

Article

# Load Profile Segmentation for Effective Residential Demand Response Program: Method and Evidence from Korean Pilot Study

Eunjung Lee <sup>1</sup>, Jinho Kim <sup>1,\*</sup> and Dongsik Jang <sup>2</sup>

<sup>1</sup> School of Integrated Technology, Gwangju Institute of Science and Technology, 123 Cheomdangwagi-ro, Buk-gu, Gwangju 61005, Korea; jkl51149@gist.ac.kr

<sup>2</sup> Korea Electric Power Research Institute, 105 Munji-ro, Yuseong-gu, Daejeon 34056, Korea; yojian2@kepco.co.kr

\* Correspondence: jeikim@gist.ac.kr; Tel.: +82-62-715-5322

Received: 26 February 2020; Accepted: 11 March 2020; Published: 13 March 2020



**Abstract:** Due to the heterogeneity of demand response behaviors among customers, selecting a suitable segment is one of the key factors for the efficient and stable operation of the demand response (DR) program. Most utilities recognize the importance of targeted enrollment. Customer targeting in DR programs is normally implemented based on customer segmentation. Residential customers are characterized by low electricity consumption and large variability across times of consumption. These factors are considered to be the primary challenges in household load profile segmentation. Existing customer segmentation methods have limitations in reflecting daily consumption of electricity, peak demand timings, and load patterns. In this study, we propose a new clustering method to segment customers more effectively in residential demand response programs and thereby, identify suitable customer targets in DR. The approach can be described as a two-stage k-means procedure including consumption features and load patterns. We provide evidence of the outstanding performance of the proposed method compared to existing k-means, Self-Organizing Map (SOM) and Fuzzy C-Means (FCM) models. Segmentation results are also analyzed to identify appropriate groups participating in DR, and the DR effect of targeted groups was estimated in comparison with customers without load profile segmentation. We applied the proposed method to residential customers who participated in a peak-time rebate pilot DR program in Korea. The result proves that the proposed method shows outstanding performance: demand reduction increased by 33.44% compared with the opt-in case and the utility saving cost in DR operation was 437,256 KRW. Furthermore, our study shows that organizations applying DR programs, such as retail utilities or independent system operators, can more economically manage incentive-based DR programs by selecting targeted customers.

**Keywords:** data analysis; demand response (DR), load profile clustering; k-means; targeting of customer

## 1. Introduction

Recently, distributed energy resources (DER) such as photovoltaic (PV), wind turbine (WT), energy storage system (ESS), and demand response (DR) have been rapidly expanded on the distribution system. Because of this trend, the power demand characteristics have been more complicated. In addition, various business models and policies as resources application increased have been created. However, the DER expansion leads to load fluctuation on the distribution system locally. The DR program has been regarded as one of the solutions to mitigate imbalance. For this reason, DR programs have recently received significant attention. Under a DR program, electricity consumers change their electricity consumption patterns in response to a time-based rate or incentive payments for the periods

when needed [1]. Utilities and/or independent system operators (ISO) manage DR programs to avoid peak demand, high prices, and variable generation of renewables.

DR programs can be divided into two types: price- and incentive-based. Price-based DR programs vary the electricity price depending on certain time conditions being met [1]. Time of use (TOU), critical peak pricing (CPP), and real time pricing (RTP) are examples of this type of DR. Meanwhile, incentive-based DR programs encourage customers to shed their load or sell back to the electricity market. In the case of incentive-based DR programs, targeting suitable customers takes priority before DR implementation [2]. According to the peak time rebate program implemented by San Diego Gas & Electric (SDG&E), targeted enrollment, which selects suitable customers to participate in incentive-based DR programs, is essential for efficient DR operation [3]. Before DR program introduction, the customer demand characteristics analysis is significant because of heterogeneous characteristics. Especially, the costs of recruiting DR customers may be considerable, as the process involves several activities such as marketing, education, and DR system support and operation. If utility companies or ISO do not select suitable customers for enrollment in the DR program, the losses caused by enrollment of inappropriate customers could be substantial. Therefore, to minimize losses, it is essential to secure a large DR capacity with a relatively small number of customers.

Before choosing suitable customers with potential in electricity consumption and similarity between peak time and event, analyzing the load profiles of customers is essential. We considered the customer targeting concept through analyzing several typical load profiles as a result of load profile segmentation. Therefore, load profile segmentation analysis should be conducted for selecting adequate customers. Various clustering methods are normally employed to perform electricity consumer segmentation. Residential electricity consumption is uncertain and variable due to various factors affecting demand, such as home appliance usage patterns, the number of family members, lifestyle patterns, customer occupations, and income levels. These factors cause residential demand to have far more variability than commercial and industrial demand [4], thus making the residential load profile segmentation problem relatively more difficult. When analyzing load profile clusters, their load patterns or characteristics are commonly applied as variables. However, in residential load profile clustering, only considering load patterns poses a number of problems such as an excessively broad spectrum of hourly consumption rates and different peak occurrence times within the same group, whereas the drawback of only considering load characteristics is that consumer patterns are not reflected accurately. To determine suitable DR participant groups, residential customers should therefore be segmented by both pattern and consumption scales.

This paper proposes a two-stage k-means model to address pattern and consumption scales. In the first stage, k-means clustering is conducted based on load characteristics, such as daily consumption and peak occurrence time. In the second stage, k-means clustering is performed based on hourly load profile of residential customers. This methodology is applied to over 800 Korean residential DR participants, for whom hourly electricity use data is available. The results reveal an appropriate segmentation methodology for DR participants. This paper contributes to the literature on load profile segmentation for targeting customers by:

- Extending the k-means clustering method to reflect all load patterns and characteristics, thus resulting in outstanding performance;
- Deriving home appliances and usage pattern data using only electricity consumption data and not any additional data such as customer information, thus making the analysis more efficient;
- Presenting load profile segmentation of Korean household electricity demand data; and
- Conducting data analysis to suitable select groups for DR.

The remainder of this paper is organized as follows. In Section 2, we illustrate the current state-of-the-art clustering methodology. In Section 3, we present the proposed two-stage k-means model, which ensures effective household load profile segmentation for targeting residential customers. In Section 4, we show the effect of targeting residential customers in the DR program and compare this

effect to the effect of opt-in enrollment in Korea. Section 5 concludes and outlines ideas for further research in this area.

## 2. Literature Review

This section presents a review of the current state-of-the-art methodology for load profile segmentation. Many studies have been performed to segment load profile accurately by applying various clustering methods. K-means, self-organizing maps (SOM), mixture models, expectation maximization (EM), and spectral clustering have been widely used as clustering methods. Among the several methods available for clustering to address load pattern segmentation, the most commonly employed are standard k-means [5–10], adaptive k-means [11,12], fuzzy k-means [13,14], and g-means [15], which is an alternative clustering model to k-means. SOM [16,17] is commonly employed by itself but has also been combined with other clustering methods such as k-means and hierarchical clustering as a hybrid model [18]. Mixture models [19,20] and EM [21] are also popular as statistical clustering methods. For DR program operation, DR customer segmentation is commonly conducted for many reasons. Spectral clustering applying information entropy based piecewise aggregate approximation is proposed for commercial demand response application being able to reflect multiscale similarities [22]. Recently, deep learning based clustering such as deep embedded clustering has become a trend for use in residential baseline estimation [23]. Each of the existing clustering methods normally used for electricity consumer segmentation has its own characteristics and is summarized in Table 1 for each characteristic. As explained in Table 1, each clustering method has its advantages in terms of data type or separation process. Although there are a lot of existing clustering methods, k-means has great strength in that it is easier than other existing models and shows good performance in various problem solving cases.

**Table 1.** Characteristics of previous clustering methods.

Clustering Method	Characteristics
k-means	<ol style="list-style-type: none"> <li>1. Easy to implement clustering model</li> <li>2. Less computation compared with other clustering methods</li> <li>3. Fast and applicable to a wide range of problems</li> <li>4. Necessary to specify the number of initial clusters</li> </ol>
SOM	<ol style="list-style-type: none"> <li>1. Excellent clustering result</li> <li>2. Easy evaluation grouped by visual inspection</li> <li>3. Necessary to specify the number of initial clusters</li> </ol>
Mixture models	<ol style="list-style-type: none"> <li>1. Ability to model a mixture of both continuous and categorical data</li> <li>2. Providing probability that a given point belongs to each of the possible clusters</li> </ol>
Spectral clustering	<ol style="list-style-type: none"> <li>1. Allowing more flexible distance metrics and performing well</li> <li>2. Necessary to specialized machines with large memory to compute full graph Laplacian matrix (quadratic/super quadratic complexities in the number of data point)</li> </ol>
Embedded clustering [24]	<ol style="list-style-type: none"> <li>1. Able to simultaneously learn feature representations and clustering assignments using deep neural networks</li> <li>2. Less sensitive to the choice of hyper parameters</li> </ol>

It also can be used to increase accuracy of customer baseline and select appropriate customers for DR. Zhang et al. [7] proposed clustering by k-means before baseline estimation, and it demonstrated improved results. In regard to addressing clustering structure issues, some studies have employed two-stage clustering methods [14,15] which are similar with the proposed methodology in this study, showing that this structure could reflect all the load factors (i.e., voltage, residential type, consumption, and pattern) better than the structures prevalent in the literature. However, load profile segmentation

for DR targeting enrollment was not performed in these studies, and they were just focused on similar patterns in groups, which has the limitation of large variation in customer daily consumption. It is hard to use the existing models as it is in this study. Therefore, we considered the two-stage methodology to reflect load characteristics affecting DR at the first stage.

Commonly, optimization methods are utilized for customer targeting in DR program, and there are some studies on this without load profile segmentation [25,26]. Kwac et al. [25] proposed solving the stochastic knapsack problem (SKP) as a means to recruit optimal customers for DR programs. Zhou et al. [26] designed an adaptive targeting method to estimate DR effects.

This paper describes a customer targeting and DR analysis model through a two-stage clustering analysis. The proposed methodology will enable the effective selection of customers for DR programs and illustrate a better DR effect than in opt-in enrollment.

### 3. Targeting Customers for Incentive DR Using a Two-stage Load Profile Clustering Method

Selecting and recruiting appropriate customers for DR programs is essential for the successful operation of incentive-based DR. DR potential can be estimated by analyzing customer load characteristics. In this study, we derived adequate customer groups for residential DR from demand data through the load profile segmentation. There are many methods for clustering such as k-means, SOM, fuzzy clustering, Gaussian Mixture Models (GMMs), and hierarchical clustering. We adopted k-means methods in view of simplicity and accuracy, and designed load profile segmentation framework as two-stage methodology considering load characteristics in the first step and load profile value in the second step.

#### 3.1. k-means

k-means is a popular method for cluster analysis in data mining that is commonly employed to study electricity demand clustering. It is a simple and robust algorithm which aims to separate  $n$  observations into  $k$  clusters [15,27]. When a dataset  $X = \{x_1, x_2, \dots, x_N\}$  (with  $x_i \in \mathbb{R}^n$ ) and  $K$  clusters  $C = \{C_1, C_2, \dots, C_K\}$  are given, each  $x_i \in X$  is assigned to exactly one cluster  $C_k \in C$ , which is characterized by a cluster centroid  $\mu_k$ . The classical k-means clustering method is performed as follows. First, the integer value  $K$  corresponding to the number of clusters is determined. Then, the initial cluster centroid set  $\{\mu_1, \mu_2, \dots, \mu_K\}$  is selected randomly. Data point  $x_i \in X$  is assigned to the closest  $\mu_k$  through distance comparison against  $\{\mu_1, \mu_2, \dots, \mu_K\}$  using the Euclidean distance. The formula for setting the data set in clusters is illustrated by Equation (1):

$$\text{cluster}(x_i) = \underset{k \in \{1, \dots, K\}}{\operatorname{argmin}} \|x_i - \mu_k\|^2 \quad (1)$$

The clustering algorithm aims to minimize the sum of squares within the groups and maximize it between the groups. The cost function  $J$  to be minimized in k-means is therefore expressed by Equation (2):

$$J = \frac{1}{N} \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (2)$$

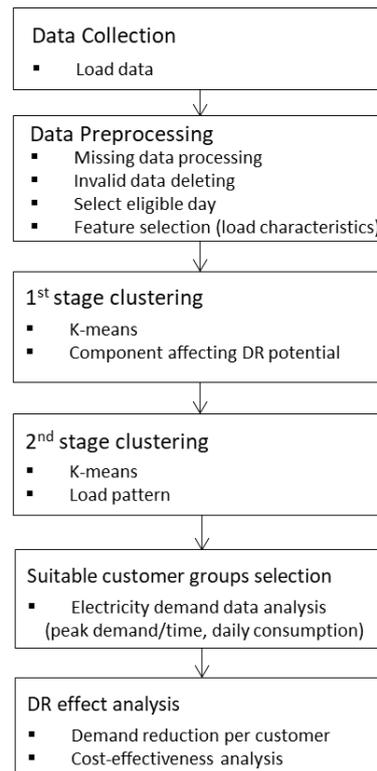
The cluster centroid set update is performed by calculating the mean data set belonging to cluster  $C_k$  as given by Equation (3):

$$\mu_k = \frac{1}{|C_k|} \sum_{x_i \in C_k} x_i \quad (3)$$

This process is repeated until the distribution of the dataset among the clusters no longer changes. In other words, cluster centroids do not change.

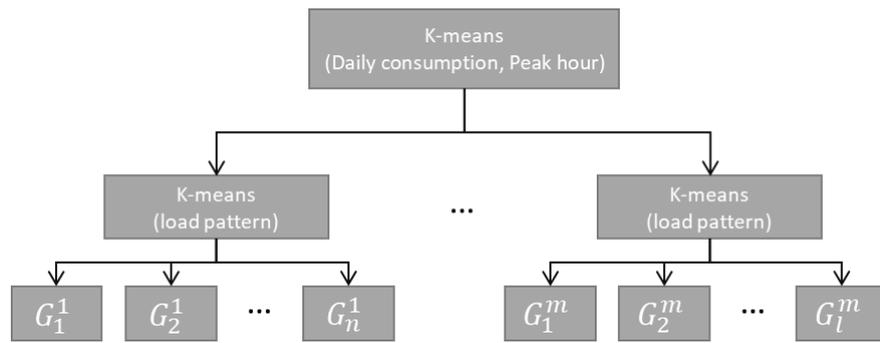
### 3.2. Methodology for Customer Targeting Based on Two-Stage Clustering Method in Efficient DR Operation

The framework used to segment customers into groups based on load profile and to determine appropriate groups for incentive-based DR program participation is depicted in Figure 1. First, load data is collected for load profile clustering. Subsequently, we perform data preprocessing comprising data selection (i.e., exclude weekends, holidays, and event days from the data) and cleansing (i.e., replace missing data and delete incomplete customer data). After data preprocessing, a two-stage load profile clustering is performed to segment residential DR customers in accordance with electricity consumption characteristics and their load profile.



**Figure 1.** Flowchart of the proposed two-stage clustering methodology and customer targeting strategy.

Load profile including information such as peak time, duration, and electricity consumption can estimate approximately how much customers can reduce their capacity, so this information could be an important factor for determining which customers can reduce the most demand during the implementation of the DR program. These characteristics should be extracted from the load profile and treated as variables in the clustering method. Therefore, the characteristics (i.e., daily consumption, peak time) are considered in the first stage of clustering. In the second stage, the classification variable is the normalized load profile. Suitable DR participation groups are then derived by analyzing the segmentation results. Distributions of peak time, average consumption, and peak demand scale could be obtained from this analysis. After selecting the target groups, a DR effect analysis is conducted to verify the effect of targeted enrollment. This analysis shows the demand reduction capacity per customer of the targeted enrollment, and these results are compared with the results obtained assuming opt-in enrollment into the DR program. When the clustering method is applied, considering many variables does not always produce reliable results. Therefore, it is necessary to include the essential variables strategically. However, if there are too many variables to segment customers well, a method to deal with this problem should be devised. In this study, we improve load profile clustering performance by applying our proposed methodology. Figure 2 explains the proposed two-stage load profile clustering algorithm.



**Figure 2.** Two-stage clustering methodology for load profile segmentation.

Before load profile segmentation, load characteristics should be found from load profile by using feature selection (being the process of selection of a subset of relevant features). Features used for cluster input variables are selected through correlation analysis. When we derive relevant features from load profile, we consider factors (i.e., daily consumption, peak time, difference between peak demand and minimum demand) affecting effective DR operation.

The next step is normalization of load characteristics for 1st stage segmentation and load profile for 2nd stage segmentation instead of using raw data. Normalization transforms the load to a number from 0 to 1 and can provide better performance by changing the value of input data. The normalization about load characteristics was conducted on the basis of each variable. On the other hand, the normalization about load profile was used in accordance with each customer. Min–max normalization was used, as illustrated by Equation (4) in the case of load profile normalization:

$$\tilde{d}_{i,t} = \frac{d_{i,t} - \min_{t \in T}(d_{i,t})}{\max_{t \in T}(d_{i,t}) - \min_{t \in T}(d_{i,t})} \quad (4)$$

where  $i$ ,  $t$ ,  $d_{i,t}$ , and  $\tilde{d}_{i,t}$  are customers, time, demand of customer  $i$  at time  $t$ , and the min–max normalization result, respectively.

After the normalization process, segmentation based on load characteristics is preceded by the k-means method before the load profile segmentation as explained in Figure 2. This process separates customers based on their consumption scale and peak times. In other words, it is a process to segment customers over a large range. The reason why these components are chosen is that consumption scale would be an indicator to estimate how much customers can reduce their demand, and the customer's peak time occurrence during an event indicates whether customers stay at home. The next step is customer segmentation based on load profiles, which is conducted for all members of each group following the first-stage clustering analysis. The effect of separation as two-stage k-means clustering is that features can be better reflected as compared to basic k-means.

The main goal of this analysis is to determine a way to produce the most significant effect with suitable customers enrolled in the DR program. To achieve this goal, we propose two standards to select customer groups with high DR potential. If a peak demand event occurs, the likelihood of customers staying in their homes is relatively high. It may be argued that the corresponding customers tend to be able to reduce their demand effectively. However, this is not an absolute indicator. In some cases, for instance, the demand of some customers could be high although peak demand time may not remain constant, or some customers may register an insignificant demand reduction although peak demand times remain constant. Therefore, we stipulate the following criteria to determine the target groups:

1. Customer groups with high demand consumption;
2. Customers groups who have similar peak demand times with an event.

After the load profile segmentation, result analysis through a boxplot chart is adopted as a method of excluding customer groups who are inappropriate customers in DR program participation.

### 3.3. Internal Evaluation of the Clustering Method

After performing customer segmentation via clustering, the accuracy the clustering result should be assessed. Evaluation methods are commonly divided into external and internal processes [28]. In external evaluation, the result is assessed by a comparison with the actual value. Internal evaluation is normally used when the data does not contain actual values; thus, the assessment is based on the idea that good results have minimum distance within clusters and maximum distance between clusters (i.e., high intracluster similarity and low intercluster similarity). Although there are many evaluation methods available, we consider only internal evaluations, since they are used to measure the goodness of clustering evaluation structure without respect to external information (i.e., labels or actual results). Among these, we use the Davies–Bouldin index (DBI) and Dunn index (DI). The DBI is an internal evaluation method to quantify clustering quality. It evaluates customer segmentation based on the similarity between clusters and is calculated as follows:

$$DBI = \frac{1}{n} \sum_{i=1}^n \max_{j \neq i} \frac{\sigma_i - \sigma_j}{d(\mu_i, \mu_j)} \quad (5)$$

where  $n$ ,  $\mu_i$ ,  $\sigma_i$ , and  $d(\mu_i, \mu_j)$  are the number of clusters, the centroid of cluster  $i$ , average distance between  $\mu_i$  and all objects in cluster  $i$ , and the Euclidean distance between  $\mu_i$  and  $\mu_j$ , respectively. Its output is a single number, and clustering algorithms with lower output values indicate better performance.

The DI is another internal evaluation method to quantify clustering quality. The indicator measures how well clusters are separated and how dense they are. It can be formulated as follows:

$$DI = \frac{\min_{1 \leq i < j \leq n} d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \quad (6)$$

where  $d(i, j)$  and  $d'(k)$  are the distances between centroids of cluster  $i$  and  $j$  and between objects within cluster  $k$ , respectively. Its output is a single number, and clustering algorithms with larger output values indicate improving performance.

### 3.4. Cost-effective Analysis

From the perspective of operation research (OR), cost-effective analysis is an important component. The effectiveness of customer targeting through load profile segmentation in DR operation is operation cost reduction. Thus, we need to identify the amount of cost variation compared with opt-in and targeting recruitment. We confirmed it by using the cost-effectiveness test which is one of the economic analysis methods usually performed before public project investment [29]. It is divided into Total Resource Cost (TRC), Program Administrator Cost (PAC), Ratepayer Impact Measure (RIM), and Participant Cost Test (PCT). We considered the PAC test to recognize the cost effect according to DR customer targeting in perspective of DR operator (i.e., utility or ISO). The cost and benefits list should be defined before economic effectiveness estimation. The list for analysis from the perspective of DR operators can be specified in Table 2.

To identify whether the utility project is appropriate for investment, each cost/benefit item should be calculated. If the cost-effectiveness test result has a positive value, it represents that the project has profit. The project is a non-profitable business in the opposite case.

Avoided energy costs is the benefit of decreasing the amount of power purchased in accordance with electricity consumption reduction. It can be formulated as follows:

$$AEB = ER \times ARU \quad (7)$$

where  $ER$  and  $ARU$  are the amount of power reduction and the unit cost of energy avoidance (i.e., average system marginal price (SMP) during DR event), respectively.

**Table 2.** Benefits and cost list for cost-effectiveness analysis in DR operation.

Index	Benefit/Cost Lists
Benefits	Avoided energy costs
	Avoided transmission and distribution cost
	Revenue gain/loss from changes in sales
Cost	Incentives paid
	DR System operation cost
	Measure and evaluation cost
	Marketing and education cost

Avoided transmission and distribution cost is the benefit reducing demand for transmission and distribution construction as a result of decreasing annual peak demand. It can be formulated as follows:

$$ATDB = PR \times (ATU + ADU) \quad (8)$$

where  $PR$ ,  $ATU$ , and  $ADU$  are peak reduction capacity in power system, unit cost of transmission construction avoidance, and unit cost of distribution construction avoidance, respectively.

In the case of the cost list, it contains the cost of revenue loss from changes in sales, incentives, DR system operation, measure, evaluation, marketing, and education. Revenue loss from changes in sales is a cost as the utility company cannot provide power to customers as an amount of DR reduction. Incentive paid cost is cost for utility companies to provide incentives to DR participants as a result of demand reduction. Measurement, evaluation, marketing, and education cost are included in DR operation cost and we assume that these costs are calculated proportionate to the number of DR customers.

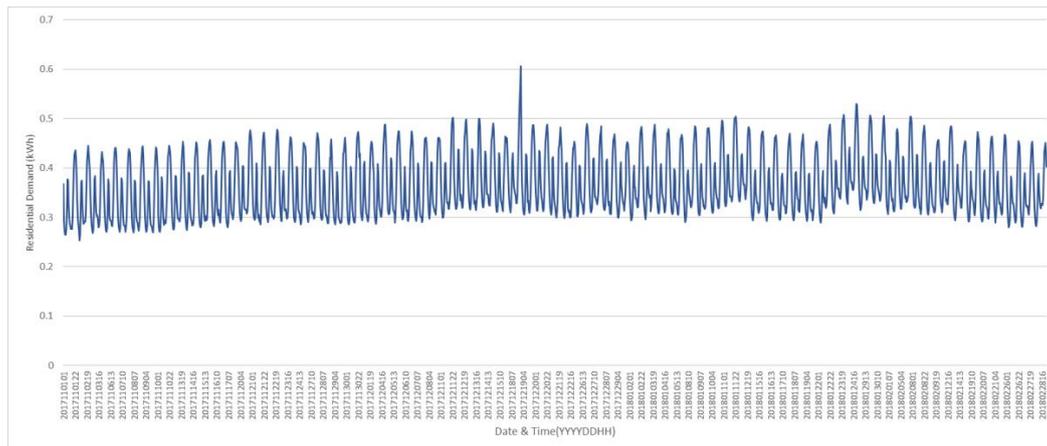
#### 4. Load Profile Segmentation for Effective DR Program Operation in Korea

DR options in Korea have mostly been unavailable to residential customers and have been implemented only for commercial and industrial customers. However, utility companies have recently attempted to attract residential customers by changing their policies and opening DR programs to them. The Korea Electric Power Corporation (KEPCO) which is a utility in Korea also conducted a peak-time rebate (PTR) pilot program from November 2017 to February 2018 in 10 events to develop an appropriate residential DR program in Korea [30]. It was performed with about 800 residential customers living in Seoul, Korea. The PTR program was designed based on incentive-based DR to mitigate peak demand by reducing participant demand in accordance with the utility's notification. After the DR event, the PTR provides incentives based on the amount of demand reduction achieved after participants receive a notification to reduce their demand. It does not have any penalty in the case of the PTR program and can make customers who pay a flat electricity price realize that the electricity price has a time-varying rate system.

Although this PTR program is designed for opt-in customers, targeted enrollment to select residential customers with high DR potential is necessary to improve the benefits of the DR program. Therefore, we analyze residential customer demand data from the PTR pilot program and apply the two-stage clustering methodology discussed in the previous section. From this study, we obtain customer clusters according to load pattern and consumption and select suitable groups for efficient DR operation through an analysis of group characteristics. Finally, we identify the actual demand reduction effect in the case of opt-in operation and targeted enrollment operation by applying residential customer data during an actual PTR event.

#### 4.1. Input Data

This study was conducted using residential demand data. We obtained residential hourly demand data in the Korea Electric Power Cooperation (KEPCO) service area where residents live. This data covered 847 residential customers, all of whom participated in the PTR pilot program. The data covered the period from November 2017 through February 2018, during which time the PTR pilot program operated from mid-January to the end of February. The PTR events occurred throughout nine days from 17:00 to 20:00. The average hourly demand from residential PTR program customers in Seoul, Korea is illustrated in Figure 3.



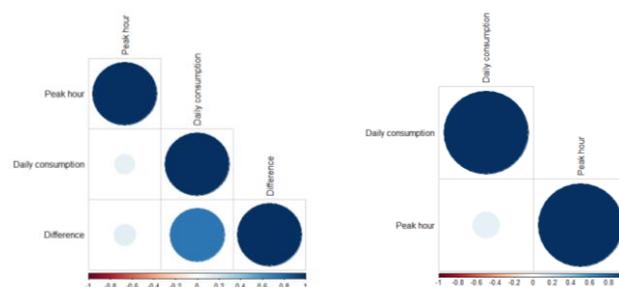
**Figure 3.** Average hourly demand for residential PTR program customers in Seoul, Korea, from November 2017 through February 2018.

#### 4.2. Data Preprocessing: Feature Selection

In the data, preprocessing, missing data imputation, deleting invalid data, selecting eligible days, and reducing dimensions are performed to obtain reasonable results. We conducted a feature selection as part of the dimension reduction for DR potential. When we select features, we consider which factor affects to DR reduction as follows:

1. Is the customer living at home and contributing to peak reduction during the DR event?
2. Are there any incentives to reduce their demand due to large usage?
3. Is the large capacity that can be reduced compared to the base load?

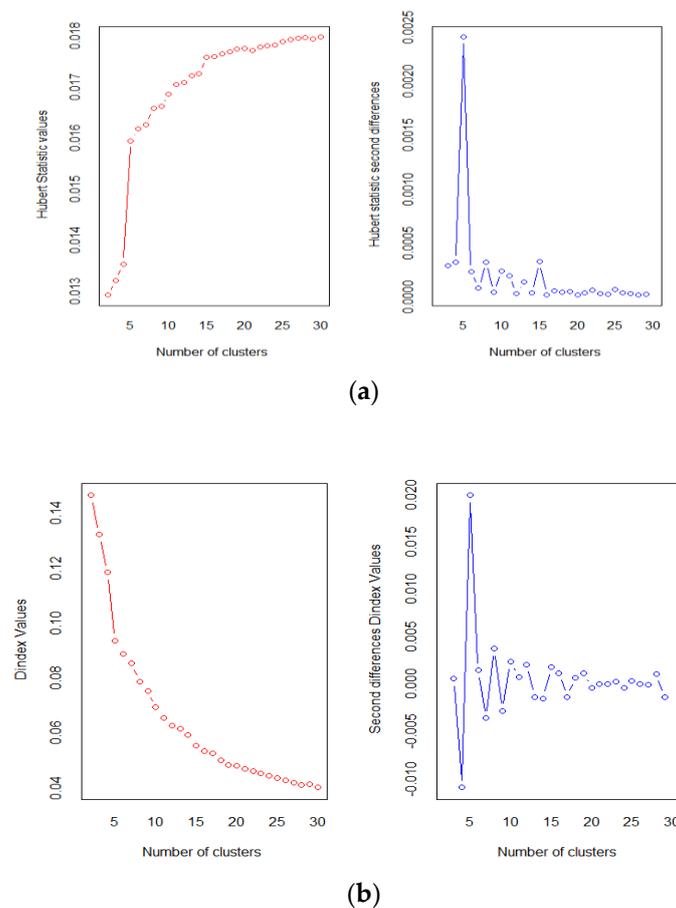
Therefore, we selected three features (i.e., daily consumption, peak hour, difference between maximum and minimum demand) from the load profile as principal factors. Deleting features through correlation analysis between these features should be processed. As a result of the correlation analysis, we selected two features (daily consumption and peak hour) for 1st stage clustering based on demand characteristics. The correlation analysis between demand characteristics' features is illustrated in Figure 4.



**Figure 4.** Feature selection from load profile characteristics of residential customers.

### 4.3. Load Profile Segmentation of Residential DR Customers

When load profile clustering is conducted for customer segmentation, it is essential to determine the optimal number of clusters in the data. We used the NbClust package in the R statistical software to estimate the number of clusters, following Charrad et al. [31]. This package provides 30 indices which determine the number of clusters in a data set and offers the best clustering scheme [32]. Hubert statistics values and Dunn index values are also provided by NbClust. These numbers provide a graphical method to determine the number of clusters. We can realize that the number of optimal clusters is situated in a peak point in their plots of second differences, indicating that the number of optimal clusters is six when first-stage clustering based on demand characteristics (i.e., daily consumption, peak demand time, and difference between peak and minimum demand) was conducted. Figure 5 shows the Hubert index and D index results.



**Figure 5.** Examination methods to determine the number of clusters by (a) the Hubert index and (b) DI index.

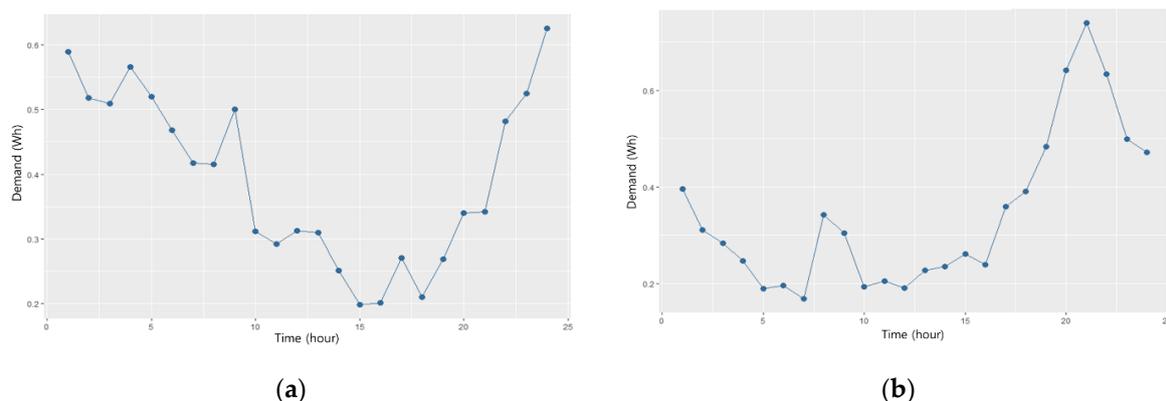
After the first-stage clustering, second-stage clustering for customer segmentation based on demand patterns was conducted. The optimal number of clusters for each of the six groups separated by demand characteristics were 3, 2, 2, 2, 2, and 2. Therefore, we separated residential customers, who participate in the PTR pilot program into 13 groups according to load patterns and consumption.

We performed load profile segmentation through our proposed method and then compared the resulting customer segmentation according to different clustering models. We examined 12 methods, including our proposed method. The remaining clustering methods are based on the fundamental k-means, SOM [16,17,33], and FCM [34–36] methodology in which the classification variables are: (1) demand characteristics, (2) load patterns, and (3) both characteristics and load patterns.

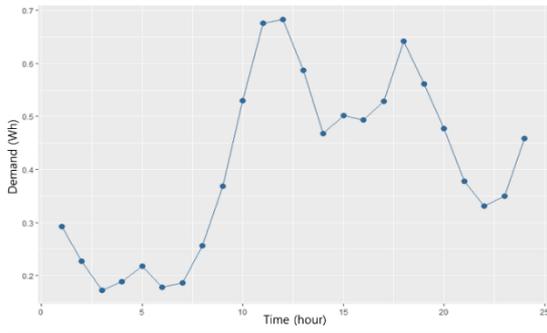
To compare the results, the internal evaluation measures Davis-Bouldin index (DBI) and Dunn index (DI) were used. The DBI and DI result of clustering methods were presented as Table 3. The proposed methodology showed the best result according to Table 3, so we conclude that our proposed methodology is indeed appropriate. We judged that the reason why the proposed methodology has a better result can be explained as follows. It is at a point that separation as two-stage clustering framework can reflect each feature impact, considering that factors affecting DR reduction in 1st stage segmentation make it so that rough load profile clustering before 2nd stage segmentation separates each feature impact by its pattern. Generally, clustering methods separate data based on the distance of input variables. Therefore, the undesirable result would be presented if a lot of input variables which can make each variable effect difficult to verify are used unnecessarily. However, the proposed methodology considers all variables by separating the clustering method into two stages. It can make an outstanding result in two-stage k-means clustering. The proposed method separates residential customers into 13 groups, with the load profiles of each group illustrated in Figure 6. Groups 1 through 13 contain 14, 15, 25, 76, 56, 38, 85, 120, 88, 98, 68, 85, and 79 customers, respectively. The load profile of the 13 groups showed morning peak, evening peak, nighttime peak, and dual morning and night peaks. Residential customers do not usually consume electricity during daytime, so these peak characteristics were consistent with residential load profiles.

**Table 3.** DR operation clustering result evaluation according to variable selection and clustering structure.

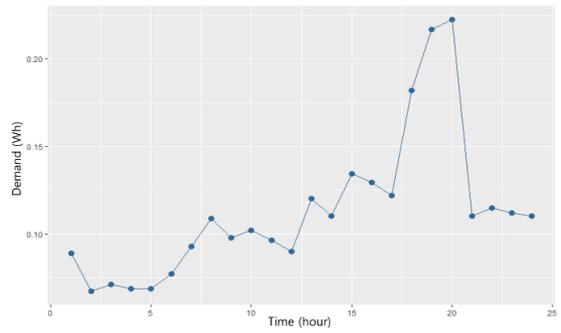
Index	DBI	DI	Ranking
One-stage k-means (Characteristics)	0.5007	1.0084	2/10
One-stage k-means (Characteristics, Pattern)	1.2108	1.2065	3/9
One-stage k-means (Pattern)	1.3493	1.2481	10/7
Two-stage k-means (Characteristics, Pattern)	0.3637	1.8804	1/1
One-stage SOM (Characteristics)	1.2456	1.2298	7/8
One-stage SOM (Characteristics, Pattern)	1.2804	1.3270	9/6
One-stage SOM (Pattern)	1.2669	1.3330	8/5
Two-stage SOM (Characteristics, Pattern)	1.8948	0.7824	11/12
One-stage FCM (Characteristics)	1.2400	1.4659	5/3
One-stage FCM (Characteristics, Pattern)	1.2340	1.5159	4/2
One-stage FCM (Pattern)	1.2437	1.4554	6/4
Two-stage FCM (Characteristics, Pattern)	2.0115	0.7922	12/11



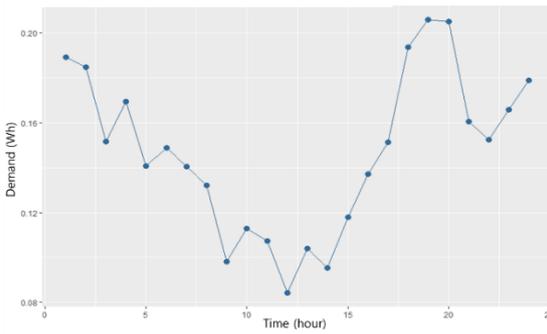
**Figure 6.** Cont.



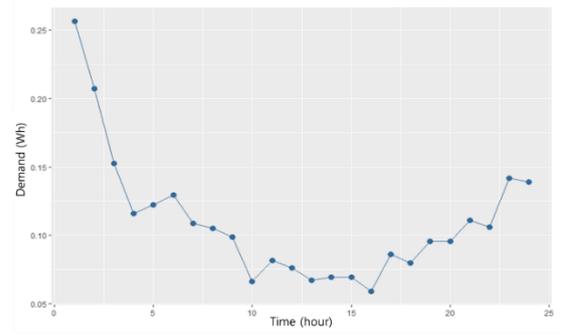
(c)



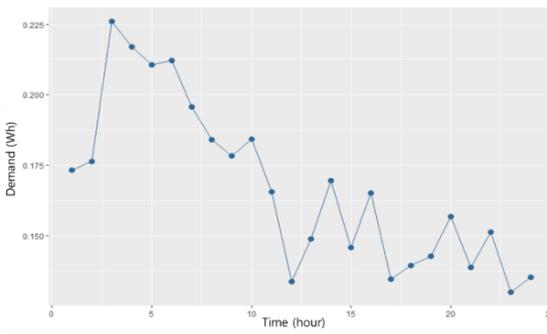
(d)



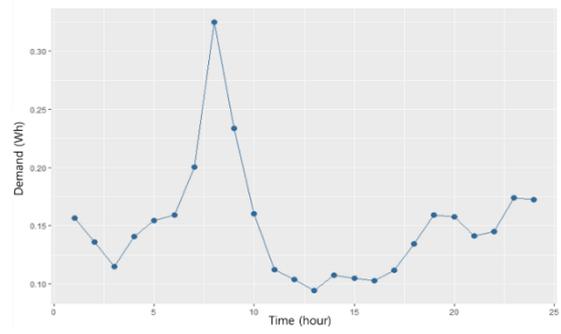
(e)



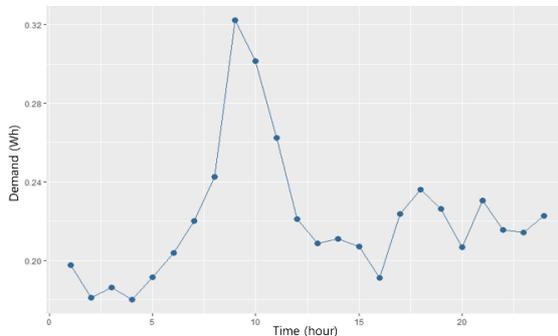
(f)



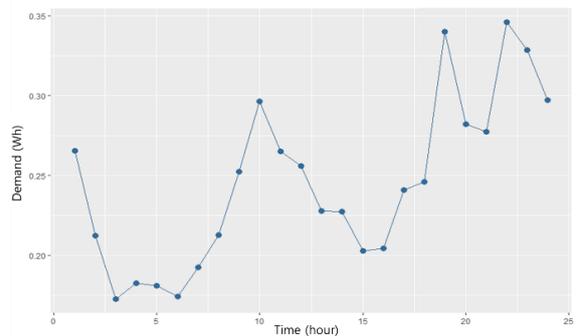
(g)



(h)

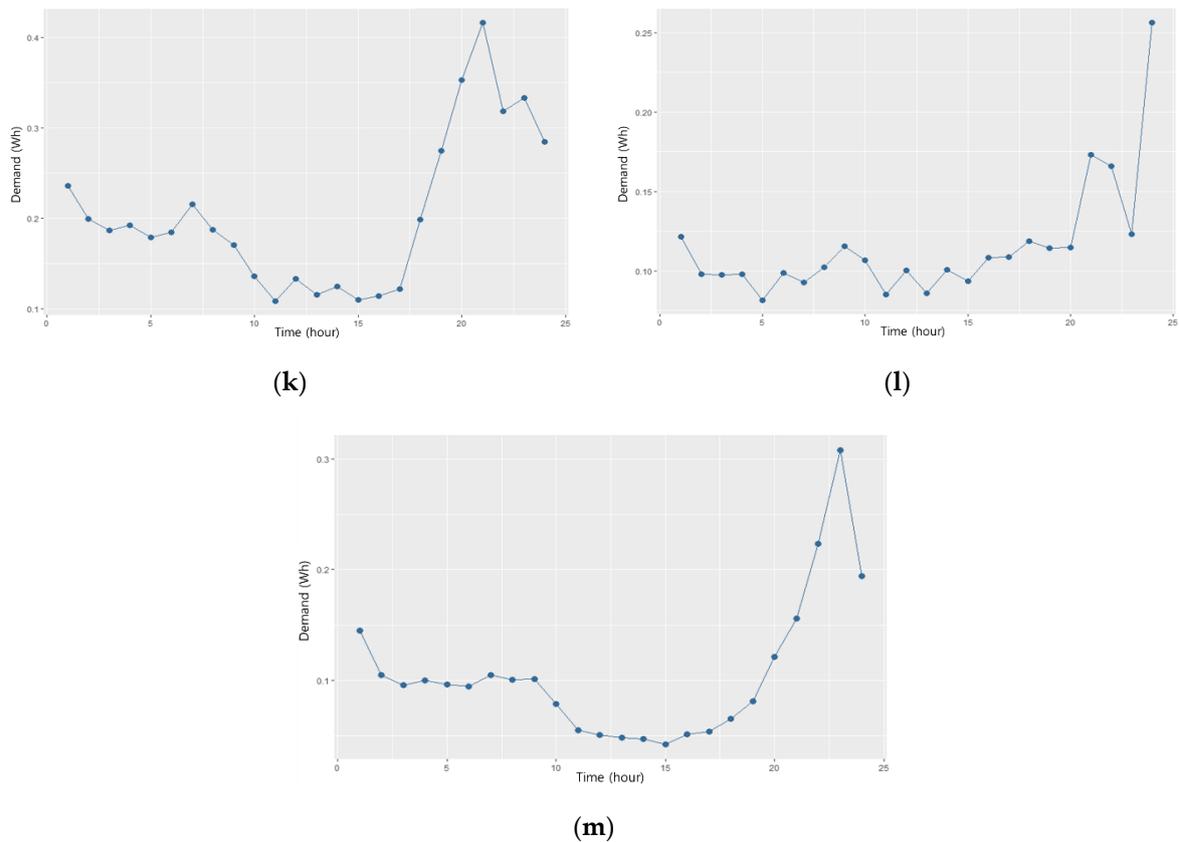


(i)



(j)

Figure 6. Cont.



**Figure 6.** Load profiles of customer segmentation results by the proposed clustering methodology: (a) group 1; (b) group 2; (c) group 3; (d) group 4; (e) group 5; (f) group 6; (g) group 7; (h) group 8; (i) group 9; (j) group 10; (k) group 11; (l) group 12; (m) group 13.

#### 4.4. Customer Targeting for DR Operation

Appropriate customer selection for DR participation by using the load profile segmentation result in the previous section is applied to study efficient DR operation. The 13 load patterns are shown in Figure 7. It can be possible to use load profile and the amount of consumption to estimate DR potential from the 13 load pattern. If customers consume little electricity, their DR participation would be inefficient to a DR operator, despite having a suitable pattern for DR (i.e., nighttime peak and dual morning and night peaks). Therefore, the DR operator considers both factors. To reflect these components, a boxplot analysis is conducted. Peak time (i.e., hour of the day) when maximum demand happens and average consumption boxplots for the 13 groups were analyzed, and they are illustrated in Figure 8.

First, we eliminated groups which experienced inconsistent peak demand occurrence times as the events. The groups not corresponding to this criterion were 6, 7, 8, and 9. Then, groups with low electricity consumption were also deleted, as they are inappropriate for economic purposes. The groups corresponding to little consumption were 4, 5, 12, and 13. We emphasize that utility companies should operate the PTR program using the remaining groups, namely, groups 1, 2, 3, 10, and 11. The total number of customers included in this targeted enrollment scenario was 220.

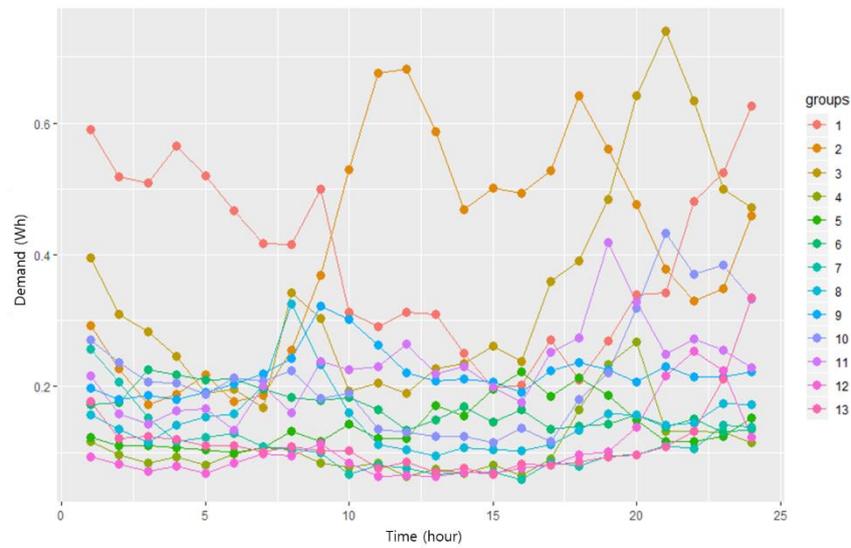
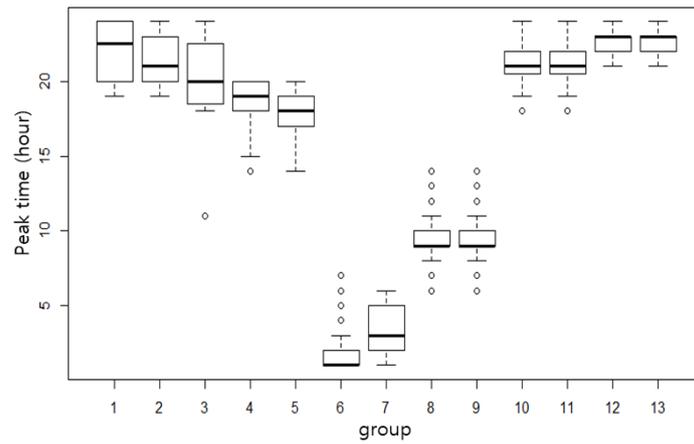
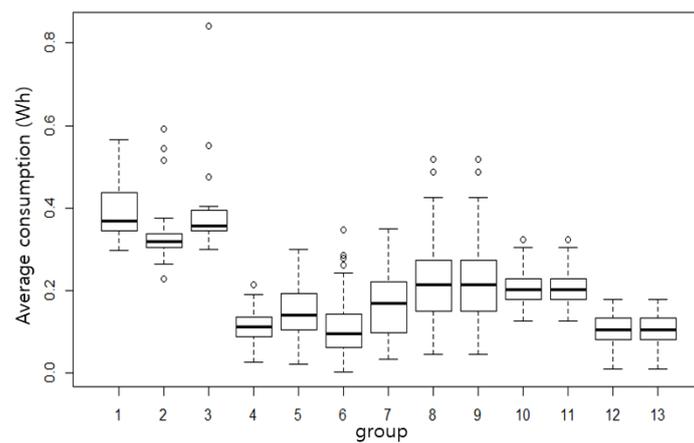


Figure 7. Load profile of customer clusters by the proposed method.



(a)



(b)

Figure 8. Boxplot of the proposed clustering method: (a) peak time, and (b) average consumption.

After finding targeting groups, calculating the amount of demand reduction was performed to identify the effect in accordance with targeting customers by using the actual PTR event data.

There were 847 DR participation customers and there were nine event days when the utility company notified residential PTR customers to reduce demand. We considered that all of residential customer (i.e., 847 customers) are participated in PTR pilot program in case of Opt-in enrollment, and targeted enrollment is attracted by a group of customers who are able to reduce their demand more than other groups during the event. There were 220 customers in the targeted enrollment group, which is different from the number of customers in the opt-in enrollment group. To compare the demand reduction for both types of enrollment, we calculated average demand reduction of event days per customer in both cases. As the customer baseline load (CBL) should be estimated for demand reduction capacity due to the DR event, we applied the Max 4 of 5 method, which has been used for the PTR program in Korea [30]. The Max 4 of 5 method estimates CBL by averaging high demand of four days among five eligible days, which means days excluding weekends, event days, and holidays. Average demand reduction for event days per customer in opt-in enrollment, targeted enrollment, and the 13 groups are illustrated in Table 4. Average demand reductions for the opt-in enrollment and targeted enrollment program were 0.2620 (kWh) and 0.3496 (kWh), and the difference between them was 0.0876 (kWh).

**Table 4.** Average demand reduction for event days per residential customers during the Peak Time Rebate (PTR) pilot program.

Groups	The Number of PTR Participants	Average Demand Reduction (kWh)
Opt-in enrollment	847	0.2620
Targeted enrollment	220	0.3496
Group 1	14	0.3646
Group 2	15	0.4177
Group 3	25	0.3402
Group 4	76	0.2690
Group 5	56	0.2365
Group 6	38	0.1914
Group 7	85	0.2362
Group 8	120	0.2862
Group 9	88	0.2544
Group 10	98	0.3129
Group 11	68	0.3126
Group 12	85	0.1985
Group 13	79	0.1990

The electricity consumption for 6~8PM was 1.3569 (kWh). The demand reduction ratio based on common demand during events was 19.31% and 25.76%, respectively. An improvement of 6.45% was observed, with targeted enrollment reduction increasing demand reduction by 33.44%, in comparison with opt-in enrollment. Thus, it is significantly more efficient to operate the DR program with customers who have larger DR potential, as defined in this study.

Additionally, we conducted a cost-effectiveness analysis for managing the DR program in two cases: residential customers who want to participate in the DR program and targeted residential customers who have large DR potential. We assume that the demand reduction of targeted customers is the same as the actual DR participants in identifying the cost-effectiveness of DR customer targeting. Economic analysis based on the California Standard Practice Manual is performed from the perspective of the DR operator [29]. There were 847 total households participating in the PTR pilot program whose total average reduction is 221.914 kWh, and 635 households (which comprise 75% of the total participants) that we determined as DR targeting participants.

Customer operation cost decreased due to the reduced number of customers, and the amount of increased benefit is 437.256 KRW, the exchange rate is 1100 KRW, marking a 108.58% benefit increase over the existing economic analysis result. The economic analysis changes by customer targeting is presented as Table 5.

**Table 5.** Cost-Effectiveness Analysis Changes by Incentive DR Targeting (Unit: KRW).

Benefits and Cost List		No Targeting Customers (Before)	Targeting Customers (After)
Benefits	Avoided energy costs	861,043	861,043
	Avoided transmission and distribution cost	11,589,254	11,589,254
	Revenue gain/loss from changes in sales	1,335,866	1,335,866
Cost	Incentives paid	4,271,845	4,271,845
	DR System operation cost	1,131,856	848,245
	Measure and evaluation cost	437,926	328,194
	Marketing and education cost	175,248	131,335
	Utility company benefit result	5,097,557	5,534,813
Difference			437,256

## 5. Conclusions

We presented an appropriate DR customer selection methodology for a Korean residential DR program to maximize the DR effect with lower customer enrollment. The proposed method showed better performance than other methods. Our method is divided into two parts. The first is customer segmentation according to load profile and consumption, and the second is targeted group selection based on two standards for DR participation. When we conducted customer segmentation, a two-stage clustering method was introduced. Customers were clustered by demand characteristics as variables in the first stage, and then segmented based on load patterns in the second stage. It can reflect more features of residential demand data than existing clustering methods, that makes better result in customer segmentation. Customer groups were classified as having higher DR potential by peak time and consumption patterns to select adequate groups having large potential in PTR program. As a result, the targeted groups were 1, 2, 3, 10, and 11 in our sample of residential customers in Korea, and their average demand reduction was 0.3496 (kWh), for an improvement of approximately 0.0876 (kWh), which increased savings by 33.44% compared to demand reduction due to opt-in enrollment. The proposed method allowed identifying enhanced DR effects. After the DR targeting demand reduction, we also conducted the cost-effective analysis of the PTR program from the perspective of the DR operator.

As a result, we observed that targeted DR capacity may be achieved with a small number of customers if targeted enrollment is implemented, which can use infrastructure and operation costs effectively. These results provide insights into the efficient use of DR in Korea. The number of customers and total DR capacity of targeted enrollment decreased compared with opt-in enrollment. However, if the number of customers who would like to participate in the DR program is high enough when the official full-scale program starts, selecting optimal customers among them would be more highly important. Therefore, the proposed method would be of great help in ensuring an efficient and economically sensible DR program in Korea.

We considered the residential customer targeting based on customer segmentation in demand response in this paper. Customer segmentation focus on the model structure to reflect features affecting demand response well. Some researches consider clustering model with heuristic algorithm in other areas, so we will apply this concept in further study.

**Author Contributions:** Conceptualization, E.L. and J.K.; data analysis, simulation, and methodology framework development, E.L.; writing, review, and editing, J.K.; supporting data collection and comments for improving the article, D.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by GIST Research Institute(GRI) grant funded by the GIST in 2020, and this work was supported by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20191210301930).

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

## References

1. U.S. Department of Energy (DOE). *Benefits of Demand Response in Electricity Markets and Recommendations for Achieving Them: A Report to the United States Congress Pursuant to Section 1252 of the Energy Policy Act of 2005*; U.S. DOE: Washington, DC, USA, 2006.
2. Wang, Y.; Chen, Q.; Kang, C.; Zhang, M.; Wang, K.; Zhao, Y. Load profiling and its application to demand response: A review. *Tsinghua Sci. Technol.* **2015**, *20*, 117–129. [[CrossRef](#)]
3. George, S.; Bode, J.; Berghman, D. 2012 San Diego Gas & Electric Peak Time Rebate Baseline Evaluation; California Measurement Advisory Council. 2012. Available online: [http://www.calmac.org/publications/SDGE\\_PTR\\_Baseline\\_Evaluation\\_Report\\_-\\_Final.pdf](http://www.calmac.org/publications/SDGE_PTR_Baseline_Evaluation_Report_-_Final.pdf) (accessed on 13 September 2018).
4. Mohajeryami, S.; Doostan, M.; Schwarz, P. The impact of Customer Baseline Load (CBL) calculation methods on Peak Time Rebate program offered to residential customers. *Electr. Power Syst. Res.* **2016**, *137*, 59–65. [[CrossRef](#)]
5. Rhodes, J.D.; Cole, W.J.; Upshaw, C.R.; Edgar, T.F.; Webber, M.E. Clustering analysis of residential electricity demand profiles. *Appl. Energy* **2014**, *135*, 461–471. [[CrossRef](#)]
6. Park, S.; Ryu, S.; Choi, Y.; Kim, J.; Kim, H. Data-driven baseline estimation of residential buildings for demand response. *Energies* **2015**, *8*, 10239–10259. [[CrossRef](#)]
7. Zhang, Y.; Chen, W.; Xu, R.; Black, J. A cluster-based method for calculating baselines for residential loads. *IEEE Trans. Smart Grid* **2016**, *7*, 2368–2377. [[CrossRef](#)]
8. Tureczek, A.; Nielsen, P.S.; Madsen, H. Electricity Consumption Clustering Using Smart Meter Data. *Energies* **2018**, *11*, 859. [[CrossRef](#)]
9. Cominola, A.; Spang, E.S.; Giuliani, M.; Castelletti, A.; Lund, J.R.; Loge, F.J. Segmentation analysis of residential water-electricity demand for customized demand-side management programs. *J. Clean. Prod.* **2018**, *172*, 1607–1619. [[CrossRef](#)]
10. Wang, F.; Li, K.; Liu, C.; Mi, Z.; Shafie-Khah, M.; Catalão, J.P. Synchronous Pattern Matching Principle-Based Residential Demand Response Baseline Estimation: Mechanism Analysis and Approach Description. *IEEE Trans. Smart Grid* **2018**, *9*, 6972–6985. [[CrossRef](#)]
11. Kwac, J.; Flora, J.; Rajagopal, R. Household energy consumption segmentation using hourly data. *IEEE Trans. Smart Grid* **2014**, *5*, 420–430. [[CrossRef](#)]
12. Kwac, J.; Tan, C.-W.; Sintov, N.; Flora, J.; Rajagopal, R. Utility customer segmentation based on smart meter data: Empirical study. In Proceedings of the 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm), Vancouver, BC, Canada, 21–24 October 2013; pp. 720–725.
13. Ozawa, A.; Furusato, R.; Yoshida, Y. Determining the relationship between a household's lifestyle and its electricity consumption in Japan by analyzing measured electric load profiles. *Energy Build.* **2016**, *119*, 200–210. [[CrossRef](#)]
14. Tsekouras, G.J.; Hatziaargyriou, N.D.; Dialynas, E.N. Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Trans. Power Syst.* **2007**, *22*, 1120–1128. [[CrossRef](#)]
15. Mets, K.; Depuydt, F.; Develder, C. Two-stage load pattern clustering using fast wavelet transformation. *IEEE Trans. Smart Grid* **2016**, *7*, 2250–2259. [[CrossRef](#)]
16. Verdú, S.V.; Garcia, M.O.; Senabre, C.; Marín, A.G.; Franco, F.G. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. *IEEE Trans. Power Syst.* **2006**, *21*, 1672–1682. [[CrossRef](#)]
17. Räsänen, T.; Ruuskanen, J.; Kolehmainen, M. Reducing energy consumption by using self-organizing maps to create more personalized electricity use information. *Appl. Energy* **2008**, *85*, 830–840. [[CrossRef](#)]
18. Räsänen, T.; Voukantsis, D.; Niska, H.; Karatzas, K.; Kolehmainen, M. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. *Appl. Energy* **2010**, *87*, 3538–3545. [[CrossRef](#)]
19. Haben, S.; Singleton, C.; Grindrod, P. Analysis and clustering of residential customers energy behavioral demand using smart meter data. *IEEE Trans. Smart Grid* **2016**, *7*, 136–144. [[CrossRef](#)]
20. Granell, R.; Axon, C.J.; Wallom, D.C. Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles. *IEEE Trans. Power Syst.* **2015**, *30*, 3217–3224. [[CrossRef](#)]
21. Labeeuw, W.; Stragier, J.; Deconinck, G. Potential of active demand reduction with residential wet appliances: A case study for Belgium. *IEEE Trans. Smart Grid* **2015**, *6*, 315–323. [[CrossRef](#)]

22. Shunfu, L.; Fangxing, L.; Erwei, T.; Yang, F.; Dongdong, L. Clustering Load Profiles for Demand Response Applications. *IEEE Trans. Smart Grid* **2019**, *10*, 1599–1607.
23. Mingyang, S.; Yi, W.; Fei, T.; Yujian, Y.; Goran, S.; Chongqing, K. Clustering-Based Residential Baseline Estimation: A Probabilistic Perspective. *IEEE Trans. Smart Grid* **2019**, *10*, 6014–6028.
24. Junyuan, X.; Ross, G.; Ali, F. Unsupervised deep embedding for clustering analysis. *Int. Conf. Mach. Learn.* **2016**, 478–487.
25. Kwac, J.; Rajagopal, R. Data-driven targeting of customers for demand response. *IEEE Trans. Smart Grid* **2016**, *7*, 2199–2207. [[CrossRef](#)]
26. Zhou, D.P.; Balandat, M.; Tomlin, C.J. Estimation and Targeting of Residential Households for Hour-Ahead Demand Response Interventions—A Case Study in California. In Proceedings of the 2018 IEEE Conference on Control Technology and Applications (CCTA), Copenhagen, Denmark, 21–24 August 2018; pp. 18–23.
27. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1979**, *28*, 100–108. [[CrossRef](#)]
28. Aghabozorgi, S.; Shirkhorshidi, A.S.; Wah, T.Y. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [[CrossRef](#)]
29. California Public Utilities Commission (CPUC). Demand Response Cost-Effectiveness Protocols. Available online: <https://www.cpuc.ca.gov/General.aspx?id=7023> (accessed on 7 August 2019).
30. Lee, E.; Jang, D.; Kim, J. A Two-Step Methodology for Free Rider Mitigation with an Improved Settlement Algorithm: Regression in CBL Estimation and New Incentive Payment Rule in Residential Demand Response. *Energies* **2018**, *11*, 3417. [[CrossRef](#)]
31. Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A.; Charrad, M.M. Package ‘NbClust’. *J. Stat. Softw.* **2014**, *61*, 1–36. [[CrossRef](#)]
32. Charrad, M.; Ghazzali, N.; Boiteux, V.; Niknafs, A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **2014**, *61*, 1–36. [[CrossRef](#)]
33. Wehrens, R.; Buydens, L.M. Self-and super-organizing maps in R: The Kohonen package. *J. Stat. Softw.* **2007**, *21*, 1–19. [[CrossRef](#)]
34. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [[CrossRef](#)]
35. Cebeci, Z.; Yildiz, F.; Kavlak, A.T.; Cebeci, C.; Onder, H. Ppclus: Probabilistic and Possibilistic Cluster Analysis. 2019. Available online: <https://CRAN.R-project.org/package=ppclus> (accessed on 17 February 2020).
36. Pereira, R.; Fagundes, A.; Melício, R.; Mendes, V.M.F.; Figueiredo, J.; Martins, J.; Quadrado, J.C. A fuzzy clustering approach to a demand response model. *Int. J. Electr. Power Energy Syst.* **2016**, *81*, 184–192. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).