Application of Occupant Behavior Prediction Model on Residential Big Data Analysis

Yunjeong Mo†
Construction Management Department
University of North Florida
Jacksonville, FL, USA
y.mo@unf.edu

Dong Zhao
School of Planning, Design and Construction
Michigan State University
East Lansing, MI, USA
dzhao@msu.edu

ABSTRACT

Occupant behavior is multifaceted, and a systematic approach is required to understand occupant behavior comprehensively. This research aims to define a structure of the relationship between energy consumption, building technology, and occupant behavior, using the Occupant Behavior Prediction Model. The model can predict and explain occupant energy usage-related activities. A machine learning approach is used to develop the model, and datasets from the American Time Use Survey (ATUS) are used to verify the model. The results show that the energy use activities with higher predictive performances are more stable and habitual compared to the ones with lower predictive performances. The prediction accuracy achieved by this model for these habitual activities reached as high as 99%. The findings imply that the building systems and control strategies need to be adjusted to accommodate habitual energy use behaviors, rather than changing the behaviors. In addition, educational interventions seem more effective on the less habitual behaviors, which often change.

CCS CONCEPTS

• Applied Computing - Physical sciences and engineering

KEYWORDS

Occupant behavior, Residential building, Energy Use Prediction, Big Data, Urban Scale Data Analysis

ACM Reference format:

Yunjeong Mo and Dong Zhao. 2021. Application of Occupant Behavior Prediction Model on Residential Big Data Analysis. In *Proceedings of The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys), Nov 17-18, 2021, Coimbra, Portugal,* 4 pages. https://doi.org/10.1145/3486611.3491121

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

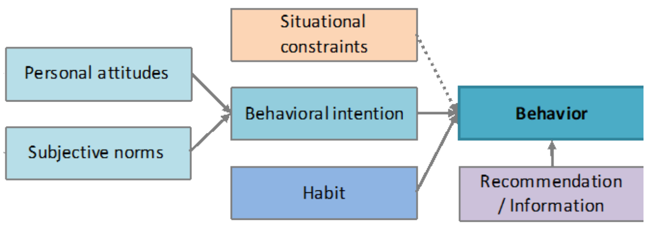
ACM ISBN 978-1-4503-9114-6/21/11…$15.00
https://doi.org/10.1145/3486611.3491121

1 INTRODUCTION

Residential building energy consumption is affected by climate, physical properties of the building, building services and energy systems, appliances in the household, occupant behavior, and the interactions among them [1]. As the building technologies grow more advanced, the energy consumption in residential buildings becomes more influenced by occupant behavior and living style, which emphasizes the need to understand occupant behavior and the relationship between occupant behavior and energy consumption. Occupant behaviors have been often studied based on socioeconomic factors, such as age, gender, marital status, number of children, employment status, and income level. However, this method has significant shortcomings, in that socioeconomic factors cannot fully explain occupants' energy consumption patterns. Even if occupants have similar socioeconomic characteristics, these similar characteristics do not guarantee similar behaviors. When an analysis only considers socioeconomic factors, the result will provide limited information [2]. Occupant behavior is associated with more than socioeconomic factors, and occupant behavior can be caused by a variety of factors [3]. It is critical to comprehensively identify not just occupant-specific characteristics like socioeconomic status and behavior hierarchy, but also external factors such as building attributes and climate. Therefore, a model is necessary to define and understand occupant behavior comprehensively [4]. This research aims to define a model of relationships between energy consumption, building technology, and energy usage-related behavior, then uses that model to explain occupant behavior. This model is applied to predict occupants' behavior and to identify how predictable and habitual each activity is. "Habitual behavior" denotes a behavior influenced by habits. This new model integrates the concept of habitual behavior and reduces the gap between energy consumption and occupant behavior.

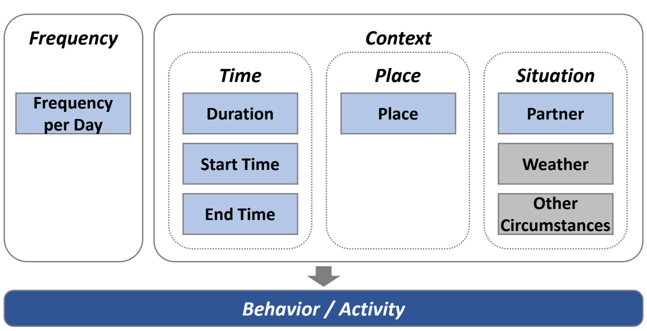
2 BACKGROUND

Behavioral routines and lifestyles are critical for energy saving because they have significant influences on daily energy use, but they are difficult to affect, changing gradually over time or not at all. Most people want to maintain their existing behavioral routines, lifestyles, and habits. Therefore, changing attitudes is easier than changing behaviors, and many studies report that building occupants' attitudes have changed to be more energyconscious, but they are unlikely to change their behaviors to match [5]. Behavioral intention and habit lead to behavior when situational constraints do not exist (Figure 1). Danner et al. [6] studied the role of habit and intention in the prediction of people's future behavior. They suggest that the frequency and stability of the context of past behavior mediate the role of intention. Intention has more influence on future behavior when habits are weak with low frequency or unstable context, while it has less influence when habits are strong with high frequency and stable context. Similarly, Triandis [7] suggested a model explaining the interaction between habit and intention in the prediction of future behavior: when a habit is stronger, the relationship between intention and behavior becomes weaker.



3 BEHAVIOR PREDICTION MODEL

The Occupant Behavior Prediction Model aims to predict occupant behavior through energy consumption data. In addition, this model can identify habitual and non-habitual behaviors, which can potentially be used for efficient building operation/control strategies, interventions/education, and so on. Unlike previous models that focused on predicting energy consumption by occupant behavior, or changing occupant behavior through intervention or education, this model investigates the reverse: predicting occupant behavior based on energy consumption.



The Occupant Behavior Prediction Model incorporates the function of habit on the formation of behavior, which is innovative in residential energy and occupant behavior studies. Existing studies [8, 9] suggested that the strength of a habit should be measured by reflecting its frequency and stability of its context. They estimated the strength of habits by multiplying a measure of past behavior frequency with a measure of context stability. This provided a habit scale, where a higher score indicates a strong habit with high frequency in a stable context, and a lower score indicates a weak or nonexistent habit with low frequency in an unstable context. Given that the contexts remain relatively stable, past choice of behavior can have more influence on the later choice of behavior [10]. Wood et al. [11] defined habits as behaviors that are performed repeatedly in stable contexts, because context stability is important for automatic responding. The components of this Occupant Behavior Prediction Model are extracted from those habitual behavior studies and used to measure the strength of habit in occupant behavior. Behaviors and activities are explained with the following main components (Figure 2).

- ***Frequency:*** Number of times an activity performed per day
- ***Context:***
  - ***Time***
    - ***Duration:*** Total minutes of an activity, from the start time to the end time
    - ***Start Time:*** Start time (HH:MM) of an activity
    - ***End Time:*** End time (HH:MM) of an activity
  - ***Place (Where):*** Physical location where an activity is performed
  - ***Situation***
    - ***Partner (Who):*** Person/people with whom an activity is performed
    - ***Weather:*** Weather conditions when an activity is performed
    - ***Other Circumstances:*** Other circumstances affecting an activity such as special events and ambient people

4 CASE STUDY

4.1 ATUS Data

Table 1. New Activity Code, Energy, Appliances

Code	Activity	Energy
AA01	Washing, dressing, and grooming	E,W,G
BB01	Interior cleaning	E
BB02	Laundry	E,W,G
BB03	Food and drink preparation	E,W,G
BB04	Kitchen and food clean-up	E,W
BB05	Heating and cooling	E,G
BB06	Gardening, ponds, pools, and hot tubs	W,G,E
BB07	Care for animals and pets	E,W
BB08	Vehicle repair and maintenance	E
CD01	Physical care for children	E,W
CD02	Physical care for/helping adults	E,W
EF01	Work for job(s)/research/homework	E
LL01	Television	E
LL02	Listening to/playing radio or music	E
LL03	General computer use	E

*** E: Electricity, W: Water, G: Gas*

The American Time Use Survey (ATUS) is an annual national survey conducted by the U.S. Bureau of Labor Statistics (U.S.

BLS) [12]. The U.S. BLS conducts the national survey on how the population allocates time in their daily lives. The ATUS assesses what (activity), where (place), and with whom (partner) a nationally representative sample of Americans spends their time on a regular day. The survey contains detailed daily activities from more than 10,000 respondents per year [13].

In this study, the ATUS 2015 data are used to examine energy usage-related behavior, focusing on habitual consumption among habitual, structural, and daily variation consumptions. The activities are defined in three tiers: the first tier has 18 overall categories of activities, the second tier has more detailed 110 subcategories under the first tier, and the third tier has the most detailed 465 categories under the first and second tiers. The energy usage-related activities are selected and reorganized from the third-tier activities (Table 1).

4.2 ML Classification

This research used a machine learning (ML) approach to understand energy usage-related behavior based on the behavior prediction model using the ATUS data. The Occupant Behavior Prediction Model was applied to predict energy usage-related activities and to identify the predictability and the habitual characteristic of each activity. For the data analysis, various packages in Python and R were used, and Support Vector Machine was used to predict occupant behavior from their energy usage pattern. SVM is based on the maximum-margin hyperplane, an algorithm used to find a special type of linear model [14]. In the dataset, the numeric variables were standardized to a mean of 0 and a standard deviation of 1. The performance of the algorithms was evaluated with Accuracy, Precision, Recall, and F1-score. Accuracy, Precision, Recall, and F1-score were used to evaluate the predictive performance of the algorithms. Accuracy is the percentage of correct predictions, or the ratio of true predictions to the total number of instances. Precision and Recall are the indexes of relevance. Precision is the ratio of correct positive predictions to all positive predictions. A low precision implies a large number of false positives. Recall is the ratio of correct positive predictions to the sum of correct positive predictions and wrong negative predictions. A low recall implies a large number of false negatives. F1-score is the harmonic mean of precision and recall.

Based on the Occupant Behavior Prediction Model, *Frequency, Duration, Start Time, End Time,* and *Partner* variables were initially selected among the available features in this dataset. The *Place* variable was excluded in the baseline algorithm selection since it only contains the values of Home (1) and Not Collected (-1). Originally, the activity file from the 2015 ATUS contained 214,429 activities from 10,905 respondents. For this study, only energy usage-related activities were selected, so 76,980 activities from 10,849 respondents remained. Since this study focuses on residential energy behaviors, those activities were narrowed down to only include the ones that happened in the respondent's home or yard. The ATUS does not collect the location and partner information for certain types of activities, such as sleeping and grooming, due to privacy concerns. Therefore, it was assumed that those activities happened alone at home [2]. This left 67,115 activities from 10,772 respondents, which this study used for the analysis. 70 percent of the whole dataset was set as the training set and the remaining 30 percent was set as the testing set.

4.3 Result

Among the features, *Partner* is a categorical variable, and *Frequency, Duration, Start Time,* and *End Time* are numeric variables. The original ordinal HH:MM format of Start Time and End Time was converted to a numeric minute format. For example, 13:10 is converted to 790 minutes. The value ranges are very different among the numeric variables, which can affect the performance of the machine learning algorithms. For example, many algorithms (such as the radial basis function (RBF) kernel of SVM) assume that all input variables/features have means of 0 and variances in the same order of magnitude. Thus, if one feature has much larger variance than the others, it might have too heavy an influence on the objective function and weaken the estimating power from other features as expected [15]. This explains the performance improvement of SVM with standardized features because the RBF kernel is used in this run. In the following steps, SVM with standardized features is further developed to improve its predictive performance.

The performance of the model for each activity was calculated as summarized in Table 2. *Washing, dressing, and grooming* (AA01) shows the highest accuracy (0.99), which means this model predicts 99% of AA01 activity correctly. The model predicts *Watching television* (LL01), *Physical care for children* (CD01), and *Food and drink preparation* (BB03) with higher performance. However, the model incorrectly predicts *Heating and cooling* (BB05), *Vehicle repair and maintenance* (BB08), and *Listening to/playing radio or music* (LL02). The number of instances of an activity is also relevant to the model's predictive performance for each activity, since the model can be trained better with more data.

Table 2: Predictive Performance of Each Activity

Code	Accuracy	Precision	Recall	F1-score	Count
AA01	0.99	1.00	0.99	0.99	4607
LL01	0.82	0.62	0.82	0.71	4900
CD01	0.65	0.63	0.65	0.64	1527
BB03	0.71	0.45	0.71	0.55	3003
BB04	0.50	0.49	0.50	0.49	1031
BB01	0.28	0.35	0.28	0.31	1028
BB07	0.14	0.47	0.14	0.22	635
LL03	0.15	0.39	0.15	0.21	1101
EF01	0.06	0.42	0.06	0.11	742
BB02	0.06	0.31	0.06	0.10	860
CD02	0.02	1.00	0.02	0.04	88
BB06	0.01	0.08	0.01	0.01	400
BB05	0.00	0.00	0.00	0.00	34
BB08	0.00	0.00	0.00	0.00	66
LL02	0.00	0.00	0.00	0.00	113

*** Counts are from the testing set, which is 30% of the whole dataset*

5 Discussion

The Occupant Behavior Prediction Model can predict occupant behavior with overall 64% accuracy for the ATUS dataset, and its accuracy can reach up to 83% for a subgroup of habitual activities. Notably, the model shows 99% accuracy for predicting washing, dressing, and grooming activity and 82% accuracy for predicting watching television activity. The multi-class classification problems are challenging, and achieving high accuracy in these compared to binary classification problems is difficult [16, 17]. The Occupant Behavior Prediction Model is applied for multi-class classification with 15 classes (15 activities). The result demonstrates high performance pertaining to multi-class classification, especially considering that the probability of correct predictions with simple statistical calculation is 6.7%.

This model can identify more habitual activities and less-habitual activities based on the prediction performance of each activity. The model was tested on the ATUS data to predict activities of the general occupants from nationally representative samples. From the results, people tend to wash, dress, and groom (AA01) as more predictable routines, and watch television in a predictable pattern. They take care of children (CD01) frequently when the children are in need of their care and help. Food and drink preparation (BB03) and kitchen and food clean up (BB04) are habitual and predictive behaviors. Interior cleaning (BB01), laundry (BB02), care for adults (CD02) or pets (BB07), general computer use (LL03), and working at home (EF01) are less predictive, meaning less habitual behavior. Heating fuel preparation (BB05), vehicle maintenance (BB08), and listening to radio/music or playing music (LL02) are very difficult to predict, and therefore they are non-habitual behaviors.

There exist some limitations to this study. The ATUS collects diary data for only one specific day from a respondent and does not ensure that it is a typical day for the respondent. Although this shortcoming is compensated for by the large number of samples collected, another study using occupants' daily records of multiple days is suggested to identify more precise and specific patterns of occupants' behavior. Also, while the ATUS records one activity at a time, multiple activities can happen concurrently in reality. For example, people may do laundry while watching television. Thus, the complexity of the activities should be considered when applying this model to another dataset.

The Occupant Behavior Prediction Model innovatively incorporated the concept of habit to predict occupant behaviors and identify habitual/non-habitual activities, while previous studies about occupant behaviors have tended to focus more on socioeconomic attributes to predict energy consumption. This novel approach explores the past habitual characteristics of the households, predicts their future behaviors, and identifies their habitual behaviors. Habitual behaviors are more difficult to change, but they are easier to predict. For these activities and behaviors, energy systems need to find efficient control strategies that are suitable for these behaviors rather than trying to change the behaviors. In contrast, less habitual behaviors, which are difficult to predict, might be easier to change, and education or intervention might be more effective on these activities. The result can be used to develop more improved occupant schedules and to set specific energy control strategies. Also, the results can be used to develop effective interventions or education for residential occupants. This model will be further applied to examine the geographical patterns of activities (horizontal analysis), and the timely patterns of activities (vertical analysis) in the following studies.

ACKNOWLEDGMENTS

This work is in part supported by National Science Foundation (NSF) No.2046374. Any opinions and conclusions in this work are of the authors and do not necessarily reflect the view of NSF.

REFERENCES

[1] Joakim Widén and Ewa Wäckelgård. 2010. A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy*, 87, 6 (2010), 1880-1892.
[2] Longquan Diao, Yongjun Sun, Zejun Chen and Jiayu Chen. 2017. Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation. *Energy and Buildings* (2017).
[3] Yunjeong Mo and Dong Zhao. 2021. Effective Factors for Residential Building Energy Modeling using Feature Engineering. *Journal of Building Engineering* (2021), 102891.
[4] Chen, Weiwei Yang, Hiroshi Yoshino, Mark D Levine, Katy Newhouse and Adam Hinge. 2015. Definition of occupant behavior in residential buildings and its application to behavior analysis in case studies. *Energy and Buildings*, 104 (2015), 1-13.
[5] Loren Lutzenhiser. 1993. Social and behavioral aspects of energy use. *Annual review of Energy and the Environment*, 18, 1 (1993), 247-289.
[6] Unna N Danner, Henk Aarts and Nanne K Vries. 2008. Habit vs. intention in the prediction of future behaviour: The role of frequency, context stability and mental accessibility of past behaviour. *British Journal of Social Psychology*, 47, 2 (2008), 245-265.
[7] Harry C Triandis. 1979. *Values, attitudes, and interpersonal behavior*. University of Nebraska Press, City.
[8] Judith A Ouellette and Wendy Wood. 1998. Habit and intention in everyday life: The multiple processes by which past behavior predicts future behavior. *Psychological bulletin*, 124, 1 (1998), 54.
[9] Wendy Wood, Leona Tam and Melissa Guerrero Witt. 2005. Changing circumstances, disrupting habits. *Journal of personality and social psychology*, 88, 6 (2005), 918.
[10] Ching-Fu Chen and Wei-Hsiang Chao. 2011. Habitual or reasoned? Using the theory of planned behavior, technology acceptance model, and habit to examine switching intentions toward public transit. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14, 2 (2011), 128-137.
[11] Wendy Wood, Jeffrey M Quinn and Deborah A Kashy. 2002. Habits in everyday life: Thought, emotion, and action. *Journal of personality and social psychology*, 83, 6 (2002), 1281.
[12] Daniel Kahneman, Alan B Krueger, David A Schkade, Norbert Schwarz and Arthur A Stone. 2004. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306, 5702 (2004), 1776-1780.
[13] U.S.BLS. 2018. *American Time Use Survey*. U.S. Bureau of Labor Statistics, City.
[14] Ian H Witten, Eibe Frank, Mark A Hall and Christopher J Pal. 2016. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
[15] Scikit-Learn. 2017. *sklearn.preprocessing.StandardScaler*. Scikit-Learn.
[16] Dewan Md Farid, Li Zhang, Chowdhury Mofizur Rahman, M Alamgir Hossain and Rebecca Strachan. 2014. Hybrid decision tree and naïve Bayes classifiers for multi-class classification tasks. *Expert Systems with Applications*, 41, 4 (2014), 1937-1946.
[17] Isabelle Guyon and André Elisseeff. 2003. An introduction to variable and feature selection. *Journal of machine learning research*, 3, Mar (2003), 1157-1182.