Where, When and Watt?

Akhil Mathur Anand Prakash Vasudevan Nambeesan Yoolhee Kim Zeal Shah
{akhilm, anandkrp, vnambees, yoolheek, zpshah}@andrew.cmu.edu
Course 12-752 Data-Driven Building Energy Management
Carnegie Mellon University (CMU), Pittsburgh PA, United States

ABSTRACT
Machine learning tools can provide useful results when working with large datasets. This project attempts at identifying power consumption usage patterns for appliances and predicting occupancy patterns in a household provided by the DRED (Dutch Energy Residential Energy) dataset. In the process of exploring the dataset, variation of power consumption of the various rooms of the household were observed. This activity allowed the observation of seasonal patterns in the power consumption. The power consumption of the individual appliances over the year also yielded interesting patterns. In addition, the variation of occupancy in the different rooms over the year is also reported. K-Means clustering from scikit-Learn was used to cluster both power consumption and occupancy datasets. This exercise yielded hourly patterns in the power consumption of the various appliances studied. The occupancy clustering provided information about the occupancy patterns of the various rooms of the household. Additionally, Decision trees from scikit-learn was used to predict occupancy based on power consumption. After cross validation, the model has report high accuracies for predicting the occupancy.

Keywords
Machine learning; occupancy prediction; appliance power prediction; K-means clustering; Decision-tree classification.

1. INTRODUCTION
With the growth of ubiquitous computing and increasing popularity of powerful cyber-physical systems, "smart homes" are becoming increasingly common. A significant feature of smart homes, in addition to improvement in lifestyle, is efficient energy consumption as well as information on the state of the household such as occupancy. In this paper, based on the energy consumption patterns of appliances in a residence, the occupancy of various rooms is predicted.

Firstly, determining the consumption pattern of a home helps energy efficiency by schedule loads or tasks. For example, if the occupant runs the washing machine during the peak grid energy demand, based on the performed exploratory data analysis, recommendations can be made to shift the operating time. Secondly, with the predicted occupancy information, the environment settings can be appropriately set (fans and HVACs can be turned off if no occupant), and hence reduce power consumption.

Accurate occupancy prediction using appliance energy consumption data mitigates the need to install privacy invading sensors like motion sensors and cameras - which also affect aesthetics of the residence.

Section 2 talks about the dataset used in this paper, Section 3 discusses the algorithms and the techniques used to achieve the above objectives – this would help in replicating the research we have done – and Section 4 contains the results and analysis of the same. We discuss cross-validation of our model in Section 5. The report is concluded with Section 6, in which we provide what could be done to take this work forward. Finally, the list of references used is provided in Section 7.

2. DRED DATA SET
This dataset was provided by DRED (Dutch Residential Energy Dataset), which is an open-access, publicly available dataset from the Netherlands. The dataset provides details on the energy consumption and occupancy of a single household located in the Netherlands [1]. The data was collected by several sensors measuring electricity, occupancy, and ambient parameters of the household. The energy consumption is collected over a period of 6 months from July 5th to December 5th 2015 for each second. The electricity monitor has aggregated energy consumption and appliance level energy consumption, as well as room-level location information of occupants.

The figure below displays the layout of the household, as well as the appliance in each of the rooms of the household.

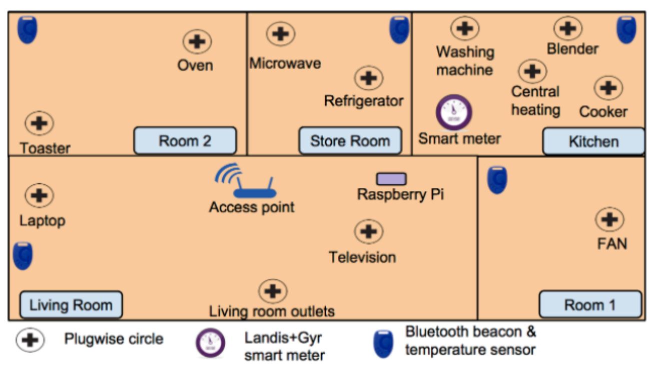


Figure 1. Room and Appliance Layout of Household [1]

3. PROPOSED APPROACH
First the data set will be described through exploratory data analysis and k-means clustering of the appliance and occupancy data. Eventually decision tree classification will be used in order to predict the occupancy of each room based on the appliance data.

3.1 Exploratory Data Analysis
For the exploratory data analysis, we can rearrange the data frame to get the power consumption of the different appliances in separate columns. We use the occupancy data provided with the data set to add to our final data frame by checking for the occupancy in every time period to represent the final occupancy in the room through 1 and 0s, representing occupancy and vacancy respectively. To do further analysis, we can resample the data frame to get daily data instead of seconds. This would allow us to

get a better representation of the consumption pattern for different appliances.

3.2 K-means Clustering Appliance Use and Occupancy

3.2.1 Appliance Power Clustering

The power consumption of each appliance in this residence is divided into two clusters. Two clusters were chosen to attempt to divide the power use based when the rooms are occupied and when they are not. Each appliance has power consumption data for every second over the 6-month period. The appliances for which energy data is available are: television, fan, fridge, laptop, electric heating element, oven, washing machine, microwave, toaster, sockets, and cooker. The results of this clustering can be found in section 4.2.1.

3.2.2 Occupancy Clustering

The occupancy of each room is defined as either zero (unoccupied) or one (occupied). The dataset provides this number for every second. For clustering this dataset to get hourly patterns over the length of the dataset, this number is averaged over the hour. Therefore, fractional occupancy is generated. This fraction will give information pertaining to the duration of the occupancy over the hour (higher number implies the room was occupied for more number of seconds of the hour). Each room in the dataset was divided into two clusters. The results of this clustering can be found in Section 4.2.2.

3.3 Regression Tree for Room-Level Occupancy Prediction Model

In the part, we try to predict occupancy in each of the rooms based on the power consumption of all the appliances in the house. We use a Regression Tree to create this prediction model and test this to find the accuracy. We will also create regressions tree models based on the power consumption only of the appliances in the room and based on the accuracy, we could decide if it is sufficiently accurate. If satisfactory accuracy is obtained, it would help us reduce the input vector space for prediction and hence could get us faster results and smaller trees.

4. RESULTS AND DISCUSSION

4.1 Exploratory Data Analysis

Through our data analysis, we try to measure and learn about the power consumption in different parts of the house.

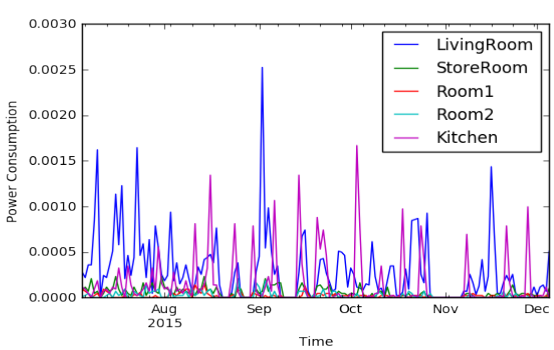


Figure 2. Variation in occupancy of rooms for the given year

The above graph shows the frequency of occupancy for each of the individual rooms throughout the year. Since the data frame provides occupancy information in the Boolean form, the mean of 1's and 0's gives us an hourly average value to compare the occupancy in different rooms. We see from the graph that the living room and the kitchen are the most occupied during most periods. The graph shows a unique un-occupancy during the month of November, for the whole house. We can correspond this with the power consumption data to make reasonable conclusions.

Figure 3 below show the power consumption of the different appliances according to their location in the house. We can identify the appliances with the highest power consumption in every room and their consumption pattern with seasonal variation.

We can identify the following appliances- laptop and television in the main room, fan in room 1, oven in room 2, fridge in the store room and the cooker in the kitchen as the highest consuming power appliances. Some common observations made with our data is the reduction in the fan power consumption with the onset of October. The refrigerator draws considerable power throughout the year and is the most consuming appliance in the house. We observe that almost all the appliances draw no power for certain part of November. Using this data, we can conclude that the house was unoccupied for this part of the month.

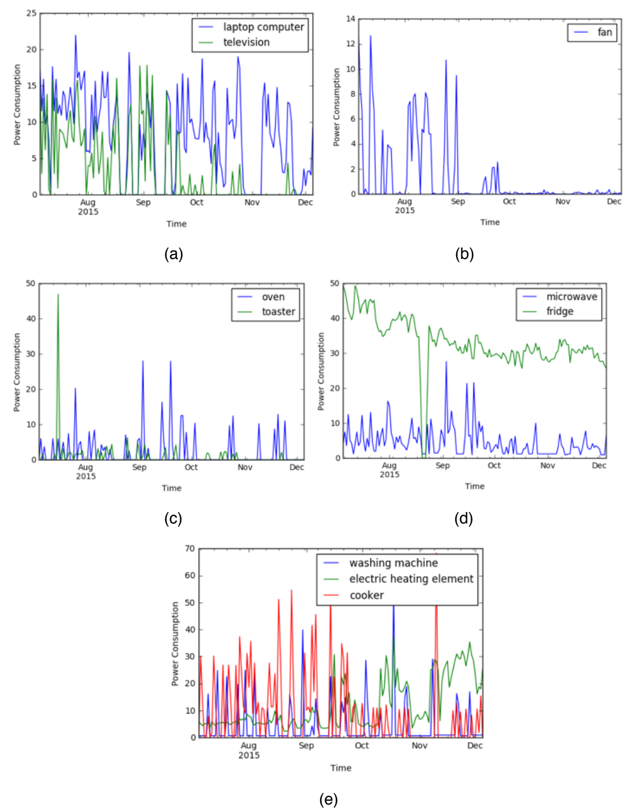


Figure 3. Variation of power consumption in (a) Living Room (b) Room 1 (c) Room 2 (d) Store Room (e) Kitchen

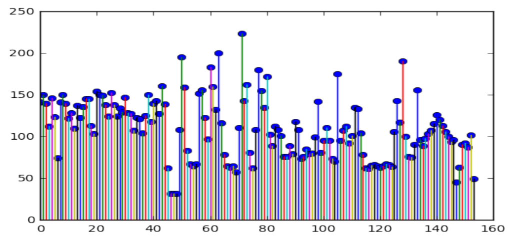


Figure 4. Mains power consumption for the year

The stem plot shown in Figure 4 displays the variation in the mains power consumption for each day of the given data set. We observe that there is a clear decrease in the overall consumption with the onset of winters, as the house consumes more natural gas for its heating requirement.

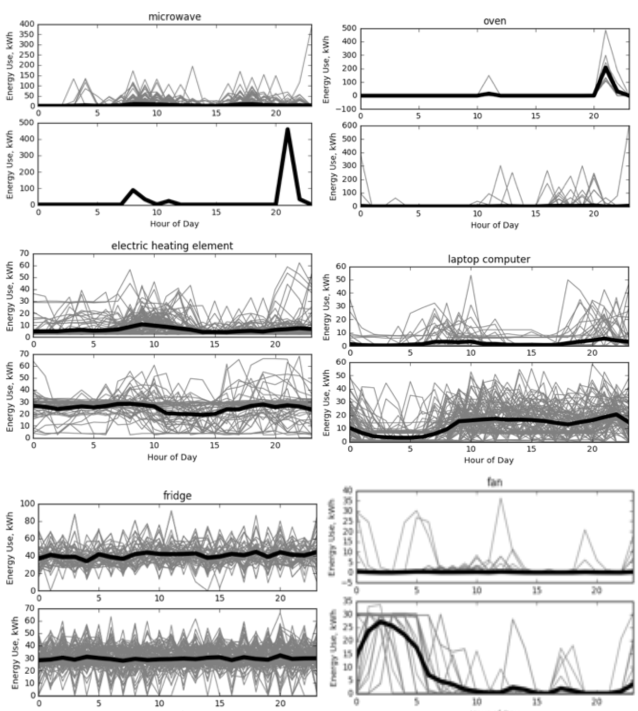
4.2 Clustering

4.2.1 Appliance Power Clustering

It was noted that as the number of clusters was increased for the power appliance data, the variance of each cluster is reduced. However, the dataset was divided only into two clusters per room as an attempt to divide the dataset into high occupancy and low occupancy clusters.

Some trends to note are that laptop is used more from the tenth hour of the day (10:00 AM) and on. The microwave is used throughout the day, but mostly around 10 AM and 6 PM. The oven is used less frequently, but is used from 6 to 8 PM. The fridge power use is constant throughout the day, as expected. The fan is most often used at night, from 12 to 5 AM. The washing machine power peaks around 10 AM to 3 PM. The toaster is almost always used in the morning around 10 AM, while the cooker peaks around both 10 AM and 8 PM.

Below are the graph results of the appliance power clustering.



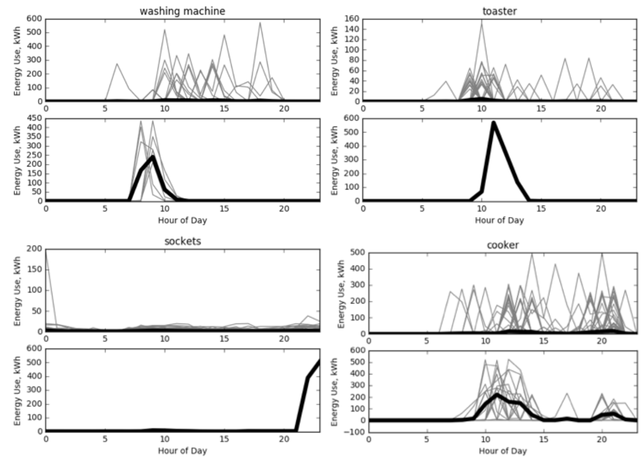


Figure 5. Appliance Power Clustering of television, fan, fridge, laptop, electric heating element, oven, washing machine, microwave, toaster, sockets, and cooker

4.2.2 Occupancy Clustering

It was noted that as the number of clusters was increased, the variance of each cluster is reduced. For simplicity, the dataset was divided only into two clusters per room. A general comment maybe made that dataset was divided into a high occupancy cluster and a low occupancy cluster. The kitchen sees high activity during lunch time (12:00 to 15:00) on the high occupancy days and more scattered activity (maybe coffee breaks) on low occupancy days. The living room was mostly occupied in the start of the day on the high occupancy days. The store room and the other two rooms have a periodic occupancy associated with them. On the high occupancy days, it may be observed that the three rooms seem to have definite activities at mid-day (around 10 :00), meaning there is a mandatory activity associated with those rooms at that time.

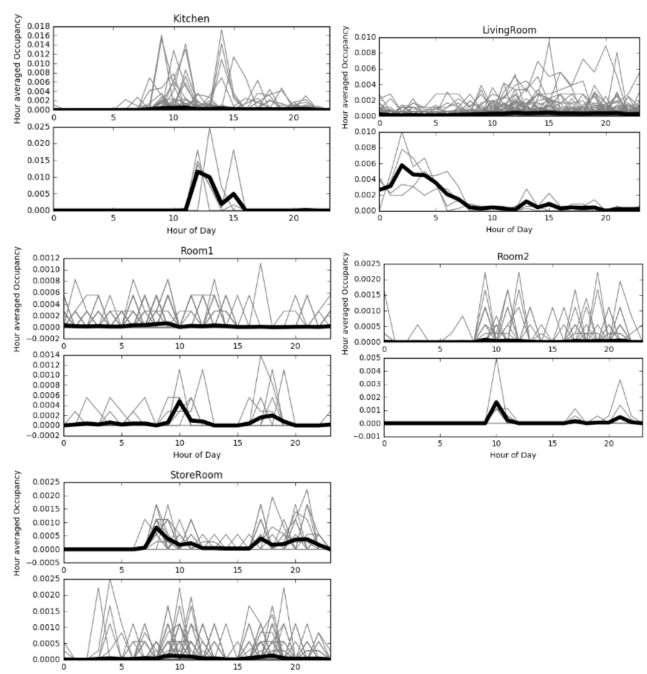


Figure 6. Occupancy Power Clustering of each Room

4.3 Occupancy Prediction Model

As mentioned in Section 3.3, we have run regression trees for occupancy prediction in each room. Table 1 shows the accuracies for occupancy prediction based on all the appliances in the house. Table 2 gives the most important features (or appliances) for this prediction in decreasing order of importance. Now after running regression trees for occupancy prediction based only on the power consumption of appliances present in each room, Table 3 has been populated. The underlined appliance is the most important one of the appliances in the list for that room.

Table 1 shows that our models have a high accuracy of prediction. Laptop is a very strong indicator of occupancy in the living room and similarly, the fan is a strong indicator of Room1 occupancy. Microwave is also a strong indicator of store room occupancy. For kitchen and room2, a combination of multiple appliances can give a good prediction of occupancy in that room.

Table 3 proves that knowing just the consumption of appliances in the room is sufficient and conclusive in predicting the occupancy of the room.

Table 1. Room level Occupancy Prediction Accuracies using all appliance consumption data

Room	Prediction Accuracy
Kitchen	99.98%
Living Room	99.97%
Store Room	99.95%
Room1	99.99%
Room2	99.99%

Table 2. Most Important Appliances for room level Occupancy Prediction

Room	Important Appliances
Kitchen	Washing Machine, Laptop
Living Room	Laptop
Store Room	Microwave, Laptop
Room1	Fan
Room2	Toaster, Laptop, Oven

Table 3. Room level Occupancy Prediction Accuracies using consumption data of appliances present in the room

Room	Prediction Accuracy	Appliances Used
Kitchen	99.98%	Heating, <u>Washing Machine</u>, Cooker, Blender
Living Room	99.97%	TV, <u>Laptop</u>, Sockets
Store Room	99.99%	Fridge, <u>Microwave</u>
Room1	99.99%	<u>Fan</u>
Room2	99.99%	Oven, <u>Toaster</u>

5. VALIDATION

We have done 10-Fold cross validation using the K-Fold package in scikit-learn. All the accuracies reported are scores after cross validation in Tables 1 and 3 of Section 4.3 as "Prediction Accuracy". The cross validation reveals high accuracy of the model predicting occupancy using appliance power data. The reason why the prediction accuracy is so high may be because there is a 50% chance of predicting the correct occupancy (1 or 0), which is a quite high probability of getting the correct prediction.

6. FUTURE WORK

We could take this work forward by considering the demand prices of electricity and predicted occupancy to automatically schedule the non-critical heavy loads during times of low demand to reduce the stress on the grid. We can pinpoint anomaly detection by studying the power consumption pattern of appliances.

7. REFERENCES

[1] "DRED Data Set." *DRED Data Set*. N.p., n.d. Web. 09 Dec. 2016.

[2] Akshay S.N. Uttama Nambi, Antonio Reyes Lua, and Venkatesha R. Prasad. 2015. LocED: Location-aware Energy Disaggregation Framework. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments* (BuildSys '15). ACM, New York, NY, USA, 45-54. DOI: http://dx.doi.org/10.1145/2821650.2821659