Exploration In Pattern of Domestic Appliance Electricity Consumption

Xingyu Wang
Carnegie Mellon University
Pittsburgh, Pennsylvania
xingyuw@andrew.cmu.edu

Jiqian Dong
Carnegie Mellon University
Pittsburgh, Pennsylvania
jiqiand@andrew.cmu.edu

ABSTRACT

In order to make full use of the big data and dig more valuable information collected by the emerging sensors, more advanced data analysis methodologies are needed. As data driven building energy management is one of the key issues in smart city and other sustainable proposals, we devoted to developing an electricity load pattern detection method to analysis the load consumption of individual domestic appliance. In this paper, we focus on a single home yearly data for individual appliance from UMass Trace Repository, which records the instantaneous power of each appliance with a time resolution of one hour and our goal is to find information of the users' pattern in that home. The data mining methods we using are K-means clustering and linear regression. As the results, we make it possible to segment the main load patterns and sub-patterns from a complex ensemble of all the appliances, as well as find out the linear relationship between temperature and usage for certain appliances.

CCS CONCEPTS

• **Computing methodologies → Machine learning**; **Learning linear models**;

KEYWORDS

User Pattern Recognition, Energy Efficiency, Clustering, Linear Regression

ACM Reference Format:

Xingyu Wang and Jiqian Dong. 2017. Exploration In Pattern of Domestic Appliance Electricity Consumption. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

As sensing technology rapidly develops, the deployment of advanced metering infrastructure (AMI) has increased greatly within the past few decades. These smart meters, advanced and powerful, has made concrete information about electricity consumption available for further research. Specifically, the load pattern for a single domestic appliance is now accessible. The load pattern and consumption information from smart meters has the potential to detecting the anomalies (finding the abnormal consumptions of appliance to avoid danger), performing as the priori information for Automatic Setup Nonintrusive Appliance Load Monitor (AS-NALM)[2], supporting commercial analysis or decision making as well as improving energy reduction recommendations.[3]

In this project, with the help of modern analysis approaches such as python and other machine learning algorithms, we are dedicating in developing a data driven method to infer the electricity consumption lifestyles of one single family. As the lifestyle can be reflected from the usage pattern of a set of appliances, we can infer the consumption habits by clustering the load features and find out the main patterns. Also, since consumption patterns for many appliances are closely related to temperature, i.e the usage of furnace may be inverse proportion to temperature, we manage to find out the relationship in between using the linear regression models.

Our main contributions can be summarized as following: 1. Segmentation: from the ensemble of all the hourly powers acquired from all the appliances, find out the main patterns and infer the family lifestyle. 2. Regression: using linear regression to further study the relationship between load pattern of each individual appliance and potential predictors such as temperature. 3. Comparison the baseline models: comparing the difference of results from linear model using two different temperatures: outdoor temperature and apparent temperature and find out whether using these different temperatures may have large effect on the final results.

2 METHOD

2.1 Dataset

The 2 tables for individual domestic appliances load consumption and temperature information we used come from the Home C datasets for 2014. Its origin is UMass Trace Repository maintained by the Laboratory for Advanced System Software. The data set for domestic appliances contains the instantaneous power values for each appliance with the time resolution of 1 hour and a total sampling time for 1 year. Even the hourly interval can be too sparse to monitor the load pattern some certain appliances that will frequently change the working states, the entire 1-year data can compensate this drawback.

2.2 Load Normalization

To obtain the daily pattern of each appliance, we group the power values by days and hours at the very beginning, and then we obtain daily load curve containing 24 values in each day. Second, since the absolute power value for different appliances are not in the same scale, which will further impact the clustering results, we

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

normalized the daily load curve. Giving a daily consumption profile $l(t)$, we decomposed the value as $l(t)$.

$$a = \sum_{t=1}^{24} l(t) \quad \text{and} \quad s(t) = \frac{l(t)}{a} \tag{1}$$

where a is the total daily demand and $s(t)$ is the normalized load pattern, which we denominate *loadshape*.

2.3 K-Means

The corresponding data mining approaches we applied are K-means for appliances clustering and linear regression for load predicting. As linear regression is the easiest and most interpretable model among statistic learning models, we use it to predict the instantaneous power.

We applied an unsupervised learning method called K-means clustering.[4] The goal of this algorithm is to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Given a set of observations $(x_1, x_2, \dots, x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into $k(\leqslant n)$ sets $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within cluster sum of squares (WCSS) (i.e. variance). Formally, the objective is to find:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 = \arg\min_{S} \sum_{i=1}^{k} |S_i| \, Var[S_i] \tag{2}$$

where μ_i is the mean of points in S_i.

When applying K-means algorithm to the load shapes, we can group the consumption different appliances by common features into several clusters. Each cluster may represent a single pattern of usage, and we denominate it as load pattern. In our experiments, we firstly mix all the daily load shapes for each individual appliance then apply the K-means algorithm to do the segmentation. The centroid of one cluster may represent one usage pattern of appliances, i.e. the constant running, the constant shutting, periodic running, or frequently running due to the users' lifestyles. These several main patterns can be illustrated by the cluster centroids.

2.4 Linear regression

In statistics, linear regression is a linear approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data.[5] The goal of this work in our project is firstly to figure out whether there exists linear relationship between individual appliance and temperature, and then test out whether adapting different kind of temperature data may influence the results of prediction.[1] We investigate the effect by implementing the linear regression in two models using different temperatures data: *a.* outdoor air temperature data; *b.* apparent temperature and then comparing the results. We define the first model as the baseline model the second model as the variant model.

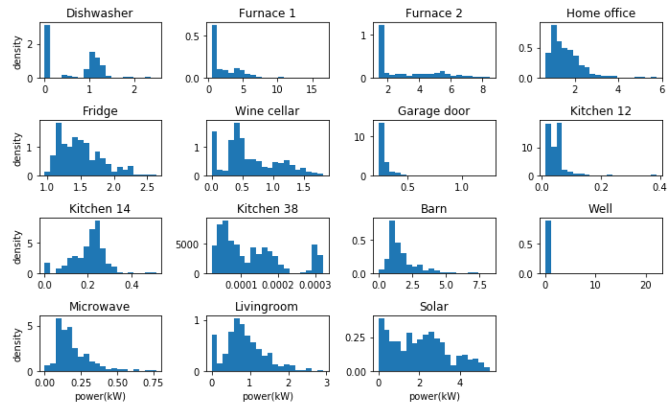

Figure 1: Normalized Histogram Of Each Appliance's Power

3 EXPERIMENTS ON DATA

3.1 Data primitive exploration

After comparing the power values of all the appliances in that home, we selected 15 typical appliances that can be used in most families and have the largest energy consumption. We treat them as our target appliances to further explore the information of load patterns. The target appliances in that house are: Dishwasher, Furnace 1, Furnace 2, Home office, Fridge, Wine cellar, Garage door, Kitchen 38, Barn, Well, Microwave, Kitchen 12, Kitchen 14, solar, living room.

In order irst the unnormalized data for these 15 appliances can be visualized by the box plotting in Fig.2 where each box depicts the daily demands of each appliance. We can find that there exist some certain prototypes for different appliances: some appliances consume electricity at a high level while some mainly remains steady. Also the Fig.1 shows the power density for each appliance, we can see the variation of power value is obvious. Therefore, as mentioned before, in order to guarantee the scales of data in clustering are the same, we need to normalize the load curve for each day.

3.2 Load patterns exploration

In order to dig out the main pattern for each appliance and find out the common features from the different appliances, we decided to implement K-means algorithm to cluster the load shapes. We use two ways of clustering to determine the main usage pattern, the first one is clustering using the average load shape for a single appliance, the second is using voting mechanism.

In first approach, to begin with, we take the average of all the load shapes for all the 15 appliances in the year and obtain 15 points, each point represents one annual average daily load shape for a single appliance. We label these points as the main load shapes for 15 appliances. Then we do the K-means clustering in this 15 points to test out how many usage patterns can be find in this points set. In experimenting, after comparing the results of different cluster number, we found that in this model the 15 load shapes can be nicely classified into 4 categories. The outcoming labels and cluster centroids of clastering are shown in Table.1 and Fig.3, respectively.

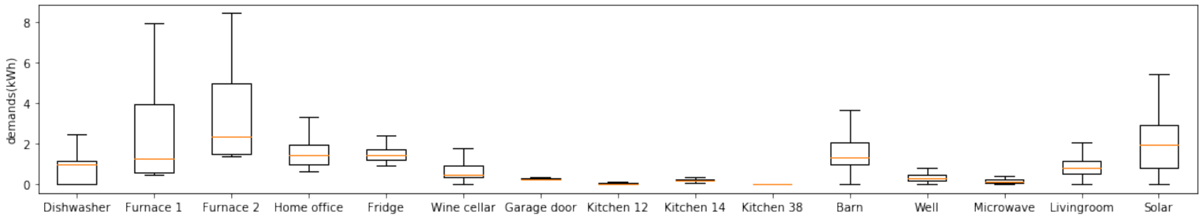

Figure 2: Boxplot Of Each Appliance's Daily Power

Table 1: Clustering By Average Load Curve

group#0:	Dishwasher, Furnace 1, Furnace 2, Home office, Fridge, Wine cellar, Garage door, Kitchen 38, Barn, Well, Microwave
group#1:	Kitchen 14, Living room
group#2:	Kitchen 12
group#3:	Solar

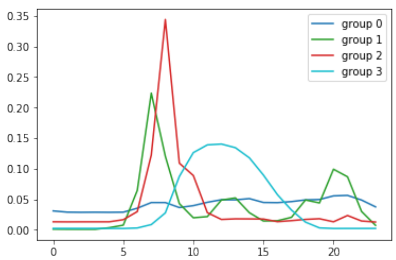

Figure 3: Clustering By Average Load Curve

Combining the Table.1 together with Fig.3, we can clearly find there are mainly 4 groups of load shapes. Each group may represent one specific load pattern that can be easily interpreted. For example, the table shows that there are 11 appliances are clustered into the group #0, and by watching the corresponding load curve of group #0, we can conclude the pattern of group #0 contains most of the appliances which can be labeled as constant shutting appliances.

Since this model only focus on the average load shape of the year, there are only 15 points taken into clustering, a lot of information such as the subpatterns can be neglected, an alternative way is addressed. In this case, instead of treating the average as the main load shape of the year, we implement K-means directly on the total 5490 (365×15) daily load shapes. By clustering the loads shape into 5 categories we can take several subpatterns into account. In order to find out the main usage pattern for each appliance, we use the voting mechanism to obtain the most probable load shape. In the voting mechanism, for each appliance, extract the group with the largest number to represent the main pattern, the rest load shapes in other groups can be defined as the sub-patterns, accordingly. The principle of voting mechanism can be explained as following: for one specific appliance, its 365 load shapes will be clustered into 5 different categories, each category will contain different number of load shapes. The group with the largest number means most of the load curves can be categorized into this pattern, thus the centroid

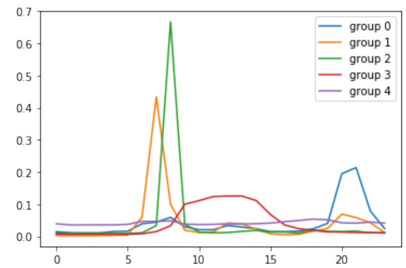

Figure 4: Clustering By Daily Load Curve Voting

of this group may represent the main pattern of this appliance. This clustering method, compared to the previous one, contains more data points and can reflect the main pattern more rationally.

The results of K-means clustering for the total load shapes are shown in Table.2 and Fig.4. Since each application. It has almost the same conclusion as the results from the previous method, the slight difference lies on the cluster for appliance Kitchen 12. In the second method, we find the most probable pattern for Kitchen 12 is in group#0, which is a constant shutting group. However, in the first method, when taking the average of all the load shapes from Kitchen 12 , it becomes no longer a constant shutting appliance. The possible reason of it is perhaps the Kitchen 12 is a high power appliance and will bring up the average load shape even if it is rarely used.

Table 2: Clustering By Daily Load Curve Voting

group#0:	*empty (sub-pattern)*
group#1:	Kitchen 14, Well, Living room
group#2:	*empty (sub-pattern)*
group#3:	Solar
group#0:	Dishwasher, Furnace 1, Furnace 2, Home office, Fridge, Wine cellar, Garage door, Kitchen 12, Kitchen 38, Barn, Microwave

We can further study the correlation between each group by using the correlation matrix, which is shown in Fig.5. If there are clear positive correlation between 2 groups, the correlation values should be positive, vice versa. The result illustrates there is a positive correlation between group #0 and group #1. That can be explained as: If the main pattern of one appliance is clustered into the group #0, it may have the sub-pattern in group #1. However, if the main pattern is clustered into group #1, it is not likely to

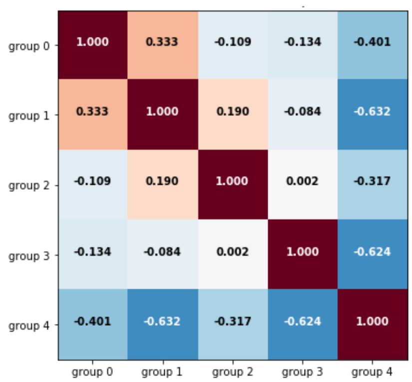

Figure 5: Correlation Between Different Patterns

have the sub-pattern in group #4, because the correlation value is negative.

3.3 Temperature effects reseach

In this part, 2 linear regression models are used respectively to explore the relationship between demands and outdoor physical temperature or apparent temperature. As the results, the temperature do have linear relationship to a certain of appliances. The normalized coefficients are plotted in Fig.6.

From the Fig.6, it is not difficult to conclude that even the absolute values of the coefficients are comparatively small, there exists certain linear relationship between power and temperature value. The relationship is intuitively rational and also can be interpreted. For example: the Wine cellar has the positive relationship to the temperature due to the need of keeping the wire in cold storage. The Furnaces have negative correlation to temperature because there is no need of warming when temperature is high.

At the same time, the effect of two different kind of temperatures on this linear regression method can be seen from Fig.6. The 2 models have the approximate coefficients, thus we can infer the two temperatures may perform similarly in this prediction. To further illustrate this facts, the predicted power density from variant model and baseline model are plotted in Fig.7. As mentioned before, the x axis represents the predicted values from the baseline model (linear regression on outdoor air temperature) and y axis represents the corresponding values from variant model (regression on apparent temperature). As the data points mainly lie on the diagonal of the figure, the performances of two models are similar.

4 CONCLUSION

Base on the results of K-means clustering, we find out the a good way of finding the usage pattern for a single family is to do cluster work on daily load shapes of individual domestic appliances. The main pattern can be extracted using voting mechanism and the correlation between clusters can illustrate the relationship between main pattern and sub-patterns. Second, from the results of linear regression, even the coefficients are rather small, we can still draw the conclusion that both outdoor temperature and apparent temperature can be used to predict the consumption of certain appliances. The effect of applying different temperatures is not apparent by comparing the predicted values.

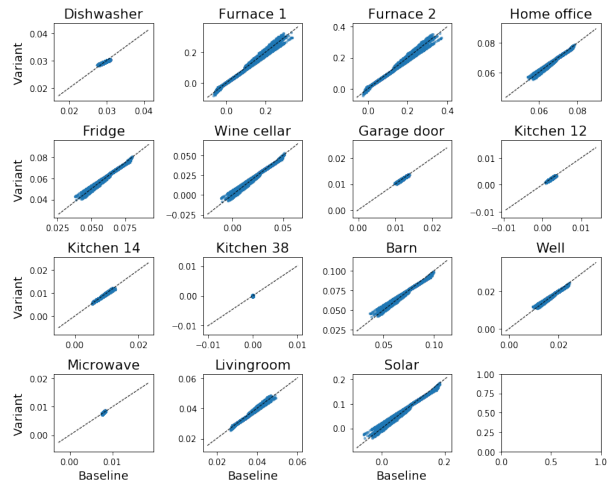

Figure 7: Consistency Examination

REFERENCES

[1] Nathan J. Addy, Sila Kiliccote, Duncan S. Callaway, and Johanna L. Mathieu. 2014. How Baseline Model Implementation Choices Affect Demand Response Assessments. *Journal of Solar Energy Engineering* 137, 2 (2014), 021008.
[2] George W Hart. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* 80, 12 (1992), 1870–1891.
[3] Jungsuk Kwac, June Flora, and Ram Rajagopal. 2014. Household Energy Consumption Segmentation Using Hourly Data. *IEEE Transactions on Smart Grid* 5, 1 (2014), 420–430.
[4] Wikipedia. [n. d.]. K-Means Clustering. ([n. d.]). Retrieved Dec 16, 2017 from https://en.wikipedia.org/wiki/K-means_clustering
[5] Wikipedia. [n. d.]. Liner Regression. ([n. d.]). Retrieved Dec 16, 2017 from https://en.wikipedia.org/wiki/Linear_regression

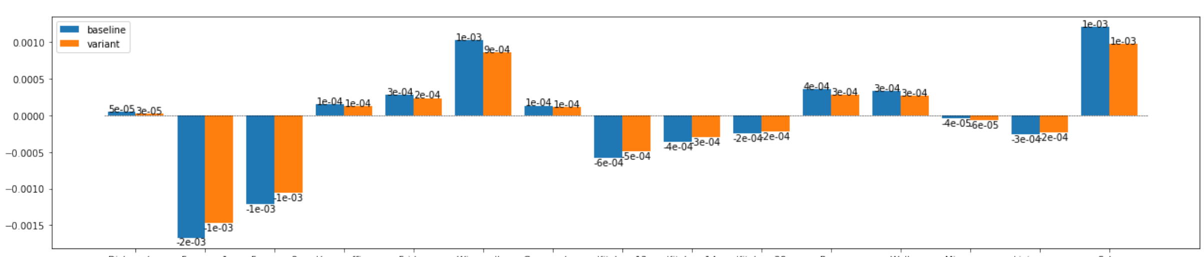

Figure 6: Normalized Coeffcient Of Each Appliance