, we see that previous override fractions are the most important predictor. At the same time, we see higher values (i.e., more overridden branches) lead to higher predicted fractions by the RF model. This intuitively makes sense; if many of the buildings previously were overridden there is no centralized intervention (e.g., by the portfolio manager) to quickly change the overridden fraction during the next hour. From the figure we also see that previous heating and cooling temperature setpoints have significant (and reversed) effects. Given that overrides are assumed to address dissatisfaction with the indoor air temperature, this too is plausible. Finally, we see the time of day to be a strong indicator (and specifically mid-range values of the feature). We saw earlier in Figure 1, how mid-range (i.e., mid day) values appeared to correlate with higher fractions of buildings being overridden.

We trained a second RF model using only the top five features based on the SHAP values. This RF model used the same hyperparameters as those selected during cross-validation described above (Section 2.4).

3.2 Model Predictions

As described in Section 2.4, following a year of training, we tested the RF models with the best performing hyperparameters and both full and reduced features sets on the most recent month of data (i.e., April 2021) which had been withheld from training. Figure 4 shows the comparison between the actual fraction of buildings with an override and the fraction predicted by the baseline and RF models

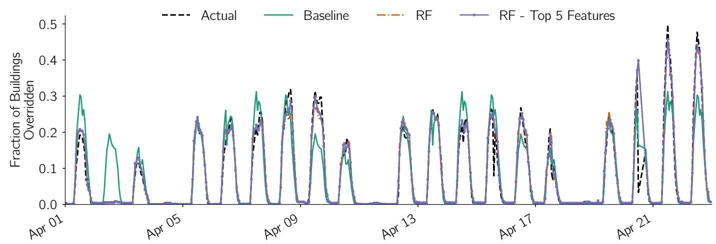

Figure 4: Predictions of the baseline and RF models with full and top five features over the testing period.

with both the full and top five features. Table 2 shows the values of the three performance metrics (Section 2.5) over the testing period.

Table 2: Prediction evaluation metrics for baseline and RF models. All metric values are unitless fractions.

Model	RMSE	MBE	MAE
Baseline	0.044	0.074	0.22
RF	0.024	-0.001	0.01
RF - Top 5 Features	0.022	-0.002	0.01

Based on visual inspection of the predictions and the tabulated metrics, we observe that our baseline method is a decent predictor. The idea that time of day and day of week are strong predictors makes sense given these buildings are occupied only during certain periods of the day and overrides would need to occur on the CT device. This is in contrast to a residential application for CTs where overrides could often be done remotely by the occupants. However, we note that both RF models are a stronger predictor of the fraction of buildings being overridden with the inclusion of its richer feature set. We also observe that the RF with the top five features performs nearly identically to the full feature set.

Unfortunately, to our knowledge, no open datasets of portfolios of CT data from SCBs exist to be able to compare our findings or methods on. This lack of open data sets continues to be a barrier to be able to compare and improve on methods from other researchers.

4 CONCLUSIONS AND FUTURE WORK

Connected thermostats, along with other IoT solutions, can provide additional insights into the performance of, currently underserved, small commercial buildings. In this work, we used data from a portfolio of 250 small commercial buildings to study users' schedule override behaviors and develop a portfolio-level prediction model. The feature-importance analysis showed strong auto-regressive tendencies. This suggests that thermal discomfort may trigger a series of manual overrides during the working hours that can hinder the efficiency of energy conservation programs and coordination across a managed building portfolio. Other important features are cooling and heating temperature setpoints. These values are currently scheduled on the portfolio level. However, the results suggest that building-specific schedules may improve the thermal comfort in the buildings by minimizing the need for overrides. However, such schedules require a better understanding of the building-level indoor conditions and user-thermostat interactions. Therefore, in the future works, we will explore the data from individual buildings and thermostats. The performance of building-specific override prediction models will be investigated in order to create more efficient schedules that can be used by the portfolio manager.

ACKNOWLEDGMENTS

This research was supported by the Office of Energy Research and Development (OERD) of Natural Resources Canada. We also thank Jared Goodman, Andre Roman, Rajendran Avadaiappan, and their team at Energy Metrics for their support.

REFERENCES

[1] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (Oct. 2001), 5–32. https://doi.org/10.1023/A:1010933404324
[2] Kristen S Cetin, Mohammad Hassan Fathollahzadeh, Niraj Kunwar, Huyen Do, and Paulo Cesar Tabares-Velasco. 2019. Development and validation of an HVAC on/off controller in EnergyPlus for energy simulation of residential and small commercial buildings. *Energy and Buildings* 183 (Jan. 2019), 467–483. https://doi.org/10.1016/j.enbuild.2018.11.005
[3] Md Monir Hossain, Tianyu Zhang, and Omid Ardakanian. 2019. Evaluating the Feasibility of Reusing Pre-trained Thermal Models in the Residential Sector. In *Proceedings of the 1st ACM International Workshop on Urban Building Energy Sensing, Controls, Big Data Analysis, and Visualization.* ACM, New York NY USA, 23–32. https://doi.org/10.1145/3363459.3363529
[4] Brent Huchuk, William O'Brien, and Scott Sanner. 2021. Exploring smart thermostat users' schedule override behaviors and the energy consequences. *Science and Technology for the Built Environment* 27, 2 (Feb. 2021), 195–210. https://doi.org/10.1080/23744731.2020.1814668
[5] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
[6] Jin Woo Moon and Seung-Hoon Han. 2011. Thermostat strategies impact on energy consumption in residential buildings. *Energy and Buildings* 43, 2-3 (Feb. 2011), 338–346. https://doi.org/10.1016/j.enbuild.2010.09.024
[7] Natural Resources Canada Office of Energy Efficiency. 2009. Survey of commercial and institutional energy use.
[8] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Mattieu Perrot, and Edouard Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[9] Marco Pritoni, Alan K. Meier, Cecilia Aragon, Daniel Perry, and Therese Peffer. 2015. Energy efficiency and the misuse of programmable thermostats: The effectiveness of crowdsourcing for understanding household behavior. *Energy Research & Social Science* 8 (July 2015), 190–197. https://doi.org/10.1016/j.erss.2015.06.002
[10] Michael Rovito, Gita Subramony, Laurentia Duffy, and Pete Savio. 2014. Advanced Thermostats for Small- to Medium-Sized Commercial Buildings. In *2014 ACEEE Summer Study on Energy Efficiency in Buildings.* 12.
[11] Marianne F. Touchie and Jeffrey A. Siegel. 2018. Residential HVAC runtime from smart thermostats: characterization, comparison, and impacts. *Indoor Air* 28, 6 (Nov. 2018), 905–915. https://doi.org/10.1111/ina.12496
[12] U.S. Energy Information Administration. 2021. Units and calculators explained: Degree Days. https://www.eia.gov/energyexplained/units-and-calculators/degree-days.php