2018

A two-level clustering strategy for energy performance evaluation of university buildings

Kehua Li
University of Wollongong, kl751@uowmail.edu.au

Zhenjun Ma
University of Wollongong, zhenjun@uow.edu.au

Duane A. Robinson
University of Wollongong, duane@uow.edu.au

Jun Ma
University of Wollongong, jma@uow.edu.au

Publication Details

Li, K., Ma, Z., Robinson, D. & Ma, J. (2018). A two-level clustering strategy for energy performance evaluation of university buildings. In K. Inthavong, C. P. Cheung, G. Yeoh & J. Tu (Eds.), Proceedings of the 4th International Conference on Building Energy & Environment (pp. 168-173). Melbourne, Australia: Conference On Building Energy & Environment.

A two-level clustering strategy for energy performance evaluation of university buildings

Abstract

This paper presents a clustering strategy to evaluate the energy performance and identify typical daily load profiles of buildings. The cluster analysis included intra-building clustering and inter-building clustering. The intra-building clustering used Gaussian mixture model clustering to identify the typical daily load profiles of each individual building. The inter-building clustering used hierarchical clustering to further identify the typical daily load profiles of a stock of buildings based on the typical daily load profiles identified for each individual building. The performance of this strategy was tested and evaluated using the two-year hourly electricity consumption data collected from 40 buildings on a university campus in Australia. The result showed that this strategy could discover the information related to building energy usage. The results obtained from this study could be potentially used to assist in decision making for energy performance enhancement initiatives of university buildings.

Keywords

clustering, two-level, strategy, performance, buildings, energy, evaluation, university

Disciplines

Engineering | Science and Technology Studies

Publication Details

Li, K., Ma, Z., Robinson, D. & Ma, J. (2018). A two-level clustering strategy for energy performance evaluation of university buildings. In K. Inthavong, C. P. Cheung, G. Yeoh & J. Tu (Eds.), Proceedings of the 4th International Conference on Building Energy & Environment (pp. 168-173). Melbourne, Australia: Conference On Building Energy & Environment.

A Two-level Clustering Strategy for Energy Performance Evaluation of University Buildings

K. Li [1], Z. Ma [1], D. Robinson [1] and J. Ma [2]

[1]Sustainable Buildings Research Centre
University of Wollongong, Wollongong, New South Wales 2522, Australia

[2]SMART Infrastructure Facility
University of Wollongong, Wollongong, New South Wales 2522, Australia

SUMMARY

This paper presents a clustering strategy to evaluate the energy performance and identify typical daily load profiles of buildings. The cluster analysis included intra-building clustering and inter-building clustering. The intra-building clustering used Gaussian mixture model clustering to identify the typical daily load profiles of each individual building. The inter-building clustering used hierarchical clustering to further identify the typical daily load profiles of a stock of buildings based on the typical daily load profiles identified for each individual building. The performance of this strategy was tested and evaluated using the two-year hourly electricity consumption data collected from 40 buildings on a university campus in Australia. The result showed that this strategy could discover the information related to building energy usage. The results obtained from this study could be potentially used to assist in decision making for energy performance enhancement initiatives of university buildings.

INTRODUCTION

With the improvement of living standards, the energy consumption of buildings has increased dramatically over the last several decades. In order to reduce building energy consumption, it is essential to have a better understanding of building energy usage behaviours. University buildings generally have high energy consumption (Chung and Rhee 2014) and are receiving increasing attention for improved energy efficiency as institutions look to reduce their energy related emissions to demonstrate sustainability initiative or comply with mandated reporting.

A number of previous studies have investigated energy consumption characteristics of university buildings. Gul and Patidar (2015), for instance, analysed the relationship between the electricity demand and the user activities in a multi-functional university building. Davis and Nutter (2010) studied the daily occupancy profiles of six types of university buildings and derived the occupancy schedules of these buildings which were then used to enhance building energy simulation. Scheuer et al. (2003) analysed the environmental impact of a university building in terms of both operating energy usage and whole life cycle energy consumption. Guan et al. (2016) analysed the features of demand load and energy consumption of university buildings in order to support the energy planning from the demand side.

Identification of typical building load profiles has been considered as an effective way to assist in understanding building energy consumption characteristics and helping the development of cost effective strategies for load shifting and peak demand control. Cluster analysis, which can find groups of objects to ensure that the objects in the same group are similar to each other but are different from those in other groups (Han et al. 2011), has been used for the identification of building typical load profiles. Miller et al. (2015) proposed a method called DayFilter to detect the underlying information from building performance data and sub-system metrics to identify potential areas for energy savings. In this strategy, the building daily load profiles were first transformed into the strings using Symbolic Aggregate approximation (SAX) and the typical daily load profiles were then identified using k-means clustering method. do Carmo and Christensen (2016) analysed the heating load profile of 139 dwellings using k-means clustering algorithm. A binary regression analysis was also performed to explore whether the difference in heating load profiles between the clusters can be attributed to building characteristics. A cluster analysis strategy to identify typical building daily load profiles based on the variation similarity of the load profiles was presented by Ma et al. (2017). The performance test of this strategy based on hourly heating energy data of 19 university buildings showed that the identified typical heating load profiles can provide information such as the peaks and troughs of the daily heating demand, daily high heating demand period and daily load variation, which can hardly be revealed with the strategy that only focused on the magnitude of the load profiles.

Different from the previous research, a two-level clustering strategy was developed in this study to identify typical daily load profiles of a stock of university buildings. Gaussian mixture model (GMM) clustering, which works well with the profiles that are complex and varied considerably, was used to identify the typical daily electricity usage profiles of each individual building and a hierarchy clustering was then used to further identify the typical daily load profiles of a stock of buildings based on the typical daily load profiles identified for each individual building. The performance of this strategy was tested and evaluated using the two-year hourly electricity consumption data collected from 40 buildings on a university campus in Australia.

STRATEGY DEVELOPMENT

Outline of the proposed strategy

The outline of the proposed clustering based strategy to examine the building energy performance is presented in Figure 1. It consisted of the intra-building clustering, inter-building clustering and a result evaluation and interpretation. Intra-building clustering is used to identify the typical daily load profiles of each individual building and remove the outliers in the daily load profiles. The time series building data was first converted into hourly electricity usage per unit floor area and segmented into daily load profiles. After removal of the daily load profiles with missing data, a Gaussian mixture model based cluster analysis was then used to cluster the daily load profiles such that the profiles in the same group are similar but different from those in other groups. The typical daily load profile in the cluster was identified using the method to be introduced in the section "Gaussian mixture model clustering" to represent all daily load profiles in this cluster.

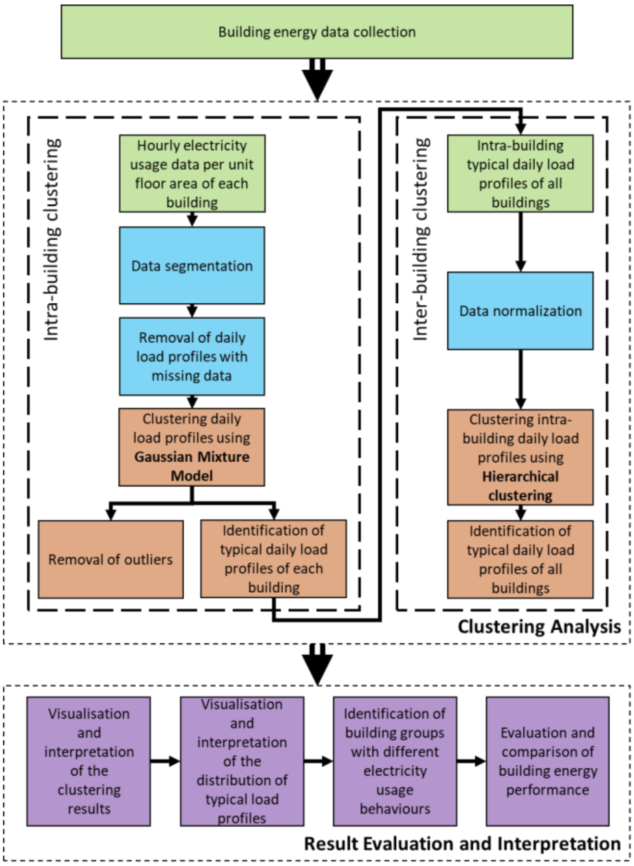

Figure 1 Outline of the proposed strategy.

In the inter-building clustering, all the typical load profiles identified for each individual building were normalized through rescaling so that the new profiles had a mean value of 0 and variance of 1 (Han 2011). Then, the Euclidean distance between each pair of the normalized profiles was calculated to determine the dissimilarity measure. A hierarchical clustering technique was then used to determine the structure and the number of the clusters. The typical daily load profiles for a stock of buildings were then formed by calculating the mean value of all the profiles in each cluster.

The results from the cluster analysis were then evaluated and interpreted to provide an overall understanding of the building energy performance and energy usage behaviours. The distribution of the typical daily load profiles was plotted as a calendar view to better understand the temporal distribution of the typical daily load profiles identified through intra-building clustering.

Gaussian mixture model-based clustering

A GMM is a weighted combination of several normal distributions which called mixture components (Figueiredo 2002). A GMM fitted with a dataset x can be presented as:

$$\begin{cases} f(x) = \sum_{i=1}^{K} \pi_i \phi(x;\ \mu_i, \sigma_i) \\ \phi(x;\ \mu_i, \sigma_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp(-\frac{(x-\mu_i)^2}{2\sigma_i^2}) \\ \sum_{i=1}^{K} \pi_i = 1 \end{cases} \quad (1)$$

where K is the number of mixture components, ϕ is the Gaussian density probability function, π_i, μ_i and σ_i are the weight, mean and standard deviation of the i^{th} mixture component, respectively.

Gaussian mixture model clustering is a technique to group a large dataset into different clusters based on GMM and has been used in medical imaging (Labeeuw and Deconinck 2013), residential electricity load modelling (Noe and Gee 2011) and microarray expression analysis (McLachlan et al. 2002).

To conduct a GMM-based clustering, a GMM with K mixture components was first fitted with the data points. In this study, each data point was one daily load profile. Once the GMM had fitted, a clustering of the daily load profiles into K clusters can be obtained in terms of the fitted probabilities of belonging to each mixture component for the daily load profiles. An outright assignment of the daily load profiles into K clusters was achieved by assigning each daily load profile to the component to which it has the highest estimated probability of belonging (McLachlan et al. 2002). The algorithm used to fit the GMM is the expectation-maximization (EM) algorithm. Another key task involved in the GMM approach to clustering is to determine the optimal number of components, K. In this study, Bayesian Information Criterion (BIC) was used for this purpose. The number of components that minimizes BIC was chosen as the number of components for the Gaussian mixture model. The mean vector of each mixture component is identified as the typical daily load profile (Singh 2010). The daily load profiles that have too small probability in all groups are classified as outliers.

Hierarchical Clustering

The inter-building clustering was achieved using an agglomerative hierarchical clustering (Han 2011). The agglomerative hierarchical clustering is a bottom-up strategy, which starts with treating each object as a separate cluster and then merges the atomic clusters into larger clusters until all objects are in a single cluster. Dissimilarity measure and linkage criteria are two important components of hierarchical clustering algorithm. In this study, the dissimilarity measure used was Euclidean Distance (ED) and the linkage criterion used was Ward's method, in which two clusters should be merged if the merge can minimize the increase in the sum of squared error. The advantages of the hierarchical clustering are that the number of clusters can be determined during the clustering process and the overall process can be represented by a tree structure graph called dendrogram, which can help to visualise the cluster structure and assist in determining the optimal number of clusters.

VALIDATION OF THE PROPOSED STRATEGY

The hourly electricity usage data of 40 buildings from the University of Wollongong (Figure 2) in Australia were collected from 2014 to 2015. The 2-year data were used to test and evaluate the performance of this proposed strategy. Table 1 summarises the building main functions and floor areas as well as the mean hourly electricity usage per square meter. The functions of the buildings included office, education room, laboratory, sports centre, student accommodation and common area such as study area/social area. From Table 1, it can be seen that the electricity consumption among different buildings varied considerably, even for the buildings with similar functions.

Figure 2 Aerial photo of University of Wollongong, Australia.

Table 1 Building information and electricity usage

No.	Main function	Building Area (m^2)	Mean unit hourly electricity usage (kWh/m^2)	Number of clusters identified
#1	O/L	5,376	0.1396	3
#2	O/L	5,439	0.0107	2
#3	O/L	13,567	0.0085	2
#4	O/L	2,349	0.0155	2
#5	O/L	3,622	0.0192	2
#6	O/L	4,071	0.0323	2
#7	O/L	1,812	0.0112	3
#8	O/L	14,725	0.0170	2
#9	O/L	889	0.0111	3
#10	O/L	7676	0.0040	5
#11	O	11,618	0.0019	2
#12	O	1,027	0.0020	3
#13	O	1,645	0.0076	2
#14	O	6,999	0.0183	3
#15	O	3414	0.0175	2
#16	O	568	0.0147	3
#17	O	11,876	0.0117	3
#18	O/E/L	2,143	0.0139	2
#19	O/E/L	6,779	0.0058	3
#20	O/E/L	2,573	0.0144	2
#21	O/E/L	3,191	0.0156	3
#22	O/E/L	5,075	0.0252	2
#23	O/E	8,345	0.0069	2
#24	O/E	4,716	0.0107	3
#25	O/E	7,374	0.0103	2
#26	O/E	6,748	0.0157	3
#27	L	1,247	0.0051	2
#28	L	445	0.0236	4
#29	L	986	0.0477	2
#30	E	1,342	0.0134	3
#31	E	380	0.0092	4
#32	E	2,198	0.0094	3
#33	S	9,384	0.0003	2
#34	S	5,801	0.0191	2
#35	C	26,125	0.0058	4
#36	C	829	0.0244	2
#37	L/C	14,874	0.0189	2
#38	L/C	5,369	0.0382	2
#39	A	5,184	0.0132	3
#40	A	15,551	0.0058	3

O: office; E: educational room; L: laboratory; C: common area; S: sports centre; A: student accommodation.

Intra-building clustering

The time series electricity usage data of each building were first processed into the electricity usage per unit square meter and then segmented to daily load profiles, and the daily profiles with missing data were considered as outliers and removed in the following study. After the completion of the data pre-processing, an average of 695 daily load profiles remained for each building.

Then, a GMM was fitted with the daily profiles of each building to identify the typical daily load profiles and outliers. The number of typical daily load profiles (i.e. clusters) identified for each building is presented in Table 1. It can be seen that more than 50% of the buildings had two typical daily load profiles and there were five clusters of Building #10.

Two buildings were selected as representatives to illustrate the intra-building clustering results. Figure 3 illustrates the typical daily load profiles identified for Building #3, which is mainly used for offices and laboratories. The red curves represented the typical daily load profile identified while the gray curves were all corresponding daily load profiles in that cluster. It can be seen that there were two typical load profiles with 226 and 480 daily load profiles respectively that were identified for this building and eight daily load profiles were considered as the outliers. In the typical daily load profile 2 (i.e. cluster 2), it can be seen that there was a clear high energy consumption period (8:00 to 17:00) during the working hours and a low energy consumption period during the rest of the day while such information cannot be observed in the cluster 1.

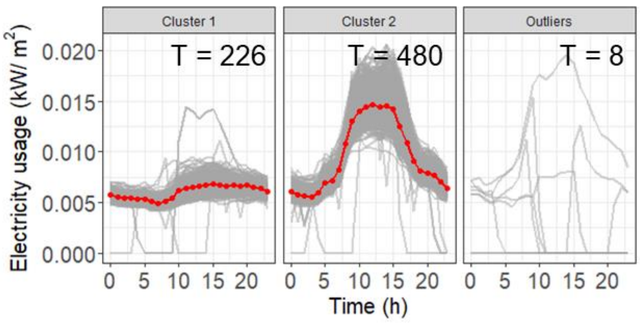

Figure 3 The typical daily load profiles and outliers of Building #3.

Figure 4 shows the distribution of the typical daily load profiles in a calendar view, in which the white blocks represented the days with missing data that were removed during the data pre-processing. It is shown that the typical daily load profile 2 represented the daily load profiles of weekdays and the typical daily load profile 1 mainly appeared on weekends and certain public holidays such as Australia Day, Easter and Labour Day.

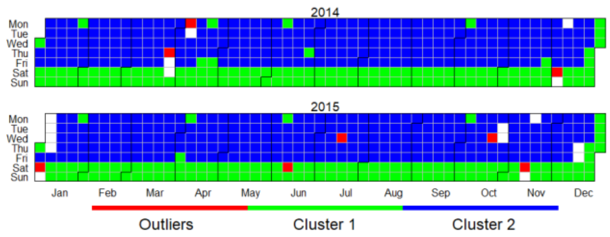

Figure 4 The calendar view of the distribution of typical daily load profiles of Building #3.

Figures 5 and 6 present the clustering result of Building #40 which is a student accommodation. There were three typical daily load profiles formed for this building and three daily load profiles were considered as the outliers. The typical daily load profile 2 represented the energy usage behavior during the semesters, in which the electricity usage started to increase at the early morning of 5.00 and experienced a small peak at 12:00. The electricity usage reached the peak at around 18:00. After that, the electricity usage decreased until the morning of the next day. The typical daily load profile 3 mainly represented the energy usage behavior during the winter holidays and session breaks. Since a considerable number of students stayed at the university during these short holidays, the trend of the typical daily load profile 3 was similar with the typical daily load profile 2 while the magnitude was smaller. The typical daily load profile 1 showed the electricity usage behavior during the summer holidays. The electricity consumption during this time period was small and relatively stable as there were only a limited number of students stayed in the building.

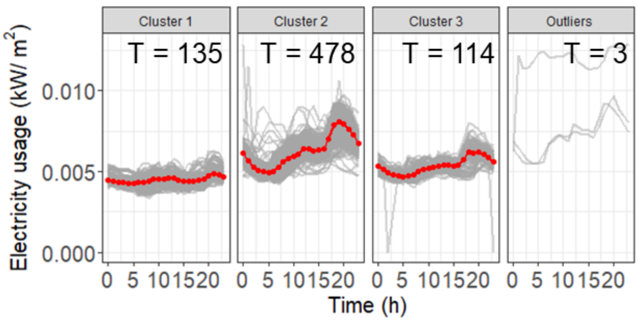

Figure 5 The typical daily load profiles and outliers of Building #40.

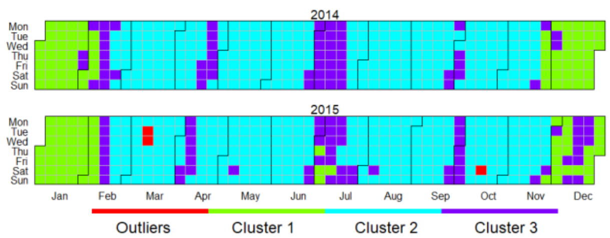

Figure 6 The distribution of typical daily load profiles of Building #40.

According to the distribution of the typical daily load profiles of each building, the buildings were classified into 4 groups. In the first group, the buildings had only one typical daily load profile during weekdays. In the second group, there were 2 typical daily load profiles which presented the daily load profiles for winter/summer period and the other time period, respectively. In the third group, there were also two typical daily load profiles but representing the electricity usage during holidays/session breaks and the session time, respectively. In the last group, there was no clear pattern for the distribution of the typical daily load profiles. The percentage of the number of the buildings in each group to the total number of the buildings with the same function is presented in Figure 7. For instance, two student accommodations were considered in this study. One of them was classified into group 2 and the other was in the group 3. Their proportion was therefore 50% each. It can be seen that most buildings had only one typical daily load profile during weekdays. The buildings used for educational rooms or offices tended to have different daily load profiles during the summer and winter seasons. The buildings used for laboratory/common areas had different daily load profiles during semesters and holidays. It is interesting to note that all laboratory buildings had no obvious pattern in terms of the distribution of typical daily load profiles, which reflected the complexity and large variation in energy consumption in such buildings.

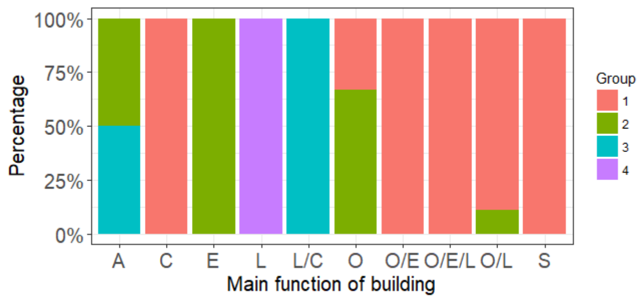

Figure 7 The distribution of the buildings with different functions.into the four groups defined.

Inter-building clustering

Through the intra-building clustering analysis, a total of 101 typical daily load profiles were identified for all 40 buildings. Since the energy consumption for university buildings mainly occurs during the weekdays, the typical daily load profiles identified for weekends were removed in the inter-building clustering analysis.

Figure 8 presents the dendrogram of the hierarchical clustering result. It can be seen that six clusters were formed when the threshold was selected as 4.5 in order to have relatively distinctive clusters while avoiding clusters with too few typical daily load profiles.

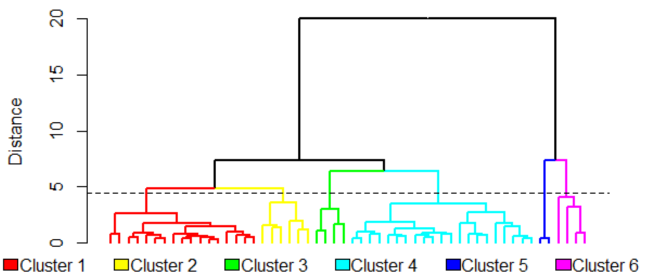

Figure 8 The dendrogram of the hierarchical clustering result.

Figure 9 shows the typical daily load profiles of all 40 buildings through inter-building clustering analysis. In the typical daily load profile 1, the boundary between working hours and off-working hours was clear and the large electricity demand occurred from 8:00 to 17:00. The typical daily load profile 3 also had a clear boundary between the working hours and off-working hours but with a longer large electricity demand period (from 6:00 to 21:00) than that of the typical daily load profile 1. In the typical daily load profiles 2 and 4, the peak demand occurred at around 12.00 but the boundary between working hour and off-working hours was unclear at morning and late afternoon, respectively. The typical daily load profiles 5 and 6 were identified in the student accommodation. Both shared the similar shape with the highest peak demand at night and the secondary peak at around midday. However, the peaks occurred at different time with different variation trends.

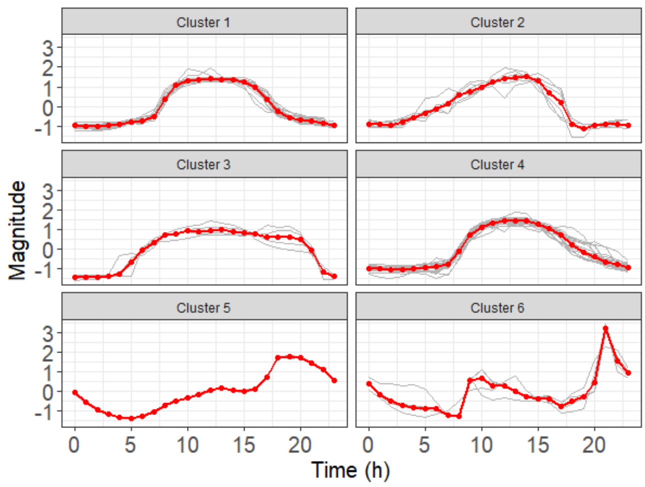

Figure 9 The typical daily load profiles identified through hierarchical clustering for all 40 buildings.

The percentage of the total number of daily load profiles of all buildings that represented by each inter-building typical daily load profile to the total number of daily load profiles of all buildings with the same function is presented in Figure 10. It can be seen that the typical daily load profile 4 accounted for a large part of the daily profiles especially for the buildings used for education rooms, common areas and laboratories. For the education rooms and common areas, the high proportion of typical daily load profiles 4 were possibly resulted by lecture/tutorial and other student activities at night. For the laboratory, the high proportion of typical daily load profiles 4 were probably due to overtime working or operating of laboratory equipment during off-working hours. The typical daily load profile 1 mainly existed in the buildings used for the office rooms. All daily load profiles of sports centres were in the typical daily load profile 3 due to its unique operating characteristics. It should be noted that some typical daily load profiles accounted for a small part of all daily load profiles. For instance, the typical daily load profile 3 in the buildings used for common areas and the typical daily load profile 4 in the buildings used for offices, which probably due to abnormal energy behaviours and might be worthwhile to investigate.

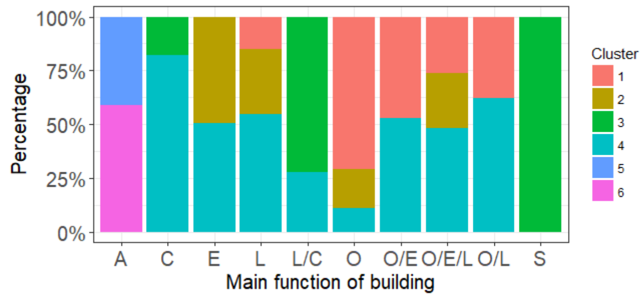

Figure 7 The distribution of typical daily load profiles in the buildings with different functions.

According to Figure 7 and 10, it can be observed that various patterns of the electricity usage could even exist in the buildings with the similar function. In this study, the buildings used for the similar functions were further classified into sub-groups based on the distribution of intra-building typical daily load profiles and the inter-building typical daily load profiles identified. The buildings used for offices and laboratories, for example, can be classified into 3 sub-groups as shown in Table 2.

Table 2 The characteristics of the three sub-groups defined for O/L buildings

Sub-group	Distribution of intra-building typical daily load profiles	Major inter-building typical daily load profile
1	one typical daily load profile in weekdays	typical daily load profile 1
2	one typical daily load profile in weekdays	typical daily load profile 4
3	second typical daily load profile in the heating/cooling period	typical daily load profile 4

Figure 11 shows the mean electricity consumption of O/L buildings that belonged to the above three sub-groups. The size of the bubbles represented the floor area of each building. It can be seen that the electricity consumption of Building #6 was much higher than the other buildings in the same sub-group, which indicated that this building might be worthwhile to further investigate for potential energy savings. The Building #8 was also worthwhile to investigate due to the same reason. Therefore, those in the sub-group with high energy consumption should be first considered in energy efficiency enhancement initiatives.

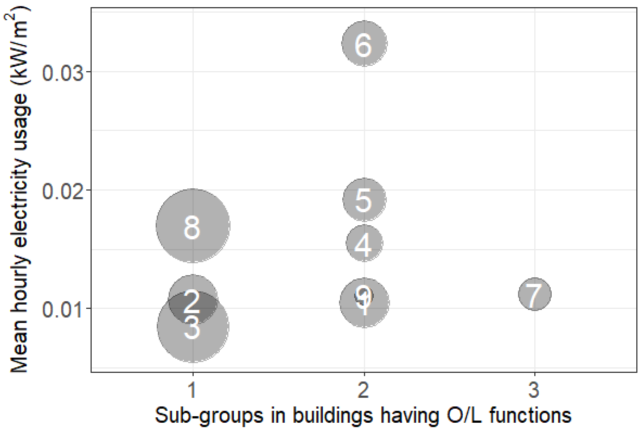

Figure 8 The energy consumption of the buildings used as offices and laboratories.

CONCLUSIONS

This paper presented a two-level clustering strategy to identify building typical daily load profiles and evaluate the energy performance of multiple university buildings. In this strategy, the typical daily load profiles of each building were first identified using Gaussian mixture model-based clustering. The identified typical daily load profiles for all individual buildings were then further clustered using a hierarchical clustering with Euclidean distance as the dissimilarity measure.

The performance of this strategy was evaluated using two-year electricity usage data collected from 40 buildings on a university campus in Australia. The result demonstrated that this strategy can discover the information related to the energy usage behaviors of the buildings. The discovered information helped to understand the building typical energy usage patterns. In the inter-building clustering, six clusters were identified to represent the major electricity usage behaviours of all 40 buildings of concern. The relationship between the major electricity usage behaviour and the function of the building was also analysed. The results from this study can be potentially used to support the decision making for energy efficiency enhancement initiatives.

REFERENCES

Chung, M. H., & Rhee, E. K. (2014). Potential opportunities for energy conservation in existing buildings on university campus: A field survey in Korea. Energy and Buildings, 78, 176-182.

Gul, M. S., & Patidar, S. (2015). Understanding the energy consumption and occupancy of a multi-purpose academic building. Energy and Buildings, 87, 155-165.

Davis, J. A., & Nutter, D. W. (2010). Occupancy diversity factors for common university building types. Energy and buildings, 42(9), 1543-1551.

Scheuer, C., Keoleian, G. A., & Reppe, P. (2003). Life cycle energy and environmental performance of a new university building: modeling challenges and design implications. Energy and buildings, 35(10), 1049-1064.

Guan, J., Nord, N., & Chen, S. (2016). Energy planning of university campus building complex: Energy usage and coincidental analysis of individual buildings with a case study. Energy and Buildings, 124, 99-111.

Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

do Carmo, C. M. R., & Christensen, T. H. (2016). Cluster analysis of residential heat load profiles and the role of technical and household characteristics. Energy and Buildings, 125, 171-180.

Miller, C., Nagy, Z., & Schlueter, A. (2015). Automated daily pattern filtering of measured building performance data. Automation in Construction, 49, 1-17.

Ma, Z., Yan, R., & Nord, N. (2017). A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings. Energy.

Figueiredo, M. A. T., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. IEEE Transactions on pattern analysis and machine intelligence, 24(3), 381-396.

Labeeuw, W., & Deconinck, G. (2013). Residential electrical load model based on mixture model clustering and Markov models. IEEE Transactions on Industrial Informatics, 9(3), 1561-1569.

Noe, A., & Gee, J. C. (2001, June). Partial volume segmentation of cerebral MRI scans with mixture model clustering. In IPMI (pp. 423-430).

McLachlan, G. J., Bean, R. W., & Peel, D. (2002). A mixture model-based approach to the clustering of microarray expression data. Bioinformatics, 18(3), 413-422.

Singh, R., Pal, B. C., & Jabr, R. A. (2010). Statistical representation of distribution system loads using Gaussian mixture model. IEEE Transactions on Power Systems, 25(1), 29-37.