

## Article

# Clustering of Load Profiles of Residential Customers Using Extreme Points and Demographic Characteristics

Hyun Cheol Jeong, Minseok Jang, Taegon Kim and Sung-Kwan Joo \*

The School of Electrical Engineering, Korea University, Seoul 02841, Korea; jeonghc93@korea.ac.kr (H.C.J.); blue7xa7@korea.ac.kr (M.J.); xorhs98@korea.ac.kr (T.K.)

\* Correspondence: skjoo@korea.ac.kr; Tel.: +82-2-3290-4820

**Abstract:** In this paper, a systematic method is proposed to cluster the energy consumption patterns of residential customers by utilizing extreme points and demographic characteristics. The extreme points of the energy consumption pattern enable effective clustering of residential customers. Additionally, demographic characteristics can be used to determine an effective extreme point for the clustering algorithm. The K-means-based features selection method is used to classify energy consumption patterns of residential customers into six types. Furthermore, the type of energy consumption pattern can be identified depending on the characteristics of residential customers. The analytical results of this paper show that the extreme points are effective in clustering the energy consumption patterns of residential customers.

**Keywords:** extreme points; demographic characteristics; customer load grouping



**Citation:** Jeong, H.C.; Jang, M.; Kim, T.; Joo, S.-K. Clustering of Load Profiles of Residential Customers Using Extreme Points and Demographic Characteristics. *Electronics* **2021**, *10*, 290. <https://doi.org/10.3390/electronics10030290>

Academic Editor: Taha Selim Ustun  
Received: 5 January 2021  
Accepted: 22 January 2021  
Published: 26 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As climate change issues such as global warming emerge, social demands for energy conservation and efficient use are increasing. In response to these demands, the South Korean government plans to reduce greenhouse gas emissions by 37% in its Until 2030 program [1]. Accordingly, they are shifting their policy from supply-side management to demand-side management. Internet of Things (IoT) devices are used to efficiently manage energy for energy consumers [2]. Advanced metering infrastructure is also being expanded and distributed in South Korea by the Korea Electric Power Corporation; this will lay the foundation for the shift to energy policies that are centered on demand management. Thus, the energy consumption of customers can be analyzed in detail by obtaining load profiles for real-time demand [3]. Identifying customers' energy consumption durations, the stability of daily energy consumption patterns over time, and actual energy use can provide customers with insight into their energy consumption behavior patterns [4]. The analysis of the customer load profile can be related to marketing and design of the load management program. For example, a customer having high energy consumption may purchase high-energy-efficient devices, and customers with constant daily energy consumption patterns respond to the message that energy needs to be saved [5].

Furthermore, various energy consumption patterns can be clustered through clustering algorithms, and the results utilized for consulting on electricity rates considering the characteristics of each group. By offering different time-of-use tariffs to customers in the different groups, the complementarity among different customers maximizes profit for the concerned retailer [6]. In addition, the clustering of time-sharing peak–valley pricing can be used to meet the various needs of customers, resulting in load shifting and energy efficiency [7,8]. Therefore, an effective analysis based on clustering of the residential customer energy consumption pattern is important for demand side management.

Several studies have proposed clustering methods by using the historical load profile of customers. Bidoki et al. analyzed load profiles of customers with several clustering methods such as K-means, weighted fuzzy mean K-means, and self-organizing map (SOM). The

comparison of clustering results obtained by these methods show that the best clustering method changes for the historical load profiles depending on the purpose of clustering [9]. Zhou et al. proposed a five-stage clustering model based on the fuzzy C-means (FCM) clustering algorithm to efficiently group smart meter data with massive, dynamic, high-dimensional, and heterogeneous characteristics [10]. Xu et al. proposed the two-stage K-means algorithm using standard K-means and integral K-means to cluster customers with energy consumption patterns. In Reference [11], customers are clustered using their energy consumption and peak time information. Choi et al. proposes a method to reflect the calendar day features in energy consumption patterns utilizing time domain mapping [12]. Density-based spatial clustering of applications with noise (DBSCAN) is used in Reference [12] to group customers with similar load profiles.

Customer clustering analysis based on load profile utilize 15 min or 1 h energy consumption data. However, it is difficult to analyze the energy consumption features with 15 min or 1 h energy consumption data since high-dimensional data are required. Therefore, various feature selection studies have been proposed to extract relevant features for the customer's energy consumption. Abubaker analyzes energy consumption data in the Tulkarm region of Palestine and proposes a new customer segmentation method by clustering customers with similar energy consumption behavior patterns. Two features are selected from energy consumption data using principal components analysis (PCA) and utilize K-means algorithms to cluster customers and obtain new customer segments [13]. Haben et al. analyzed customer behavior affecting the factors of peak and variation of energy demand via analysis of smart meter data of customers. Smart meter data of customers is used to analyze energy consumption over a period of time that affects the peak. A finite mixture model is used to identify the behavior of energy consumption by seasonal and temporal characteristics [14]. Chico et al. proposes an algorithm for effectively clustering the energy consumption pattern by applying the probabilistic approach to the support vector machine (SVM) to select features for outliers in the energy consumption pattern. The proposed method can be used to help suppliers come up with demand response options relevant to customers by adding new clusters that differentiate outliers [15]. Wang et al. proposes a clustering method that measures Kullback–Liebler distances to obtain various features of energy consumption behavior, which uses the clustering technique by fast search and find of density peaks (CFSFDP) to cluster customers [16]. Reference [16] introduces a method to determine the appropriate demand response according to the type of energy consumption behavior of the customer. However, it is difficult to understand how the features obtained by PCA are physically related to energy consumption characteristics.

In this paper, the extreme points of energy consumption patterns are used in combination with demographic characteristics to obtain effective features of a typical residential customer while maintaining the physical meaning of those features. In this study, the extreme points are designed to reflect the difference in energy consumption between two time periods. In addition, the energy consumption of residential customers is correlated with the demographic characteristics. This paper presents an effective feature selection method for clustering the load profiles of residential customers using extreme points and demographic characteristics.

Additionally, previous clustering studies of energy consumption customers use load profile from specific regions, or some studies use load profile from 18 cities, but not more than 600 customers [9–16]. In this paper, however, clustering analysis of energy consumption customers is carried out using load profiles from 2041 households with various demographic characteristics, such as area, number of household members and salary level in seven regions.

The remainder of this paper is organized as follows. Section 2 analyzes the features of residential customers' energy consumption patterns. Section 3 describes the method for clustering energy consumption patterns. Finally, Sections 4 and 5 present the results and conclusions.

## 2. Selection of Extreme Points of the Energy Consumption Pattern with Demographic Characteristics

In this paper, a clustering analysis method utilizing important features of the residential customer energy consumption pattern is proposed. The load profile of residential customers is obtained using data collected between 1 June 2018 and 31 May 2019. Z-score ( $Z_{n,t}$ ) is computed for load profile over a period time in order to detect outliers based on 95% of the confidence interval ( $-1.96 < Z_{n,t} < 1.96$ ). Once an outlier is detected, the load profile of the customer is replaced with the average energy consumption of the same time period in the same season. The Z-score is calculated as follows [17]:

$$Z_{n,t} = \frac{L_{n,t} - \mu_{total,t}}{\sigma_{total,t}} \quad (1)$$

where  $L_{n,t}$  is the data of customer  $n$  at time  $t$ ,  $\mu_{total,t}$  is the total customer average data at time  $t$ , and  $\sigma_{total,t}$  is the total customer standard deviation data at time  $t$ .

In addition, the features of energy consumption patterns are compared among customers by normalizing the load profile to effectively cluster residential customers. The data are normalized to a maximum value of 1 and minimum value of 0 using minimum–maximum normalization. The load profile is normalized as follows [18]:

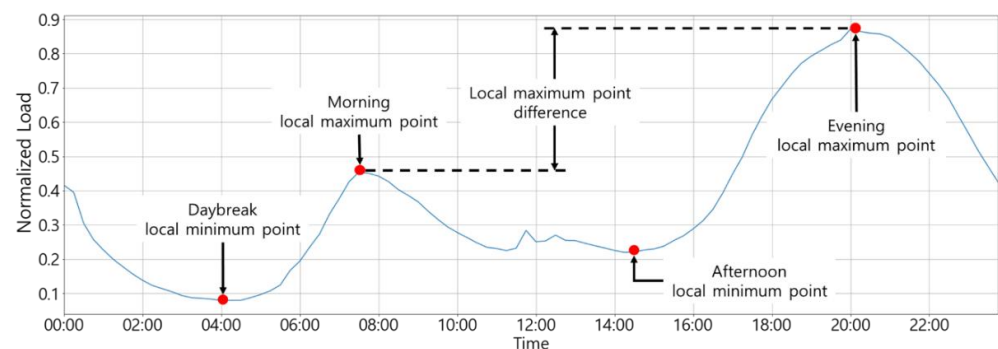
$$NL_{n,t} = \frac{L_{n,t} - \min(L_n)}{\max(L_n) - \min(L_n)} \quad (2)$$

where  $NL_{n,t}$  is the normalized data of customer  $n$  at time step  $t$ ,  $L_{n,t}$  is the data of customer  $n$  at time  $t$ , and  $L_n$  is the load profile of customer  $n$ .

### 2.1. Load Profile Analysis

Figure 1 shows the normalized average annual energy consumption pattern of the residential customers. The extreme point of the time-series data is defined as follows [19]:

- Local maximum point:  $NL_{n,t-1} - NL_{n,t} > 0$ ,  $NL_{n,t} - NL_{n,t+1} \leq 0$ .
- Local minimum point:  $NL_{n,t-1} - NL_{n,t} < 0$ ,  $NL_{n,t} - NL_{n,t+1} \geq 0$ .



**Figure 1.** Normalized average load profiles of residential customers.

There are four extreme points in the load profile: daybreak local minimum point ( $LP_{min,daybreak}$ ), morning local maximum point ( $LP_{max,morning}$ ), afternoon local minimum point ( $LP_{min,afternoon}$ ), and evening local maximum point ( $LP_{max,evening}$ ). In addition, the difference between the morning local maximum point and the evening local maximum point ( $LP_{max,diff}$ ) is used to differentiate the time when the customer's energy consumption is concentrated. Therefore, the extreme points represent the energy consumption of residential customers by time period, and these can be compared to each other. Thus, the extreme points are expected to be effective in clustering the energy consumption patterns of residential customers.

## 2.2. Feature Selection

Residential customers are clustered by selecting features that can effectively analyze the characteristics of each cluster. Feature selection is undertaken by analyzing the correlation between monthly total energy consumption and customer demographic data: area, number of household members, salary level, age, and education level. Figure 2 shows the distribution of customers by regions and demographic data.

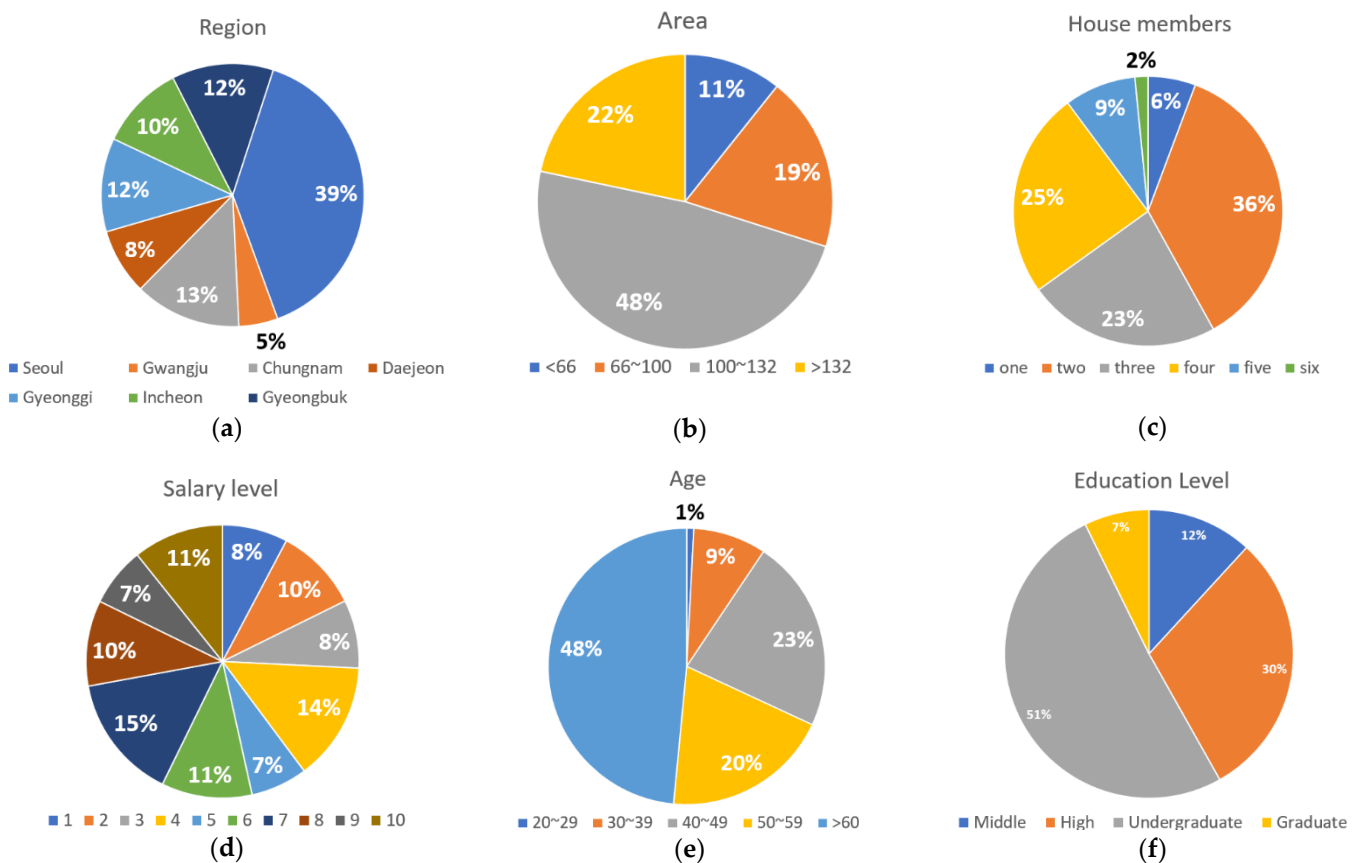


Figure 2. Distribution of customers by region and demographic data.

Correlation analysis is carried out by computing Spearman's correlation coefficient because demographic data are ranking data, and monthly total energy consumption is continuous data. Spearman's correlation coefficient  $R_s$  is calculated according to the following equation [20]:

$$d_n = L_{rank,n} - D_{rank,n} \quad (3)$$

$$R_s = 1 - \left( \frac{6 \sum_{n=1}^N d_n^2}{N(N-1)} \right) \quad (4)$$

where  $d_n$  denotes the rank data difference of customer  $n$ ,  $L_{rank,n}$  is the rank total monthly energy consumption of customer  $n$ ,  $D_{rank,n}$  is the rank demographic data of customer  $n$ , and  $N$  is the total number of customers;  $0 < R_s < \pm 0.4$  indicates weak correlation,  $\pm 0.4 \leq R_s < \pm 0.6$  moderate correlation, and  $\pm 0.6 \leq R_s \leq \pm 1$  strong correlation [21].

Table 1 shows the correlation between the monthly total energy consumption and demographic data. The monthly total energy consumption shows a moderate correlation with the area and number of household members, and weak correlation with salary level and household head information (age and education level). Therefore, the features are selected by utilizing the data of area and number of household members.

**Table 1.** Spearman’s correlation coefficient between energy consumption and demographic data.

Classification	Demographic				
	Area	Household Members	Salary Level	Household Head	
				Age	Education Level
$R_s$	0.4	0.4	0.3	0	0.2

Tables 2 and 3 show the average extreme value and range of each attribute. The range for each attribute is relatively large in  $LP_{max,diff}$ ,  $LP_{min,daybreak}$ , and  $LP_{max,morning}$ . Therefore, these attributes were selected as features for clustering residential customers.

**Table 2.** Average feature value and range by area attribute.

Classification	Area (m <sup>2</sup> )				Range
	<66	66~100	100~132	>132	
$LP_{min,daybreak}$	0.0898	0.0462	0.0417	0.0198	0.0700
$LP_{max,morning}$	0.7340	0.6590	0.6820	0.7015	0.0750
$LP_{min,afternoon}$	0.1431	0.1552	0.1486	0.1715	0.0284
$LP_{max,evening}$	0.9199	0.9318	0.9362	0.9298	0.0163
$LP_{max,diff}$	−0.1859	−0.2728	−0.2542	−0.2283	0.0869

**Table 3.** Average feature value and range by number of household members.

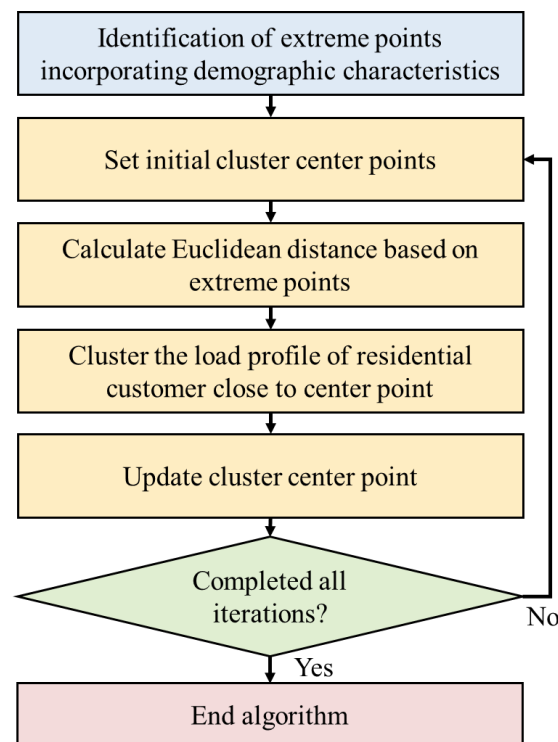
Classification	Number of Household Members						Range
	1	2	3	4	5	6	
$LP_{min,daybreak}$	0.1064	0.0553	0.0366	0.0241	0.0316	0.0284	0.0824
$LP_{max,morning}$	0.7284	0.7102	0.6646	0.6702	0.6707	0.6133	0.1150
$LP_{min,afternoon}$	0.1258	0.1304	0.1699	0.1719	0.1783	0.1892	0.0634
$LP_{max,evening}$	0.8962	0.9059	0.9570	0.9531	0.9504	0.9427	0.0609
$LP_{max,diff}$	−0.1678	−0.1957	−0.2924	−0.2829	−0.2797	−0.3293	0.1615

### 3. Clustering Method for Residential Customer Energy Profile Using Extreme Points and Demographic Characteristics

For the load profiles of residential customers, the difference in energy consumption between customers is small and the range of energy consumption is wide. In this study, clustering of load profiles of residential customers is carried out by computing the distance from the cluster center point in the residential customer energy consumption data. In this paper, therefore, the simple and effective K-means grouping algorithm is adopted to classify the load profiles of residential customers by computing distances from the center point of an individual group [13]. It attempts to compute the Euclidean distance between individual load profile set  $I = \{i_1, i_2, \dots, i_n\}$  and the center point ( $c_{y,centroid}$ ) of group set  $C = \{c_1, c_2, \dots, c_k\}$ . The load profiles of residential customers are clustered by minimizing the sum of the Euclidean distances of the group set [22]. The clustering function taken from Reference [22] follows:

$$\operatorname{argmin}_{c_y} \sum_{x=1}^n \|i_x - c_{y,centroid}\|^2 \quad (5)$$

Figure 3 shows the procedure of the proposed clustering method incorporating extreme points and demographic characteristics. Initially, the center point of an individual group is arbitrarily selected. Subsequently, the Euclidean distance between the individual load profile set I and the initial center point of group is computed. Next, the input data  $i_x$  are clustered at the cluster center point closest to  $i_x$ . All input data are distributed to each cluster set. Then, the cluster center point is updated with the average value of the data in each cluster set. This process iterates a set number of times.



**Figure 3.** Clustering method using extreme points and demographic characteristics.

The K-means algorithm is applied to determine the optimal number of groups. The elbow method based on distance is used to determine the optimal value  $k$  [22]. Other methods for determining the optimal value  $k$  include silhouette score, Davies–Bouldin index using cohesion, and separation [23,24]. There is little separation between residential customer load profiles. In this paper, therefore, the elbow method is adopted to compute the distance from the cluster center point to residential customer load profile in order to determine the optimal value  $k$ . The elbow method calculates the sum of squared error (SSE) to determine the optimal value  $k$ , which is used herein. The SSE is calculated as follows [22]:

$$SSE = \sum_{y=1}^k \sum_{i \in c_y} \|i - c_{y,centroid}\|^2 \quad (6)$$

where  $k$  is the number of clusters,  $c_y$  is the  $y$ -th cluster,  $i$  is the data in  $c_y$ , and  $c_{y,centroid}$  is the center point of the  $y$ -th cluster.

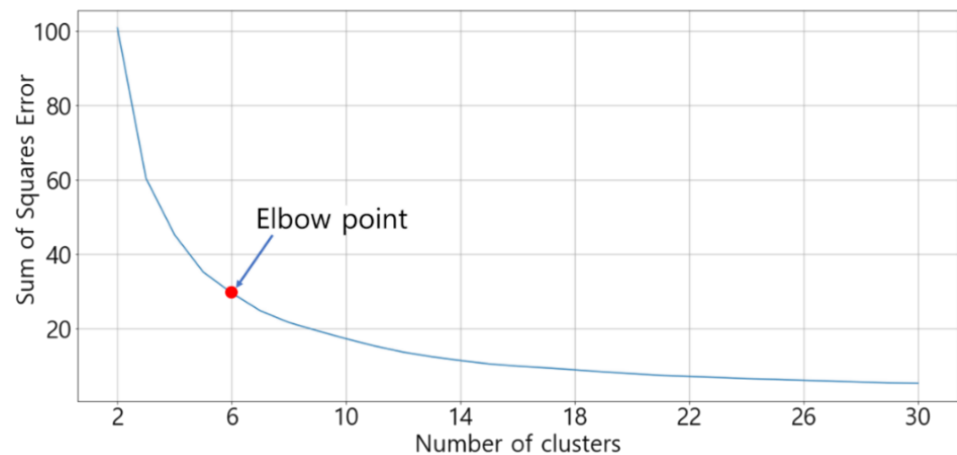
As  $k$  increases, the cohesion increases because the number of data in the cluster decreases. Thus, the SSE decreased. When the SSE nears the optimal value  $k$ , the rate of decrease of the SSE is reduced. Therefore, the SSE shows a flat graph after reaching  $k$ . Since the graph is elbow-shape, this method is called the elbow method.

This study calculates the SSE  $k$  from 2 to 30 clusters. Figure 3 shows the SSE graph. As a result of using the elbow method, the optimal  $k$  for residential customer clustering is determined to be six.

#### 4. Results

As a result of clustering, the energy consumption of residential customers shows six patterns. The energy consumption pattern for each cluster is shown in Figure 4.





**Figure 4.** Sum of squares error (SSE) of residential customers clustered with K-means.

Figure 5a shows the cluster that represents a typical energy consumption pattern for residential customers. Figure 5b shows the energy saving cluster, which saves energy in the morning. Figure 5c shows the M-pattern cluster with similar energy consumption in the morning and the evening. Most residential customers in the M-pattern cluster are office workers. It is recognized in the M-pattern cluster that energy consumption increases before leaving for work. Energy consumption of residential customers within M-pattern cluster is relatively low since residential customers within the M-pattern cluster do not normally stay at home during daytime. Figure 5d shows the morning concentrated cluster with energy consumption concentrated in the morning. In general, residential customers within the morning concentrated cluster tend to use high energy consumption appliances such as washer machines and vacuum cleaners in the morning, resulting in high energy consumption in the morning. Figure 5e shows the evening concentrated cluster with energy consumption concentrated in the evening. In contrast, residential customers within the evening concentrated cluster tend to use high energy consumption appliances in the evening, resulting in high energy consumption in the evening. Figure 5f shows the owl cluster with relatively large energy consumption at daybreak compared to the other clusters. The owl cluster shows patterns of energy consumption behavior that consistently use energy until daybreak, like owls, and have very low energy consumption during the daytime.

Table 4 shows the results of the analysis of the characteristics of each cluster: monthly total energy consumption, area, and number of household members. In addition, the energy-consuming cluster and energy-saving cluster have the same area of  $100\text{ m}^2\sim 132\text{ m}^2$ . However, the number of household members shows different characteristics; the energy consumption cluster has two members, and the energy-saving cluster has four members. The M-pattern cluster, the morning concentrated cluster, and the evening concentrated cluster show similar characteristics in area and number of household members. However, differences in energy consumption behavior patterns result in different monthly total energy consumption. The owl cluster tends to have an area of less than  $66\text{ m}^2$  and single-person households. Their monthly total energy consumption is the smallest at  $230.24\text{ kWh}$  because the area and the number of household members are smaller than those of the other clusters.

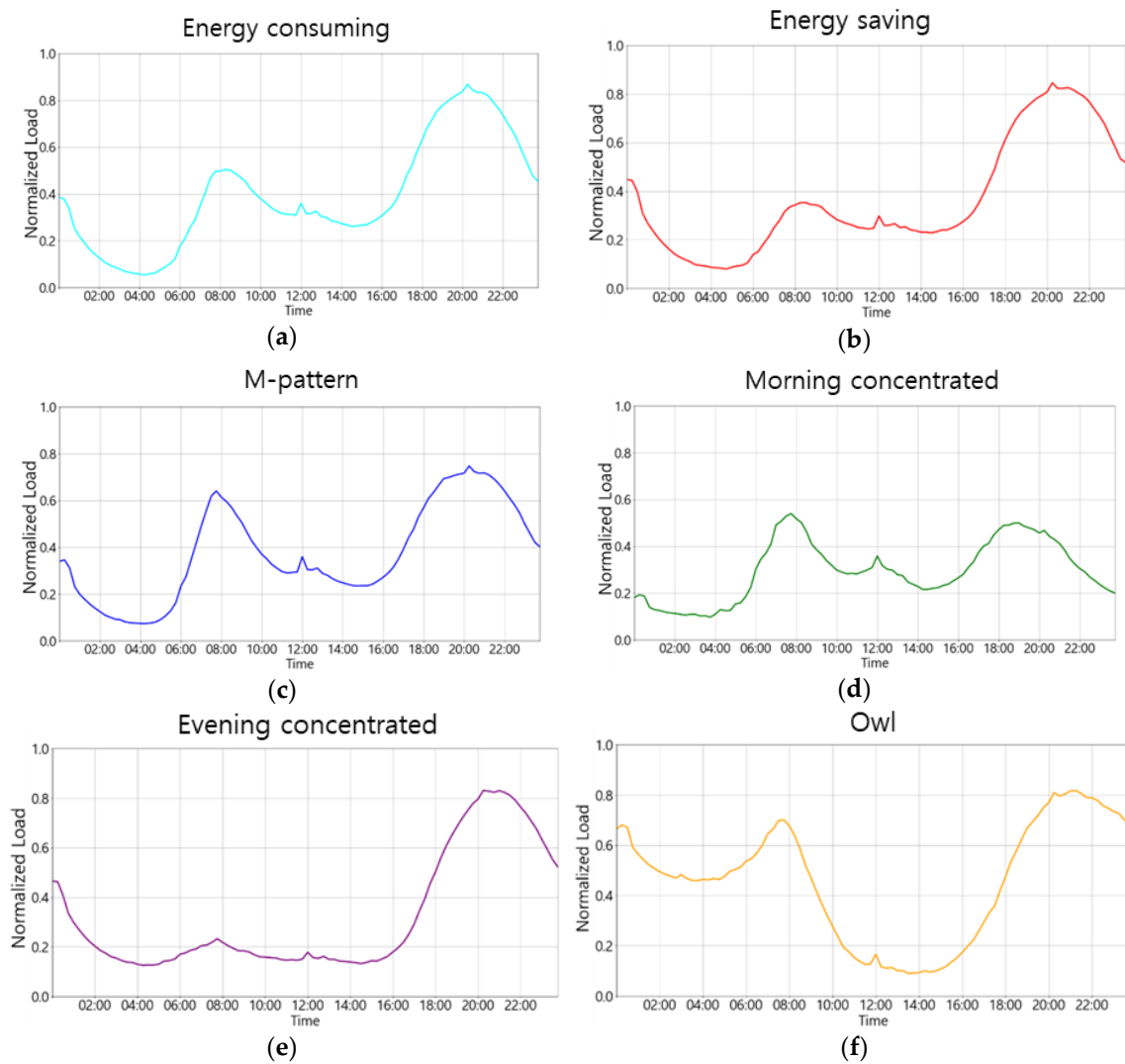


Figure 5. Results of residential customer clustering utilizing K-means.

Table 4. Monthly total energy consumption and average demographic data by cluster.

Classification	Energy Consumption [kWh]	Area (m <sup>2</sup> )	Number of Household Members
Energy consuming	273.7	100~132	2
Energy saving	268.77	100~132	4
M-pattern	263.64	100~132	2
Morning concentrated	260.91	100~132	2
Evening concentrated	251.24	100~132	2
Owl	230.24	Less than 66	1

### 5. Discussion

In this paper, extreme points of load profiles of residential customers are used to analyze the differences in energy consumption between two time periods. The load profiles of residential customers are grouped using extreme points and demographic data. The demographic characteristics of customers in each cluster are also analyzed.

The typical energy consumption pattern of a residential customer has four local extreme points. Each local extreme point appears at daybreak, morning, afternoon, and



evening. Thus, the local extreme point is effective in analyzing the differences in energy consumption over a period of time between customers.

Among the demographic data of residential customers, the number of household members and the area are found to be closely correlated with energy consumption. In this paper, therefore, the number of household members and the area information are used to determine the extreme points that best represent the demographic characteristics of residential customers. The K-means algorithm is applied for grouping the load profiles of residential customers. In this study, six patterns of load profiles of residential customers are obtained by the proposed method. This result shows the difference in monthly energy consumption and demographic characteristics according to the energy consumption pattern of each cluster. Unlike energy consumption patterns of other groups, for instance, the owl group, which represents relatively large energy consumption at daybreak, has an area of less than 66 m<sup>2</sup> and the number of household members is one, showing demographic characteristics of a single household. Therefore, customer clustering analysis based on extreme point and demographic data effectively analyze the energy consumption patterns and demographic properties.

## 6. Conclusions

This paper proposes a clustering method for residential customers using extreme points in their energy consumption patterns and demographic data. The energy consumption patterns show extreme points at daybreak and in the morning, afternoon, and evening. This feature enables effective clustering. In addition, demographic data are used to select effective extreme points for the clustering of residential customers using the K-means method.

The residential customers show six types of energy consumption patterns, and the total monthly energy consumption, area, and number of household members are analyzed in each cluster. Furthermore, energy consumption patterns can be correlated with the characteristics of residential customers. Therefore, the extreme points are effective in clustering the energy consumption patterns of residential customers.

In addition, the clustering results obtained by the proposed method can be used to study the billing and load impact. Further research needs to be directed toward the development of personalized residential electricity rates for different demographic characteristics of customers.

**Author Contributions:** H.C.J. performed the research and wrote the paper. M.J. helped with the clustering algorithm and analysis. T.K. helped with the original draft and revised the paper. S.-K.J. provided guidance for the research and revised the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Korea Institute of Energy Technology Evaluation and Planning, and the Ministry of Trade, Industry & Energy of the Republic of Korea (No. 20181210301430). This research was supported by the Basic Research Program through the National Research Foundation of Korea (NRF) funded by the MSIT (No. 2020R1F1A1075872).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ministry of Trade, Industry and Energy. *8th Basic Plan for Supply and Demand of Electric Power (2017~2031)*; Ministry of Trade, Industry and Energy: Sejong-si, Korea, 2017; p. 12.
2. Lloret, J.; Tomas, J.; Canovas, A.; Parra, L. An integrated IoT Architecture for Smart Metering. *IEEE Commun. Mag.* **2016**, *50*, 50–57. [[CrossRef](#)]
3. Lee, Y.; Hwang, E.; Choi, J. A Unified Approach for Compression and Authentication of Smart Meter Reading in AMI. *IEEE Access* **2019**, *7*, 34383–34394. [[CrossRef](#)]
4. Smith, B.A.; Wong, J.; Rajagopal, R. A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting. In *2012 ACEEE Summer Study on Energy Efficiency in Buildings*; American Council for an Energy-Efficient Economy: California City, CA, USA, 2012.

5. Kwac, J.; Flora, J.; Rajagopal, R. Household energy consumption segmentation using Hourly Data. *IEEE Trans. Smart Grid* **2014**, *5*, 420–430. [[CrossRef](#)]
6. Yang, J.; Zhao, J.; Wen, F.; Dong, Z. A Model of Customizing Electricity Retail Prices Based on Load Profile Clustering Analysis. *IEEE Trans. Smart Grid* **2019**, *10*, 3374–3386. [[CrossRef](#)]
7. Wang, S.; Wu, S.; Liu, T.; Li, J.; Lu, Y. Power user of classification strategy of multi-view clustering. In Proceedings of the 2016 16th International Symposium on Communications and Information Technologies (ISCIT), Qingdao, China, 26–28 September 2016.
8. Du, P.; Huang, D.; Liu, Z.; Zhang, H.; Zhang, J.; Yan, Q. Power customer classification based on OCHNN-Kmeans and price setting of TOU. In Proceedings of the 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), Chengdu, China, 4–7 June 2020.
9. Bidoki, S.M.; Kohan, N.M.; Gerami, S. Comparison of several clustering methods in the case of electrical load curves classification. In Proceedings of the 16th Electrical Power Distribution Conference, Bandar Abbas, Iran, 19–20 April 2011; pp. 1–7.
10. Zhou, K.L.; Yang, S.L.; Shen, C. A review of electric load classification in smart grid environment. *Renew. Sustain. Energy Rev.* **2013**, *24*, 103–110. [[CrossRef](#)]
11. Xu, S.; Barbour, E.; González, M.C. Household segmentation by load shape and daily consumption. In Proceedings of the ACM SigKDD, Halifax, NS, Canada, 13–17 August 2017; pp. 1–9.
12. Choi, J.; Lee, I.W.; Cha, S.W. Cluster Analysis to Preprocess the Building Power Usage Data without Domain Knowledge. *J. Electr. Eng. Technol.* **2020**, *15*, 685–692. [[CrossRef](#)]
13. AbuBaker, M. Data Mining Applications in Understanding Electricity Consumers' Behavior: A Case Study of Tulkarm District, Palestine. *Energies* **2019**, *12*, 4287. [[CrossRef](#)]
14. Haben, S.; Singleton, C.; Grindrod, P. Analysis and clustering residential customers energy behavioral demand using smart meter data. *IEEE Trans. Smart Grid* **2016**, *7*, 136–144. [[CrossRef](#)]
15. Chicco, G.; Ilie, I.-S. Support vector clustering of electrical load pattern data. *IEEE Trans. Power Syst.* **2009**, *24*, 1619–1628. [[CrossRef](#)]
16. Wang, Y.; Chen, Q.; Kang, C.; Xia, Q. Clustering of electricity consumption behavior dynamics toward big data applications. *IEEE Trans. Smart Grid* **2016**, *7*, 2437–2447. [[CrossRef](#)]
17. Patel, V.R.; Mehta, R.G. Impact of outlier removal and normalization approach in modified k-means clustering algorithm. *Int. J. Comput. Sci. Issues* **2011**, *8*, 331.
18. Khare, N.; Devan, P.; Chowdhary, C.L.; Bhattacharya, S.; Singh, G.; Singh, S.; Yoon, B. SMO-DNN: Spider Monkey Optimization and Deep Neural Network Hybrid Classifier Model for Intrusion Detection. *Electronics* **2020**, *9*, 692. [[CrossRef](#)]
19. Du, T.; Qu, S.; Hua, Z. A Novel Timing Series Calculation Algorithm based on Statistical Extremum for Process Object. In Proceedings of the 9th International Conference on Computer and Automation Engineering (ICCAE '17), Sydney, Australia, 18–21 February 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 94–98.
20. Akoglu, H. User's guide to correlation coefficients. *Turk. J. Emerg. Med.* **2018**, *18*, 91–93. [[CrossRef](#)] [[PubMed](#)]
21. Hyde, K.; Maier, H.; Colby, C. Incorporating uncertainty in the PROMETHEE MCDA method. *J. Multi-Criteria Decis. Anal.* **2003**, *12*, 24–259. [[CrossRef](#)]
22. Zeng, J.; Wang, J.; Guo, L.; Fan, G.; Zhang, K.; Gui, G. Cell Scene Division and Visualization Based on Autoencoder and K-Means Algorithm. *IEEE Access* **2019**, *7*, 165217–165225. [[CrossRef](#)]
23. Hong, T.; Chen, C.F.; Wang, Z.; Xu, X. Linking human-building interactions in shared offices with personality traits. *Build. Environ.* **2020**, *170*, 106602. [[CrossRef](#)]
24. Sari, P.K.; Purwadinata, A. Analysis Characteristics of Car Sales in E-Commerce Data Using Clustering Model. *J. Data Sci. Its Appl.* **2019**, *2*, 19–28. [[CrossRef](#)]