Clustering Appliance Energy Consumption Data for Occupant Energy-Behavior Modeling

Poorvesh Dongre, Asma Aldrees, and Denis Gračanin
Department of Computer Science, Virginia Tech
Blacksburg, Virginia, USA
{poorvesh,aaldrees,gracanin}@vt.edu

ABSTRACT

Energy consumption of buildings varies significantly across buildings with similar functions and locations. Occupant behavior is one of the most significant sources of uncertainty related to energy consumption in buildings. A deeper understanding of occupant energy behavior can help in designing personalized behavior intervention strategies to save energy and predict energy consumption. This paper uses the Pecan Street dataset to cluster building occupants based on the energy they consume for each appliance in the household, and then developed load profiles for each of the clusters.

CCS CONCEPTS

• **Information systems** → *Temporal data*; • **Computing Methodologies** → *k-means Clustering, Machine Learning*.

KEYWORDS

Occupant behavior, Home-appliances, Data-driven methods, Clustering

ACM Reference Format:
Poorvesh Dongre, Asma Aldrees, and Denis Gračanin. 2021. Clustering Appliance Energy Consumption Data for Occupant Energy-Behavior Modeling. In *BuildSys '21: The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 4 pages.

1 INTRODUCTION

One of the most important factors that impact building energy consumption is occupant behavior [5]. Having a better understanding of occupant energy-behavior and providing suitable behavioral interventions can result in energy savings. However, occupant energy-behavior is complex and dynamic; it is influenced by several factors such as lifestyle, social and economic status, etc. There are various methods to model occupant energy-behavior. Data-driven methods to model energy-behavior are becoming increasingly popular due to the availability of various energy-related datasets.

The dataset used in this research is a time-series energy consumption data of appliances at residential buildings in NY state provided by Pecan Street Institute [8]. Our research intends to use the contextual features in this dataset (energy consumed by appliances and occupant's demographics in a household) to create a model of occupant's energy-behavior and predict the energy consumption of occupants with similar demographics.

Understanding how energy consumption is affected when the household is equipped with different appliances and is used by occupants of varying background can help in designing intervention strategies to save energy. In this paper, we perform clustering on the time-series dataset to cluster occupant ids with similar energy consumption characteristics and develop load profiles for each cluster.

The results of our preliminary work show that occupants have similar energy consumption patterns except for a few outliers that show very high energy consumption. Future work will include predicting the load profile of an occupant based on their demographics.

2 BACKGROUND

Occupant energy behavior modeling and simulation have long been studied by researchers for energy conservation purposes. Having a better understanding of occupant energy consumption patterns can help load-serving entities (LSE) to develop effective behavioral intervention strategies that can force the occupants to save energy [7].

Norouziasl et al. did a comprehensive literature review of energy-related occupant modeling and simulation tools and techniques [6]. They show that data-driven methods are getting increasingly popular in predicting occupant behavior patterns and estimating building energy consumption. Clustering was used by Yu et al. to examine the effects of occupant behavior on the building energy consumption [10]. Diao et al. used American time use survey (ATUS) data, Residential energy consumption survey (RECS), and weather data for occupant activity recognition and clustered occupants in categories based on their everyday activites [5].

Azaza and Wallin used two clustering techniques (k-means and self-organizing maps) to identify occupant groups having high energy consumption and enable tailored dynamic pricing plans [1]. These clustering techniques were also used by Causone et al. to cluster the daily electric load profiles in different meaningful groups [3]. Wen et al. used principal component analysis (PCA), k-means, and shape-based clustering to cluster the electric load profiles [9].

The use of energy time-series data to cluster occupants can be a challenge. Energy consumption characteristics of occupants can vary across time. A possible solution, as given by Khosrowpour et al. can be segmenting energy dataset based on different times of the day and then apply clustering to categorize occupants. But the use of unsupervised machine learning (ML) techniques such as clustering gives the clusters with no labels of the energy consumption characteristics. Das et al. propose a segmentation method

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

ACM ISBN 978-1-4503-9114-6/21/11.
https://doi.org/10.1145/3486611.3491129

that uses the k-means algorithm to find the number of cluster and subsequently use supervised classification and graphical lasso to label these clusters [4].

3 APPROACH

In our approach to model occupant energy consumption behavior, we manually segment the dataset into sub-datasets based on different days (holidays and weekends). Each of these sub-datasets is then clustered using the k-means algorithm to group occupants with similar energy consumption patterns. We used the elbow method using inertia to choose the optimum number of clusters in the k-means algorithm. The elbow method is one of the most popular methods to determine the optimal value of k [2]. We calculate the inertia for each value of k and sketch an elbow plot. The optimal value of k is determined at the point corresponding to a drastic change in the rate of reduction in inertia score.

Next, the clusters with their respective centroids are projected on a 2D space using Multi-Dimensional Scaling (MDS). MDS is a form of non-linear dimensionality reduction technique that can be used to visualize the level of similarity among individual cases in a dataset.

Lastly, the energy consumption profile of every cluster is shown using a parallel plot. Observations from this plot are used to label the clusters and respective data-items. A supervised ML technique will be used on the labelled dataset to predict the energy consumption of occupants based on their demographics.

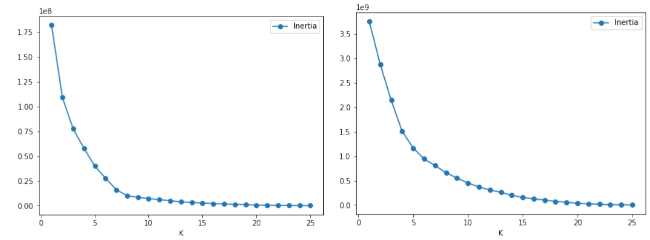
Our approach differs from other clustering-based energy-behavior modeling approaches in the following ways: (1) The dataset used captures energy consumed by individual appliances in a household (2) The dataset is manually segmented based on holidays and weekends as these days capture more energy consumption characteristics than other days (3) We apply unsupervised clustering on each segment and find occupant clusters in both segments (4) A parallel plot is used to show the energy consumption profile of each cluster in both segments and observations from the parallel plot are used to label the clusters.

4 RESULTS

The elbow plots for holidays and weekends dataset obtained by applying k-means clustering is shown in Figure 1. The optimal number of clusters as observed from both elbow plots is 8, as shown in Figure 1. The clusters along with its respective centroids as a result of the MDS exercise are shown in Figure 2. The load profiles of each cluster for holidays and weekends are shown in Figure 3. It is also important to note that some occupants may belong to different clusters during each segment.

4.1 Labelling Energy Consumption Behavior

Observation of clusters and centroids, show that most cluster in holidays and weekends have either one or two occupants. There is only one cluster, Cluster 0 (purple colored), in both segments that has the highest number of occupants in it. Examining the load profiles and clusters together reveal that Cluster 5 (yellow colored) has two occupants in the holiday segment and consume significant energy using "furnace". Similarly, Cluster 1 (blue colored) has two occupants in the weekend segment and show a peak for "furnace".



(a) Holidays (b) Weekends

Figure 1: Elbow method plot: a) during holidays; b) during weekends.

Occupants in Cluster 0 show moderate energy consumption during holidays and weekends but consume more energy using "jacuzzi", "office", and "solar". Based on the observed energy characteristics, Cluster 5 can be labelled as the "holiday heat users", Cluster 1 can be labelled as the "weekend heat users", and Cluster 0 can be labelled as the "regulars".

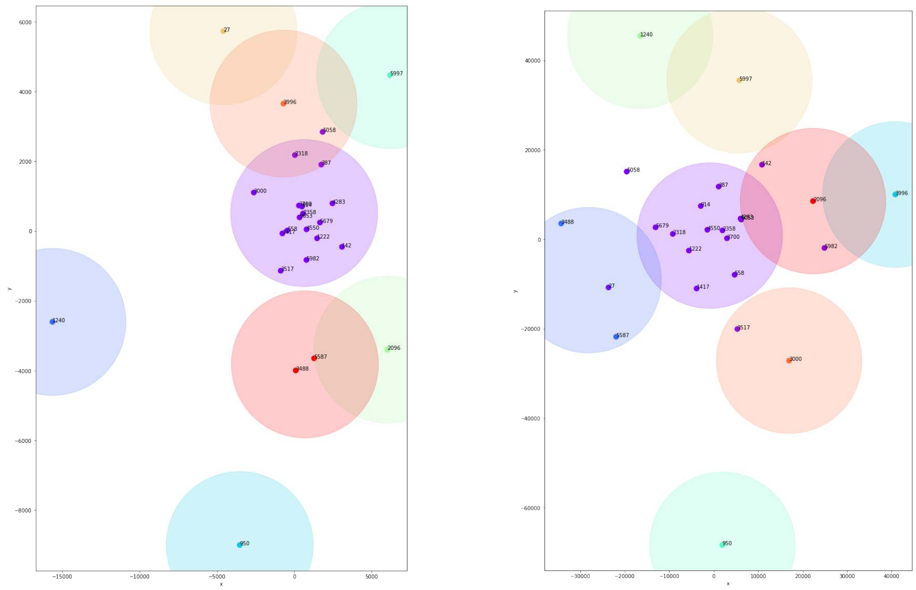
Among all outliers, Occupant-ids (1240, 950, 5997) can be labelled as critical anomalies. This is because apart from being the only occupant in that cluster they also show atypical energy consumption levels on both holidays and weekends. A detailed description of the energy consumption behavior of these occupants is given below:

- Occupant-id 1240: has the lowest energy consumption in the holidays, while had the highest energy consumption during the weekends. According to the load profile of this occupant, we could infer that they might not spend the holidays in the household and usually travel or spends most of the time out.
- Occupant-id 950: has the highest energy consumption during the holidays. While, they had the lowest energy consumption during the weekends. Regarding the significant gap of the energy consumption throughout both segments, we could infer that they might have more visitors during the holidays.
- Occupant-id 5997: has a high energy consumption during the weekends, and a very low energy consumption during the holidays.

4.2 Estimating Appliance Energy Consumption

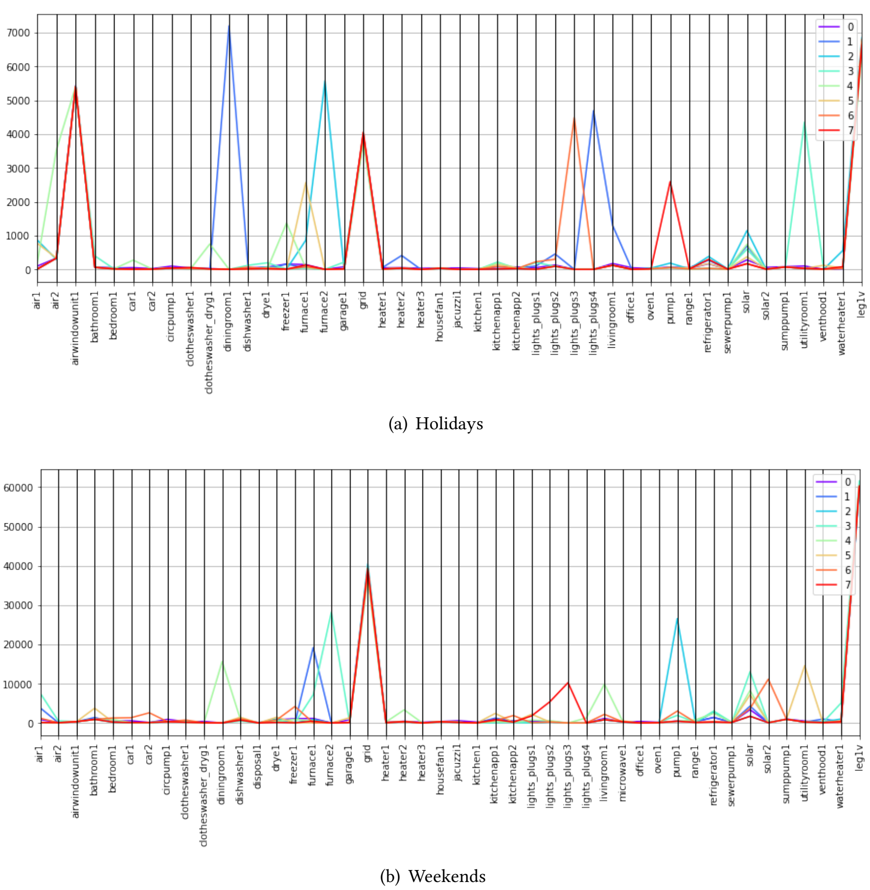
Observation of load profiles, Figure 3, indicate that during holidays, "dining room" recorded the highest energy consumption, followed by "furnace" and "air window unit". Also, "light plugs" and "grid" were used a lot. During the weekends, "grid" consumed most of the energy, followed by "furnace" and "pump".

We also evaluated the appliances based on their dependence on the voltage measure of the overall energy consumption in the household. These appliances are ranked starting with the top appliances that affect the overall energy consumption in the household. We used the Ordinary Least Square Regression Model to choose the appliances that are more dependent on the household voltage measure. The results of this analysis are illustrated in Table 1. We provided the top ten important appliances that affect the household voltage measure and significantly affect the overall energy consumption.



(a) Holidays (b) Weekends

Figure 2: Occupant clusters and centroids over segments: a) during holidays; b) during weekends.



(a) Holidays

(b) Weekends

Figure 3: Energy consumption profiles of occupant clusters for different time segments: a) during holidays; b) during weekends.

Table 1: Top ten appliances affecting the household energy consumption: a) during holidays; b) during weekends.

Holidays		Weekends	
Appliance	**Importance**	**Appliance**	**Importance**
heater3	0.055	airwindowunit1	0.181
lights-plugs1	0.052	Furnace2	0.070
sumppump1	0.051	heater3	0.049
airwindowunit1	0.044	lights-plugs1	0.048
solar2	0.032	Office1	0.039
furnace2	0.024	sumppump1	0.037
solar	0.023	lights-plugs3	0.036
heater1	0.017	solar2	0.025
diningroom1	0.015	heater1	0.023
lights-plugs2	0.014	kitchen1	0.022

By comparing the appliances affect on the overall household energy consumption, and the actual usage of these appliances by occupants, we found a significant gap in the energy consumption that should be addressed. For instance, in the holidays, "heater3" is the top appliance in affecting the overall energy while no occupants were using it, see Figure 3 (a). Additionally, in weekends, "airwindow unit1" and "heater3" have shown significant impact on the household energy consumption while they are barely used by the occupants, see Figure 3 (b). From the above analysis, we infer that there might be an energy leakage or the appliance might consume higher energy than their actual usage by occupants. This analysis calls to pay more attention to the appliances type and production, due to their significant effect on household energy consumption. While energy consumption can be enhanced by tracking occupant energy-behavior, appliances' performance evaluation is another important criterion that should be addressed to enhance the overall energy consumption activities and eliminate any potential waste of energy.

5 DISCUSSION

Clustering time-series energy consumption data is widely used to model occupant energy behavior. But clustering being an unsupervised ML technique cannot provide labels to the formed clusters and has to be combined with other methods to get these labels.

This research combines k-means clustering with a parallel plot to categorize building occupants based on their energy consumption. The objective is to use the energy-behavior model to design tailored intervention strategies for occupants with high energy usage and predict the energy-behavior of occupants with similar demographics. Our preliminary results from clustering energy data over holidays and weekends show that there are 8 clusters. Most of these clusters have one or two occupants in them and show anomalous energy consumption patterns. These occupants showed anomalies during both holidays and weekends. Therefore, personalized intervention strategies can be designed for such occupants to encourage them to save energy. The cluster with maximum occupants did not show any significant peak in the load-profile plot of clusters.

Overall, results showed a significant variation in energy consumption among appliances during holidays and weekends. These results could help energy providers, appliance producers, and building managers to look at the appliances with high energy consumption and investigate the occupant consumption behavior while using these appliances.

6 CONCLUSION

Future experimental research involves the application of supervised ML techniques on the labelled dataset to predict the energy-behavior of an occupant based on demographic information such as household income, the average age of residents in a household, etc. The design of personalized intervention strategies for occupants with very high energy usage is also an area of exploration. Further analysis can be done to understand the distribution of energy consumed by appliances over time and model occupant energy-behavior for various time periods. A similar analysis during Covid-19 might lead to different and interesting results. It is expected to see much higher energy use in these households during Covid-19 (between mid-March and December 2020) and more use of kitchen, living room, and office appliances in the same period since residents are spending most of their time inside their houses.

REFERENCES

[1] Maher Azaza and Fredrik Wallin. 2017. Smart meter data clustering using consumption indicators: responsibility factor and consumption variability. *Energy Procedia* 142 (2017), 2236–2242.
[2] Purnima Bholowalia and Arvind Kumar. 2014. EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications* 105, 9 (2014), 17–24.
[3] Francesco Causone, Salvatore Carlucci, Martina Ferrando, Alla Marchenko, and Silvia Erba. 2019. A data-driven procedure to model occupancy and occupant-related electric load profiles in residential buildings for energy simulation. *Energy and Buildings* 202 (2019), 109342.
[4] Hari Prasanna Das, Ioannis C Konstantakopoulos, Aummul Baneen Manasawala, Tanya Veeravalli, Huihan Liu, and Costas J Spanos. 2019. A novel graphical lasso based approach towards segmentation analysis in energy game-theoretic frameworks. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 1702–1709.
[5] 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* 147 (2017), 47–66.
[6] Seddigheh Norouziasl, Amirhosein Jafari, and Yimin Zhu. 2021. Modeling and simulation of energy-related human-building interaction: A systematic review. *Journal of Building Engineering* (2021), 102928.
[7] IP Panapakidis, MC Alexiadis, and GK Papagiannis. 2012. Load profiling in the deregulated electricity markets: A review of the applications. In *2012 9th International Conference on the European Energy Market*. IEEE, 1–8.
[8] Pecan Street. 2019. Pecan Street Dataport. https://dataport.pecanstreet.org/
[9] Lulu Wen, Kaile Zhou, and Shanlin Yang. 2019. A shape-based clustering method for pattern recognition of residential electricity consumption. *Journal of cleaner production* 212 (2019), 475–488.
[10] Zhun Yu, Benjamin CM Fung, Fariborz Haghighat, Hiroshi Yoshino, and Edward Morofsky. 2011. A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and buildings* 43, 6 (2011), 1409–1417.