Energy Proceedings
Vol 55, 2025
ISSN 2004-2965

Mapping Heat Electrification to Socioeconomic Segments of Consumers for Net-Zero Energy Transition#

Shangyuan Liu[1], Weiqi Hua[1]

1 School of Engineering, University of Birmingham, Birmingham, UK

(Corresponding Author: sxl1403@student.bham.ac.uk)

ABSTRACT

The decarbonisation of the built environment is a critical strategy for addressing climate change and decarbonisation of heat plays a significant role in this process. However, the UK's progress in electrifying heating with heat pumps is significantly behind that of other European countries such as Finland and Norway. This study used clustering analysis to investigate UK consumers' energy consumption patterns and their correlation with socioeconomic characteristics to identify suitable households for heat pumps. The study optimised K-means outlier removal (K-MOR) using genetic algorithms (GA) to reveal five typical energy consumption patterns consistent with a classification of residential neighbourhoods (ACORN) socioeconomic segments. The findings of our study indicate that the highest energy consumption pattern requires about 3 times the heating demand of the lowest pattern. Notably, 50.9% of households exhibit a middle-high load pattern, among which affluent households demonstrate higher heat pump adoption potential, while 14.44% of lower-income households face greater barriers to heat decarbonisation.

Keywords: Clustering Analysis, Energy Consumption Patterns, Electrification of Heat, Heat Pump, Socioeconomic Factors.

1. INTRODUCTION

The UK government has introduced a series of strategies and roadmaps for decarbonisation to achieve the net-zero emissions target by 2050 [1]. Heat pumps have been highlighted as a key measure of electrification of heating for net-zero transitions. However, current adoption rates are considerably low compared to EU countries, with only 2.13 heat pumps sold per 1,000 households in 2022 [2]. This is primarily attributed to high installation costs and elevated electricity prices. While the government has introduced various schemes and grants, there remains a crucial need for precise targeting of these incentives to households and communities requiring heating system upgrades. In this context, data-driven technologies are essential for analysing energy consumption data and locating regions and households that need support.

Data-driven technologies are able to process large amounts of data quickly and use machine learning methods to analyse and reveal complex energy consumption patterns. Previous studies employed various clustering algorithms including K-Means [3], DTW [4], spectral clustering [5], PAM, [6] and Markov cluster [7] to analyse electricity and gas load data, revealing energy usage trends across user groups. The Z-Score [8] method was used frequently to clean abnormal load data, but it is not valid for non-Gaussian-distributed data. Silhouette coefficients, Calinski-Harabasz index, and Davies-Bouldin index evaluation metrics were used to evaluate the performance of clustering algorithms [9].

Gaur et al. [10] emphasised that developing supportive policies for different consumer groups is crucial for the large-scale deployment of heat pumps. Few et al. [11] also discussed the impact of geographical regions and socioeconomic conditions on cost-effective heat pump deployment strategies. Therefore, many studies often overlook important socioeconomic factors that play a key role in determining the economic capability of consumers and the feasibility of the widespread adoption of heat pumps.

To address these challenges, this paper presented a clustering analysis approach based on a classification of residential neighbourhoods (ACORN) data and Energy Demand Research Project (EDRP) data [12] to map heat electrification to socioeconomic segments of consumers. Our work offers two primary contributions that distinguish it from existing research:

- This study proposed an adaptive load profile clustering method using a genetic algorithm (GA) to optimise K-means outlier removal (K-MOR) [13]. Compared with existing methods, it effectively

This is a paper for the 16th International Conference on Applied Energy (ICAE2024), Sep. 1-5, 2024, Niigata, Japan.

mitigated the impact of anomalous load profiles on the clustering results while avoiding local optimum problems to reflect household electricity and gas demand differences.

- Using the Hungarian algorithm and ACORN classifications, this study identified five typical energy load patterns and investigated socioeconomic impacts on energy load patterns. Results indicate that 14.44% of households have low incomes but high energy use, making them reliant on grants to install heat pumps.

The paper is structured as follows. Section 2 describes the used data and clustering algorithm as well as the approach of mapping socioeconomic segments to the clustering results. Section 3 presents the results and analysis. Finally, the conclusion is drawn in Section 4.

2. METHODOLOGY

This section introduces the proposed consumer clustering framework that primarily targets residential customers in Great Britain (GB), categorising households by energy demand levels. It also investigates how socioeconomic status affects adoption of heat pumps.

2.1 Data description and processing

EDRP is a comprehensive residential energy consumption data collection initiative undertaken in GB, which gathered half-hourly electricity and gas demand data from over 60,000 households. This study focused on households that use dual fuels (both electricity and gas) during the winter months (December, January and February), which are characterised by higher energy demand to capture typical consumption patterns. Missing values and samples with fewer than 48 half-hourly recordings per day were excluded, remaining 8,103 households. Then, Min-Max normalisation was used to ensure the data is in the same scale and facilitating faster convergence for the algorithm [14].

2.2 Clustering algorithms

Based on analyses of consumption patterns, this study used five clusters to categorise consumers into distinct groups characterised by daily energy demand profiles. Five clusters ranging from low to high usage patterns in energy consumption levels could effectively distinguish the inherent variations in different residential energy consumption behaviours.

The presence of outliers would affect clustering performance. Therefore, the K-MOR algorithm was employed to mitigate the impact of outlier loads on clustering performance. In addition, we proposed integrating GA's adaptive search strategy to optimise clustering centroids, mitigating the sensitivity of K-Means to initial centroids and the problem of premature local convergence.

The objective of the K-Means is to minimise the sum of the squared distances between data points and their corresponding cluster centroids as

$$J = \min_{u_{i,l}, \boldsymbol{c}_i} \sum_{i=1}^{n} \sum_{l=1}^{k} u_{i,l} \|\boldsymbol{x}_i - \mathbf{c}_l\|^2 \ , \tag{1}$$

where $\|\boldsymbol{x}_i - \boldsymbol{c}_l\|^2$ represents the distance between the data point $\boldsymbol{x}_i$ and geometric centroid of cluster $\mathbf{c}_l$, $u_{i,l}$ is a binary variable that describes the relationship between data points and clusters, $l \in \{1, 2, \ldots, \mathrm{k}\}$ represents the index of the cluster to which the data point $\boldsymbol{x}_i$ is assigned, and $\|\cdot\|$ is the Euclidean norm.

GA optimised the K-Means algorithm by iteratively evolving cluster centroids using selection, crossover, and mutation to refine clustering solutions. The pseudo-code of the above operations is given in **Algorithm 1**.

Algorithm1: Pseudo-Code of Genetic K-Means Clustering

Input: $\boldsymbol{X}$ (dataset), k (number of clusters), *N* (population size), *G* (number of generations), $\boldsymbol{\delta}$ (Mutation vector)

Output: Optimal cluster centroids

Initialise population $\mathbf{C}_{\text{current}} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_N\}$
for generation to G **do**
  for each $\mathbf{C}_i$ in $\mathbf{C}_{\text{current}}$ **do**
    Compute fitness $f(\mathbf{C}_i)$
  Select parent centroids ($\mathbf{C}_{\text{parents}}$) based on $f(\mathbf{C}_i)$
  Generate offspring ($\mathbf{C}_{\text{offspring}}$) by crossover
  Perform mutation on offspring using $\boldsymbol{\delta}_i$
  $\mathbf{C}_{\text{nextPopulation}} = \mathbf{C}_{\text{parents}} \cup \mathbf{C}_{\text{offspring}}$
end

Step 1: Fitness serves as the criterion for distinguishing the quality of individuals in the population. Individuals with higher fitness have a greater likelihood of survival. For a cluster centroid configuration denoted as $\mathbf{C}_i$, which is a matrix consisting of k cluster centroids, the fitness function is designed as shown in Eq. (2) [15].

$$f(\mathbf{C}_i) = \frac{1}{1+J} \ . \tag{2}$$

Here, constructing the fitness function based on the objective function J of K-Means. Since a smaller distance and J value indicate the larger the fitness value and better clustering quality.

Step 2: According to the fitness value select a portion of individuals as parents. The probability of selecting the ith set of cluster centroids (chromosome) is

$$P(\mathbf{C}_i) = \frac{f(\mathbf{C}_i)}{\sum_{j=1}^{N} f\left(\mathbf{C}_j\right)}, \tag{3}$$

where N is the population size [15].

Step 3: Crossover combines parts of two parent chromosomes to produce offspring ($\mathbf{C}_{\text{offspring}}$)

$$\mathbf{C}_{\text{offspring}} = \{\boldsymbol{c}_{p1}, \dots, \boldsymbol{c}_{pm}, \boldsymbol{c}_{q(m+1)}, \dots, \boldsymbol{c}_{qk}\}, \quad (4)$$

where $\boldsymbol{c}_{pi}$ and $\boldsymbol{c}_{qi}$ are vectors of size d, which represent cluster centroids in a d-dimensional feature space; k is the number of clusters; m is the crossover point that decides the split between the parents.

Step 4: Mutation operations are performed by introducing small random perturbations to the positions of cluster centroids. For a selected mutation cluster centroids denoted as $\mathbf{C}_i$, and we define $\mathbf{C}_i'$ as the new cluster centroid after the mutation operation, which can be expressed as:

$$\mathbf{C}_i' = \mathbf{C}_i + \boldsymbol{\delta}\,, \quad (5)$$

where $\boldsymbol{\delta}$ is small random perturbation vector of size d.

Step 5: The current population is updated by combining the selected parents and the generated offspring.

Repeat steps 1-5 to improve cluster centers and find a better solution until reaching the defined maximum number of iterations, typically set as 100.

2.3 Mapping to socioeconomic data

To define five typical load patterns based on consumption levels, the Hungarian algorithm [16] was used to match electricity and gas load clusters while minimising the cost matrix ($\mathbf{C}_{\text{cost}}$) that has a size of $k \times k$, k is the number of clusters. It can be represented as

$$\mathbf{C}_{\text{cost}} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1k} \\ d_{21} & d_{22} & \dots & d_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ d_{k1} & d_{k2} & \dots & d_{kk} \end{bmatrix} \quad (6)$$

where elements d_{ij} represent the distance between electricity load cluster i and gas load cluster j. The clusters were matched by minimising the distance value in the $\mathbf{C}_{\text{cost}}$, which quantified their similarity. This process results in five typical load patterns.

The geographical and demographic information of EDRP dataset contains basic ACORN categories for each household. Households from the five predefined load patterns were linked to their respective ACORN categories using unique identifiers in both datasets. These categories were initially in numerical codes and were converted to descriptive labels according to the classification system for interpreting socioeconomic segments intuitively. However, the "Not Private Households" category was excluded due to its sample size being significantly smaller than others. In this study, therefore, households were classified into five groups based on the ACORN: Affluent Lifestyles, Mature Prosperity, Comfortable Communities, Social Extension, and Low Income Living.

To quantify the socioeconomic composition of each load pattern, the proportion of each ACORN category was calculated. This was done by dividing the number of households in each ACORN category by the total number of households in the respective load pattern. Additionally, the daily heating demand of 5 patterns was evaluated to highlight differences in energy consumption across socioeconomic segments of consumers.

3. RESULTS AND DISCUSSION

A comparative analysis of our proposed clustering algorithm against GMM, K-Means, DPMM, and DTW clustering methods reveals its superior performance. As illustrated in Fig. 1, when the number of clusters is five, our algorithm consistently outperforms other methods across three key performance metrics.

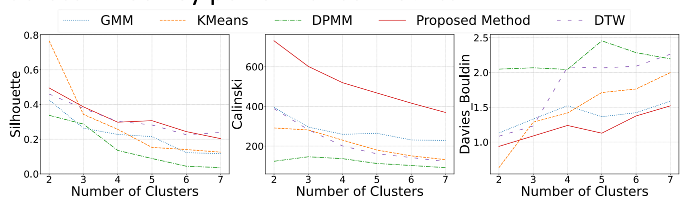

Fig. 1 Evaluation indicators comparison of algorithms

After applying the GA K-Means Clustering algorithm, the daily electricity and gas load consumption was divided into five clusters. The clustering results are shown in Fig. 2 and Fig. 3 where different colours are used to distinguish centroids for electricity and gas clusters. Comparative analysis of these figures shows that the double peaks of gas load profiles display greater intra-day variability and higher peak-to-mean ratios [17] than electricity profiles. Electricity load profiles across most clusters show a consistent feature, the peak primarily occurring at 6 pm.

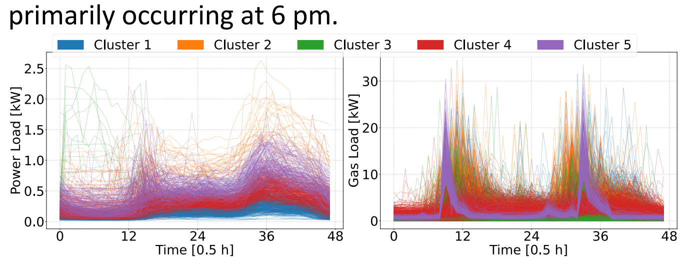

Fig. 2 Clustering result of electricity and gas daily load profiles

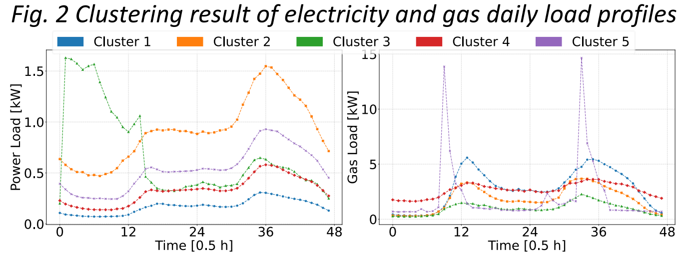

Fig. 3 Cluster centroid profiles of power and gas daily load

In the left plot of Fig. 3, cluster 1 and cluster 4 exhibit similar low electricity demand levels. In contrast, cluster 2 and cluster 3 show higher consumption levels but differ in usage patterns. Cluster 3 shows distinct peaks at night with the highest half-hourly power demand reaching about 1.6 kWh, which may represent households heating with electric radiators with storage. Cluster 5 demonstrates moderate electricity consumption levels and has an average daily demand of 0.53 kWh.

The gas load clustering results in the right plot of Fig. 3, cluster 1 and cluster 2 show moderate gas usage, with two peaks at 6.30 am and 5 pm. Cluster 3 and cluster 4 display low gas usage, and cluster 4 has slight fluctuations. However, cluster 5 has pronounced peaks in the morning and evening, potentially indicating households using advanced smart control systems. These systems may be programmed to maintain a minimum temperature and trigger intense heating when this threshold is reached.

Analysis of household energy consumption patterns reveals complex relationships with socioeconomic segments. Five typical load patterns were obtained by matching the results of electricity and gas load clustering. These patterns stand for various levels of energy consumption, increasing gradually from Pattern 1 to Pattern 5. The daily heating demand for each pattern is formulated by Eq. (6).

$$Q_{\text{Heat}} = \sum_{\text{t}=1}^{48} \left(P_{\text{t}}^{\text{Elec}} \times \eta^{\text{Elec}} + P_{\text{t}}^{\text{Gas}} \times \eta^{\text{GB}} \right) \times \Delta\text{t} \qquad (6)$$

where $P_{\text{t}}^{\text{Elec}}$ and $P_{\text{t}}^{\text{Gas}}$ denote the electricity and gas loads at time t, η^{Elec} and η^{GB} represent the heating efficiency of the electrical heating appliance and gas boiler, with values of 100% and 95% [18]. Fig. 4 reveals the daily heating demand of the highest load pattern is 3.1 times higher than the lowest load pattern.

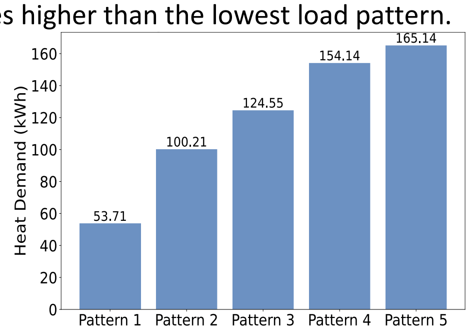

Fig. 4 Daily heating demand of households in five load patterns

Fig. 5 shows the distribution of the 5 socioeconomic segments characteristics in each pattern. High-energy-consuming households (Pattern 4 and Pattern 5) account for 50.9% of the total and are ideal candidates for heat pumps. It is because the efficiency benefits of heat pumps are maximised in the face of constant and substantial heat demand, and such households can achieve long-term energy cost savings through heat pump adoption.

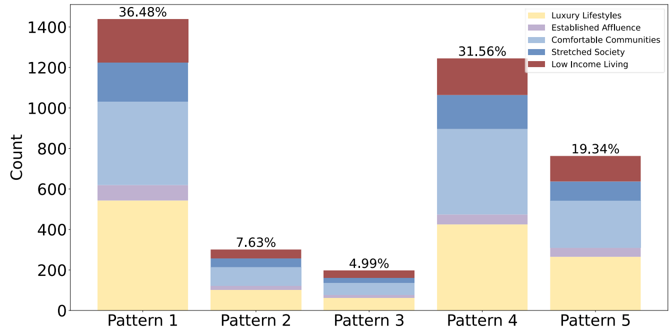

Fig. 5 Mapping the distribution of socioeconomic segments in five load patterns

However, as energy demand increases, the proportion of 'low-income living' households in the corresponding load pattern also gradually rises. Notably, 14.44 % of middle-low-income households exhibit a higher energy demand. In the UK, low-income households have a weekly net income of less than £300 after net of taxes and housing costs, whilst this figure approaches £1,000 for high-income households [19]. Therefore, low-income households may struggle to afford the initial installation cost of heat pumps, which requires a targeted grant of £5,000 per property towards heat pump [1].

4. CONCLUSIONS

This study proposed a novel approach to identifying suitable households for deploying heat pumps by integrating clustering techniques with socioeconomic segments. The proposed method integrated GA to optimise K-Means within the K-MOR framework, demonstrating superior performance in profiling energy consumption patterns compared to conventional approaches. The findings indicate that the highest load pattern's daily heating demand is about 3.1 times that of the lowest load pattern, where 50.9% of households exhibit a middle-higher-load pattern and are suitable for installing heat pumps. However, 14.44 % of households are in urgent need of capital grants to help property owners overcome the upfront cost of installing heat pumps. These insights provide a solid basis for establishing targeted heat pump deployment strategies towards net-zero targets.

In our future research, we plan to quantify the potential emission reductions from widespread heat pump adoption across different socioeconomic segments and assess whether the existing electrical infrastructure can support the increased power demand.

ACKNOWLEDGEMENT

This work was supported by the EPSRC Innovation Launchpad Network + Researchers in Residence through the project of "Digitally-Enabled Flexibility Assessment of Multi-Energy Systems Toward Net-Zero Transition" (RIR35231118-1), and The British Academy through the project of "Assembling Digital Solutions for Inclusive and Net-Zero Transition of Community Energy Systems (ADvaNCE)" (PPHE24\100207).

REFERENCE

[1] HM Government. *Net Zero Strategy: Build Back Greener*. UK, HM Government, 2021. Available: www.gov.uk/government/publications/net-zero-strategy. [Accessed: 10-Aug-2024]

[2] EHPA Market Statistics. "Number of households in 2022," Eurostat. Available: https://stats.ehpa.org. [Accessed: 10-Aug-2024]

[3] Semmelmann, L., Jaquart, P., & Weinhardt, C. (2023). Generating synthetic load profiles of residential heat pumps: a k-means clustering approach. *Energy Inform., 6*(Suppl 1), 37.

[4] Ray, G. L., & Pinson, P. (2019). Online adaptive clustering algorithm for load profiling. *Sustain. Energy Grids Netw.*, 18.

[5] Lin, S., Li, F. F., Tian, E., Fu, Y., & Li, D. (2019). Clustering load profiles for demand response applications. *IEEE Trans. Smart Grid, 10*(2), 1599-1607

[6] 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, 134*, 90-102.

[7] Wang, Y., Chen, Q., & Kang, C. (2016). Clustering of electricity consumption behavior dynamics toward big data applications. *IEEE Trans. Smart Grid, 7*(5), 2437-2447.

[8] Canet, A., Qadrdan, M., Jenkins, N., et al. (2022). Spatial and temporal data to study residential heat decarbonisation pathways in England and Wales. *Sci. Data*, 9, 246.

[9] Shorewala, V. (2021). Anomaly detection and improvement of clusters using enhanced k-means algorithm. In *Proc. 2021 5th Int. Conf. Comput. Commun. Signal Process. (ICCCSP)*, 115-121.

[10] Gaur, A. S., Fitiwi, D. Z., & Curtis, J. (2021). Heat pumps and our low-carbon future: A comprehensive review. *Energy Research & Social Science, 71*, 101764.

[11] Few, S., Djapic, P., Strbac, G., et al. (2024). A geographically disaggregated approach to integrate low-carbon technologies across local electricity networks. *Nat. Energy*, 9, 871–882.

[12] AECOM Building Engineering. (2010). *Energy Demand Research Project: Early Smart Meter Trials, 2007–2010*.

[13] Guojun Gan, & Michael Kwok-Po Ng. (2017). k-means clustering with outlier removal. *Pattern Recognit. Lett., 90*, 8-14.

[14] Zhang, X., Ding, T., & Qu, M. (2023). An algorithmic approach for inner max-min model under norm-2 type uncertainty set in data-driven distributionally robust optimization. *IEEE Trans. Power Syst.*, 38(2), 1755-1758

[15] Islam, M. Z., Estivill-Castro, V., Rahman, M. A., & Bossomaier, T. (2018). Combining K-means and a genetic algorithm through a novel arrangement of genetic operators for high-quality clustering. *Expert Syst. Appl*, 91, 402-417.

[16] Goldberger, J., & Tassa, T. (2008). A hierarchical clustering algorithm based on the Hungarian method. *Pattern Recognit. Lett.*, 29(11), 1632-1638.

[17] Makhadmeh, S. N., Khader, A. T., Al-Betar, M. A., Naim, S., Abasi, A. K., & Alyasseri, Z. A. A. (2019). Optimization methods for power scheduling problems in smart homes: Survey. Renew. Sustain. Energy Rev., 115, 109362.

[18] Barma, M. C., Saidur, R., Rahman, S. M. A., Allouhi, A., Akash, B. A., & Sait, S. M. (2017). A review on boilers energy use, energy savings, and emissions reductions. *Renew Sust Energ Rev*., 79, 970-983.

[19] Ray-Chaudhuri, S., Waters, T., Wernham, T., & Xu, X. (2023). Living standards, poverty and inequality in the UK: 2023. Inst. Fiscal Stud.