LBNL-xxxxx

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

Data Mining of Space Heating System Performance in Affordable Housing

Xiaoxin Ren[1], Da Yan[1], Tianzhen Hong[2]

[1]School of Architecture, Tsinghua University, Beijing 100084, China

[2]Building Technology and Urban Systems Division Energy Technologies Area

May 2015

This is an article published in the journal of Building and Environment, July 2015.

Disclaimer

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, or The Regents of the University of California.

Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.

Please cite this report as follows:
Ren, X., Yan, D., Hong, T. Data Mining of Space Heating System Performance in Affordable Housing. Building and Environment, 89: 1-13, 2015. LBNL-xxxxx.

Acknowledgement

This work was sponsored by the United States Department of Energy (Contract No. DE-AC02-05CH11231) and the China Ministry of Housing and Urban - Rural Development and the Ministry of Science & Technology (Grant No. 2010DFA72740-02) under the U.S.-China Clean Energy Research Center for Building Energy Efficiency. It is also part of the research of Annex 66, Definition and Simulation of Occupant Behavior in Buildings, under the International Energy Agency Energy in Buildings and Communities Program.

Data Mining of Space Heating System Performance in Affordable Housing

Xiaoxin Ren[1], Da Yan[1], Tianzhen Hong[2,*]

[1]*School of Architecture, Tsinghua University, Beijing, China*

[2]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California, CA 94720, USA*

*Corresponding Author. Email: thong@lbl.gov, Phone: (510) 486-7082, Fax: (510) 486-4089

Abstract

The space heating in residential buildings accounts for a considerable amount of the primary energy use. Therefore, understanding the operation and performance of space heating systems becomes crucial in improving occupant comfort while reducing energy use. This study investigated the behavior of occupants adjusting their thermostat settings and heating system operations in a 62-unit affordable housing complex in Revere, Massachusetts, USA. The data mining methods, including clustering approach and decision trees, were used to ascertain occupant behavior patterns. Data tabulating ON/OFF space heating states was assessed, to provide a better understanding of the intermittent operation of space heating systems in terms of system cycling frequency and the duration of each operation. The decision tree was used to verify the link between room temperature settings, house and heating system characteristics and the heating energy use. The results suggest that the majority of apartments show fairly constant room temperature profiles with limited variations during a day or between weekday and weekend. Data clustering results revealed six typical patterns of room temperature profiles during the heating season. Space heating systems cycled more frequently than anticipated due to a tight range of room thermostat settings and potentially oversized heating capacities. The results from this study affirm data mining techniques are an effective method to analyze large datasets and extract hidden patterns to inform design and improve operations.

Keywords: data mining, space heating, affordable housing, occupant behavior, building simulation, clustering, decision tree

1 Introduction

Building energy consumption has steadily increased since 2008, representing 20% to 40% of the total primary energy use in developed countries. In particular, energy consumption from the buildings sector has grown at a faster rate than the industrial and transportation sectors [1]. This upward trend in energy use becomes difficult to rebuke, due to the growth in population, increasing demands for building services and comfort levels, and the rise of occupied hours spent inside buildings. Parker et al. [2] studied energy use choices in experimental, low-income homes which were designed with high Seasonal Energy Efficiency Ratio (SEER) air conditioners, reflective roofing, solar water heaters and energy efficient lighting and appliances. They demonstrated that very different energy usage patterns were prevalent in the individual homes, especially with regard to cooling choices. In a follow up study conducted in 2012, Parker et al. [3] verified the Home Energy Saver (HES) suite for online simulation, by conducting a detailed year-long study to analyze the influence of occupant behavior on building energy use (Figure 1). The homes studied exhibited a 3-fold variation in measured energy use, with variations at the end-use level conferred as even larger. For the space heating energy use, House 9 consumed a yearly cumulative of 1,467 kWh, the most lavish user among the 10 houses, while House 9's total energy use was the least.

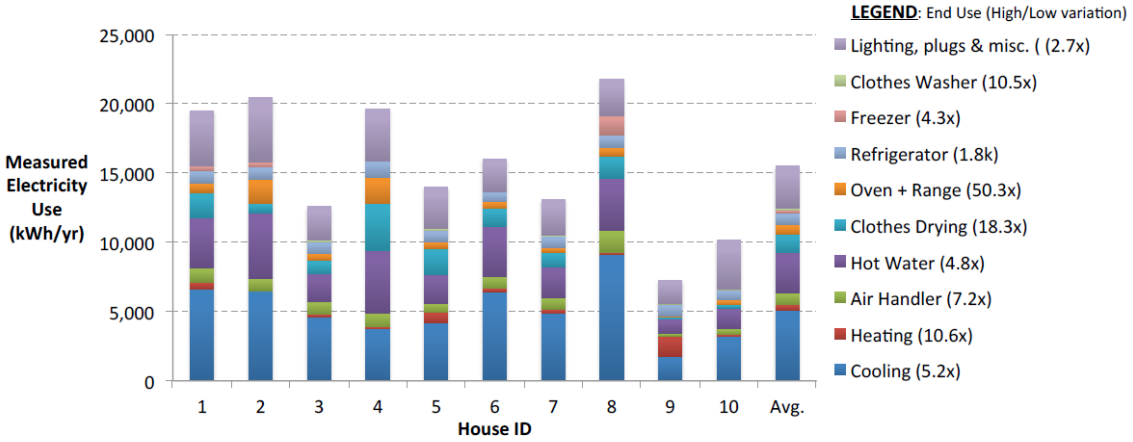


Figure 1 Measured total and end-use annual energy for the Homestead Cohort [3]

Occupant behavior has long been known to have a large impact on residential energy use [4]. Schipper et al. [5] indicated that approximately 50% of energy use in homes comes from the intrinsic building shell, equipment, lighting and electronics, while the remaining half comes from the occupant's interaction with the building components listed in Figure 1. Two significant international collaborations under the International Energy Agency Energy in Buildings and Communities Program, focused on occupant behavior research, these programs include Annex 53 titled "Total Energy Use in Buildings" [6] and Annex 66 titled "Definition and Simulation of Occupant Behavior in Buildings" [7]. Wang et al. [8] developed a building occupancy model based on the Markov chain [9] to simulate occupant movement in a building. Peng et al. [10] developed a quantitative method to describe occupant behavior in residential buildings. Ren et al. [11] developed a conditional probability model to represent air-conditioning use in residential buildings based on Wang's [8] and Peng's [10] studies.

As one of the critical occupant behaviors in residential buildings, space heating and thermostat use in buildings have gained much attention due to their significant impact on energy

consumption. Several space heating models have been developed for energy consumption and control scheme [12-14]. Peffer et al. [15] reviewed studies on how people use thermostats in homes, finding that nearly half do not use the programming features of the thermostats and suggesting that further research is needed to design thermostats which can provide more comfortable and economical indoor environments.

Recently, more sensor data on occupant behavior and building systems operations has become available. However, it is still challenging for designers and operational managers to effectively analyze the data to extract valuable information to support their decision making. Data mining can be a powerful tool to address such challenges. Data mining is the analysis step of the Knowledge Discovery in Databases (KDD) process [16]. In the 1990s, the modern data mining techniques improved, but were mainly used in areas such as economics, sociology, computer science and a few engineering subjects. In his book, Hand [17] demonstrated that the analysis of large observation datasets, using data mining methods, could reveal hidden data relationships, enabling the data to be summarized in novel ways which provided insight for decision making.

In recent years, data mining has gained popularity in the building science area, including occupant behavior, fault detection [18], building automation systems [19] and building energy performance [20-22]. Data mining is an effective technique to gain new knowledge from big data, finding the new relevance from a nonobvious context. For example in occupant behavior studies, the literature suggests that both psychology and physiology influences are important and hard to describe quantitatively. Data mining can be used to overcome this challenge, resulting in knowledge gained. Therefore, the data mining approach shall be generally used in new research areas to discovery new knowledge, from established areas. Especially with the rapid development of information technology, massive amounts of data will be available in the building sector. This presents an excellent opportunity for data mining applications to discover new science.

In occupant behavior research area, Yu [23] developed a new methodology for examining the influences of occupant behavior on building energy consumption using a clustering analysis approach. Another study from Yu et al. [24] used a methodology for discovering knowledge though a data mining approach to makes inferences from building operational data. Luhr et al. [25] and Dong et al. [26] used data mining to reveal occupant behavior trends, with a specific focus on smart homes and intelligent buildings, respectively. Wu and Clements-Croome [27] conducted a case study to understand the indoor environment through the mining of the sensory data. Motta-Cabrera and Zareipour [28] used data association mining to identify the lighting energy waste patterns in educational institutes. D'Oca and Hong [29] used data mining approaches to develop individualized occupancy schedules and to discover patterns of window opening and closing behavior in office buildings [30]. Zhao et al. [31] used office appliance power consumption to investigate different occupant behavior patterns and schedules for use in the modeling of building energy simulation.

Previous studies on space heating and thermostat use lack quantitative methods to analyze large measurement data sets, to identify and understand the diversity of system operation mode and occupant behavior. This paper aims to use data mining methods to analyze space heating performance in an affordable housing complex consisting of multifamily apartments and to provide insights into the following questions:

- What is the typical comfort demand of space heating in the affordable housing?
- Are there any unique patterns of room temperature profiles?
- How do space heating systems usually operate?

- Is there any link between room temperature profile, system operation and space heating energy use?
- What are the strategies to improve design and operation of the space heating system in the affordable housing?

2 Dataset and Previous Analyses

2.1. Dataset

The dataset used in this study was provided by Fraunhofer USA (www.fraunhofer.org). The experimental data was collected in an affordable housing complex in Revere, Massachusetts, which is near Boston. The complex consisted of multifamily apartments, as showed in Figure 2, with heating supplied by individual gas furnaces per unit. Natural gas was only used for heating and no other appliances. Tenants pay metered gas and electricity bills every month. In the experiment, the apartments measured were chosen through an opt-out process, with only rental units included. 82 apartments were measured, but 62 had complete and available data which were valid to use. The experiment lasted from Jan 11 to Mar 21, 2012. Pre-programmed thermostats were installed in each apartment, with default settings shown in Figure 3. Settings were similar to those defined in the ASHRAE 90.2 Standard [32]. The occupants were free to adjust their thermostat settings during the experiment period. The indoor temperature data was taken at 10 min intervals with the sensor placed in the living room. The gas consumption was recorded weekly. The time of furnace on/off state changes was also recorded. Thermostats settings were recorded at the end of the experiment to see if the occupants changed the settings. Based on these measurement data, the following analysis was done. In fact, room humidity is another important factor on thermal comfort besides the room dry bulb temperature in understanding the thermal demand of occupants. The experiment was taken in the winter near Boston where the air humidity does not change noticeably. Additionally, the apartment was heated only (with no humidity adjustment) and therefore the room temperature represents indirectly the thermal demand of the apartment. The room temperature data was collected with the thermostat installed on the wall of the living room at the same height of human body. The thermostat temperature represents the living room temperature which was used to control the operation of the central heating system for the apartment.

Figure 2 Front view of the studied affordable housing complex

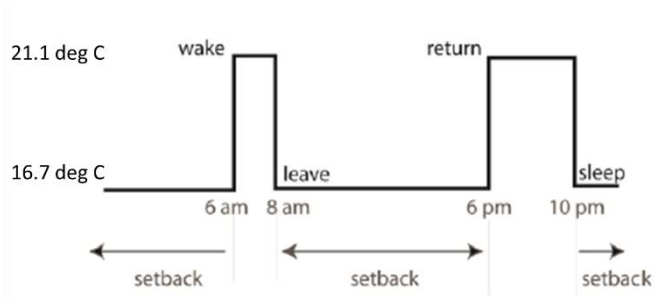


Figure 3 Default settings of the thermostat

2.2. Previous Analysis

Urban and Gomez [33] analyzed the space heating energy use of each apartment, based on floor location and orientation, as shown in Figure 4. The results showed that the space heating energy consumption was higher in lower building floors, but the orientation of the apartment had little impact.

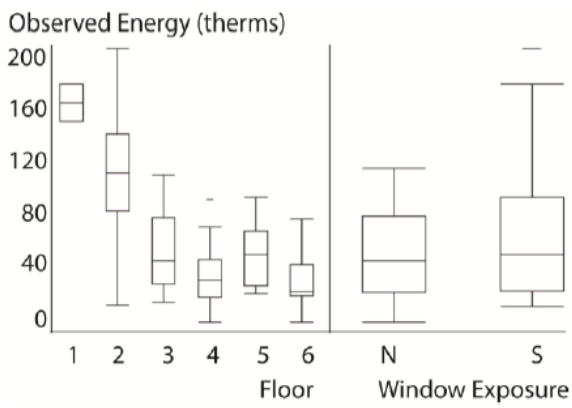


Figure 4 Observed heating energy consumption [19]

Four types of thermostat use are classified, with groupings assigned by visual inspection of the graphs of measured room temperatures. Thermostat settings were grouped into two patterns, the fixed set point pattern representing a permanent hold setting, and the scheduled pattern representing a program of a predictable manual routine. Patterns were then divided into frequent and infrequent pattern, based upon the frequency of manual overrides, including temporary overrides, manual adjustments, turning the furnace off, or opening a window.

Another study also from Urban et al. [34] combined the experiment results taken in 2012 and 2013. In the first year's experimental testing, the preprogrammed set points (Figure 5) was 62°F (16.6°C) for the leave (8:00 am to 6:00 pm) and sleep (10:00 pm to 6:00 am) periods, and 70°F (21.1°C) for the wake (6:00 am to 8:00 am) and return (6:00 pm to 10:00 pm) periods. Notably, these settings are similar to those used in the ASHRAE 90.2 Standard. While in the second year, the base temperature set point was raised from 62°F (16.6°C) to 68°F (20°C) for the leave and sleep periods. Results from the second year experiment suggest more occupants kept the default settings, probably due to the modest setback compared with the more aggressive setback in the

first year. The modest setback actually led to more energy saving because the frequency of the occupant override was reduced.

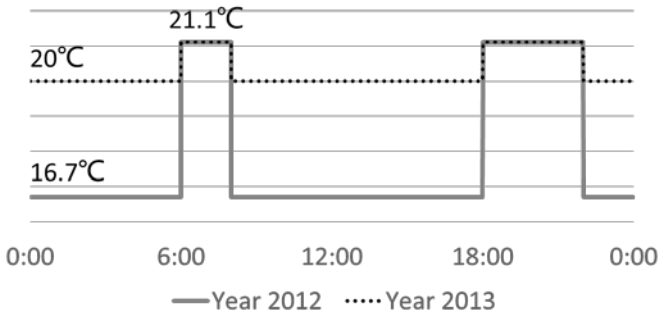


Figure 5 Default settings of the thermostat in the two years

Urban's studies [33, 34] provided interesting and valuable information to help understand how occupants adjust thermostat settings and associated space heating energy use. However most of his results came from manual analysis or visual inspection, which is labor intensive and may not scale up for large dataset applications. Furthermore, the accuracy of results could lack the needed scientific basis to justify findings.

3 Methodology

3.1. Data Mining

Data mining is a process of the Knowledge Discovery in Databases (KDD), which includes five steps: 1) Selection; 2) Pre-processing; 3) Transformation; 4) Data Mining; and 5) Interpretation / Evaluation, as shown in Figure 6. Selection is to choose what kind of data can be used in the KDD process. Pre-processing is to preliminary edit, wiping off mistakes and indicating data gaps, of the selected data. After the pre-processing, Transformation is to transform the data into some format easy for data mining. Data mining is the key process for data analysis, which will be used in this paper. Interpretation/evaluation is to evaluate the results from the data mining.

Figure 6 The data mining process [35]

Data mining was used for two types of analyses, predictive and descriptive. The predictive analysis employs classification techniques to find a model of a particular class attribute as a

function of other attributes, so that previously unprocessed records can be assigned a class as accurately as possible. The decision tree approach [36] is a predictive analysis, which uses a decision support tool such as a tree-like graph or decision model mapping possible consequences. This includes chance event outcomes, resource costs, and utility. The decision tree is used in this study to establish a flowchart-like tree structure to segregate heating energy consumption data of different apartments into various predefined classes depending on the temperature demand and building features.

The descriptive analysis uses clustering and association rule techniques. Clustering analysis [37] is the process of grouping data objects into clusters so that objects in the same cluster have high similarity, while objects in different clusters have low similarity. The K-means approach, common in clustering computation and also used in our study, calculates the dissimilarity between data objects using the Euclidian distance. Clustering analysis is used here for grouping the room temperature data into clusters based on level and variability to represent certain patterns of temperature demand of the apartments. Several notable software tools have been developed for data mining, include RapidMiner Studio 6.0, which was used in our study [38].

3.2. Technical Approach

Figure 7 shows the technical approach used in our study where the dataset of the room temperature, the space heating system operation and the gas consumption is mined to identify and understand typical use patterns. The room temperature of the 62 apartments can be clustered into specific demand patterns. The heating operation was used to infer the house and system performance. The demand and the system can be associated to the heating energy consumption with a decision tree model.

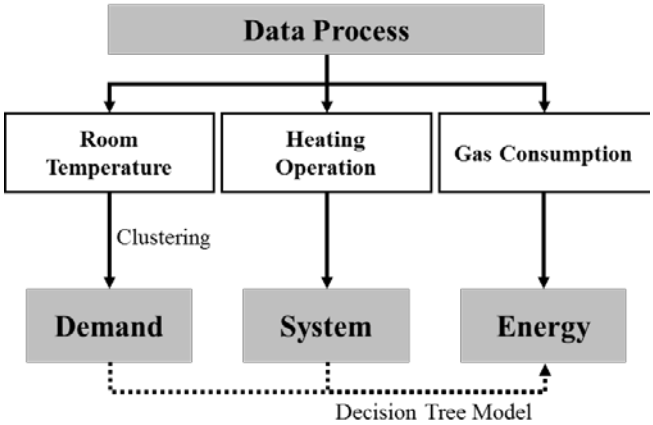


Figure 7 Technical Approach

4 Results

4.1. Room Temperature

4.1.1. Room temperature profiles

The room temperature data has been recorded in the 62 apartments, at 10-minutes intervals, from Jan 11 to Mar 21, 2012. The room temperature data are profiled in Figure 8. F1 to F6 represents the six floors of the building, and color represents the room temperature in the apartments. A

vertical line striation shows the 24 hour temperature in a single day. Great variations in room temperatures, for the 62 apartments, can be observed.

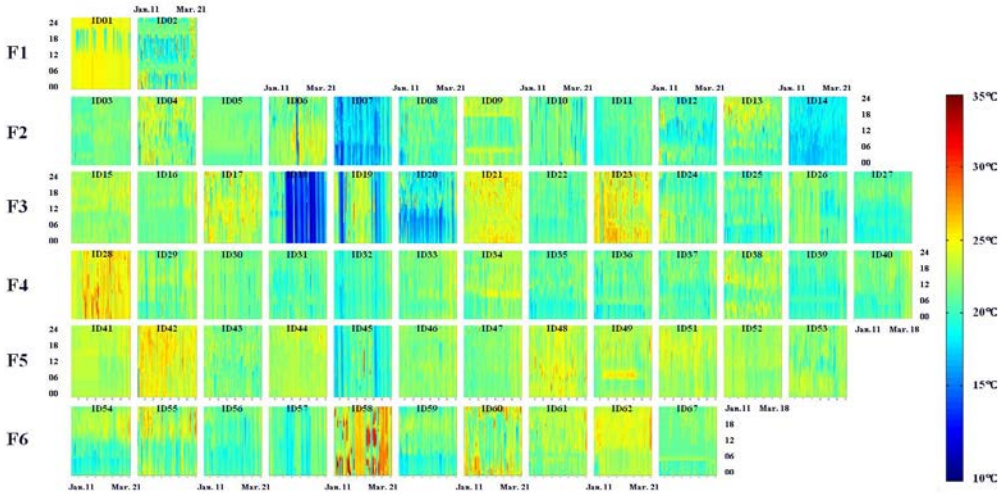


Figure 8 Profiles of room temperatures

4.1.2. Room Temperature Clusters

Clustering analysis is used to identify typical room temperatures patterns, which represents diverse heating demands. The attributions used are the hourly average room temperature from 0:00 am to 23:00 pm on weekday, the hourly standard deviation from 0:00 am to 23:00 pm on weekday during the 71-day study period. Similar data on weekends are also used. These attributes form a 24 × 4 dimensional clustering, which has a detailed description of the room temperature level, variation among days, daily schedule and differences between weekdays and weekends.

The number of the clusters needs to be predefined before the clustering analysis. Previous studies [39] recommended the number of clusters, K, as a local optimum, satisfying $2 \leqslant k \leqslant \sqrt{n}$. In this case, k was between 2 and 7. Using the Davies Bouldin Index [40] (DBI), the performance of the clustering was evaluated, with the lower DBI demonstrating clearer clustering performance. Figure 9 shows the DBI result from the performance operator in RapidMiner. Though only a small difference was identified, k=6 proved to be the optimal value chosen in the following analyses.

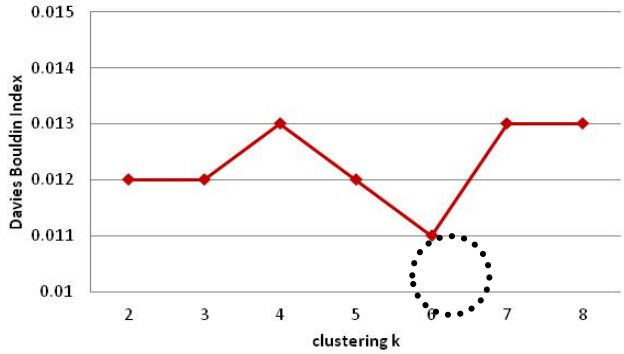
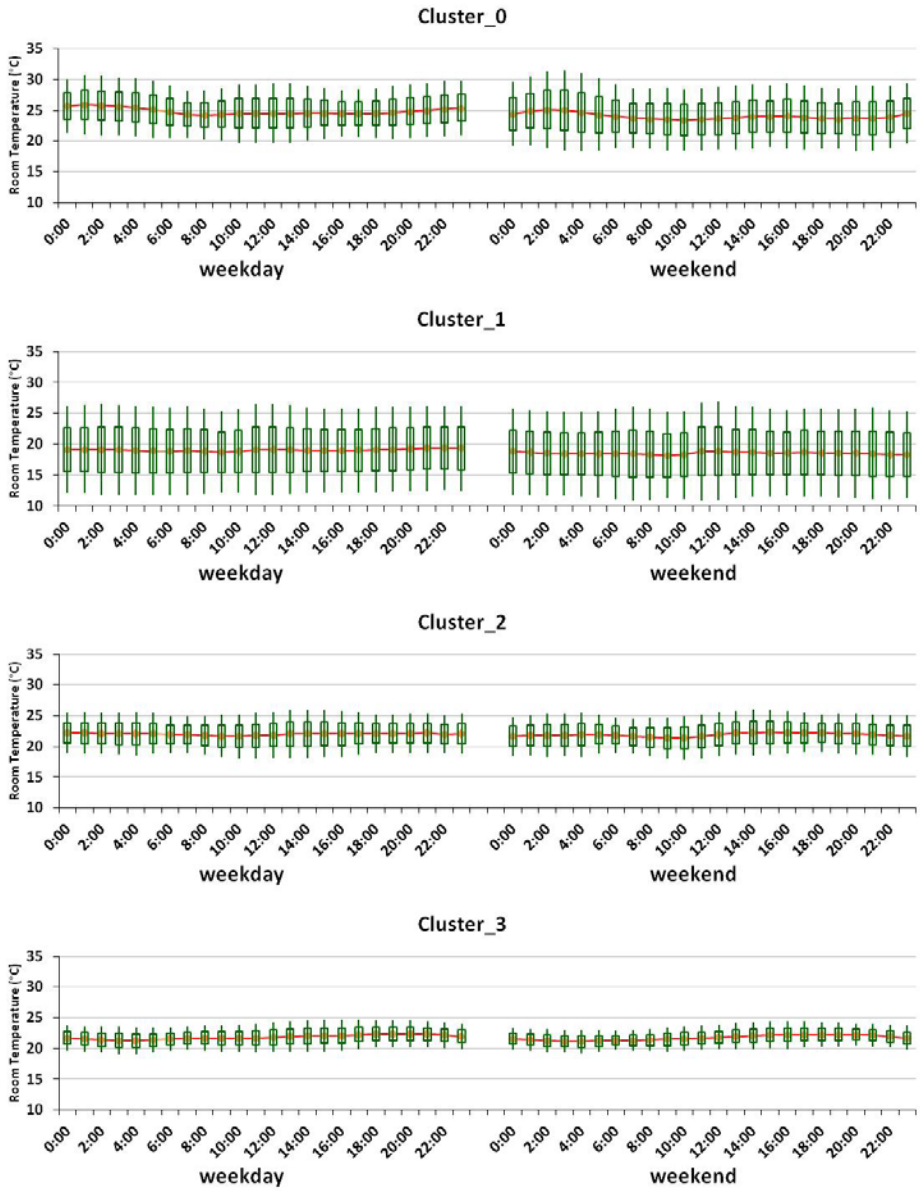


Figure 9 DBI with different k value in clustering

Figure 10 shows the results from the clustering by RapidMiner. The red lines show the hourly average room temperature, the green boxes show the variations of one standard deviation from the average values. The upper and lower ends show the range of two standard deviations. From the graphs, the basic clustering results demonstrate that the daily 24 hour variation of room temperature for all the six clusters was not so noticeable, with no significant difference between weekdays and weekends. Instead, main differences in the clusters were observed in the average room temperature level and the variations among days.



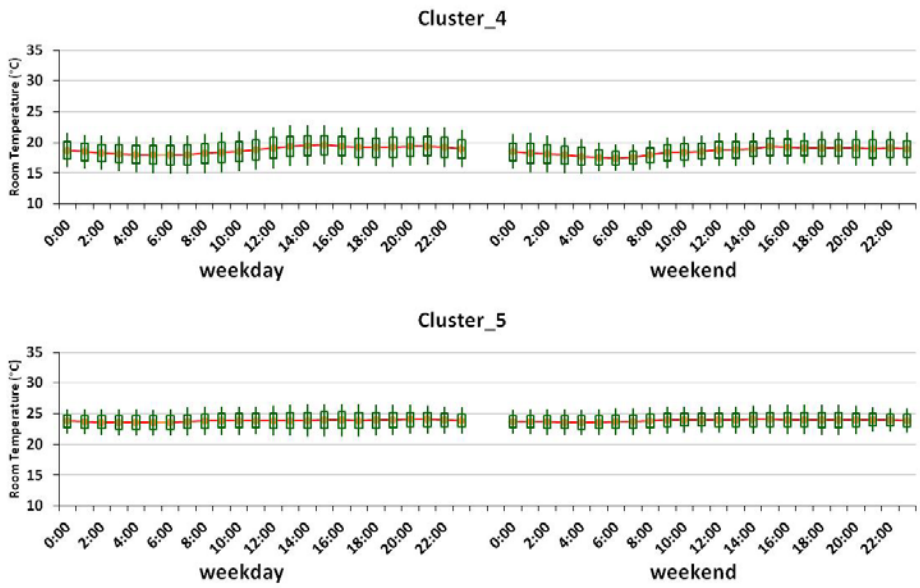


Figure 10 Room temperature clustering result

Table 2 shows the number of apartments belonging to each cluster. Most apartments were grouped into cluster 2, cluster 3 and cluster 5. Only eight apartments were not included in these three main clusters.

Table 1 The number of apartments in each cluster

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
# Apartments	3	2	12	26	3	16

To visualize the main features of the six clusters, all six clusters are profiled in a single synthetic graph, shown in Figure 11. The clusters are located in a coordinate of average room temperature and standard deviation among days. In Figure 11, the sizes of the data points represent the number of apartments belonged to each cluster, and the color shows the standard deviation of room temperature across 24 hours of the day. The different geometrical shapes signify each different cluster. Starting from bottom to top and left to right, the following can be inferred (Figure 11):

(i) Cluster 4 with three apartments, is defined as “Low T, Stable” which represents a low average room temperature level and minor variations across the days. It can also be seen that there is no apparent 24 hour change within a day; (ii) Cluster 1 with two apartments, is defined as “Low T, Varied” which represents low average room temperature level and large variations across the days. The 24 hour change within a day has a wider range than the other clusters; (iii) Cluster 3, the biggest cluster with 26 apartments, is defined as “Mid T, Stable”, with middle level and stable temperature. The 24 hour change is considered as middle level; (iv) Cluster 2, with 12 apartments, is defined as “Mid T, Varied”, with a middle level and varied temperature. There is also no apparent 24 hour change; (v) Cluster 5 with 16 apartments, is defined as “High T, Stable” with high level and stable temperature. The 24 hour change is also unnoticeable; (vi) Cluster 0, with 3 apartments, is defined as “High T, Varied”, with high level and varied temperature. There is noticeable 24 hour change in this cluster.

For the features of different clusters, descriptive names have been given such as LV (Low T, Varied) for cluster 1, MS (Mid T, Stable) for cluster 3, and so on, for identification purposes.

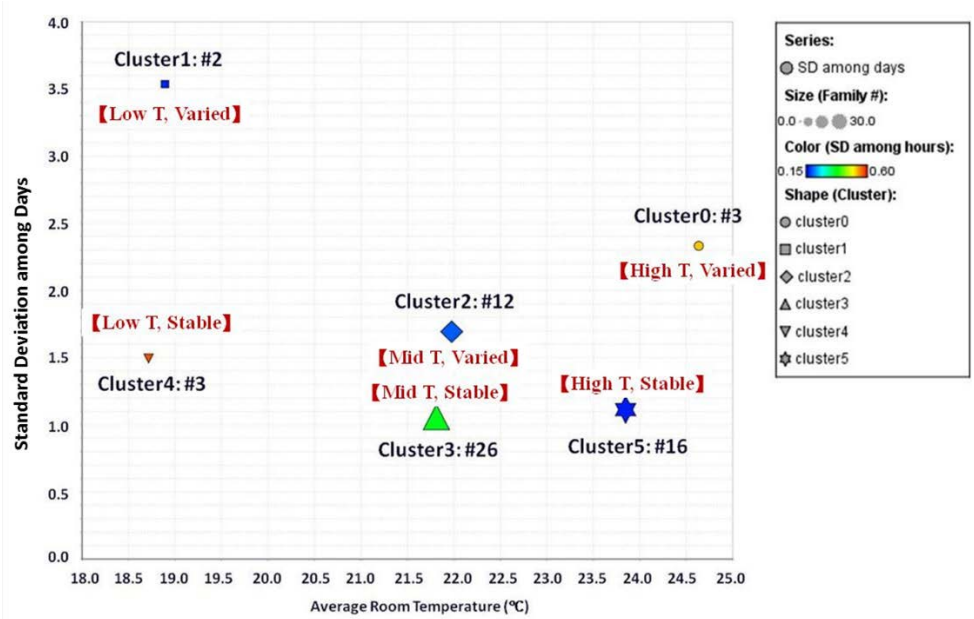


Figure 11 Clustering result synthetic graph

Figure 12 shows the rearrangement of the 62 apartments according to the six clusters of room temperatures. The six clusters represent the diverse and heterogonous space heating demand in the real affordable housing.

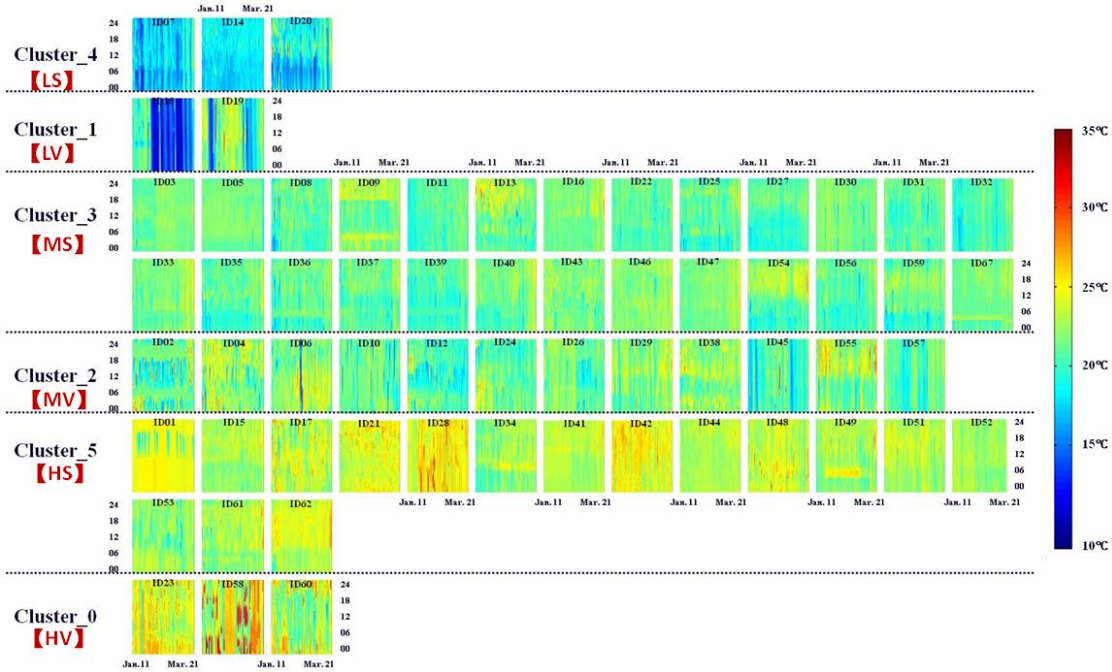


Figure 12 Six clusters of room temperature profiles

From the graph of profiled clusters, the room temperature can be seen to have specific features within one cluster. Each cluster can be considered as a pattern, with specific heating use demand.

The six patterns are meticulously defined with the 24 × 4 attributes, which describe the average room temperature level and the standard deviation among days in the heating season, for both weekdays and weekends. In this way, a stochastic algorithm can be used to generate a random profile of room temperatures in every hour, which can infer the thermostat setting adjustment. With the certain patterns of thermostat schedule and proportional distribution among patterns, the model can serve as input for space heating in building energy simulations. This can reflect the complexity of occupant behavior in space heating usage in reality.

4.2. Heating system operation

In order to analyze the heating system operation, the ON/OFF state change time of the furnace recorded data is used. Figure 13 shows the statistics of the probability distribution of the number of operation cycles in a day and the average run time per cycle. Most apartments on average cycled 10 to 15 times per day, and the average run time per cycle was mostly between 7 and 8 minutes. The two parameters multiplied together to provide the heating system operation time per day, are shown in Figure 14. It can be seen that the operation time ranged between 50 and 150 minutes per day, excluding outliers.

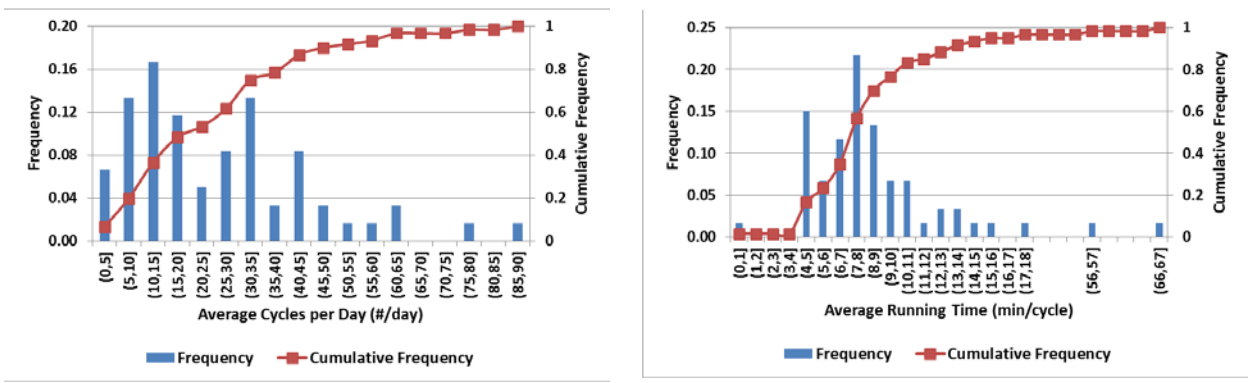


Figure 13 Statistics of basic heating operation parameters

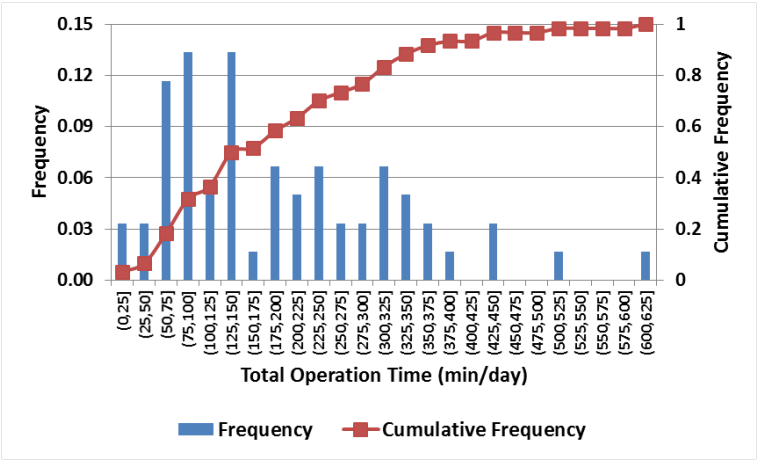


Figure 14 Statistics of heating system operation time

Figure 15 combines the heating system operation information with the six clusters into an integrated graph. The width of the column represents the continuous ON time per cycle, and the

height of the column represents the frequency of the ON/OFF cycle per day. The scatter shows the heating operation time, equivalent to Width times Height. The apartments in the graph are rearranged into clusters, with no obvious rule among them. The heating operation is not only related to the occupant demand, but also the house and system performance, as well as occupant's heating operation habits.

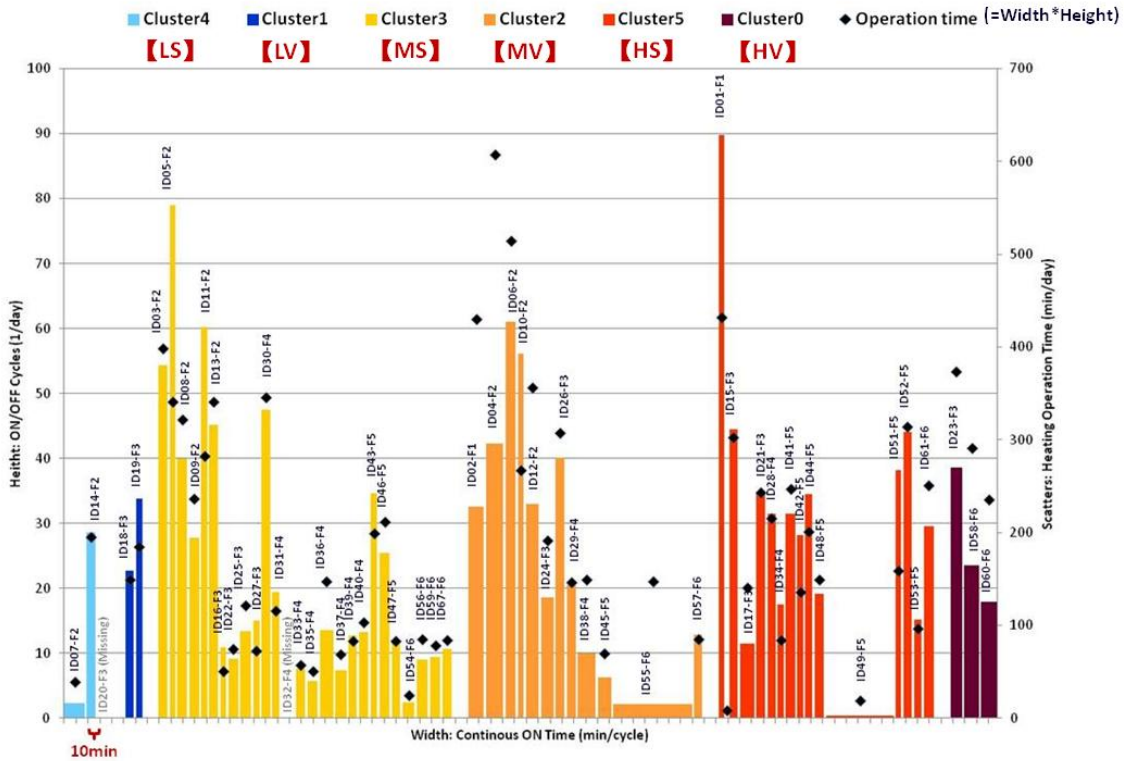


Figure 15 Heating system operation condition

Without the need to consider the impact by the occupant demand and operation habit, the pure physical characteristics of the heating system operation can be conferred from the parameter of ON/OFF ratio, defined as Time ON divided by Time OFF during the heating systems steadily ON period. Figure 16 shows such ON/OFF ratios in a 4 hour period. The ON/OFF ratio is influenced only by the house envelope, heating system performance and the status of operable windows.

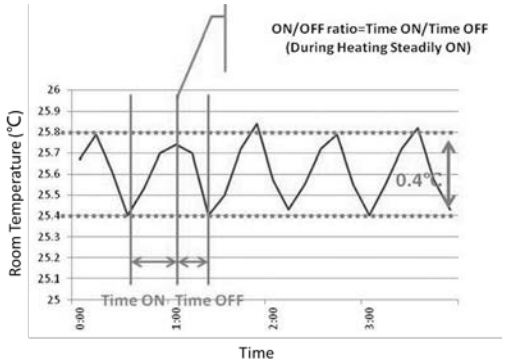


Figure 16 ON/OFF ratio calculation

The ON/OFF ratios, on one of the coldest week from Jan 16 to 22, are shown in Figure 17. Most apartments have the ON/OFF ratio between 0.6 and 0.7.

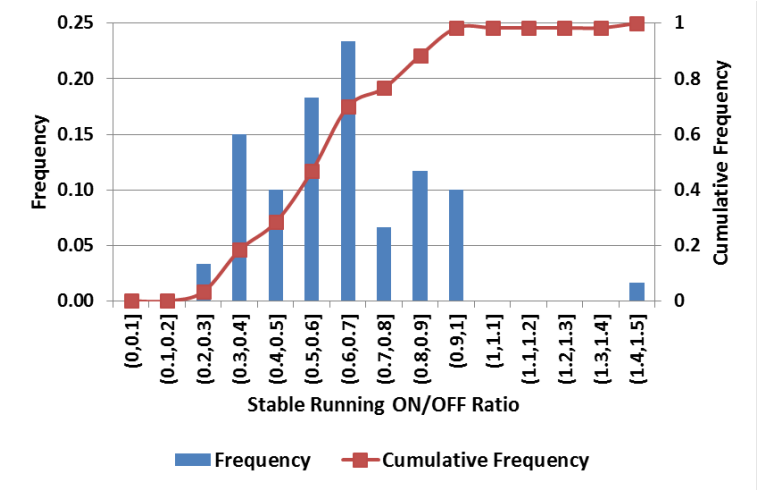


Figure 17 ON/OFF ratios during a steady heating period

Figure 18 shows the ON/OFF ratio of all 62 apartments. The ON/OFF ratio represents the physical performance of heating systems. The ON/OFF ratios can be equidistantly classified into R1-R5, as shown in Table 2. R1 has the lowest ON/OFF ratio representing a heating system with capability to raise the room temperature quickly, which can be due to adequate system heating capacity, better thermal insulation, more internal heat gains, or less window opening heat loss. On the other hand, R5 represents the opposite. Most apartments fall into the R2 and R3 classes.

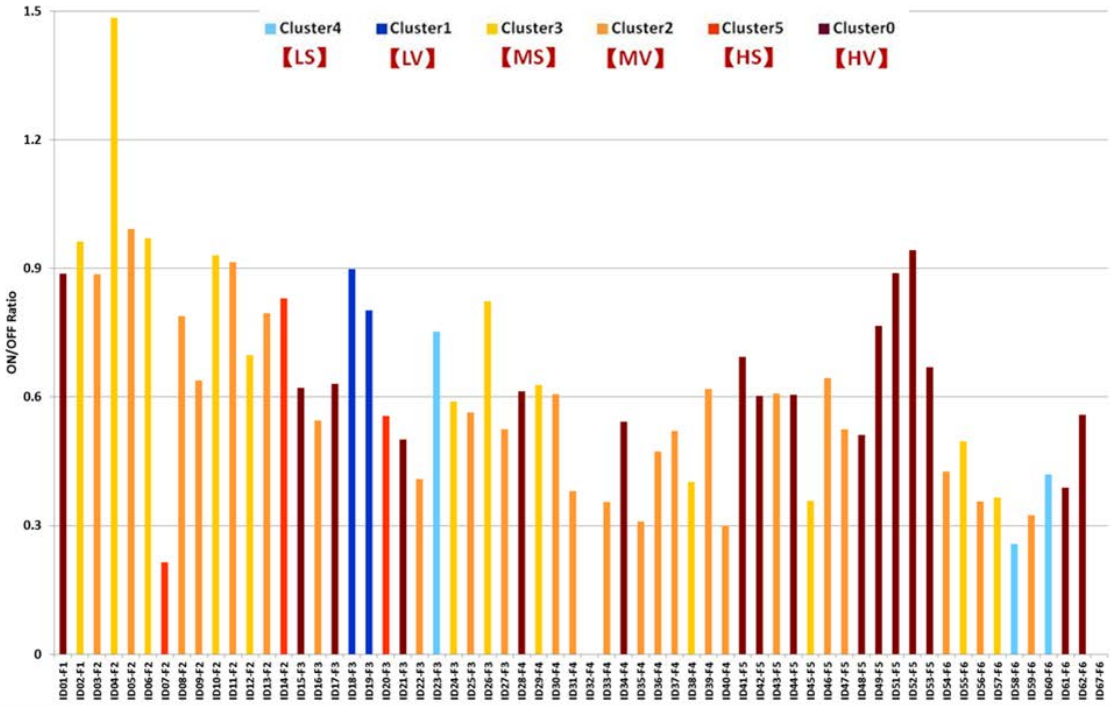


Figure 18 ON/OFF ratio of the 62 apartments

Table 2 ON/OFF ratio classification

ON/OFF ratio	Class	# Apartments
[0, 0.3)	R1	2
[0.3, 0.6)	R2	26
[0.6, 0.9)	R3	25
[0.9, 1.2)	R4	6
[1.2, 1.5)	R5	1

4.3. Gas consumption

Figure 19 shows the weekly gas consumption of the heating system by the six clusters. Within a cluster, apartments are arranged from low to high floors. As the analysis done by Urban mentioned above [20, 21], apartments on lower floors consume more heating energy. However, no obvious differences are observed among clusters, which may be due to the fact that the energy consumption of the heating system is not only related to the room temperature, but also related to the house and system performance.

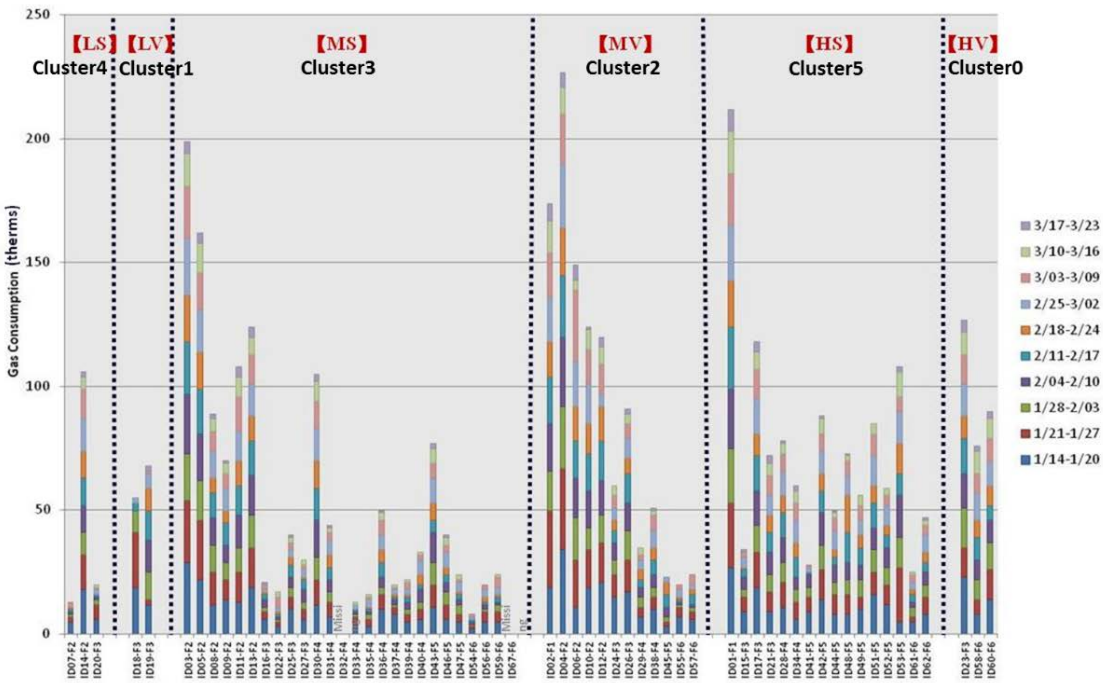


Figure 19 Weekly heating gas consumption

The heating system gas use can also be divided into equidistant classes, as shown in Table 3. Class A represents the lowest level of gas consumption, between 0 and 0.5 therms per day, while Class G represents the highest level of gas consumption, between 3.0 and 3.5 therms per day. The first three classes captured the gas consumption performance of most of the apartments.

Table 3 Heating system gas consumption classification

Gas Consumption (therms/day)	Class	# Apartments
[0, 0.5)	A	21
[0.5, 1.0)	B	14
[1.0, 1.5)	C	12
[1.5, 2.0)	D	7
[2.0, 2.5)	E	3
[2.5, 3.0)	F	2
[3.0, 3.5)	G	1

4.4. Decision tree

As shown in the analysis above, the room temperature represents the comfort demand, and the ON/OFF ratio represents the house and system performance. These are two independent factors of the heating energy use. A decision tree model was used to associate the comfort demand and the house and system performance, with the logic approach shown in Figure 20. The energy use class was predicted by the attributes of comfort demand cluster and house & system performance class.

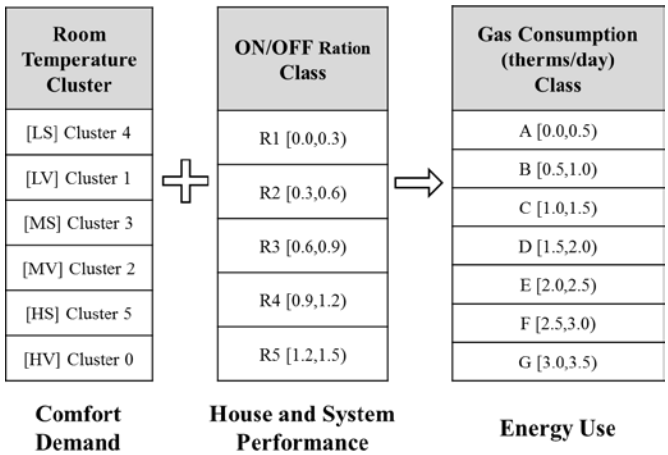


Figure 20 Logic approach of the decision tree model

The decision tree model results are shown in Figure 21. Due to the limited amount of samples, model validation was not conducted. The decision tree shows how the energy use is influenced by the comfort demand and house and system performance. For higher comfort demand temperatures, higher energy use is needed, and a varied demand may consume more energy than the stable demand. For apartments with the same comfort demand, if the house and system performed better, the energy use is lower.

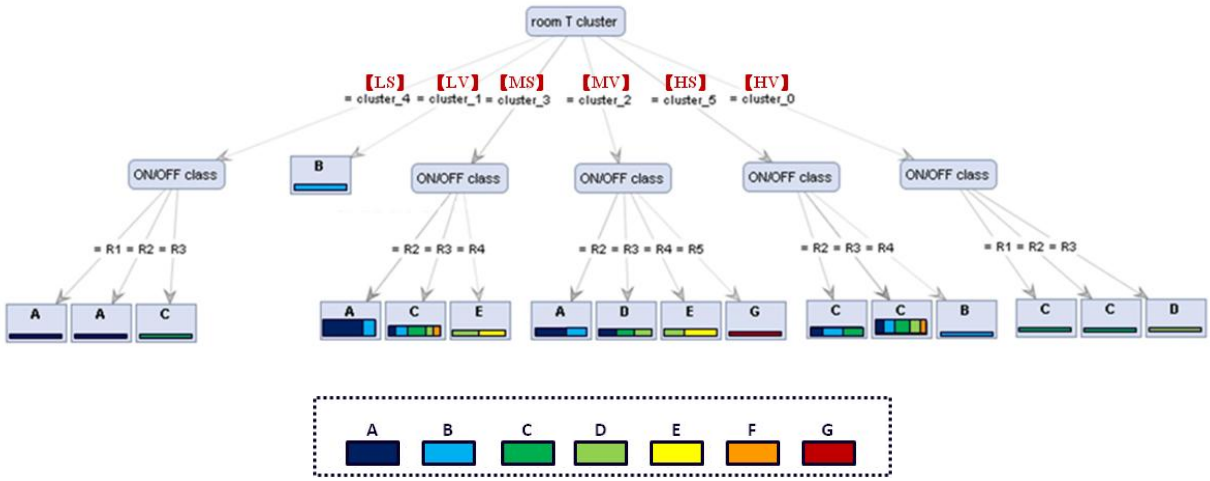


Figure 21 Decision tree model

The manuscript of the decision tree can be described as the following:

Room T cluster = cluster_4
ON/OFF class = R1: Energy use class is A {A=1, B=0, C=0, D=0, E=0, F=0, G=0}
ON/OFF class = R2: A {A=1, B=0, C=0, D=0, E=0, F=0, G=0}
ON/OFF class = R3: C {A=0, B=0, C=1, D=0, E=0, F=0, G=0}
Room T cluster = cluster_1
B {A=0, B=2, C=0, D=0, E=0, F=0, G=0}
Room T cluster = cluster_3
ON/OFF class = R2: Energy use class dominated by A {A=11, B=3, C=0, D=0, E=0, F=0, G=0}
ON/OFF class = R3: C {A=1, B=2, C=3, D=1, E=0, F=1, G=0}
ON/OFF class = R4: E {A=0, B=0, C=0, D=1, E=1, F=0, G=0}
Room T cluster = cluster_2
ON/OFF class = R2: A {A=3, B=2, C=0, D=0, E=0, F=0, G=0}
ON/OFF class = R3: D {A=1, B=0, C=1, D=1, E=0, F=0, G=0}
ON/OFF class = R4: E {A=0, B=0, C=0, D=1, E=2, F=0, G=0}
ON/OFF class = R5: G {A=0, B=0, C=0, D=0, E=0, F=0, G=1}
Room T cluster = cluster_5
ON/OFF class = R2: C {A=1, B=2, C=2, D=0, E=0, F=0, G=0}
ON/OFF class = R3: C {A=2, B=2, C=3, D=2, E=0, F=1, G=0}
ON/OFF class = R4: B {A=0, B=1, C=0, D=0, E=0, F=0, G=0}

Room T cluster = cluster_0
ON/OFF class = R1: C {A=0, B=0, C=1, D=0, E=0, F=0, G=0}
ON/OFF class = R2: C {A=0, B=0, C=1, D=0, E=0, F=0, G=0}
ON/OFF class = R3: D {A=0, B=0, C=0, D=1, E=0, F=0, G=0

Based on the decision tree model, a quick prediction of the heating energy use can be inferred to support the design of the heating system in affordable housing, according to the combinations of comfort demand clusters and heating system operations (ON/OFF ratio).

5 Discussion

In building simulation, for the input of the space heating schedule, the typical assumption is that the schedule of different occupants is fixed and homogeneous. In reality, the schedule of occupants varies significantly and is heterogeneous. The fixed uniform schedule used in the simulation maybe far from the reality (measured data), leading to incorrect decision making in technology evaluation and the energy use.

In this study, the space heating performance in affordable housing was analyzed, with typical patterns identified from the measurement data. The space heating usage schedule (thermostat settings) can be inferred from the room temperature patterns, which is a more accurate representation for the building energy simulation. The result are not entirely universal, but the data mining method in this study is considered generic enough to help discover various use patterns in different residential spaces.

For affordable housing other than ordinary residence (i.e. non-affordable housing), provides unique features can be identified. The mined patterns show the space heating use had no obvious 24 hour changes during the day for most apartments. Additionally, no obvious differences between weekdays and weekends, occurred. This may be due to the fact tenants of the affordable housing were low-income families, with one or even none of the occupants working full time. With minor to no variation, the results suggest the house was occupied all day long from morning to night, weekdays and weekends. Among the findings in this study the room temperature level and variations among days are the main differences among the six use patterns and an indication of use patterns. As in the previous study by Fraunhofer, the pre-programed setting with the energy savings schedule did not work well in the affordable housing. To defend the low temperature setting during daytime, when most occupants stay home in these apartments, the temperature settings could be changed by residents to a higher level, maybe even higher than

necessary. So the energy consumption was even higher with this energy-saving pre-program than what would occur with a modest setback schedule, used in the second year.

In the design of the space heating system in affordable housing, the unique features shall be considered. Discoveries got from the analysis include specific patterns in space heating use. Most occupants kept the heating on for nearly 24 hours a day, but some occupants turned off the heating when leaving the house or when the outdoor temperature elevated. Considering the 24 hour users, the capacity of the heating system should be controlled to a suitable level. As shown in the heating system operation results, the heating system in some apartments cycled a lot when the system was on, and the continuous running time in one cycle was quite short. The frequent cycling decreases the efficiency of the heating system, and leads to a higher energy consumption. Alternatively, the decrease of the capacity of the heating system may bring another problem. For the occupants who turn off their heating system when not needed, a limited capacity of heating system will lead to a longer waiting time for the room temperature to reach the comfort set point. This implies the occupants have to remain uncomfortable for a longer period of time after they turn on heating. This discomfort caused by such waiting period may direct some of these occupants into the pattern of a 24-hour user (i.e. never turning off the heating system). This could result in an increase of the energy consumption of heating system. The key point is that the system design should include the integration of the system operation and consider the influence of the varied use patterns. Furthermore the optimizing of the control strategy is another important point, for example a wider dead zone in thermostat control may help the problem of too frequent cycling of the space heating systems.

6 Conclusions

This study used data mining methods to analyze the space heating performance of 62 apartments in an affordable housing complex. The main conclusions are: (1) Data mining is a powerful tool to analyze big datasets of building performance. In this study data mining methods help understand household comfort demand and space heating operations by identifying and visualizing use patterns from measured data; (2) Clustering approach reveals six patterns of room temperature demand, with most households aggregated in three of the six patterns; (3) Decision tree associates comfort demand and system performance with heating energy consumption, which can be used for a quick prediction in designing the heating system of an affordable housing; and (4) The room temperature patterns can inform thermostat settings, which can be used in building energy simulations to better represent diversity of occupant comfort demand. An important aspect of this study demonstrated the successful use of data mining techniques to gain knowledge about occupant behavior, which can inform simulations of energy use of low-income housing.

Acknowledgment

This work was sponsored by the U.S. Department of Energy (Contract No. DE-AC02-05CH11231), and the China Ministry of Housing and Urban-Rural Development and the Ministry of Science & Technology (Grant No. 2010DFA72740-02) under the U.S.-China Clean Energy Research Center for Building Energy Efficiency. The work was also supported by the Twelfth Five –Year National Science and Technology Support Program “Research on a framework to support energy efficiency technologies in buildings” (Grant No. 2012BAJ12B00).

The authors appreciate Fraunhofer for sharing the dataset and answered our questions. This work is part of the research in IEA EBC ANNEX 66, Definition and Simulation of Occupant Behavior in Buildings.

References

[1] Perez-Lombard L, Ortiz J, Pout C, Universidad de Sevilla. A review on building energy consumption information. Energy and Buildings 2008; 40(3):394-396.

[2] Parker DS, Mazzara MD, Sherwin JR, Florida Solar Energy Center. Monitored energy use patterns in low-income housing in a hot and humid climate. ESL-HH-96-05-38. 1996.

[3] Paker DS, Florida Solar Energy Center. Accuracy of the home energy saver energy calculation methodology. ACEEE Summer Study on Energy Efficiency in Buildings. 2012.

[4] Socolow R, Sonderegger R, Princeton University. The twin rivers program on energy conservation in housing: four year summary report Princeton University. Report No.32 Center for Environmental Studies, 1976.

[5] Schipper LJ, Bartlett S, Hawk D, Vine E. Linking lifestyles and energy use: a matter of time? Annual Review of Energy 1989; 14:273-320.

[6] International Energy Agency Energy in Buildings and Communities Programme, Annex 53: Total Energy Use in Buildings: Analysis and Evaluation Methods. 2008-2013.

[7] International Energy Agency Energy in Buildings and Communities Programme, Annex 66: Definition and Simulation of Occupant Behavior in Buildings. www.annex66.org. 2013-2017.

[8] Wang C, Yan D, Jiang Y, Tsinghua University. A novel approach for building occupancy simulation. Building Simulation 2011; 4:149–167.

[9] Norris, James R. Markov chains. Cambridge University Press, 1998.

[10] Peng C, Yan D, Wu RH, Wang C, Zhou X, Jiang Y. Quantitative description and simulation of human behavior in residential buildings. Building Simulation 2012; 5:85–94.

[11] Ren X, Yan D, Wang C, Tsinghua University. Air-conditioning usage conditional probability model for residential buildings. Building and Environment 2014; 81:172-182.

[12] Hallert MY, Paavilainen J, Konersmann L, Haberl R, Droscher A, Frank E, Bales C, Streicher W. A unified model for the simulation of oil, gas and biomass space heating boilers for energy estimating purposes. Part I: Model development. Journal of Building Performance Simulation 2011; 4(1):1-18.

[13] Chingcuanco F, Miller EJ. A microsimulation model of urban energy use: Modeling residential space heating demand in ILUTE. Computers, Environment and Urban Systems 2012; 36(2):186-194.

[14] Engvall K, Lampa E, Levin P, Wickman P, Ofverholm E. Interaction between building design, management, household and individual factors in relation to energy use for space heating in apartment buildings. Energy and Buildings 2014; 81:457-465.

[15] Peffer T, Pritoni M, Meier A, Aragon C, Perry D. How people use thermostats in homes: A review. Building and Energy 2011; 46:2529-2541.

[16] Fayyad U, Piatetsky S. From Data Mining to Knowledge Discovery in Databases. Gregory: Smyth, Padhraic. 1996.

[17] Hand DJ, Mannila H, Smyth P. Principles of data mining. 2001.

[18] Du Z, Fan B, Jin X, Chi J. Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. Building and Environment 2014; 73:1-11.

[19] Xiao F, Fan C. Data mining in building automation system for improving building operational performance. Energy and Buildings 2014; 75:109-118.

[20] Kusiak A, Zeng Y, Xu G. Minimizing energy consumption of an air handling unit with a computational intelligence approach. Energy and Buildings 2013; 60:355-363.

[21] Hong T, Koo C, Park S. A decision support model for improving a multi-family housing complex based on CO2 emission from gas energy consumption. Building and Environment 2012; 52:142–151.

[22] Tang F, Kusiak A, Wei X. Modeling and short-term prediction of HVAC system with a clustering algorithm. Energy and Buildings 2014; 82:310-321.

[23] Yu Z, Fung BCM, Haghighat F, Yoshino H, Morofsky E. A systematic procedure to study the influence of occupant behavior on building energy consumption. Energy and Buildings 2011; 43(6):1409-1417.

[24] Yu Z, Haghighat F, Fung BCM, Zhou L. A novel methodology for knowledge discovery through mining associations between building operational data. Energy and Buildings 2012; 47:430-440.

[25] Luhr S, West G, Venkatesh S. Recognition of emergent human behavior in a smart home: A data mining approach. Pervasive and Mobile Computing 2007; 3(2):95-116.

[26] Dong B, Andrews B. Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. Building Simulation Conference 2009.

[27] Wu S, Clements-Croome D. Understanding the indoor environment through mining sensory data—A case study. Energy and Buildings 2007; 39:1183-1191.

[28] Motta-Cabrera D, Zareipour H. Data association mining for identifying lighting energy waste patterns in educational institutes. Energy and Buildings 2013; 62:210-216.

[29] D'Oca S, Hong T. A data-mining approach to discover patterns of window opening and closing behavior in offices. Building and Environment 2014; 82:726-739.

[30] D'Oca S. Hong T. Occupancy schedules learning process through a data mining framework. Energy and Buildings 2015; 88:395-408.

[31] Zhao J, Lasternas B, Lam KP, Yun R, Loftness V. Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. Energy and Buildings 2014; 82:341-355.

[32] ANSI/ASHRAE/IESNA 90.2 Standard, Energy Efficient Design of Low-Rise Residential Buildings. 2007.

[33] Urban B, Gomez C. A case for thermostat user model. 13th Conference of International Building Performance Simulation Association. 2013.

[34] Urban B et al. Improving Programmable Thermostats with Less Aggressive Default Setbacks, BECC, 2013.

[35] Fayyar UM, G. Piatetsky-Shapiro G, P. Smyth P, R. Uthurusamy R. Advances in knowledge discovery and data mining. American Association for Artificial Intelligence, 1996.

[36] Quinlan JR. Simplifying decision trees. International Journal of Man- Machine Studies 1987; 27(3):221-234.

[37] Bailey K. Numerical Taxonomy and Cluster Analysis. Typologies and Taxonomies 1994; Page:34.

[38] Independent comparative study by the Chemnitz University of Technology, The International Data Mining Cup. 2007.

[39] Gu H. Analysis and application of k-value optimization based on K-means algorithm. Natural Science Journal of Hainan University 2009; 27(4). (in Chinese).

[40] Zalik KR, Zalik B. Validity index for clusters of different sizes and densities. Pattern Recognition Letters 2011; 32(2):221-234.