

Article

Electricity Customer Clustering Following Experts' Principle for Demand Response Applications

Jimyung Kang ^{1,2} and Jee-Hyong Lee ^{1,*}

¹ Department of Electrical and Computer Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Korea; E-Mail: jmkang@skku.edu

² Korea Electrotechnology Research Institute, 111 Hanggaul-ro, Sangnok-gu, Ansan 15588, Korea

* Author to whom correspondence should be addressed; E-Mail: john@skku.edu;
Tel.: +82-31-290-7981; Fax: +82-31-299-4687.

Academic Editor: Neville Watson

Received: 15 September 2015 / Accepted: 20 October 2015 / Published: 27 October 2015

Abstract: The clustering of electricity customers might have an effective meaning if, and only if, it is verified by domain experts. Most of the previous studies on customer clustering, however, do not consider real applications, but only the structure of clusters. Therefore, there is no guarantee that the clustering results are applicable to real domains. In other words, the results might not coincide with those of domain experts. In this paper, we focus on formulating clusters that are applicable to real applications based on domain expert knowledge. More specifically, we try to define a distance between customers that generates clusters that are applicable to demand response applications. First, the k -sliding distance, which is a new distance between two electricity customers, is proposed for customer clustering. The effect of k -sliding distance is verified by expert knowledge. Second, a genetic programming framework is proposed to automatically determine a more improved distance measure. The distance measure generated by our framework can be considered as a reflection of the clustering principles of domain experts. The results of the genetic programming demonstrate the possibility of deriving clustering principles.

Keywords: electricity customer clustering; load profile; Genetic programming; demand response

1. Introduction

A smart grid, which is the combination of an electricity network and information network, provides many new applications in traditional electricity generation, distribution, and consumption areas. There can be many potential applications such as adaptive tariff, load prediction, and demand response [1–5]. In these applications, the customer clustering in which similar users are grouped together is considered to be an essential step. When a lot of customers exist in smart grid, it is certain that there are many customers who share similar characteristics and load patterns. This similarity raises the needs for clustering. A customer clustering can be considered to be a procedure that makes a few distinct clusters from a huge number of users. By using 10 representative user groups instead of 100,000 individual users, applications can operate in a more effective and faster way. For example, tariff design requires customer clustering because tariffs cannot be targeted to every individual customer [1,2]. Also, in the field of demand response, a requests for the load shedding can be targeted not to a customer, but to a group of users who will provide similar responses to the load shed request, in order to derive the maximum load shed effect with limited resources [3,4]. The prediction of day-ahead energy consumption could also be improved utilizing customer clustering before actual prediction stage [5]. Furthermore, new applications that need clustering will arise because clustering is considered to be a pre-processing stage in many data analysis scenarios [5]. As the data gathered in the smart grid increases, the importance of clustering will increase because huge amounts of data will need to be reduced in a reasonable way. Decades ago, electricity customer clustering was performed only based on pre-assigned contract types such as household, manufacturer, and school. The clustering of customers is, however, now possible based on real-time energy consumption patterns because of the richness of data in smart grids. Most medium- to high-voltage use customers are already equipped with smart meters that measure power consumption every 15 min, 96 times each day. These daily consumption data are widely accepted as standard input features in customer clustering because they are the most attractive option for characterizing electricity customers. In this study, we also consider daily consumption data as basic input features for customer clustering.

The clustering of electricity customers has been studied by many research groups. Many basic clustering methods such as hierarchical algorithms, the k -means method, and follow-the-leader method are applicable to this topic. These basic clustering methods have been compared with each other in the literature [6,7]. Other clustering methods have also been applied to the clustering of electricity customers. For example, Hino *et al.* [8] proposed use of the Gaussian mixture model (GMM) in clustering. In their work, energy consumption graphs were interpreted as a mixture of Gaussian distributions, and the distance between load profiles was defined by the K-L divergence of the Gaussian mixture distributions. In the literature, other various machine learning techniques such as the self-organizing map (SOM), neural network, support vector clustering, dynamic time warping (DTW), and latent Dirichlet allocation (LDA) have also been applied to cluster electricity customers [9–16]. Furthermore, the transformation of input data to other domains such as frequency domain has been proposed in the literature [17]. Other statistical characteristics such as off-peak factor and night impact factor are also proposed as features for customer analysis [18]. A combination of one or more clustering algorithms was also proposed [19–21]. The combination was usually implemented with additional clustering stages before the final clustering process was activated. Each work which is

presented above has showed that its proposed clustering method could provide better clustering results compared to their previous works in the view of a few general clustering performance indexes.

A common limitation of these previous works is that they focused on general purpose clustering without any consideration of real applications. For example, if customer clustering is needed for a demand response, one of the previous approaches can be utilized. There is, however, no guarantee that experts in the demand response domain will agree with the clustering result determined by general purpose clustering approaches. Figure 1 shows an example of customer clusters that we performed with an agglomerative hierarchical clustering method, which is known to produce outstanding clustering results [7]. The resulting clusters of the hierarchical clustering method consist of a single dominant cluster and many small ones. Despite a promising general cluster performance index from literature [7], this is definitely not the desired result, as a demand response strategy cannot be designed with these clusters. With various numbers of clusters, unacceptable clusters are still generated. This simple experiment shows the limitation of previous works. Since the previous approaches did not consider the requirements of real applications, cluster results might be meaningless groups of scattered points in a hyperplane. The data used in this analysis are the energy consumption patterns in southern Gyeonggi Province in Korea which are also used in the experiment section.

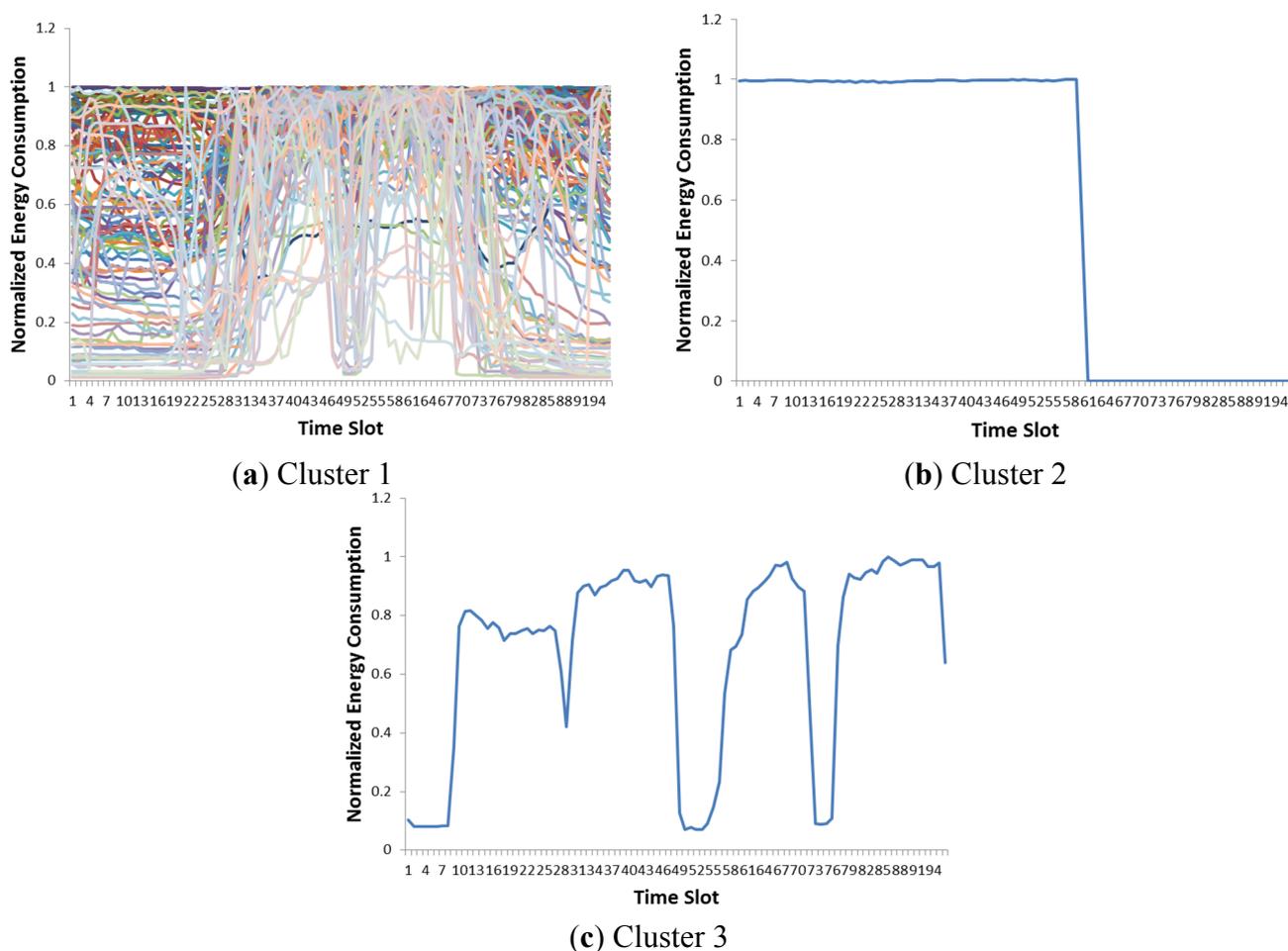


Figure 1. Hierarchical clustering result with the single linkage method (number of clusters, $k = 3$). Hierarchical clustering provides a dominant cluster and other small clusters. (a)–(c) Clusters 1–3.

In this paper, we attempt to develop a clustering method that is applicable to real applications reflecting expert knowledge. More specifically, we propose a good distance measure between customers that fits the customer clustering, especially in demand response applications, based on expert knowledge. This new distance measure, which considers the time-shift characteristic of customers, will be explained and verified in the following sections. Furthermore, we also propose a genetic programming (GP) framework that automatically identifies the clustering principle of the experts. Specifically, we try to derive the ideal distance measure between two customers from expert knowledge using GP.

For validation of the clustering results, we do not use traditional performance indexes, such as CDI (Clustering Dispersion Indicator) [22], VRC (Variance Ratio Criterion) [23], WCBCR (Ratio of Within Cluster sum of squares to Between Cluster variation) [20], SI (Scatter Index) [24], MIA (Mean Index Adequacy) [22], or DBI (Davies-Bouldin Index) [25]. Although these indexes are widely accepted in clustering, they are not proficient in specific applications such as electricity load profile clustering. They do not consider domain knowledge and only focus on the internal structure of nodes. For these reasons, we focus on an external validity index such as entropy, which compares the clustering answer with pre-assigned, ground-truth clusters. Regarding ground-truth clusters, we set up two types of pre-determined clustering answers determined by domain experts and based on the assumption of virtual customers, respectively. The concept of virtual customers will be described in Section 3.

The contributions of this paper can be summarized as follows.

1. We propose the formation of clusters that are similar to those defined by the experts. The clustering result preferred by experts can be quite different from that generated by most of the clustering methods. To our knowledge, this is the first study to focus on real domain expert knowledge. Furthermore, the problem of inapplicable clustering, shown in Figure 1, which was the motivation for our study, is presented.

2. The k -sliding distance, which is a new distance measure, is proposed and validated for electricity customer clustering. This distance is designed based on the time-shift characteristic of loads and is designed to produce an improved clustering result, especially in demand response applications. Furthermore, the k -sliding distance can be utilized with any clustering method because it is a distance measure between nodes rather than a clustering method.

3. We propose a GP framework that is formulated to automatically derive the clustering principle of experts. Specifically, we provide an improved distance measure between two customers from expert knowledge using GP. The results with extensive input features show strong support for the proposed k -sliding distance, at least in a demand response application.

The rest of this paper is organized as follows. In Section 2, a novel k -sliding distance is proposed. We verified the performance of the k -sliding distance with extensive experiments in Section 3. The clustering principle of the demand response experts is analyzed using the GP framework and is validated by experiments in Sections 4 and 5, respectively. The conclusion is presented in Section 6.

2. *k*-Sliding Distance

The proposal of this new distance is based on the time-series pattern of electricity consumption data. Most of the previous studies considered these data as a set of independent variables, which is not valid in real applications. Although 96 independent input dimensions in each customer could be a basic approach for clustering, this is not sufficient because they are interrelated in the time domain. The DTW (Dynamic Time Warping) distance, which is specially designed for time-series data, is an attractive candidate and a simple result of electricity customer clustering with the DTW distance is presented by Iglesias and Kastner [16]. The result, however, only provided the possibility of applying the time-series analysis to electricity customer clustering.

Real electricity applications, especially the demand response application in this paper, have unique viewpoints in defining similarity. Figure 2, for example, shows a few representative samples that must be considered. In Figure 2a–c, there are two customers that are assumed to be similar in our demand response application. In Figure 2a, a small delay in the time dimension can be ignored. Additionally, in Figure 2b, a small difference in usage duration can be accepted. Two customers, who show similar severe fluctuations, as shown in Figure 2c, are also expected to be very similar. In contrast, the two customers in Figure 2d are not similar. Customer #1 can reduce its electricity usage because a sharp drop exists in the pattern. Customer #2 is assumed to be in a different category because it has flat energy usage.

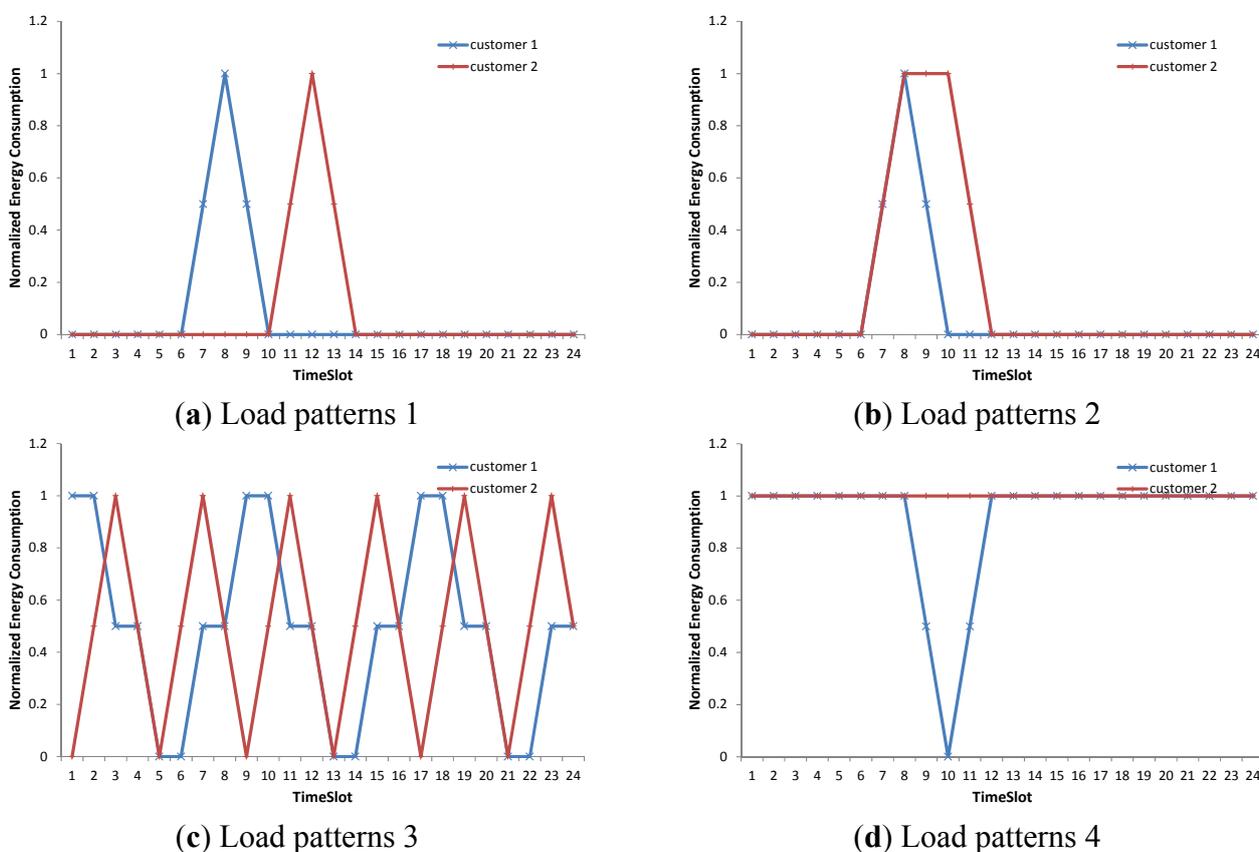


Figure 2. Representative samples for calculation of similarity. Two customers in (a–c) are considered as they are similar. Customers in (d) are not similar. (a)–(d) Load patterns 1–4.

It is certain that the Euclidean distance based on the 96 independent dimensions cannot distinguish these characteristics because it does not consider time-related features. For example, by definition, it provides a long distance in the case shown in Figure 2c.

The k -sliding distance, which we propose here, is an alternative distance measure between two time-series nodes. The definition of the k -sliding distance is presented in Equation (1).

$$Dist^{ij} = \sqrt{\sum_d [(Dist_d^{ij})^2]} \quad (1)$$

where $Dist^{ij}$ is the proposed distance between nodes i and j . $Dist_d^{ij}$ is the distance in dimension d between nodes i and j .

The distance between node i and node j is basically the squared sum of the distance in each dimension, which is calculated using Equation (2).

$$Dist_d^{ij} = \max(\min_{s=-k\dots k} |u_d^i - u_{d+s}^j|, \min_{s=-k\dots k} |u_d^j - u_{d+s}^i|) \quad (2)$$

where k is the sliding limit. u_d^i is the energy usage of node i in dimension d .

The underlying concept of this distance measure is that time can be shifted to some extent, *i.e.*, k time slots. For example, a peak in electricity usage is allowed to float in the k time window without any loss of similarity. This is especially important in a demand response application because its focus is on the peak time and load reducibility. We first consider customer i as a pivot on which to slide the usage pattern of customer j . The distance in dimension d is defined as the minimum distance in the k -sliding window. Because there are two possible pivots, we now have two measured distances in each dimension. The final k -sliding distance in dimension d is defined as the maximum of the two measured distances. The reason why we use the maximum of the two distances can be clearly seen in Figure 2d. If we use customer #2 as the pivot in the 10th time slot, the distance can be zero. If we use customer #1 as the pivot in the 10th time slot, the distance is 1. In the example, we use the larger value as the distance because an asymmetric distance at a certain dimension indicates low similarity.

The difference between the k -sliding distance and the DTW distance is stated here in order to understand the concept of the k -sliding distance. The DTW distance only focuses on the well-shaped, time-warping scenario, *i.e.*, cases Figure 2a,b. In contrast, the k -sliding distance can handle complex, mixed scenarios as well as time-warping scenarios. It only focuses on the nearby similar usage values without considering any time-dependent sequences. Simply, the k -sliding distance looks for nearby points, while the DTW distance looks for matching time sequences.

Table 1 shows the distance measured in the above examples. The Euclidean distance is not acceptable, because it does not consider any time-related aspects. The DTW distance can produce a desirable result only with time-warping scenarios such as those in Figure 2a,b. In more complex scenarios, such as that in Figure 2c, DTW generates undesirable results. On the other hand, the k -sliding distance provides a small value in the cases of Figure 2a–c and provides a large value in Figure 2d, which is the expected result.

The k -sliding distance is not a clustering method but a distance measure, like the Mahalanobis distance. Any existing methods such as k -means and hierarchical clustering can be used with the proposed k -sliding distance. Furthermore, the proposed distance can be applied to any consecutive

series data, as well as pure power consumption values. For example, the change rate of electricity consumption (*i.e.*, slope of graph) extracted from original data can also be considered as input for the k -sliding distance. Therefore, the proposed distance measure can be easily applicable to other features and can be easily applied to other clustering algorithms.

Table 1. Measured distances in the example scenarios.

Distance Type	Figure 2a	Figure 2b	Figure 2c	Figure 2d
Euclidean distance	3	1.5	6	1.5
DTW (Dynamic Time Warping) distance	0	0	3	3
4-sliding distance	0	0	0.25	1.5

3. Verification of k -Sliding Distance

In this section, we experimentally verify the performance of the k -sliding distance.

3.1. Experimental Setup

We used electricity consumption data of one month (September 2011) in high- or medium-voltage use customers who reside in Gyeonggi Province, Korea. The interval of metering is 15 min, which results in 96 samples per day. Weekends and holidays are not included, and load profiles are normalized to their peak values. Two data sets from the northern part of Gyeonggi Province and the southern part of Gyeonggi Province are available as two independent clustering sets. There are 121 customers in southern Gyeonggi Province and 114 customers in the northern part of the province. The data set contains various kinds of customers such as schools, office buildings, and shopping centers. These data are used in the following sections. There are no missing values in the data because the meter reading system we used here is a well-managed test-bed system. Outliers are not eliminated in the experiment because outlier detection itself is not an easy task.

There are two scenarios in the experiments. The first scenario, which is the main scenario in the experiments, is the clustering of customers based on Representative Load Profile (RLP). In this scenario, one month of data for each customer is averaged in order to generate a single RLP composed of 96 dimensions. Based on the graph representation of the generated RLP for each customer, two demand response experts are asked to create clusters based on their knowledge. Figure 3 presents the example clustering result from an expert using the data from southern Gyeonggi Province.

The first expert proposed 10 clusters, as shown in Figure 3, while the other expert suggested 8 clusters. The experts consider many characteristics such as overall shape, fluctuations, peak time, level of consumption, and possibility of energy reduction. The principle of clustering in these experts, however, cannot be clearly stated in the form of mathematic equations because it resides in experts' minds. Furthermore, data themselves affect each expert's individual clustering decision. In other words, the expert clustering principle is highly dependent on the provided data sets. Even when it has been decided that two nodes in a certain data set are not similar, we cannot guarantee that they are still not considered to be similar after new nodes are added to the data set. However, there are still some principles involved in the decision.

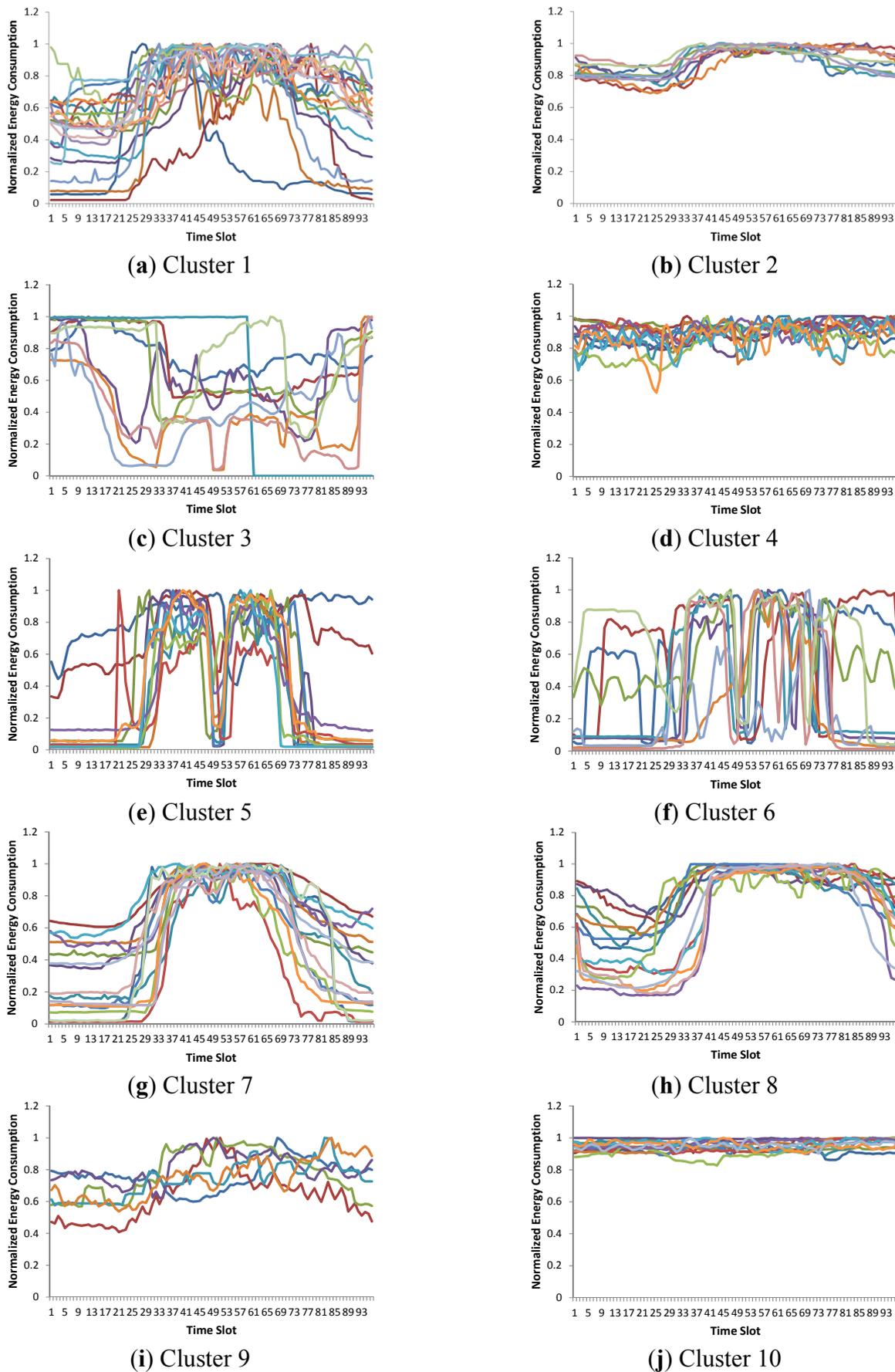
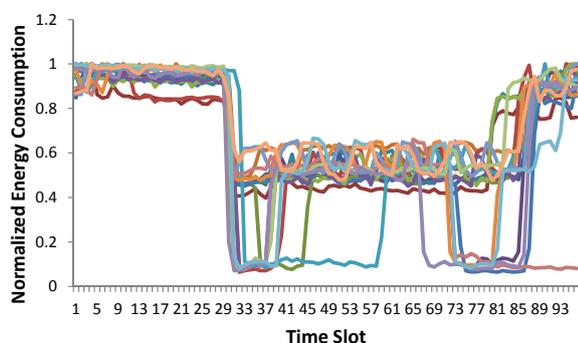
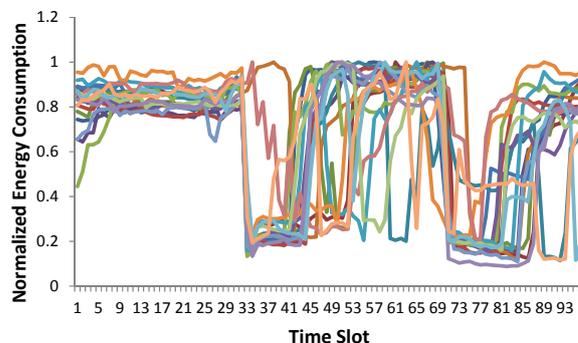


Figure 3. Clustering result by an expert with the data from southern part of Gyeonggi Province. (a)–(j) Clusters 1–10.

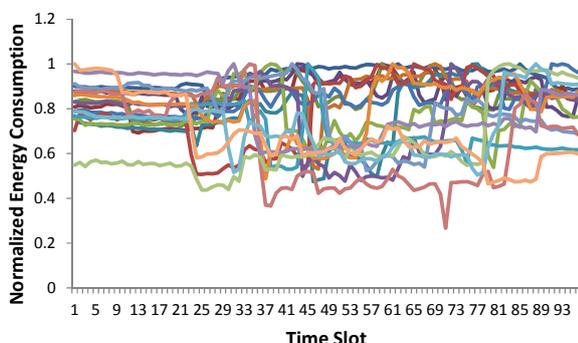
The second scenario is the clustering of virtual customers. In this scenario, every daily load profile of customers is considered to be a virtual customer. For example, 10 customers with 18 working days yield 180 virtual customers. After making clusters with these 180 virtual customers, it will be verified whether virtual customers who are generated from the same real customer are grouped together in one cluster or not. Although it is not a normal scenario in smart grid, we can validate that the proposed distance considers not only experts' viewpoint but also real characteristics of energy consumption pattern because it is certain that virtual customers (daily load profiles) from a customer share common consumption characteristics. Therefore, a correct clustering with virtual customers can be meaningful for a real scenario. The daily load profiles of the 10 randomly selected customers are presented in Figure 4. We chose 10 customers because it is a reasonable number of clusters to show the different patterns of load profiles. Highly coordinated daily load profiles in Figure 4 indicate that it is valid to utilize a single-day profile as a virtual customer.



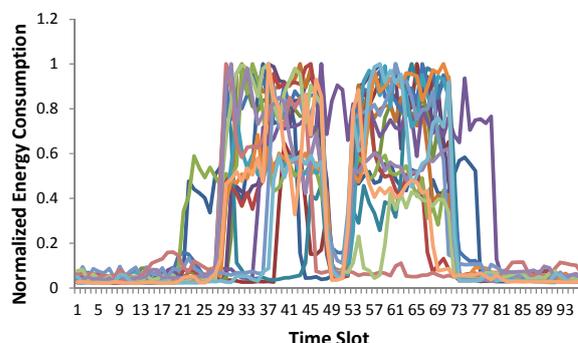
(a) Customer 1



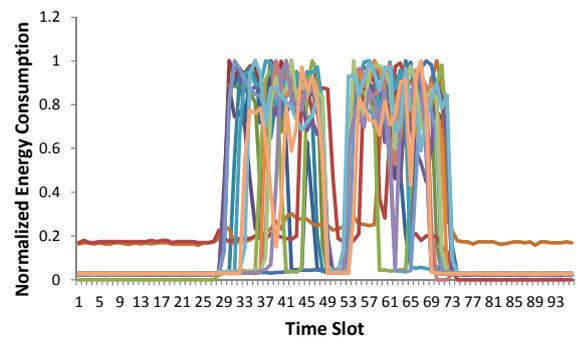
(b) Customer 2



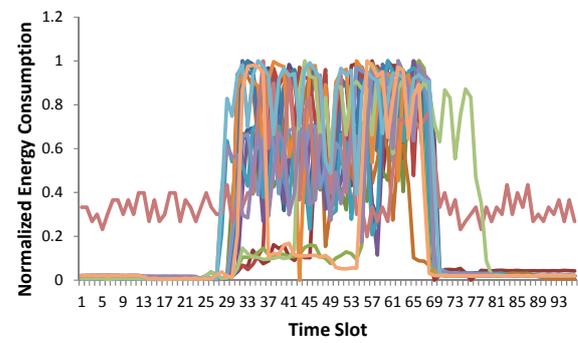
(c) Customer 3



(d) Customer 4



(e) Customer 5



(f) Customer 6

Figure 4. Cont.

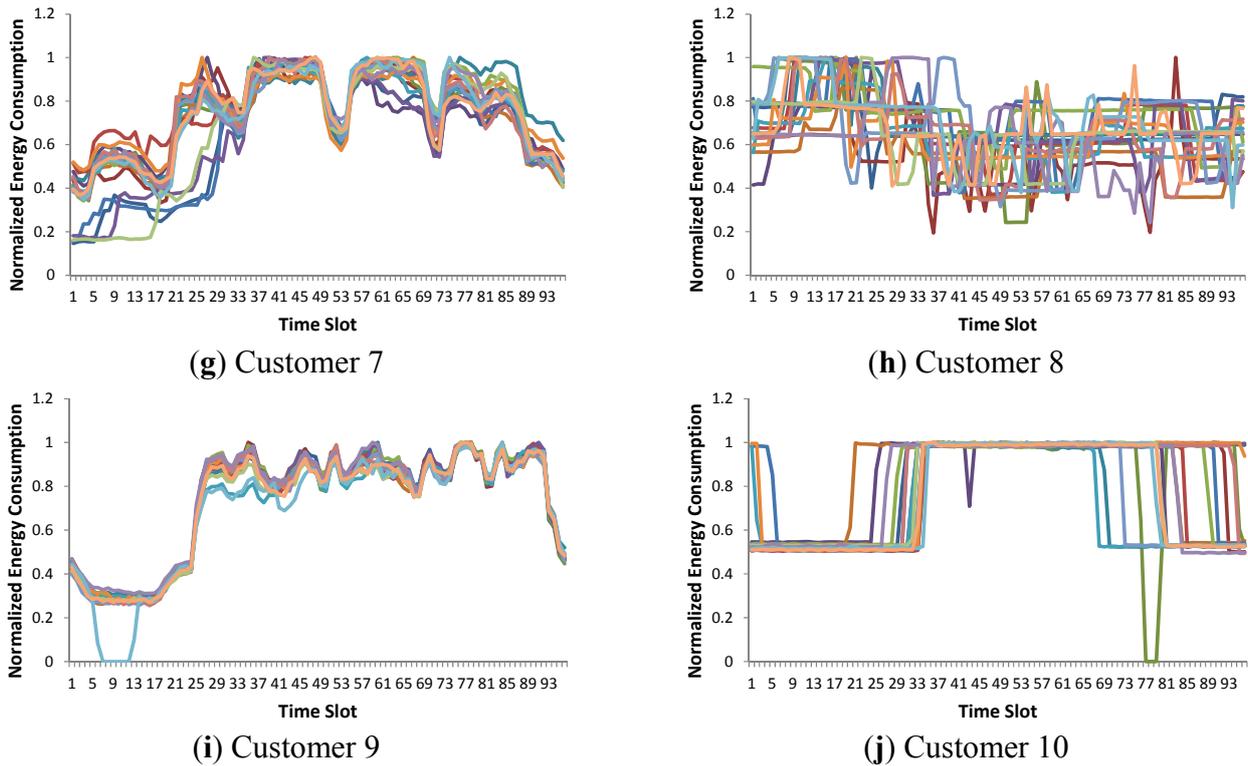


Figure 4. Daily profiles of 10 randomly selected customers. Each line is considered to be a virtual customer. There are a total of 180 virtual customers. (a)–(j) Customers 1–10.

The index of focus in this paper is the entropy of clustering [26]. An external clustering validity index, entropy, is defined in Equation (3).

$$Entropy = \sum_i \frac{|C_i|}{N} \sum_j -\frac{|L_j \text{ in } C_i|}{|C_i|} \log \frac{|L_j \text{ in } C_i|}{|C_i|} \tag{3}$$

where C_i is the i th resulting cluster, N is the number of total nodes, L_j is the load profile of expected cluster j , and $|X|$ is the size of set X .

Entropy represents the similarity of two clustering results: the clustering result and expected result. Entropy is small when the clustering result is similar to the expected result. The smaller is the entropy, the more similar are the clusters. Other external indexes such as purity and Rand index are also available [26]. We only provided an entropy index in this paper because the result does not show a remarkable difference compared to the entropy in our experiments. The expected cluster is the experts’ clustering result in the first scenario. In the second scenario, each expected cluster is composed of a single customer’s daily profiles, as shown in Figure 4.

The CDI [22], which is a commonly used clustering performance index defined in Equation (4), is also considered for the purpose of comparison. It prefers long inter-cluster distance and short intra-cluster distance. The smaller is the index value, the better is the result of clustering. We ignored the other indexes mentioned in Section 1 because the focus is not on these indexes. Furthermore, the results of other internal validity indexes are, to some extent, similar to that of the CDI index in our simulation and in the literature [7].

$$CDI(k) = \frac{1}{d(C)} \sqrt{\frac{1}{k} \sum_k d^2(C_k)} \quad (4)$$

where $d(C)$ is the average distance between cluster centers, $d^2(C_k)$ is the squared average distance between nodes in cluster k .

The experimental result, which has randomness (*i.e.*, k -means clustering), is the average of 100 executions. The clustering methods that we used are hierarchical clustering with average linkage and k -means clustering. The reason why we analyze k -means clustering and hierarchical clustering is that k -means is the most common algorithm in clustering methods and hierarchical method seems to be one of the most promising methods on electricity customer clustering in the literature [7]. Other various clustering algorithms are not presented here because we are focusing on the distance between nodes, not the clustering algorithm. We included the GMM distance [8] as a baseline because it defines the distance between nodes. Ward's linkage [27] is used with the result of the GMM, as proposed in [8].

3.2. Experimental Result

The scenario involving the RLP (Representative Load Profile) of each customer is conducted first. The CDI indexes of the southern and northern regions are presented in Figures 5 and 6, respectively. The entropy index is presented in Figures 7–10. As the entropy index is based on the experts' results, the experiment is conducted for two regions (the southern and northern parts of Gyeonggi Province) and for two experts, resulting in the production of four graphs. We changed the number of clusters from 3 to 11 in all of the experiments.

The first observation is that the hierarchical clustering methods consistently produce promising results in terms of CDI index (Figures 5 and 6) and consistently produce poor results in terms of entropy index (Figures 7–10), regardless of the number of clusters. Based on the experiments, we conclude that hierarchical clustering cannot be applied to electricity customer clustering in a demand response because of the extremely poor entropy index, which is more important than the internal validity index in the real world.

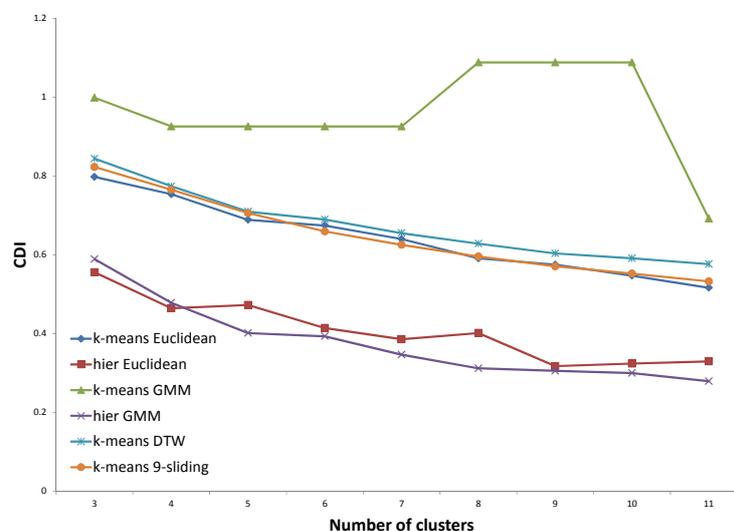


Figure 5. CDI (Clustering Dispersion Index) result for the southern part of Gyeonggi Province.

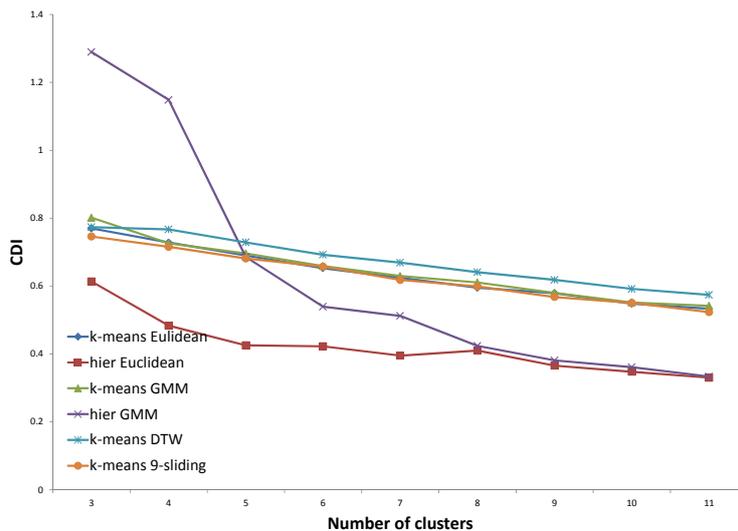


Figure 6. CDI result for the northern part of Gyeonggi Province.

The second observation is that the proposed k -sliding distance with k -means clustering provides remarkably improved entropy compared to the Euclidean distance or GMM distance in all cases. When we compare it to the DTW distance, the proposed distance is superior in most of the cases. The DTW distance provides slightly lower entropy than the k -sliding distance in a few exceptional cases, as shown in Figure 8. In the other three cases, *i.e.*, those in Figures 7, 9 and 10, the proposed distance always outperforms the DTW distance. The experimental results imply that the experts actually consider the k -sliding distance to be an important clustering principle even though they do not mention it explicitly. They think that the energy usage time associated with this measure might drift. It is obvious that the time series characteristic of data must be considered because k -sliding and DTW measures show an improved index compared to the Euclidean distance measure. On average, the overall entropy improvement of the proposed k -sliding distance is 9.2% and 2.6% compared to the Euclidean distance and DTW distance, respectively. The improvement is quite meaningful when we consider that entropy is calculated based on the log function. We set parameter k in the k -sliding distance to 9 in the above experiments, and the selection of k is based on the experiment presented in Figure 11.

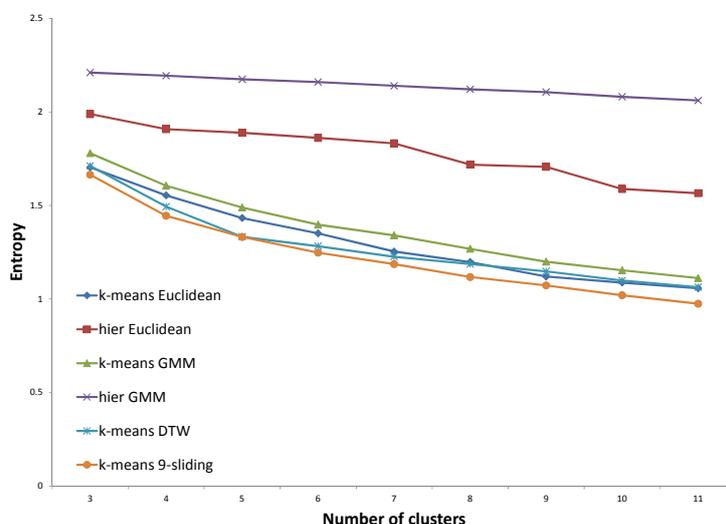


Figure 7. Entropy results for the southern part of Gyeonggi Province according to expert #1.

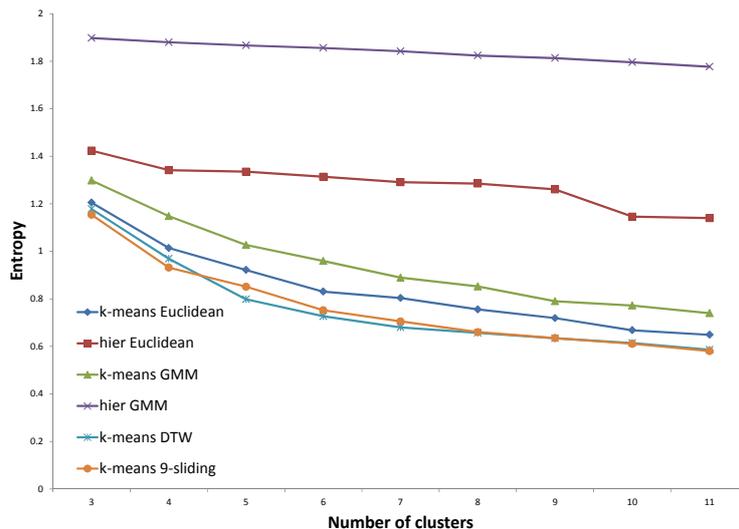


Figure 8. Entropy results for the southern part of Gyeonggi Province according to expert #2.

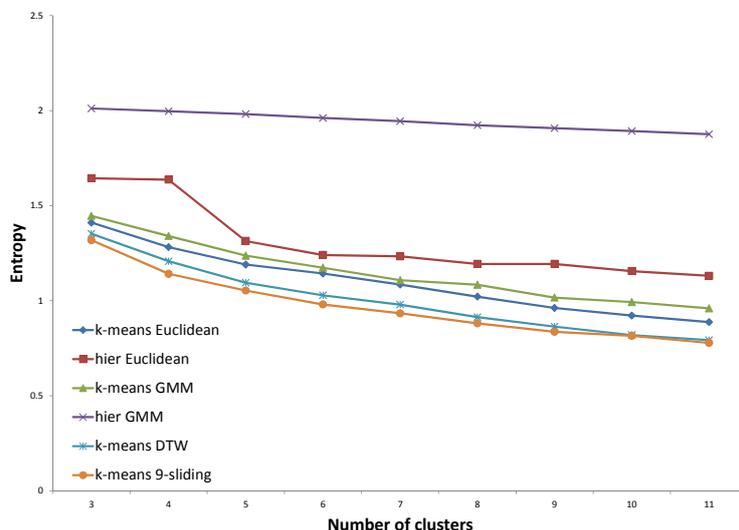


Figure 9. Entropy results for the northern part of Gyeonggi Province according to expert #1.

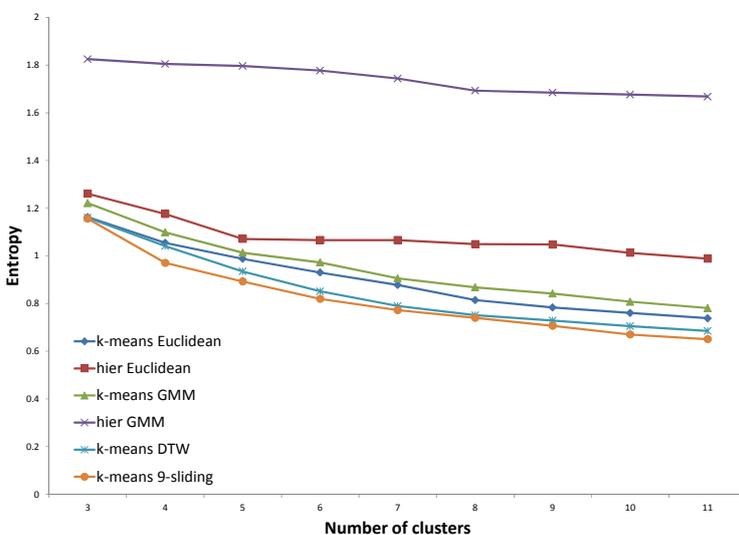


Figure 10. Entropy results for the northern part of Gyeonggi Province according to expert #2.

In Figure 11, we tested the effect of k in the k -sliding distance using the data from the southern part of Gyeonggi Province. The results show the best performance in most cases when we set k to 9 or 11. In the other three cases, the best performance was achieved with a k value of 7, 9 or 11. If we change the k value according to the data set, we can further improve the result because the optimal value is somewhat dependent on the data set and the number of clusters. We, however, fixed k to 9 in the experiments for simplicity, and the result still shows a much-improved result in most of the scenarios. Nine slots can be interpreted to represent 2 h and 15 min because each slot represents 15 min. In other words, we can assume that the experts consider two and half hours to be an acceptable time drifting limit.

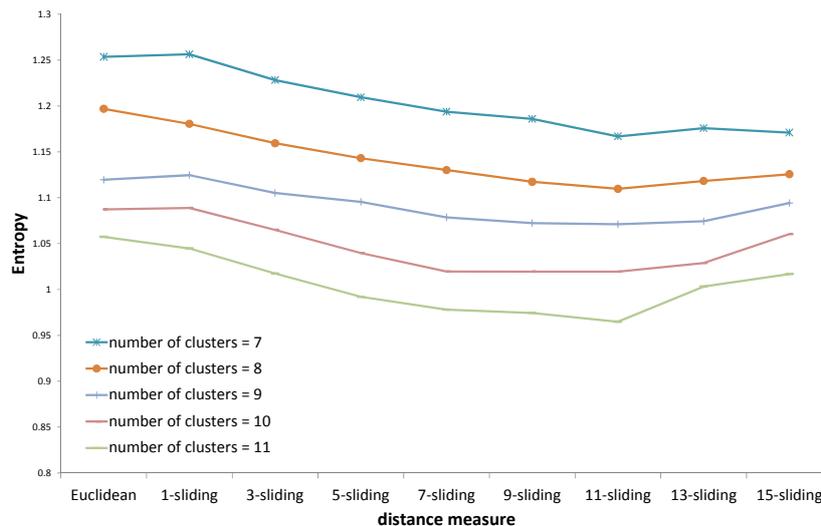


Figure 11. k -sliding distance with various values of k .

The CDI and entropy results of the second scenario, which is composed of virtual customers, are presented in Figures 12 and 13. The second scenario is another important experiment because an expert might be biased in his or her decision. Virtual customers have an inherent clustering answer because they are daily profiles from real customers, *i.e.*, they are independent of experts.

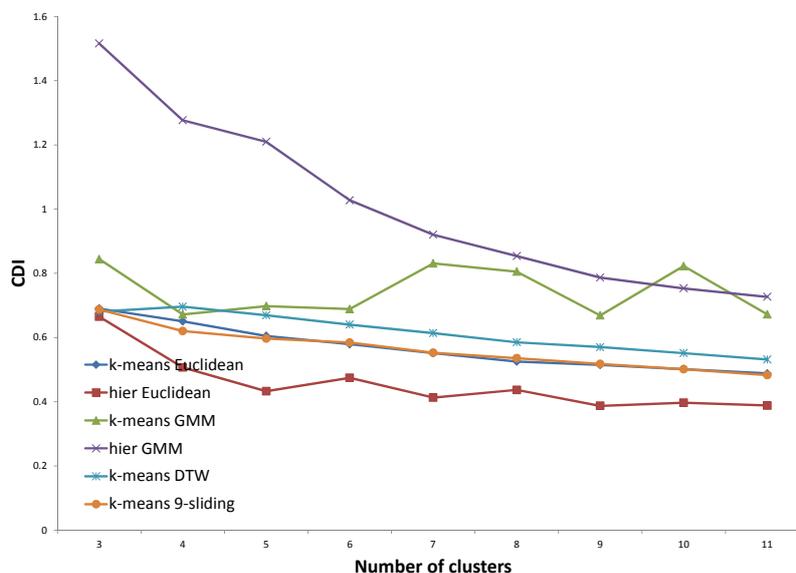


Figure 12. CDI results of virtual customers.

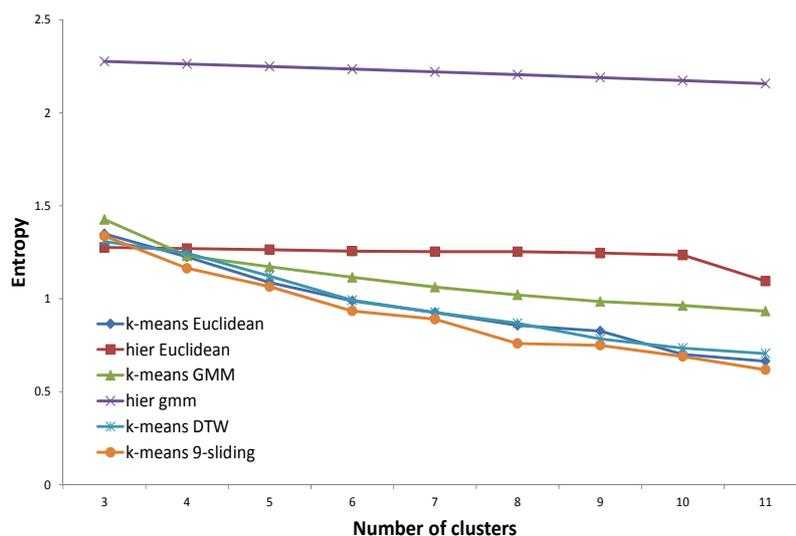


Figure 13. Entropy results of virtual customers.

In Figures 12 and 13, the performance of the second scenario is not different from that of the first scenario. The CDI result of the proposed distance, shown in Figure 12, delivers relatively poor performance compared to the hierarchical method with Euclidean distance similar to Figures 5 and 6. It seems that the hierarchical clustering method may generate a few large clusters and many small ones, which can score highly with CDI index even though it cannot be applicable to real applications as we mentioned in Section 1. The proposed k -sliding distance, however, showed the best performance in terms of entropy index regardless of the number of customers (Figure 13). Considering the results of the first and second scenarios, we conclude that the k -sliding measure is meaningful for real applications, at least for the demand side management.

4. Finding Clustering Principle Using Genetic Programming

In Sections 2 and 3, we proposed and verified a new distance measure that is suitable for demand response applications. Next, we want to know the clustering principles of experts based on the collected data. The principles of clustering can be translated to the definition of the distance measure between two nodes because experts utilize their own combined distance measure. This distance must include many characteristics such as average usage, peak time, and fluctuations. If we know the way to formulate these clustering principles, we can utilize the framework for any application in a smart grid.

More specifically, in this section, we design a GP framework that derives a distance measure from an expert or a group of experts in a demand response application. Although we focus on demand response in this study, this framework can also be applied to other applications. The goal of GP is to produce the distance measure between two nodes that optimizes entropy. The GP technique looks for a tree that represents a distance equation. The tree will automatically evolve to determine the best entropy value with the help of crossover and mutation.

The example tree that is used in this paper is provided in Figure 14. A leaf node consists of distance options and weights. Examples of distances are the proposed k -sliding distance and the Euclidean distance. A non-leaf node consists of a binary operator such as summation, subtraction, maximum, and minimum. The generated distance measure shown in Figure 14 is the sum of the k -sliding distance in

the frequency domain with a weight of 0.3 and the Euclidean distance in the time domain with a weight of 0.8. A total population of 100 and 50 generations are considered in the GP. The depth of the tree was limited to three in order to avoid over-fitting and in order to easily analyze the results. Tournament selection is used as the selection method, and the parameters related to crossover and mutation are applied in an empirical manner. The *k*-means algorithm is used for clustering.

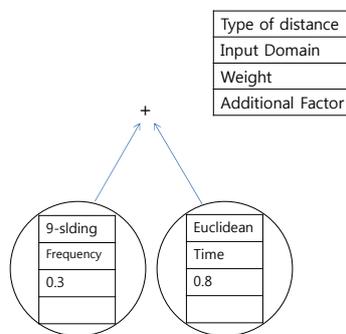


Figure 14. Example tree of genetic programming. The presented distance measure is the sum of the *k*-sliding distance in the frequency domain with a weight of 0.3 and the Euclidean distance in the time domain with a weight of 0.8.

The design of the GP framework is based on the idea that the optimal distance does not consist of a single component. The distance considered by the experts might be the combination of many distances. In order to include various possible distances, we considered input domains such as time, frequency, and slope, as well as distance options such as Euclidean, Pearson correlation, and *k*-sliding. Weights and a few additional factors are also included. The parameters of each node in GP are described in Table 2.

Table 2. Parameters of nodes in genetic programming.

Node Type	Parameters
Leaf (Distance between two customers)	Distance Option Distance between statistical values, (average, standard deviation), Euclidean distance, DTW distance, Pearson correlation coefficient distance, <i>k</i> -sliding distance
	Input Domain Time domain, Frequency domain, Slope domain
	Weight 0~1
	Additional Factors Unity operator: sin, cos, square, 1/x, square root, Peak parameter: peak-weight, peak-penalty
Non-Leaf	Summation, Subtraction, Maximum, Minimum, Average, Multiplication

Peak-weight: additional weight included if energy usage is close to its maximum; peak-penalty: additional weight included if energy usage is close to its minimum.

5. Assessment of Distance According to Genetic Programming

Unlike ordinary clustering, both training and test data are needed because the result of the GP must be verified by another test set. This is because GP inherently uses the result in the training phase.

We used two scenarios for verification of the GP framework. The first one is determination of the principle of each expert. In this scenario, both test and training data are generated by the same expert. In other words, the answers of expert #1 for the southern region are input into the GP as training data,

and the resulting tree is used as the distance in the *k*-means clustering for the northern region. As a result, the entropy for the northern region is analyzed. There are four training data sets available, those of two provinces and two experts.

The second scenario is designed to find a common principle of both experts. In this scenario, the answers from both experts #1 and #2 for the southern region are input into the GP as a training set. The resulting tree is used as the distance in the northern region. A scenario using the northern region as a training set is also possible.

From Figures 15 to 18, each tree resulting from the first scenario is presented. The first common observation is that none of the trees include the Euclidean distance. This result indicates that people do not see the load pattern in the hyper dimension as does a computer.

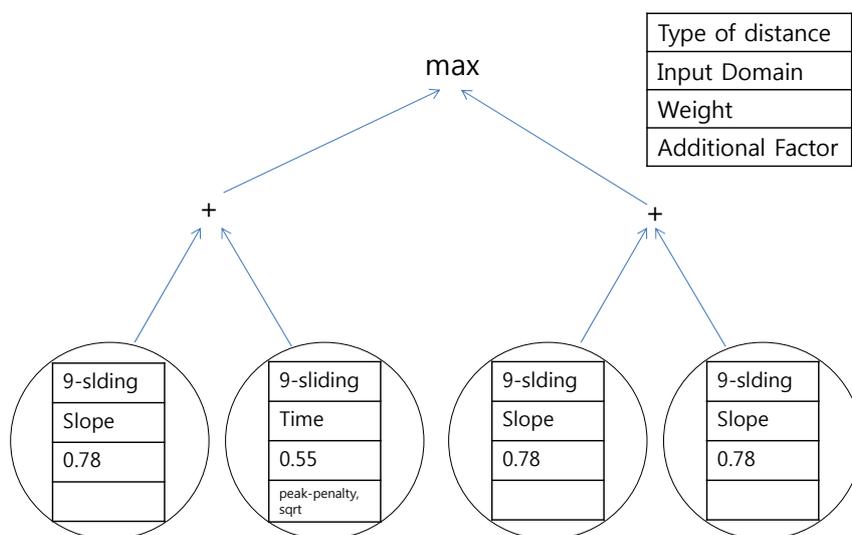


Figure 15. Proposed tree for the southern region using the data of expert #1.

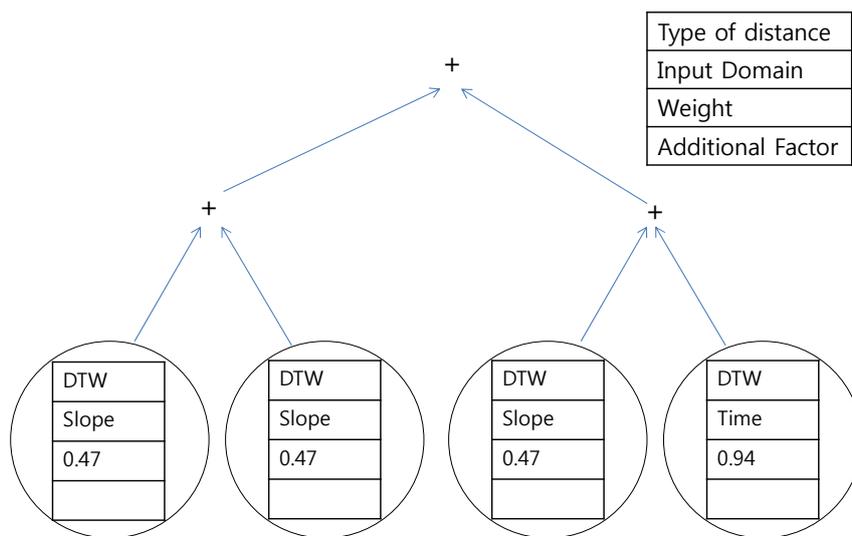


Figure 16. Proposed tree for the southern region using the data of expert #2.

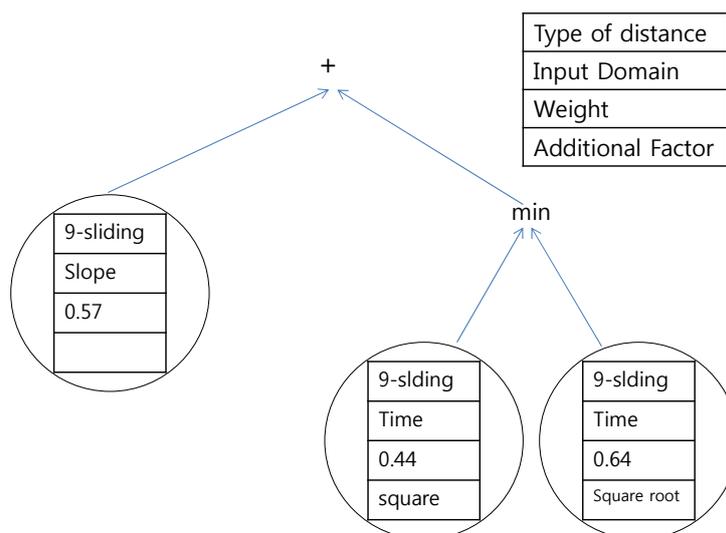


Figure 17. Proposed tree for the northern region using the data of expert #1.

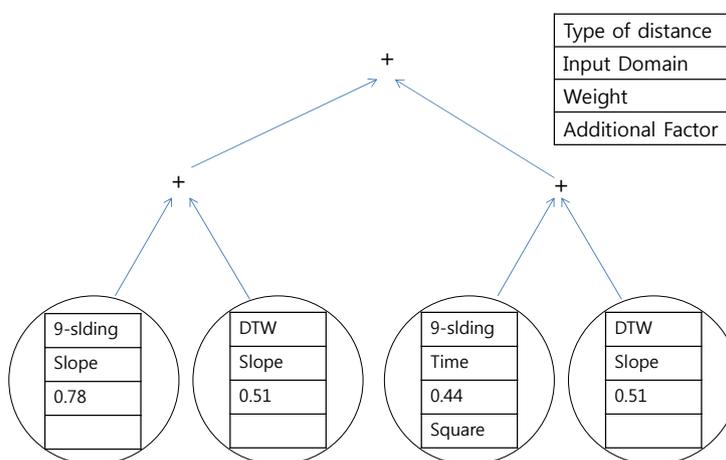


Figure 18. Proposed tree for the northern region using the data of expert #2.

The second observation is that the results are mostly composed of the *k*-sliding distance. In most cases (Figures 15, 17 and 18), *k*-sliding was the main concern in the tree structure. Exceptionally, DTW is suggested in Figure 16. The explanation for this is that expert #2 specifically focused on the DTW distance in the southern region.

Another important observation is that the frequency domain input is not utilized at all. The data in time domain was mostly selected by GP. The slope domain, which denotes the sequence of usage differences, also plays a role in the distance selection according to our experiment. From the results, we conclude that the *k*-sliding distance in the time domain and slope domain is actually meaningful to an expert’s decision.

The trees resulting from the second scenario, which looks for the common principles of both experts, are presented in Figures 19 and 20. The results are similar to those of the first scenario except that all of the distances are composed of the *k*-sliding distance. A combination of the *k*-sliding distance in the time domain and the slope domain is generated as the most suitable distance.

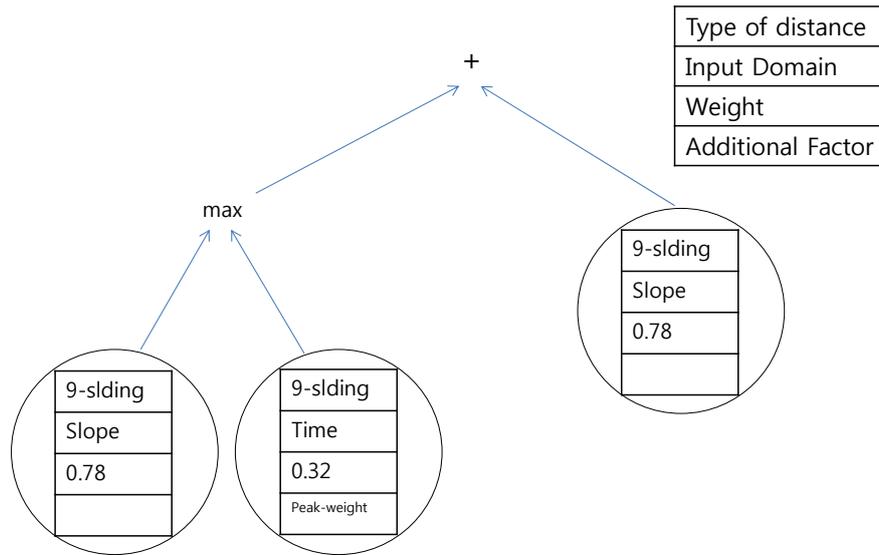


Figure 19. Proposed tree for the southern region using the data of experts #1 and #2.

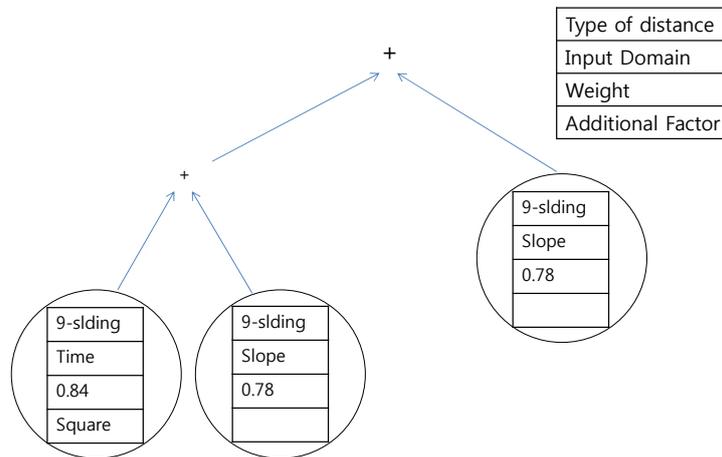


Figure 20. Proposed tree for the northern region using the data of experts #1 and #2.

The entropy results of the distances proposed by the GP are presented in Figures 21–24 for each data set.

Five distance options are presented in each figure. The Euclidean distance and *k*-sliding distance are presented as baselines. Three cases are available for GP verification. The first one is the distance that is trained with the same expert’s answer. The second case considers the distance trained with the other expert’s answer. The distance trained with both experts’ answers is considered in the third case. The distance proposed by the same expert’s decision will illustrate the principle of an expert’s clustering decision, and that proposed by both experts’ decisions will illustrate the principle of the expert group. The optimized entropy in the process of GP is not presented here because we need to verify the result on a test set, even though the optimized entropy was much smaller than all other results of the test data.

In Figure 21, all three distances proposed by the GP provided improved results compared to that of baselines. Specifically, the distance proposed by the same expert’s decision shows consistently lower entropy values. In Figure 22, the distance trained with the same expert constantly shows improved entropy results compared to those of the baselines. The distance trained with the other expert shows slightly improved results, while the distance with both experts provides similar results compared to

those of the k -sliding distance. In Figure 23, the distance trained with the same expert clearly provides an improved entropy performance, demonstrating 16.1% and 4.7% improvement compared to the Euclidean distance and k -sliding distance, respectively. The distances with the other expert and both experts, however, provided slightly degraded performances compared to those of the k -sliding distance.

In Figure 24, for the northern region using the data of expert #2, all three distances determined by GP provided a degraded performance compared to that of k -sliding. Even the distance trained by the same expert’s answer does not show an improved performance. This exceptional case stems from the fact that the generated distance consists of only DTW distances, according to the tree presented in Figure 16. The result could be considered to be over-fitting for a special case. This could also happen if the expert’s clustering principle drifts according to the input data set. Either way, from this result, we conjecture that expert #2 has a relatively fluctuating principle compared to expert #1.

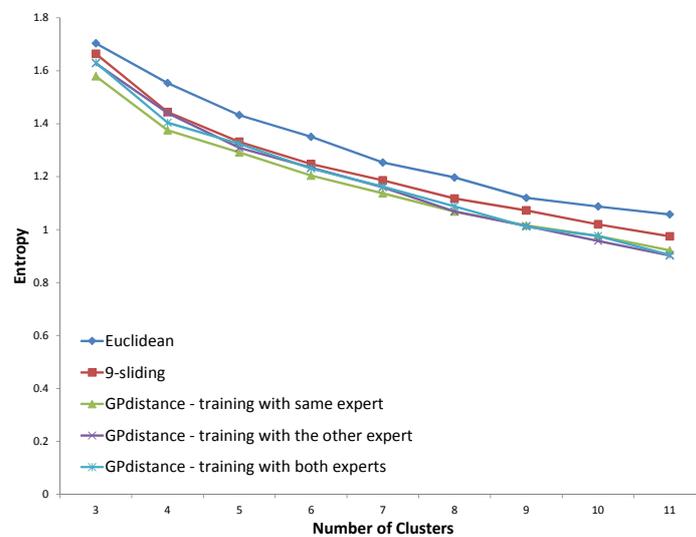


Figure 21. Entropy results for the southern region using the data of expert #1.

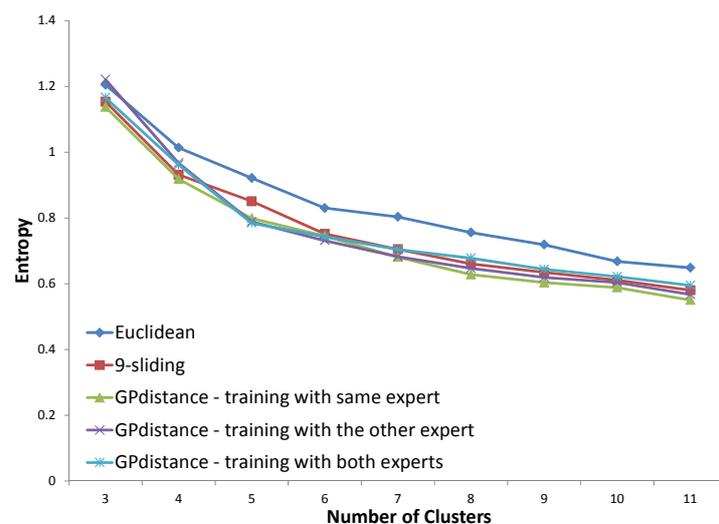


Figure 22. Entropy results for the southern region using the data of expert #2.

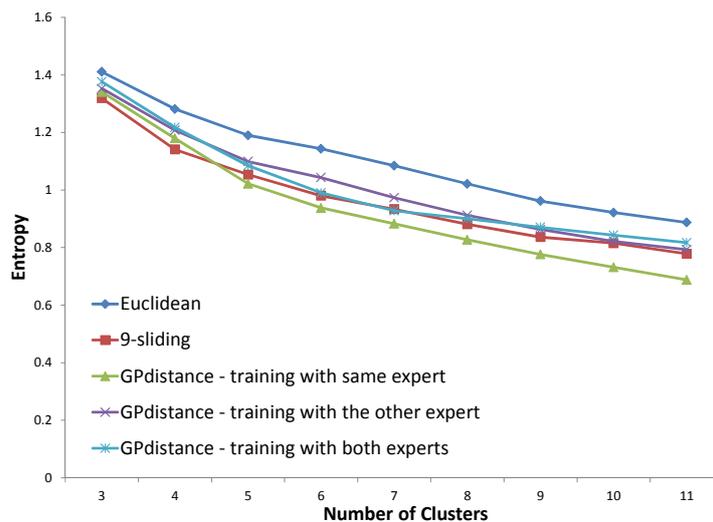


Figure 23. Entropy result for northern region using the data of expert #1.

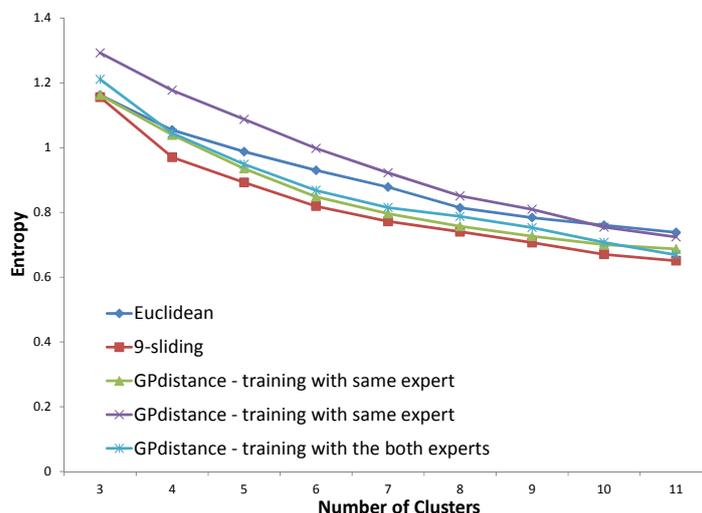


Figure 24. Entropy results for the northern region using the data of expert #2.

Despite the exceptional results shown in Figure 24, the average improvement in GP distance trained with the same expert was 11.2% compared to the Euclidean distance and 2.2% compared to the k -sliding distance. When we required the use of at least seven clusters, the improvement increased to 13.3% and 3.4% compared to the Euclidean distance and k -sliding distance, respectively. This means that too few clusters cause an erroneous result because the number of clusters determined by experts is approximately 10. Based on the results, we conclude that the GP can clearly identify the clustering principle of an expert.

The average improvement in GP distance that is trained with both experts' answers was 8.7% and -1.5% compared to the Euclidean distance and k -sliding distance, respectively. The clustering principle of the two joint experts was also found because it provides an improved result compared to the Euclidean distance. The entropy, however, is not improved compared to the k -sliding distance. There are two possible reasons for this result. The first is that the generated distance mainly consists of the k -sliding distance (Figures 19 and 20). Because the component of the proposed distance is the k -sliding distance, a considerably improved result cannot be expected when we compare it to the

k -sliding distance. The second reason is that the two experts have slightly different clustering principles according to their individual knowledge. This is verified by the fact that the distance trained with the other expert's answer usually does not provide decreased entropy (Figures 22–24).

From the results, we conclude that the clustering principle can be derived from a single expert. This, however, does not imply that the clustering principle of a group of experts cannot be derived. Actually, the results show that a clustering principle actually exists for a certain expert. We expect that a more sophisticated derivation of the clustering principle can be determined for a group of experts, even though the result was not improved compared to that of the proposed k -sliding distance in our experiment due to a small number of experts.

In summary, in all cases involving two provinces and two experts, the principle found with the GP was mainly composed of the k -sliding distance and DTW distance in the time domain and slope domain. The result of the GP showed the possibility of finding clustering principles because it showed an improved result compared to the Euclidean distance in all of the cases and compared to the proposed k -sliding distance in most of the cases.

6. Conclusions

Although the clustering of electricity customers is essential to various applications, the definition of good clustering is ambiguous in real scenarios. This ambiguity was the motivation for this research. This study can be considered a turning point in customer clustering for real applications.

In this paper, we moved our focus from clustering itself to real applications. A novel k -sliding distance was proposed for clustering and was verified through experiments. The results strongly suggest that the k -sliding distance is a good distance measure for electricity customer clustering, at least in the field of demand response.

A genetic programming framework was also proposed to derive clustering principles. Although the design of an optimal genetic programming framework itself is an important topic for another independent research project, we hereby verified that the derivation of clustering principles is possible with the proposed simple genetic programming framework for real applications such as demand response.

From the result we described in this paper, it would be interesting to apply the clustering to application scenarios. We are planning to verify the effect of the proposed clustering in a real demand response system in the near future.

Acknowledgments

This research was supported by Korea Electrotechnology Research Institute (KERI) Primary research program through the National Research Council of Science & Technology (NST) funded by the Ministry of Science, ICT (Information Communication Technology) and Future Planning (MSIP) (No. 15-12-N0101-61).

This research was also supported by the ICT R & D program of MSIP/IITP (B0101-15-0559, Developing On-line Open Platform to Provide Local-business Strategy Analysis and User-targeting Visual Advertisement Materials for Micro-enterprise Managers).

Author Contributions

Jimyung Kang mainly conducted this research. Jee-Hyong Lee advised on this research as an advising professor.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Chicco, G.; Ionel, O.M.; Porumb, R. Electrical load pattern grouping based on centroid model with ant colony clustering. *IEEE Trans. Power Syst.* **2013**, *28*, 1706–1715.
2. Chen, C.; Kang, M.; Hwang, J.; Huang, C. Synthesis of power system load profiles by class load study. *Int. J. Electr. Power Energy Syst.* **2000**, *22*, 325–330.
3. Albadi, M.; El-Saadany, E. A summary of demand response in electricity markets. *Electr. Power Syst. Res.* **2008**, *78*, 1989–1996.
4. Conejo, A.J.; Morales, J.M.; Baringo, L. Real-time demand response model. *IEEE Trans. Smart Grid* **2010**, *1*, 236–242.
5. Jin, C.H.; Pok, G.; Lee, Y.; Park, H.-W.; Kim, K.D.; Yun, U.; Ryu, K.H. A SOM clustering pattern sequence-based next symbol prediction method for day-ahead direct electricity load and price forecasting. *Energy Convers. Manag.* **2015**, *90*, 84–92.
6. Chicco, G.; Napoli, R.; Pigliione, F. Comparisons among clustering techniques for electricity customer classification. *IEEE Trans. Power Syst.* **2006**, *21*, 933–940.
7. Chicco, G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy* **2012**, *42*, 68–80.
8. Hino, H.; Shen, H.; Murata, N.; Wakao, S.; Hayashi, Y. A versatile clustering method for electricity consumption pattern analysis in households. *IEEE Trans. Smart Grid* **2013**, *4*, 1048–1057.
9. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Carro, B.; Sánchez-Esguevillas, A. Classification and clustering of electricity demand patterns in industrial parks. *Energies* **2012**, *5*, 5215–5228.
10. López, J.J.; Aguado, J.A.; Martín, F.; Muñoz, F.; Rodríguez, A.; Ruiz, J.E. Hopfield-K-Means clustering algorithm: A proposal for the segmentation of electricity customers. *Electr. Power Syst. Res.* **2011**, *81*, 716–724.
11. Chicco, G.; Ilie, I.S. Support vector clustering of electrical load pattern data. *IEEE Trans. Power Syst.* **2009**, *24*, 1619–1628.
12. Kang, J.; Lee, J.-H. Clustering of electricity customers using latent dirichlet allocation. In Proceedings of the International Conference on Electrical Engineering, Jeju, Korea, 15–19 June 2014.
13. 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.
14. Tsekouras, G.J.; Kotoulas, P.B.; Tsirekis, C.D.; Dialynas, E.N.; Hatziaargyriou, N.D. A pattern recognition methodology for evaluation of load profiles and typical days of large electricity customers. *Electr. Power Syst. Res.* **2008**, *78*, 1494–1510.

15. Marques, D.Z.; de Almeida, K.A.; de Deus, A.M.; da Silva Paulo, A.R.G.; da Silva Lima, W. A comparative analysis of neural and fuzzy cluster techniques applied to the characterization of electric load in substations. In Proceedings of the IEEE/PES Transmission and Distribution Conference and Exposition: Latin America, Sao Paulo, Brazil, 8–11 November 2004; pp. 908–913.
16. Iglesias, F.; Kastner, W. Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns. *Energies* **2013**, *6*, 579–597.
17. Carpaneto, E.; Chicco, G.; Napoli, R.; Scutariu, M. Electricity customer classification using frequency-domain load pattern data. *Int. J. Electr. Power Energy Syst.* **2006**, *28*, 13–20.
18. Ramos, S.; Duarte, J.M.; Duarte, F.J.; Vale, Z. A data-mining-based methodology to support MV electricity customers' characterization. *Energy Build.* **2015**, *91*, 16–25.
19. Panapakidis, I.P.; Papadopoulos, T.A.; Christoforidis, G.C.; Papagiannis, G.K. Pattern recognition algorithms for electricity load curve analysis of buildings. *Energy Build.* **2014**, *73*, 137–145.
20. Tsekouras, G.J.; Hatziargyriou, 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.
21. Zhang, X.; Liu, J.; Du, Y.; Lv, T. A novel clustering method on time series data. *Expert Syst. Appl.* **2011**, *38*, 11891–11900.
22. Chicco, G.; Napoli, R.; Postolache, P.; Scutariu, M.; Toader, C. Customer characterization options for improving the tariff offer. *IEEE Trans. Power Syst.* **2003**, *18*, 381–387.
23. Calinski, T.; Harabasz, J. A Dendrite Method for Cluster Analysis. *Commun. Stat. Simul. Comput.* **1974**, *3*, 1–27.
24. Pitt, B.D.; Kitschen, D.S. Application of data mining techniques to load profiling. In Proceedings of the 21st IEEE International Conference Power Industry Computer Applications, 16–21 May 1999, Santa Clara, CA, USA; pp. 131–136.
25. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, *1*, 224–227.
26. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008.
27. Ward, J.H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.