Electricity, gas, and water load profile analysis of a Canadian household: A case study

Yue Liu
Student ID: 32043465
FIT5128
Supervisor: Hao Wang
Word Count: 4648

Contents

Electricity, gas, and water load profile analysis of a Canadian household: A case study

Yue Liu
Faculty of Information Technology
Monash University
Melbourne, Victoria, Australia

ABSTRACT

With the development of smart meter, energy load data can be collected with finer resolution and transmitted in real time. Such improvement encourages many research directions to gain insights from the data. However, many existing studies only focus on analyzing one type of energy/utility, which might neglect any potential relationship between the energies. In this study, a Canadian dataset is retrieved consisting electricity, gas, and water load data recorded every minute. K-Means clustering algorithm is employed to discover the load patterns of the energy load data. The results are analyzed with seasonal variation to form narration about the household's behavior pattern. In addition, the relationship between the clusters is explored, which indicates a sign of association between certain water and gas load patterns.

1 INTRODUCTION

Smart meter is widely applied across the world, replacing conventional meter. There are over 97 million residential smart meters installed in the US by the end of 2021 [5]. The emergence of smart meter allows the consumption data to be collected automatically with finer resolution (e.g., every minute or every 15 minutes) and transmitted in real time. With the increasing load data size, the rich information hidden within the load data can be extracted to provide insights on several aspects. The major research directions supported by the load data are load analysis, load forecasting, and load management [18]. For load analysis, an example study is to analyze the daily load profiles and the similarities between them, as well as seasonal and irregular changes in the time series load data [4]. On the other hand, load forecasting may apply algorithms such as long short-term memory (LSTM) recurrent network to train the model and predict future energy consumption [8]. As for load management, the focus could be performing load scheduling optimization which allows the total cost of residential power consumption to be minimized [14]. In addition, it is equally essential to have metrics to evaluate the performance of the smart meter data analytics [11]. Such applications of the load data all contribute to improving the efficiency and sustainability of both the energy demand side and supply side. In detail, the research insights can inspire governments to construct better energy related policies and help energy suppliers to understand their customers better.

1.1 Motivation

The fundamental of the applications of the load data requires the researchers to have a thorough understanding of the data. It is rather common to have studies dedicated to electricity load profiling with diverse algorithms [18]. However, this limits the researchers to form narrative only based on electricity load patterns. Similarly, there are existing studies only focused on natural gas or water consumption. This is, the field of research is lack of analysis of all three energies/utilities within the same study. Analyzing the three load patterns at the same time could be beneficial to explore any potential relationships between any two combinations of the data. Furthermore, the result of the clustering may enhance the understanding of the consumer's behavior pattern.

The remaining of this thesis is organized as follows. Section 2 will illustrate the background and related work, with the problem formulation at the end. Section 3 focuses on the methodology of this thesis, including data description, data pre-processing, details of the selected clustering method. Section 4 is the analysis of the clustering result with visualization providing a thorough demonstration of the load data. Section 5 concludes this thesis with limitations and future work directions.

2 BACKGROUND AND RELATED WORK

This section introduces the background and the related work of the thesis. The problem formulation is included at the end of this section.

Electricity load profile analysis is relatively common in the field of study. A multi-layered clustering is proposed in [2] for electricity load profiling. In the first layer, K-Means clustering is applied to make sense of data locality, where clusters with small size are recognized as outlier/abnormal power consumption behavior. Later, the more representative data is processed to find the global electricity load profile. This multi-layer approach is beneficial when the dataset is large since the first layer reduces the computational cost for the second layer. Similarly, another study also proposed two-layer clustering [19]. The first layer employs a hierarchical clustering algorithm, and the second layer employs the fuzzy C-Means clustering, which produces 4 final clusters as electricity load patterns. In the study of [17], the authors compared K-Means clustering with Ward's agglomerative hierarchical clustering and found that Ward's method does not scale well to large datasets and is more likely to produce more even-sized clusters. From the K-Means clustering, 4 electricity clusters are obtained. During the analysis of the 4 clusters, it is discovered that electricity load pattern is affected by seasonal variation. Specifically, the electricity demand in winter is higher than that in summer. It is explained in the study that such observations may be related to the fact that households tend to spend more time at home during winter.

Gas load profiling is often achieved by employing clustering algorithms. For a study conducted with data from Algeria, K-Means clustering algorithm is selected to discover gas load patterns [9]. The elbow method is introduced to find the optimal number of gas clusters and returns 5 being the value for K. In the load profile analysis, it is noted that the gas clusters are influenced by seasonal variation since there is only one dominant cluster during winter. In [13], the gas load profiling is conducted on a city and town level with a Greek dataset. The study employs a two-stage clustering where K-Medoids clustering is compared with hierarchical agglomerative clustering. It is concluded in the study that hierarchical agglomerative clustering is recognized as the suitable clustering solution for the dataset.

The existing studies on water load profiling are more focused on domestic hot water (DHW) instead of water in general. Nonetheless, there are findings that can be related to general water load profiling. It is observed in a study based in Swiss that the DHW consumption shows no significant decrease during weekend, and the correlation with outdoor temperature is not strong [3]. On the contrary, [15] discovered that the consumption of shower water is under the influence of seasonal variation, which results in longer duration shower when the temperature is lower. Addressed in [6], DHW profiles are mixtures of random and patterned draw events. Generally, there is a peak in the morning and a peak in the evening for residential buildings, which indicates the specific timing for some daily activities. The findings in [1] coincide with the above findings. The authors found that during weekdays, there is a morning peak between 7:00 and 9:00, and an evening peak between 20:00 and 22:00. For weekends, the morning peak tends to shift 2 to 3 hours later compared to weekdays. For the mentioned studies that are associated with DTW, the analysis is barely conducted with clustering algorithms. It is common to group the time series data by some conditions and then compute the average to produce representative load profiles. For instance, the data was grouped by the number of people in the household and day of week (weekday, weekend, total day) before load profiling analysis [1].

It is not common to have studies focusing on load profiling of more than one utility. A study based in Singapore explores the load profiling of both natural and electricity [7]. The findings imply that the load profile analysis should be associated with the local geographical conditions. For instance, it is mentioned that Singapore is a tropical country, therefore, the households have high electricity consumption due to air conditioning and show less variation in the load patterns when season changes. In the same study, a shift between electricity and gas load patterns in the evening during weekdays was discovered. For non-residential buildings, the load profiling of electricity and heat (heat pump) consumption discovered that the load profiles are dependent on outdoor temperature, time of day, and type of day (weekday, weekend, etc.) [10].

From the above studies, it is addressed how the analysis of each utility load pattern is conducted with depth, which provides many important insights on load pattern analysis. Meanwhile, it is much less common to find studies that analyze multiple utility load patterns of residential buildings. The potential reason for such observation might be the difficulty retrieving all three utility load data from the same households with high data recording frequency. Thus, this thesis will aim to explore the load profiles of all three utility of the same household and to identify any potential relationships between the three utility load profiles.

3 METHODOLOGY

This section will discuss data pre-processing and the selected clustering algorithm for load profiling – K-means clustering algorithm.

3.1 Data Description

The load data is obtained from the Harvard Dataverse [16]. It records the utility load of a house on minutely level from April 2012 to March 2014. After pre-processing, the load data of electricity, water, and gas are concentrated every 15 minutes, with 729 days on record. It is mentioned in [12] that the data was collected from a house in Vancouver, Canada. The house has two families. The first family consists of a male who is a full-time university student, a female who is self-employed, and a child who is between the age of 5 and 6 and a full-time elementary student. The second family consists only a male who is full-time employed.

3.2 Data Pre-processing

As the major task of this thesis being dependent on discovering the daily load patterns, the obtained load data is first grouped by data collection dates. It is mentioned in the above section that the data was collected on a minute basis, which generates an enormous amount of data entries per day for all three utilities. This is to say, the computational cost of performing any clustering algorithm on the raw temporal data is very high. In addition, due to the volatile nature of the load data, performing clustering algorithm on the raw data might result in having many excessive clusters, which makes it hard to interpret generalized profiles. Therefore, the load data of the three utilities is aggregated every 15 minutes for better clustering outcome. Since the thesis focuses on three different types of utilities, which are of completely different scales, it is essential to transform the data into the same scale for analysis. By normalizing the load data, the shape and the trend of the load pattern are retained, with the value being in the range [0, 1]. The normalization process is defined as follows. Firstly, it is essential to obtain the range of the selected dataset. This is achieved by computing the difference between the maximum and the minimum of the dataset. Then, each entry subtracts by the minimum, and then divided by the range of the dataset. After pre-processing, the load data of electricity, water, and gas are normalized and aggregated every 15 minutes, which means 96 records per day, with 729 days on record.

3.3 K-Means Clustering

K-Means clustering is one of the most popular unsupervised machine learning algorithms. It is designed to partition the dataset into K distinct clusters where every data point is part of one and only one cluster. The partitioning process is achieved by minimizing the within-cluster sum of squares (WCSS). In terms of metric of the K-Means clustering, there exist options such as Euclidean distance and dynamic time warping (DTW). Since time shifts in the load patterns are of interest of this thesis, Euclidean distance is selected to be the metric of the algorithm.

K-Means clustering is considered to be convenient to implement in practice. However, this algorithm is sensitive to the initialization of centroids, which means that different initializations might result in different clusters. There is also another disadvantage of K-Means clustering, which is the K value must be manually determined before the partitioning process. Thus, to find the optimal K for the given dataset, two approaches are chosen – the elbow method and the Silhouette score. The elbow method is an iterative method that runs K-Means clustering on the dataset with selected K values. Then it is possible to plot the selected K values against the corresponding WCSS. The optimal K value should be the one that is at the elbow position of the curve. As for the Silhouette score, it is a metric to evaluate the clustering result, ranging from -1 to 1. Additionally, -1 stands for the clusters being poorly matched, while 1 stands for the opposite.

The mentioned approaches are applied to electricity, gas, and water load data separately, to determine the most suitable K values. Unfortunately, for all three utilities, it is difficult to distinguish the elbow from the curves. As for the result of the Silhouette score, it is presented that all K values (ranging from 3 to 14 with an increasing step of 1) have similar scores between 0.02 and 0.10, regardless of the utility type. Therefore, the

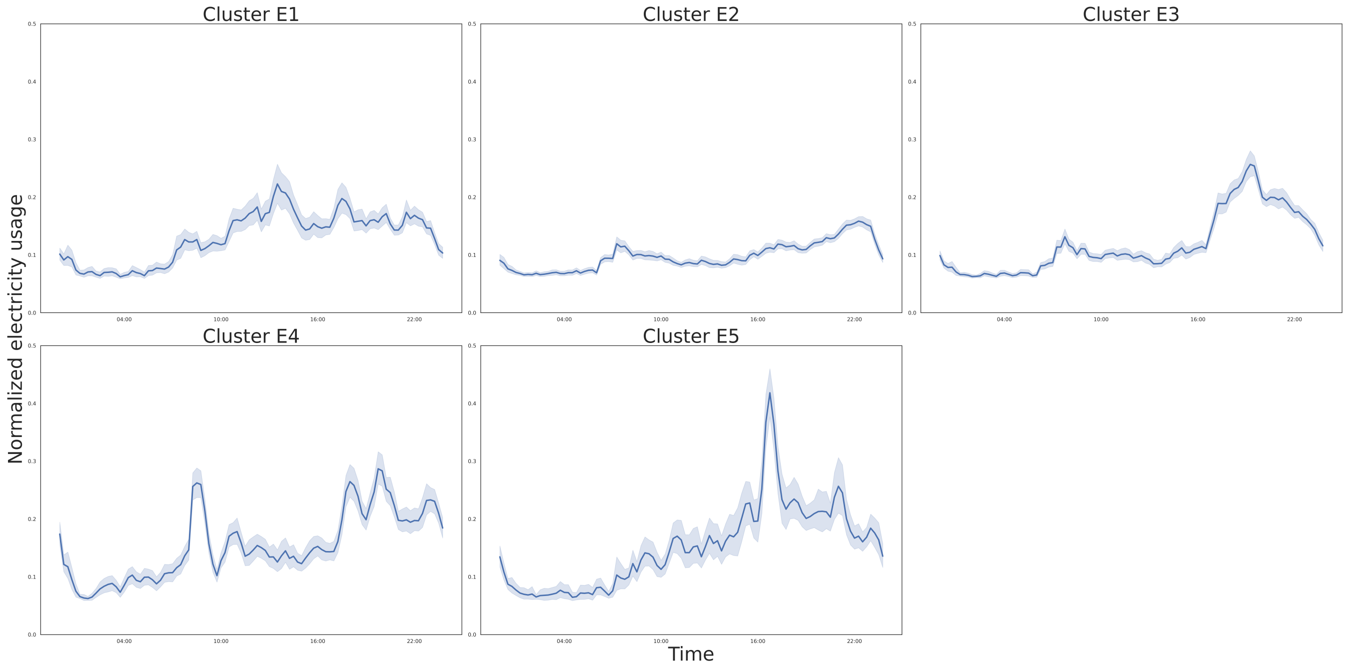

Figure 1: Five Electricity Clusters from K-Means.

final numbers of clusters are determined via analyzing the cluster size as well as the visualization of the clustering. More information is provided in the following section.

4 RESULTS

This section aims to discuss the results of K-Means clustering of the load patterns of electricity, gas, and water. To make the comparison more convenient to analyze, the data used for plotting is normalized. Additionally, the clusters will be discussed with the consideration of the seasonal changes. Finally, the occurrence of the combinations of the three utility load patterns will be explored to discover any potential relations.

4.1 Load Pattern Clustering

K-Means clustering is applied to the three types of utility load data (electricity, gas, and water). The Silhouette coefficient calculation is employed to determine the number of clusters of the three types of utility load data separately. However, the result of the calculation does not provide a clear indication. Therefore, the final numbers of clusters are chosen based on the visualization of the clustering and the distribution of the clustering. The details of each of the utility types will be presented in the following paragraphs.

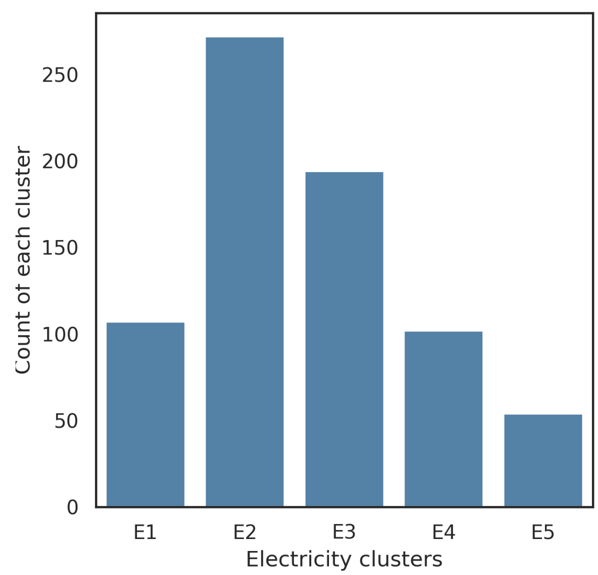

Figure 2: Electricity Cluster Size.

For electricity load data, the final number of the clusters is five. Figure 1 demonstrates the result of the clustering. In this plot, there are five subplots, each representing an electricity cluster, namely E1 to E5. In each subplot, the shaded area is the range of all normalized electricity load data at corresponding time point. The line is generated by computing the mean of the normalized electricity load data at each time point, which provides an overview of the shape of the cluster. In addition, figure 2 illustrates the count of each cluster, where E2 being the cluster with the largest size. As can be observed from figure 1, all electricity load patterns are distinct. E2 and E3 have similar peak hours in the morning, which might imply that the residents of the house tend to start their day around 7:00 and then some of them might leave the house for work or study. Later in the day, they return to the house at around 15:00 and end the day between 21:00 and 22:00. However, it is worth noting that the peak from E3 is much higher than that from E2. Since the peak of E3 is around 19:00, it is possible that this is due to the usage of kitchen appliances such as electric oven. On the other hand, E1 displays a relatively high electricity consumption during daytime which might suggest that the residents are present in the house during the day. In E4, the pattern suggests that the resident(s) are active throughout the day. They conduct in-house activities that consume a large amount of electricity, which locates the spike around 9:00. Cluster E5 implies that the residents start the day around 8:00 and stay in the house for indoor activities. The load pattern reaches its peak at around 15:00, which is also the peak out of all the clusters. Referring to figure 2, E5 is a relatively rare pattern to be observed.

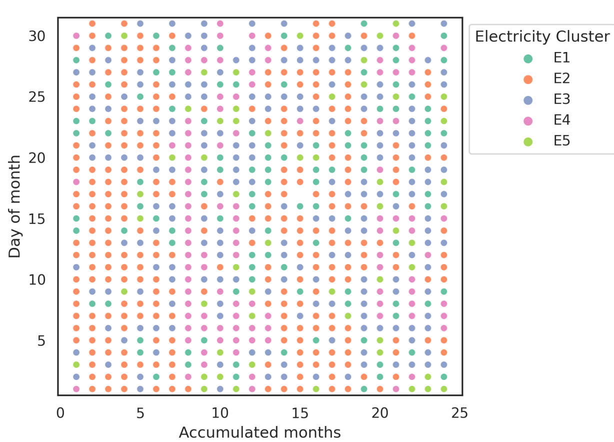

Figure 3: Distribution of the Five Electricity Clusters over the 24 Months Span.

Figure 3 shows how the five electricity clusters are distributed throughout the 24-month period that starts in April. The x-axis represents the accumulated months of the data collection period, while the y-axis represents the day of month. Overall, there is not a single cluster that is dominant throughout the entire period. From figure 3, it can be observed that in the first six months (Apr to Oct) the frequency of having E2 is relatively high compared to the other clusters, which suggests that E2 is the major load pattern for spring and summer. Therefore, it is likely that the highest peak in E3 is due to the cooling or heating system in the house when the weather is warmer or colder after residents return home. Alternatively, the peak could also be associated with cooking activity, where high electricity-consuming kitchen appliances are used. Moving along to the months with lower outdoor temperatures, E4 is the only dominant load pattern in November and the frequency of E4 remains higher than other months until February. With E4 having the second highest peak among all the clusters, this observation indicates that the peak may be related to the heating system in the house. As for E5, despite the low occurrence count, this pattern still appears almost every month throughout all seasons. Thus, E5 implies certain regulated behaviors that the residents perform on monthly basis. Such behavior could be house cleaning that requires electric appliances such as laundry dryer or vacuum machine.

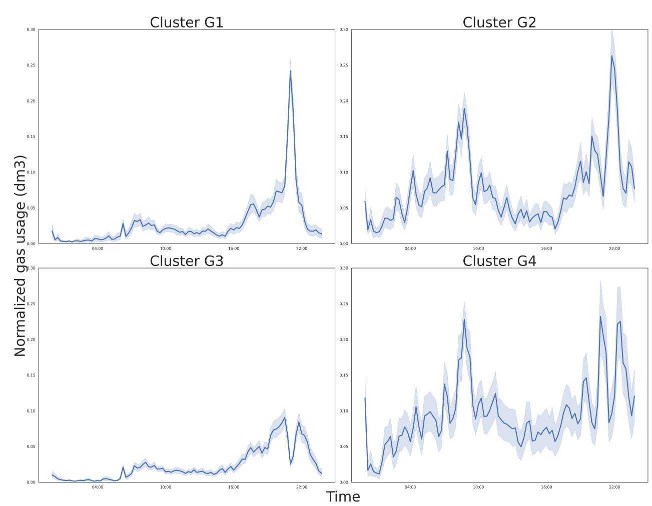

Figure 4: Four Gas Clusters from K-Means.

For gas load data, the final number of the clusters is four. Figure 4 demonstrates the result of the clustering.

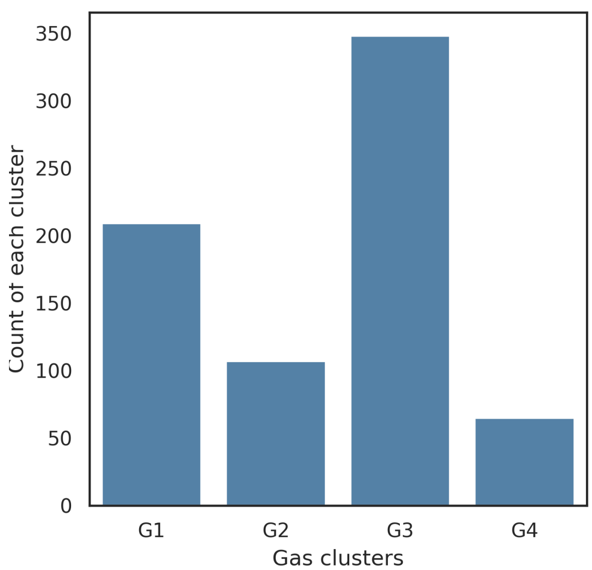

Figure 5: Gas Cluster Size.

In this plot, there are four subplots, each representing a gas cluster, namely G1 to G4. In each subplot, the shaded area is the range of all normalized gas load data at corresponding time point. The line is generated by computing the mean of the normalized gas load data at each time point, which provides an overview of the shape of the cluster. In addition, figure 5 illustrates the count of each cluster, where G3 being the cluster with the largest size. The final four clusters share similarities to a certain degree. They can be roughly grouped into two classes, where G1 and G3 being the more stable class and G2 and G4 being the less stable class. Looking at both G1 and G3, the first spike is around 7:00, as it marks the start of a day where the residents might be using hot water or gas cooking stove. Later, the gas consumption stays relatively stable and low for both clusters before 16:00, as some residents might be away from the house. After 16:00, the gas consumption starts to rise rapidly, which might indicate the return of the residents. Between 18:00 and 20:00, the rise in both G1 and G3 implies the usage from the gas stove for preparing dinner. For G2 and G4, the load pattern throughout the day is much less stable compared to G1 and G3, with the trend spiking and dropping constantly. This is a sign that the spiking might have a connection with the heating system in the house. Both clusters have secondary peaks at around 9:00 and major peak(s) around 22:00. However, it is unclear regarding the peak around 9:00. In comparison, there are two major peaks in G3 and G4 at around 21:00, but only one major peak in G1 and G2. Considering there are two families in the house, it is possible that the first peak in G3 and G4 is related to the family with a child, since the family might have an earlier sleep schedule. Meanwhile, the second peak might be related to the single-member family. Since the water heater in the house runs on gas, it is reasonable to link the peak of all gas load patterns around 21:00 with hot water usage.

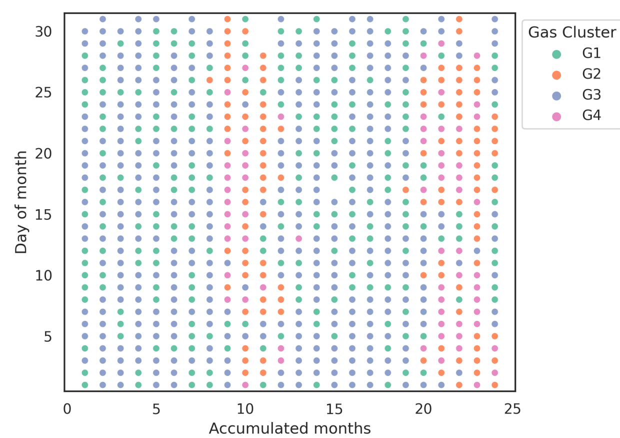

Figure 6: Distribution of the Four Gas Clusters over the 24 Months Span.

Figure 6 shows how the four gas load patterns are distributed throughout the 24-month period. It could be observed that G1 and G3 are rather scattered across the entire period while G2 and G4 are concentrated at specific months. There is a clear transformation from the combination of G1 and G3 to the combination of all the clusters shown in the plot. Combined with the analysis from above, it is highly possible that G2 and G4 are directly influenced by the house heating system during winter, which explains the spiking and dropping throughout the day. Once the outdoor temperature increases, the occurrence of G2 and G4 decreases drastically. This is to say, G1 and G3 are more dominant during spring, summer, and autumn.

For water load data, the final number of the clusters is three. Figure 7 demonstrates the result of the clustering. In this plot, there are three subplots, each representing a water cluster, namely W1 to W3. In each subplot, the shaded area is the range of all normalized water load

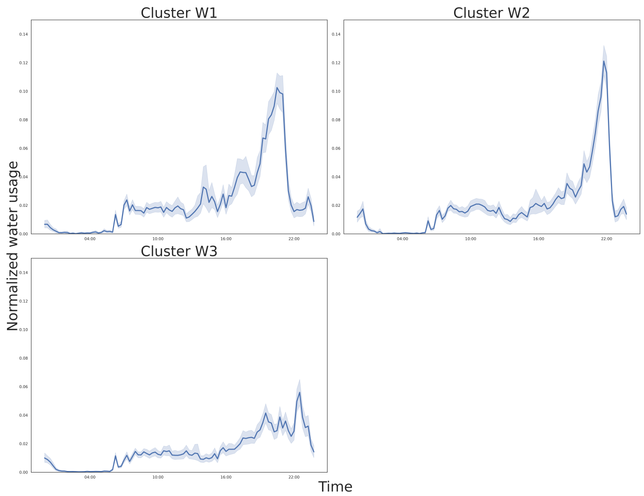

Figure 7: Three Water clusters from K-Means.

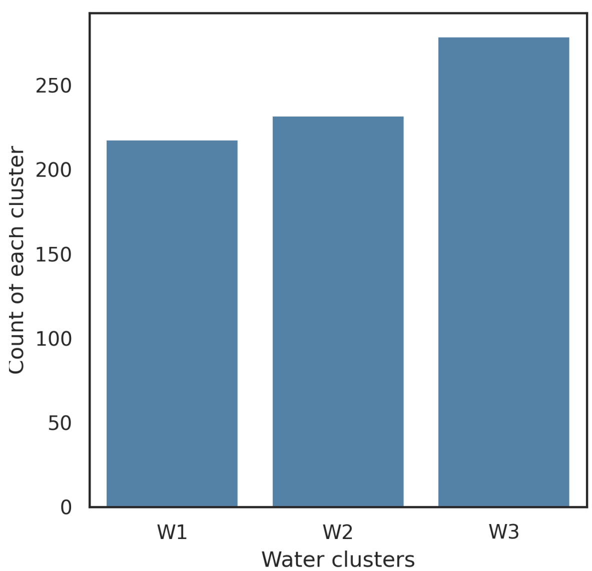

Figure 8: Water Cluster Size.

data at corresponding time point. The line is generated by computing the mean of the normalized water load data at each time point, which provides an overview of the shape of the cluster. In addition, figure 8 illustrates the count of each cluster, where W3 being the cluster with the largest size. In W1, the first spike of the load pattern appears around 6:00, which indicates that at least one resident is awake and starts the day. From 6:00 onward, the plot illustrates that some resident(s) stay in the house during the day. Between 16:00 and 18:00, there is a slow rise in the pattern that could be due to residents preparing their dinner. Starting from 18:00, the load pattern climbs and reaches its peak at roughly 20:00, then drops sharply. This period might be the time when all residents start to use water in the bathroom and get ready for rest. After 22:00, there is a small, which is likely to be related to the use of bathroom as well. Compared to W1, W2 displays a relatively similar pattern in the daytime but with a sharper peak before 22:00. Unlike W1, there are two small spikes after 22:00. This observation implies that the residents in the house have a regulated pattern of waking up before and after midnight and the water usage is likely to be related to the bathroom. W3 is more distinct from W1 and W2. The major difference between them is the load pattern between 18:00 and 22:00. As shown in W3, there are several sharp peaks in that period. In addition, the water usage during peak hours at night is much lower than that from W1 and W2. Unlike W1 and W2, it is interesting to notice that the highest peak of W3 occurs after 22:00.

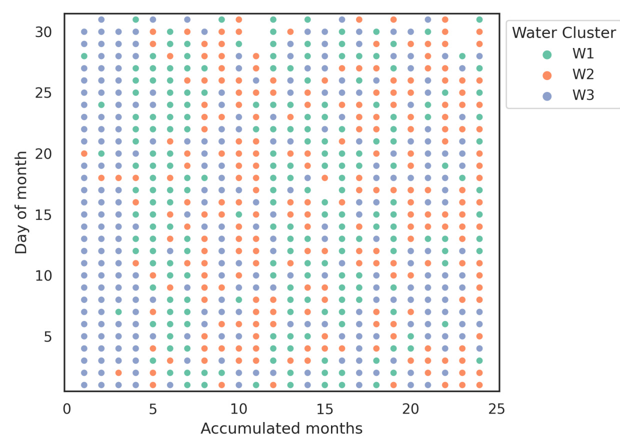

Figure 9: Distribution of the Three Water Clusters over the 24 Months Span.

Figure 9 shows how the three water load patterns are distributed throughout the 24-month period. It is interesting to discover that W3 is dominant in the first three months of the data collection period. Next, W1 begins to have an increasing frequency in the following months until the end of October. Then the trend transforms to a roughly equally distributed W2 and W3 with only few W3 during the winter months. To sum up, W1 tends to be more observable in any other season but winter. However, the occurrence pattern of W2 remains undiscovered.

4.2 Potential Relationship among the Clusters

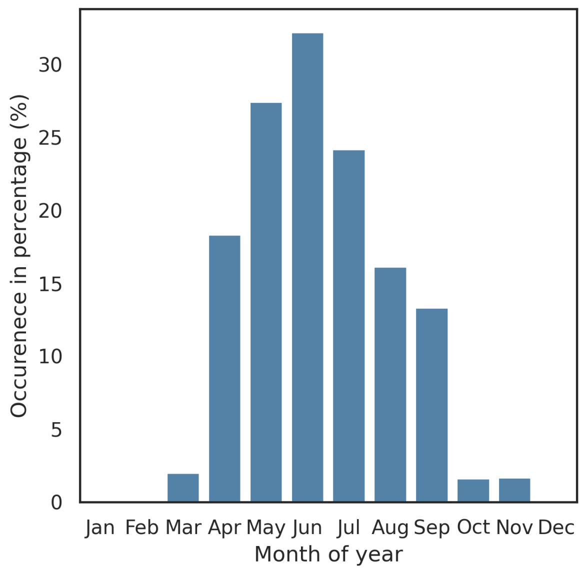

Figure 10: Coverage of E2W3G3 in percentage.

To further discover the potential relationship among the three utility load patterns, figure 12 aims to visualize the distribution of the occurrence of all possible combinations of the clusters. As can be observed, the occurrences of most of the combinations lie around 10. There are also ten combinations that have the count around 30. The combination with the largest size is E2W3G3, with the count being 84. The occurrence of E2W3G3 takes over 11.5% of the total count. Figure 10 demonstrates the coverage of E2W3G3 in percentage across the year.

As shown in this figure, this combination is frequent during spring and summer. Specifically, the occurrence in June takes up to 32%. The combinations with the lowest count are E2W1G2, E2W1G3, and E4W1G4, with the counts being 0. Analyzing the result in a utility-wise manner, it can be noted that G2 and G4 have relatively low occurrences in the top-count combinations in figure 12, while W1, W2, and W3 are more evenly distributed. For electricity load patterns, E2, E3 and E4 are all visible in the top-count combinations except for E1.

To analyze any potential relationship, the top three cluster combinations (E2W3G3, E2W1G1, and E2W1G3) are selected for plotting, which can be found in figure 11. Throughout the visualization of the selected combinations, it can be observed that the trend of load pattern of gas coincides with that of water most of the time, regardless of the clusters. Such observations have high potential to be related to the fact that the hot water system in the house runs on gas. Thus, when the trend of load pattern of water displays a sign of peaking and that of gas remains flat or decreasing, it is reasonable to assume that the residents use cold water instead of hot water at the time. Furthermore, when the trends of both load patter do not coincide, it is likely that the gas is used for not only hot water, but also gas stove and house heating. For instance, it could be difficult to discover any patterns between gas load and water load when the gas load cluster is G2 or G4, as the two load patterns are particularly influenced by the house heating system. Gas and electricity load patterns are more likely to share similar signs of spiking and dropping between 16:00 and 20:00. However, there are a few minor noticeable delays with the signs. The extent of the spiking and dropping is not related between the two load patterns.

There is no obvious all-time relationship between all three utility load patterns. The trends of all three load patterns are potentially increasing or decreasing at some specific time points. In terms of the extent of increasing or decreasing of all three load patterns, it is rather challenging to conclude the result despite having the most frequent combinations. A possible explanation is that electricity load is determined by many home appliances while the other two are determined by more limited and specific home appliances. Therefore, there are unexpected sharp spikes in electricity load pattern, comparing to gas and water load patterns.

5 CONCLUSION AND FUTURE WORK

This thesis aims to explore the residential load patterns of electricity, gas, and water, with the consideration of seasonal variation. Furthermore, with the established load patterns, this thesis attempts to investigate any potential relationships between electricity, gas, and water load patterns. The methodology designed for the above tasks is K-Means clustering using Euclidean distance. By implementing the methodology on the real residential load data, five, four, and three clusters are

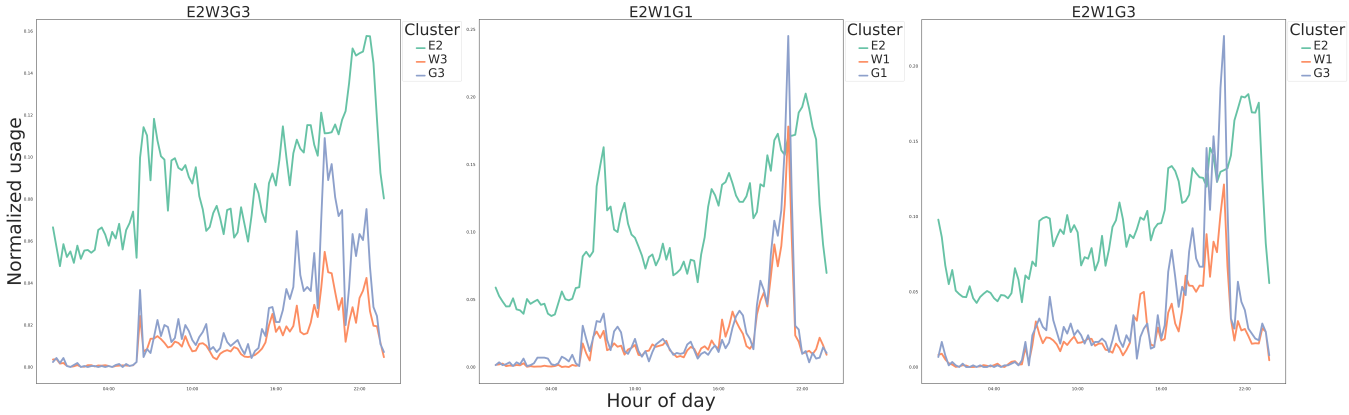

Figure 11: Coverage of E2W3G3 in percentage.

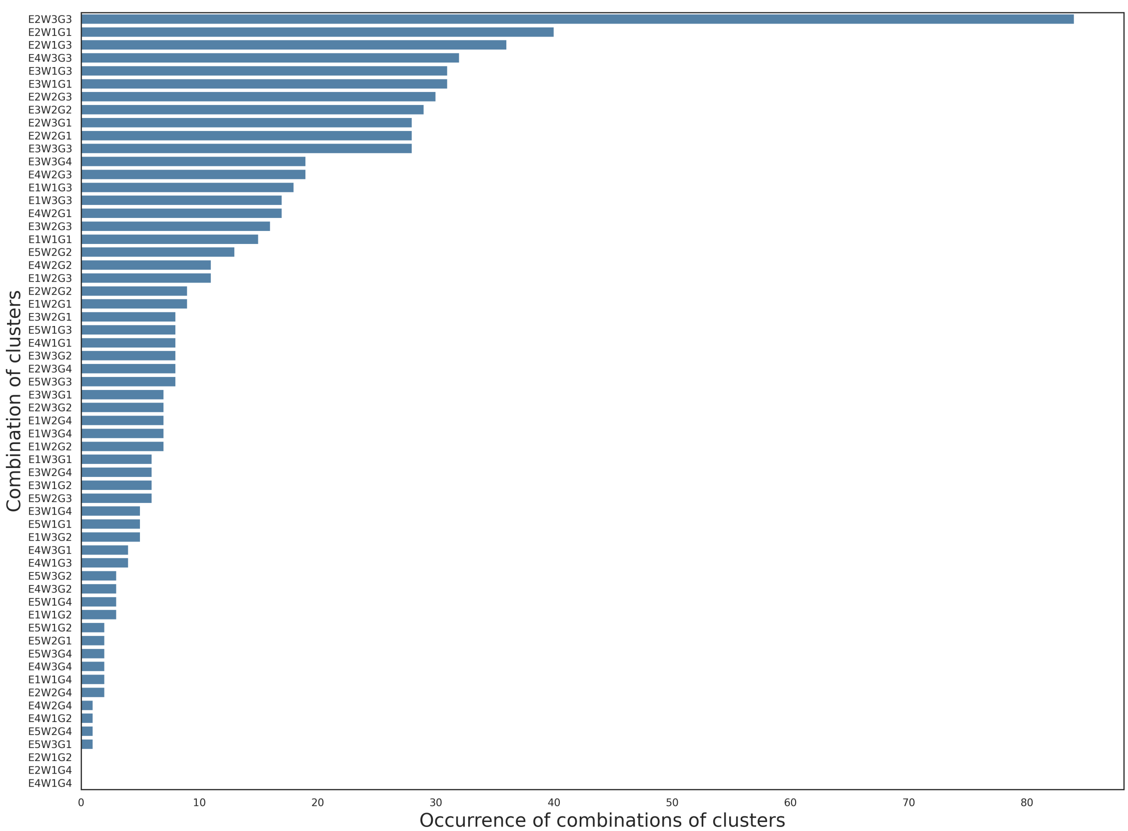

Figure 12: Count of All Cluster Combinations.

captured to represent electricity, gas, and water load patterns respectively. Based on the observations of the visualization of the clustering result, it is possible to analyze the behavior patterns of the residents with the temporal scale. Additionally, seasonal variation is also considered when analyzing the result and is found to have profound impact on gas load patterns specifically. On the other hand, visualization of the common cluster combinations implies that some gas load patterns are closely related to water load patterns. Due to the limitations of the availability of the load data, the explanation of the result might be restricted. Specifically, it is challenging to analyze the residents' behavior at certain sharp spikes and drops, when the water and gas load data is not recorded on appliance-level. With available dataset, the load pattern of gas and water could be decomposed to appliance-level, which might further illustrate resident's behavior pattern and provide more detailed assumptions on the spiking and dropping, as well as seasonal variations.

REFERENCES

[1] Kaiser Ahmed, Petri Pylsy, and Jarek Kurnitski. 2016. Hourly consumption profiles of domestic hot water for different occupant groups in dwellings. *Solar Energy* 137 (2016), 516–530.

[2] Omar Y Al-Jarrah, Yousof Al-Hammadi, Paul D Yoo, and Sami Muhaidat. 2017. Multi-layered clustering for power consumption profiling in smart grids. *IEEE Access* 5 (2017), 18459–18468.

[3] Juan de Santiago, Osvaldo Rodriguez-Villalón, and Benoit Sicre. 2017. The generation of domestic hot water load profiles in Swiss residential buildings through statistical predictions. *Energy and Buildings* 141 (2017), 341–348.

[4] Grzegorz Dudek, Anna Gawlak, Miroslaw Kornatka, and Jerzy Szkutnik. 2018. Analysis of smart meter data for electricity consumers. In *2018 15th International Conference on the European Energy Market (EEM)*. IEEE, 1–5.

[5] EIA. 2022. How many smart meters are installed in the United States, and who has them? https://www.eia.gov/tools/faqs/faq.php?id=108&t=3 Last accessed 31 May 2023.

[6] Elena Fuentes, L Arce, and Jaume Salom. 2018. A review of domestic hot water consumption profiles for application in systems and buildings energy performance analysis. *Renewable and Sustainable Energy Reviews* 81 (2018), 1530–1547.

[7] Payal Gupta, Thaw Tar Thein Zan, Justin Dauwels, and Abhisek Ukil. 2018. Flow-based estimation and comparative study of gas demand profile for residential units in Singapore. *IEEE Transactions on Sustainable Energy* 10, 3 (2018), 1120–1128.

[8] Qiang Jiang, Jia-Xiong Zhu, Min Li, and Hai-Yin Qing. 2018. Electricity power load forecast via long short-term memory recurrent neural networks. In *2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC)*. IEEE, 265–268.

[9] Oussama Laib, Mohamed Tarek Khadir, and Lyudmila Mihaylova. 2019. Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks. *Energy* 177 (2019), 530–542.

[10] Karen Byskov Lindberg, Steffen J Bakker, and Igor Sartori. 2019. Modelling electric and heat load profiles of non-residential buildings for use in long-term aggregate load forecasts. *Utilities Policy* 58 (2019), 63–88.

[11] Xiufeng Liu, Lukasz Golab, Wojciech Golab, Ihab F Ilyas, and Shichao Jin. 2016. Smart meter data analytics: systems, algorithms, and benchmarking. *ACM Transactions on Database Systems (TODS)* 42, 1 (2016), 1–39.

[12] Stephen Makonin, Bradley Ellert, Ivan V Bajić, and Fred Popowich. 2016. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014. *Scientific data* 3, 1 (2016), 1–12.

[13] Ioannis P Panapakidis, Aggelos S Bouhouras, Georgios C Christoforidis, and Constaninos Parisses. 2019. A Clustering Based Methodology for Natural Gas Demand Analysis. In *2019 IEEE 13th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*. IEEE, 1–6.

[14] Md Atiqur Rahman, Ishmam Rahman, and Nur Mohammad. 2020. Demand side residential load management system for minimizing energy consumption cost and reducing peak demand in smart grid. In *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*. IEEE, 376–381.

[15] Kumudu Rathnayaka, Hector Malano, Shiroma Maheepala, Biju George, Bandara Nawarathna, Meenakshi Arora, and Peter Roberts. 2015. Seasonal demand dynamics of residential water end-uses. *Water* 7, 1 (2015), 202–216.

[16] Makonin Stephen. 2016. AMPds2: The Almanac of Minutely Power dataset (Version 2). https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FFIE0S4&version=&q=&fileAccess=&fileTag=&fileSortField=&fileSortOrder= Last accessed 4 May 2023.

[17] Gianluca Trotta. 2020. An empirical analysis of domestic electricity load profiles: Who consumes how much and when? *Applied Energy* 275 (2020), 115399.

[18] Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang. 2018. Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid* 10, 3 (2018), 3125–3148.

[19] Lin Xiqiao, Wu Wanlu, Zeng Bo, Yan Xu, Han Shuai, and Qin Lijuan. 2018. Analysis of large-scale electricity load profile using clustering method. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*. IEEE, 1–5.