Two-Stage Clustering of Household Electricity Load Shapes for Improved Temporal Pattern Representation

MILAD AFZALAN$^{1}$, FARROKH JAZIZADEH$^{2}$, AND HODA ELDARDIRY$^{1,2}$, (Member, IEEE)

$^{1}$Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA
$^{2}$The Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA

Corresponding author: Hoda Eldardiry (hdardiry@vt.edu)

This work was supported in part by Virginia Tech’s Open Access Subvention Fund (VT OASF).

ABSTRACT With the widespread adoption of smart meters in buildings, an unprecedented amount of high-resolution energy data is released, which provides opportunities to understand building consumption patterns. Accordingly, research efforts have employed data analytics and machine learning methods to segment customers based on their load profiles, which help utilities and energy providers promote customized/personalized targeting for energy programs. Existing energy consumption segmentation techniques use assumptions that could reduce clusters’ quality in representing their members. Therefore, in this paper, we investigated a two-stage clustering method for capturing more representative load shape temporal patterns and peak demands through a cluster merging approach. In the first stage, load shapes are clustered (using classical clustering algorithms) by allowing a large number of clusters to accurately capture variations in energy use patterns, and cluster centroids are extracted by accounting for limited shape misalignment within the range of Demand Response (DR) timeframes. In the second stage, clusters with similar centroids and power magnitude ranges are merged using Complexity-Invariant Dynamic Time Warping. We used three datasets consisting of $\sim250$ households ($\sim15000$ profiles) to demonstrate the efficacy of the framework, compared to baseline methods, and discuss the impact on energy management. The proposed investigated merging-based clustering also increased correlation between cluster centroids and the corresponding members by 3-9% for different datasets.

INDEX TERMS Clustering methods, energy segmentation, demand response, smart meters, energy management, two-stage clustering.

NOMENCLATURE

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_i$</td>
<td>Centroid of $C_i$</td>
</tr>
<tr>
<td>$\mu_i'$</td>
<td>Temporary centroid of $C_i$</td>
</tr>
<tr>
<td>$|C_i|$</td>
<td>Number of load shapes in $C_i$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Maximum cluster size parameter</td>
</tr>
<tr>
<td>$C_i$</td>
<td>Cluster $i$</td>
</tr>
<tr>
<td>$CH/CHI$</td>
<td>Calinski-Harabasz Index</td>
</tr>
<tr>
<td>$CIDTW$</td>
<td>Complexity-Invariant Dynamic Time Warping</td>
</tr>
<tr>
<td>$corr(x, \mu)$</td>
<td>Correlation between $x$ and $\mu$</td>
</tr>
<tr>
<td>$corr_i$</td>
<td>Average correlation coefficient for cluster $i$</td>
</tr>
<tr>
<td>$CVI$</td>
<td>Cluster Validation Index</td>
</tr>
<tr>
<td>$d(i,j)$</td>
<td>Distance between load shapes $i$ and $j$</td>
</tr>
<tr>
<td>$DBA$</td>
<td>Dynamic Time Warping Barycenter Averaging</td>
</tr>
<tr>
<td>$DBI$</td>
<td>Davies-Bouldin Index</td>
</tr>
<tr>
<td>$DTW$</td>
<td>Dynamic Time Warping</td>
</tr>
<tr>
<td>$HC$</td>
<td>Hierarchical Clustering</td>
</tr>
<tr>
<td>$K$</td>
<td>Number of clusters</td>
</tr>
<tr>
<td>$K'$</td>
<td>Initial cluster number (for overpopulation)</td>
</tr>
<tr>
<td>$N$</td>
<td>Total number of load shapes</td>
</tr>
<tr>
<td>$SIL$</td>
<td>Silhouette Index</td>
</tr>
<tr>
<td>$SOM$</td>
<td>Self Organizing Map Clustering</td>
</tr>
<tr>
<td>$T$</td>
<td>Total number of samples in a daily profile</td>
</tr>
<tr>
<td>$t$</td>
<td>A sample in a daily profile</td>
</tr>
<tr>
<td>$WAC$</td>
<td>Weighted Average Correlation</td>
</tr>
<tr>
<td>$WCSS$</td>
<td>Within Cluster Sum of Squared Error</td>
</tr>
<tr>
<td>$x$</td>
<td>Daily profile</td>
</tr>
</tbody>
</table>

The associate editor coordinating the review of this manuscript and approving it for publication was Pierluigi Siano.
I. INTRODUCTION

Demand response (DR) mechanisms help electricity providers maintain distribution reliability, reduce generation costs and environmental concerns, and increase the utilization of renewable energy. Residential sector accounts for ∼40% of electricity consumption in the U.S. [1], making it a proper candidate for DR engagement [2]. However, a challenging problem for the efficient implementation of DR is tied to the highly varied energy load shapes across buildings [3], which highlights the importance of learning and predicting the behavior of customers. Advanced metering infrastructure (AMI) and smart meters have been adopted to provide fine-grained (hourly or sub-hourly) energy data for advanced analytics to enhance the efficiency of power network operations. Electricity load shapes segmentation from smart meters, through clustering techniques, could inform decision-makers about different patterns of energy use towards improving resource allocation in power systems [4], [5].

With the wide adoption of smart meters nationwide [6], and the availability of hourly and sub-hourly data, segmentation methods have been used to find similar patterns of electricity consumption behavior of customers. This task can be applied in different capacities to improve power system operations including implementation of DR programs [7], load forecasting [8], tariff determination [9], and renewable integration [10], [11]. The primary objective of the customer segmentation task is to transform a large database, might end up in wrong clusters that are not representative of the entire customer base.

As the review of the literature shows, different efforts have mainly adopted a variety of clustering techniques and considered their output to be self-evident from an application’s standpoint, without further verification that the resultant clusters properly reflect the actual energy use behaviors of different consumers. In other words, averages of all profiles associated with different clusters have been treated as representations of distinct energy use behaviors for the sample community, although the average representation may not properly represent the actual shapes of profiles in each cluster. This assumption could result in two problems: 1) the output of the segmentation can be less reliable, and customers will be assigned to the clusters that do not reflect their energy use behavior realistically, and 2) some of the representative energy use behaviors, which have a considerable frequency in a database, might end up in wrong clusters that are not representative of their shape. As a result, specific energy use characteristics that are important to utilities and decision-makers in the process load balancing and planning, including the distribution of relative peak demands across the load shapes and the magnitude of peaks, may not be well reflected in the segmentation outcome.

In this paper, seeking to evaluate cluster representation, we investigated a cluster merging approach for identifying representative electricity load shapes that account for improved compatibility between cluster centroids and their contributing load shapes while capturing both temporal variations and peak values of energy use profiles. To this end, we have investigated a framework for two-stage clustering of daily electricity load shapes by initially overpopulating the set of clusters to preserve the accuracy and further merging similar clusters by accounting for cluster shape alignment. To reduce the number of clusters (some of which are similar) into representative patterns, a Dynamic Time Warping (DTW) based merging approach is implemented. We have utilized this framework to show the possibility of improving the quality of clusters for representing their comprising load shapes. The investigation has been conducted using real-world smart meter data and leveraging various evaluation metrics to measure cluster compactness and correlation analysis.

The rest of the paper is organized as follows: section II presents the related work, section III defines the problem, section IV discusses the proposed approach, section V presents the results, and sections VI and VII present a discussion and conclude the paper.

II. RELATED WORK

The concept of load shapes segmentation or customer segmentation has received increasing attention with the proliferation of smart meters. Initial efforts [12], [13] have leveraged clustering techniques for classifying customers based on their load shape patterns. Different algorithms, including K-means [7], Self-Organizing Map (SOM) [14], hierarchical clustering (HC) [15], and Expectation-Maximization (EM) [16] have been investigated for the task of customer segmentation across tens to hundreds of households [7], [17].

Several recent efforts in household electricity segmentation have focused on addressing specific applications. In a class of studies, the use of big data and the efficiency of clustering at the city-scale have been investigated [18], [19]. Therefore, due to the increasing size of datasets, dimensionality reduction techniques such as Principal Component Analysis (PCA) and Symbolic Aggregate Approximation (SAX) [5] have been the subject of research in a group of studies in addition to feature extraction from load shapes considering selected attributes of interest (such as peak distributions or critical timeframes) [10], [20]. Furthermore, model-based techniques, which account for the phase shifts and time lags have been proposed to mitigate the impact of noisy and unequal time-series with incomplete information in real-world scenarios [21]. In some recent attempts, investigation of segmentation on disaggregated (i.e., individual load level) data, including Air Conditioning (AC) loads, has been carried out for DR applications [22], [23].

Several previous studies have looked into the integration of two-stage clustering approaches for electricity load profiling [7], [29]–[31]. In [29], a two-stage clustering was used for the classification of customers first by calculating the representative daily load shapes and then by assigning the
customers to dominant frequent clusters. Kwac et al. [7] introduced a two-stage adaptive k-means approach for large-scale datasets to constrain the distance between load shapes and cluster centroids within a certain distance. In [30], a two-stage clustering based on fast wavelet transformation was used first to extract typical load shapes for each user and then group similar ones in the second stage. The work proposed in [31] used a two-stage approach to both accounts for detecting the outlier load shapes while achieving the cluster compactness.

As a core assumption in household electricity segmentation and its potential applications, the clustering output (both in terms of cluster quality and the number of clusters) has been assumed to reasonably represent the energy behavior patterns of the entire customer base. Such an assumption has been considered as self-evident [24], investigated through common cluster validation indices (CVIs) [5], [15], [25], or justified through empirical inspections [18]. The application of common CVIs has been deemed to be viable for quality assessment of the clustering outcome at the first sight. Examples of typical CVIs include Bayesian Information Criterion (BIC) [26], Silhouette index [27], and Davies–Bouldin index (DBI) [28]. However, as outlined in [18], such generic statistical indicators for model selection will not necessarily work for the electricity segmentation task. In other words, proper metrics that indeed measure the compatibility of individual load shapes with their representative clusters have been mainly ignored in the evaluation processes as outlined in [11].

As the trends in the literature show, prior research on household electricity segmentation has not emphasized on assessing the compatibility of load shapes with their representative cluster centroids (beyond using classical cluster validation metrics). As such, the nuances of distinct temporal patterns reflected on individual load shapes in each cluster might be masked by classical clustering methods and internal validation techniques. Therefore, in this work, we focused on a segmentation framework to investigate the possibility of improving clusters’ representation for distinct temporal patterns and peak demand magnitude through a two-stage process of segmentation and cluster merging. To this end, we have shown the applicability of cluster merging using DTW-based distance measures for improved and fine-grained representation of load shape clusters.

**TABLE 1. Characteristics of the datasets.**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Location</th>
<th>Number of households</th>
<th># of daily profiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>Austin, TX</td>
<td>129</td>
<td>7535</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>Austin, TX</td>
<td>100</td>
<td>5676</td>
</tr>
<tr>
<td>Dataset 3</td>
<td>Boulder, CO</td>
<td>31</td>
<td>1790</td>
</tr>
</tbody>
</table>

**III. PRELIMINARIES AND PROBLEM STATEMENT**

Due to the highly varied patterns of energy consumption across different households on a daily basis, the task of segmentation can become challenging. In this section, the problem statement is presented through a quantitative case study. We first demonstrated how solely relying on common cluster validation indices (CVIs) for segmentation can lead to coarse-level representation of load shapes. We further highlighted the importance of recovering distinct load shapes from the clustering process.

**A. DATASET DESCRIPTION**

In this study, we used the data from the Pecan Street Project [41], which is an ongoing campaign in energy-efficiency initiatives through aggregate and appliance-level instrumentation of residential buildings with metering devices. Here, we used a set of datasets collected from residential buildings in Austin, TX, and Boulder, CO, during July and August 2015. Table 1 presents the characteristics of the datasets. The electricity use data has a resolution of one sample per 15 minutes. Therefore, each daily profile contains 96 data points. The duration of data collection for each dataset was 60 days. For compiling the datasets, consecutive days, regardless of weekdays or weekends, were considered, and profiles with missing values were excluded from the analysis. Since the distinction of power intensity was of importance in the clustering task as well, no power normalization was performed in this work. Therefore, both temporal consumption habits and the power intensity were aimed to be considered for load profile comparison. Three datasets were considered to impose adequate variations in the energy consumption styles. Since the households were primarily located in Austin, TX, we divided the households in that geographical location into two datasets to have multiple datasets of comparable size.

Figure 1 shows the distribution of total daily energy consumption for all observations in different datasets. The median energy consumption for dataset 1, dataset 2, and dataset 3 is 44 kWh, 42kWh, and 17kWh, respectively.

**B. PROBLEM STATEMENT DESCRIPTION**

To demonstrate the problem in a generic form, a common clustering outcome on daily profiles has been presented. Figure 2 shows the clustering results on a sample of ~8000 daily load shapes (separate from the aforementioned datasets) clustered into five groups using the K-means algorithm, which is commonly applied in the literature for the datasets of similar size and nature. In Figure 2(a), each subplot shows the temporal shape of each cluster, and the vertical axis is the power magnitude (in kW). The red line is the centroid of each cluster, which is obtained by averaging all daily profiles (in the order of hundreds or thousands) associated with the cluster. In Figure 2b, several examples of daily load profiles associated with clusters 1 and 4 are shown. In the literature, the segmentation results have been typically presented similar to what has been shown in Figure 2a, without further investigation of cluster quality. Therefore, the validity of clustering results was assumed as self-evident, and no further investigations were carried out to show how representative each cluster is to its associated members. However, as shown in Figure 2b, examples of load shapes for cluster #1 are retrieved that do not resemble the shape of its centroid. Similarly, several load shapes in cluster #4 are presented (Figure 2b) that do not
FIGURE 1. Distribution of energy consumption among all profiles.

FIGURE 2. Segmentation of ~8000 load shapes using K-means: a) five clusters representing the entire dataset, b) examples of individual daily profiles for clusters 1 and 4 that are misclassified.

TABLE 2. The definitions of common Cluster Validation Indices (CVIs).

<table>
<thead>
<tr>
<th>CVI</th>
<th>Definition</th>
<th>Selection criterion</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davies Bouldin Index (DBI)</td>
<td>$\frac{1}{k} \sum_i \max_{j \neq i} \left{ \frac{\sum_{x \in C_i} d(x, \mu_j) + \sum_{x \in C_j} d(x, \mu_i)}{d(\mu_i, \mu_j)} \right}$</td>
<td>Minimum</td>
<td>[28]</td>
</tr>
<tr>
<td>Silhouette (SIL)</td>
<td>$\frac{1}{K} \sum_i \left{ \frac{1}{\max{b(i), a(i)}} \sum_{x \in C_i} d(x, \mu_i) \right}$</td>
<td>Maximum</td>
<td>[27]</td>
</tr>
<tr>
<td>Calinski-Harabasz Index (CHI)</td>
<td>$\sum_{i=1}^{K-1} \sum_{x \in C_i} d(x, \mu_i) / (K-1)$</td>
<td>Maximum</td>
<td>[32]</td>
</tr>
<tr>
<td>Within Cluster Sum of Squared (WCSS)</td>
<td>$\sum_{i=1}^{K} \sum_{x \in C_i} |x - \mu_i|^2$</td>
<td>Elbow</td>
<td>[33]</td>
</tr>
</tbody>
</table>

Notation: $K$: number of clusters; $C_i$: Cluster $i$; $\|C_i\|$: Number of load shapes in $C_i$; $\mu_i$: centroid of $C_i$; $N$: total number of load shapes; $a(x) = \frac{1}{\|C_i\|} \sum_{j \in C_i, i \neq j} d(i, j)$, $b(x) = \min_{k \neq i} \frac{1}{\|C_k\|} \sum_{j \in C_k} d(i, j)$.

C. INVESTIGATION OF TYPICAL CVIs

Segmentation is commonly performed using a small number of clusters. This is intuitive to ensure the ease of interpretation as a primary criterion for segmentation. To identify and justify the number of clusters, cluster validation indices (CVIs) are commonly used. Here, to show the challenge of using typical CVIs and show the complexity in clustering patterns, we used several well-known CVIs; namely, the Davies-Bouldin Index (DBI) [28], Silhouette (SIL) [27], Calinski Harabasz Index (CHI) [32], and Within Cluster Sum of Squared (WCSS) error [33]. CVIs' definitions have been provided in Table 2.

We measured the CVIs for four common clustering methods (Self-Organizing Map (SOM), K-means, Fuzzy c-means (FCM), and hierarchical clustering (HC)). To choose the number of clusters ($K$), CVI is measured over a range of $K$ values, and a selection criterion (Table 2) is used on CVI values.
We compared the range of $K = 5, 10, 15, 20, \ldots, 120$ to have a reasonable estimation of low to high number of clusters. We used 1 cluster increments in the commonly used range of 5 to 10 (i.e., 5, 6, 7, 8, 9, 10), and 5 cluster increments for $K$ more than 10 till $K_{\text{end}} = 120$ (i.e., 10, 15, 20, \ldots, 120). For each value of $K$, we clustered the data and measured CVIs. Figure 3 presents the CVIs for Dataset 1, with each subplot representing one CVI. The trend of increase/decrease for all 4 CVIs (each subplot) was almost consistent across datasets. Therefore, here, we have presented the interpretations only for Dataset 1.

![FIGURE 3. CVI for different clustering algorithms.](image)

DBI metric uses a min-rule for choosing the proper $K$. As the results show, for all methods, a low value of $K = 5$ leads to low DBI values. It must be noted that DBI values for fuzzy c-means showed to be very high. Therefore, we did not present its values in the first subplot to avoid masking the variation trends for other methods. SIL metric uses a max-rule for selecting the $K$. Similar to the previous case, the lowest value of $K = 5$ is estimated. CHI metric uses a max-rule for selecting the $K$. We select $K = 5$ for this metric. However, for WCSS, which follows the elbow criterion for determining the $K$, $K = 30$ is a reasonable value for hierarchical, SOM, and K-means. When it comes to algorithms, for DBI and SIL metrics, K-means shows a marginal improvement over SOM and hierarchical clustering. For the CHI metric, SOM and K-means offer equally better performance. For the WCSS metric, hierarchical clustering has the lowest values and seems to have a better performance. Therefore, there is no consistency in selecting a superior approach by considering combinations of CVIs and clustering algorithms. Nonetheless, to visually evaluate the performance of common algorithms based on optimum CVIs, an assessment of the clustering results is presented in the next section. Table 3 shows optimum $K$ values, inferred based on CVIs for each dataset investigated in the following section.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$K$ (Dataset 1)</th>
<th>$K$ (Dataset 2)</th>
<th>$K$ (Dataset 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC, SOM</td>
<td>30</td>
<td>25</td>
<td>25</td>
</tr>
</tbody>
</table>

**TABLE 3. K values for different datasets and algorithms inferred by using the optimum CVIs.**

### D. COMPARISON OF CLUSTERING RESULTS

From cases in Table 3, Figure 4 presents the clusters for dataset 1 as an example demonstration. Each subplot is one cluster, and the red curve is the cluster centroid, averaged over all profiles. The frequency value above each subplot shows the occurrence rate of the cluster within the entire dataset. In Figure 4(a), with $K = 5$, four out of five clusters have peak demand around 18:00, but with different peak magnitudes. Cluster 5 seems to be an outlier due to its considerable high peak usage in addition to its very low frequency (less than 1% of the data). In Figure 4(b) and Figure 4(c), with a higher population of clusters, more distinct energy patterns were revealed, while their presence has been masked by selecting a lower number in Figure 4(a). Examples include clusters 4, 11, and 17 in Figure 4(b), and clusters 8, 19, and 21 in Figure 4(c). Furthermore, a comparison between hierarchical clustering and SOM with an equal number of clusters (Figure 4(b) and Figure 4(c)) show that hierarchical clustering results in identifying clusters with very low frequency (4 clusters with a frequency of less than 0.5%) while SOM forms clusters with higher frequency. As shown in Figure 4, although CVI has been used to select $K = 5$, using such a low number will result in a very coarse-level representation of load shape groups. Specifically, the cluster centroids, which altogether are supposed to encompass the energy use behavior of the community, might not reflect the temporal shape/peak demands of their associated members.

Given the variations in the nature of CVIs, using different metrics results in selecting different $K$ values for different algorithms. Generic CVIs might not work well for this domain-specific problem because several groups of energy patterns with distinct loads may not be recovered without selecting a large number of clusters, while their shapes/magnitude characteristics could be of interest to energy planners/utilities. Specifically, utilities could leverage the resulting segmentation output, and based on the requirement of a DR program (such as implementation time or peak demand), identify potential customers for DR engagement. The following section introduces an approach to investigate solutions to this problem.

### IV. TWO-STAGE CLUSTERING ON HOUSEHOLD ELECTRICITY LOAD SHAPES

As discussed in Section III-D, different CVIs and algorithms could result in different load shape patterns, and optimizing the number of clusters using CVIs could result in a low number of clusters, with a potential of losing representative load shapes. On the other hand, selecting a high number
FIGURE 4. Cluster comparison for dataset 1: a) K-means, $K = 5$, b) HC, $K = 30$, c) SOM, $K = 30$. The vertical axis is Power (kW) and the horizontal axis is Time (hr).

of clusters could improve the representation of various load shapes. However, it could affect the interpretation of the outcome due to the presence of multiple correlated clusters. To account for this trade-off, we have investigated a two-stage
clustering approach that initially overpopulates clusters to improve the representation of load shapes with respect to their respective cluster centroids and further merge the closely related ones based on their temporal shape to enhance the interpretability.

The general framework is shown in Figure 5. The objective was to evaluate the impact of reducing the initial cluster library with fine-grained representation without losing the essential information in load shape patterns/peak demands. In the first stage, a large number of clusters is generated, which are then transformed with a time-series averaging technique (instead of direct calculation of the mean values) to represent cluster centroids. In the second stage, pairwise distances of clusters are calculated with a distance measure that accounts for shape alignment between cluster centroids to merge similar ones. Each component of this framework is described in the following subsections.

A. FIRST STAGE: INITIAL CLUSTER REPRESENTATION

1) OVERPOPULATION OF CLUSTERS

To provide a diverse set of load shapes that captures a wide selection of the possible profiles, a representative of the actual distribution of the profiles in the entire dataset is needed. Therefore, in the first stage, through a benchmark clustering, an initial value of \( K' \) for the number of clusters is considered (\( K' > K \)). \( K' \) can be selected by measuring the WCSS error, such that increasing \( K' \) do not cause a considerable change in the WCSS. To this end, the elbow curves for WCSS are used for the selection of \( K' \). Using any of the classical clustering techniques, \( K' \) groups are populated.

2) CLUSTER REPRESENTATION

Since each cluster may include thousands of load profiles, a proper representation for each cluster that resembles its content is required. The most intuitive way for cluster representation is the component-wise arithmetic averaging (simple averaging of each sample across all observations in a cluster). However, this approach to averaging may result in a centroid that is dissimilar to any of its associated time-series in a given cluster [34]. To this end, we employed the Dynamic Time Warping Barycenter Averaging (DBA) technique [35]. DBA is a time-series averaging technique that preserves the nuances of variations in individual profiles. In contrast to conventional time-series averaging, which may result in centroids that differ from the original time-series, DBA uses an expectation-maximization approach by refining the medoid of each group through finding the best set of alignments within each group through iterations. The barycenter averaging also refers to associating each component of the time series to its corresponding components of other observations. In this context, the corresponding components of other observations are selected based on the Dynamic Time Warping (DTW) distance matrix.

In each iteration, the following steps are taken [35]:

1) Measuring the DTW distance between each profile \( (x \in C_i) \) and the temporary average centroid \( (\mu'_i) \) to find the association of each sample \( (t \in 1, 2, \ldots, T) \) of the centroid \( (\mu_i) \) with respect to samples \( (t \in 1, 2, \ldots, T) \) of individual profiles within cluster \( i \).

2) Updating each sample of the centroid as the barycenter of samples associated with it from the previous step.

The barycenter function is defined as:

\[
barycenter(\xi_1, \xi_2, \ldots, \xi_m) = (\xi_1 + \xi_2 + \ldots + \xi_m)/m \tag{1}
\]

where \( \xi_i \) is a sample from a given profile \( x \) and \( m \) is the length of the vector of associated (matched) samples in a warping path.

Using previous iteration average centroid \( (\mu'_i) \), the \( t \)-th sample of the current iteration average centroid \( (\mu_i) \) is defined as:

\[
\mu_i(t) = barycenter(assoc(\mu'_i(t))) \tag{2}
\]

Here, \( assoc \) function associates each sample of the \( \mu_i \) to the samples (one or more) of the profiles based on calculating a DTW distance.

Considering \( K' \) initial clusters, DBA is applied to the content of each one to obtain the centroids \( \mu_i \). To show the impact of DBA versus conventional averaging, we presented a set of clusters in Figure 6. As can be seen, DBA could enhance the representation of centroids in each cluster by emphasizing the peaks and valleys reflected in profiles.

![Figure 5](image-url) Two-stage clustering framework. Benchmark clustering is an arbitrary clustering technique (e.g., k-means). Acronyms: DBA (Dynamic Time Warping Barycenter Averaging); CIDTW (Complexity-Invariant Dynamic Time Warping).

![Figure 6](image-url) DBA averaging versus conventional averaging.
B. SECOND STAGE: CLUSTER MERGING

In this stage, the clusters that are highly similar in shape and magnitude are merged to reduce the final size of the cluster library. Given the initial cluster library size of $K'$ in the first stage, the objective is to reduce the library size to $K$. We used an iterative merging process to transform the dataset from $K'$ to $K$ clusters, in which $K$ was interpreted from the WCSS elbow curve in a benchmark clustering process. In each iteration, the matrix of similarity measure between cluster $i$ and $j$ is constructed, and the closest pairs are merged ($K' \rightarrow K' - 1$). The process is continued till $K' \rightarrow K$. Considering the nature of load profile datasets, it is important to employ a robust measure for merging the time-series that accounts for the inherent small time shift in similar load shapes while differentiating larger temporal differences. Figure 7 shows an example of two household load shapes that have similar energy use patterns (double demand peaks in the morning and evening) but are relatively different in the time of peak demand (~1-hour difference in peak timing). Given the small shift in peak time, DR plans consider these load shapes as potential targets for peak hours in the same DR window. Commonly-used similarity metrics such as Euclidean distance fail to capture similarity and return a large distance in such cases while DTW-based [45], [46] or the KSC-based measures [47], [48] account for capturing these similarities in load shapes. In this work, we employed the Complexity-Invariant Dynamic Time Warping (CI-DTW) as the distance measure [36]. CI-DTW is a DTW-based distance measure that is invariant to the complexity of time-series (e.g., number of peaks and valleys). Therefore, it avoids matching pairs of simple objects that are subjectively apart from those with more complex patterns with similar shapes [36]. Given that DTW uses the width of an adjustment window for matching point by point in time-series, its values could be adjusted in a relaxed or strict way based on the duration requirements for DR programs (i.e., the DR time frame). To investigate the efficacy of the merging approach, in this study, we have implemented the legacy DTW distance. To investigate the efficacy of the merging approach, in this study, we have implemented the legacy DTW distance. To investigate the efficacy of the merging approach, in this study, we have implemented the legacy DTW distance. To investigate the efficacy of the merging approach, in this study, we have implemented the legacy DTW distance. To investigate the efficacy of the merging approach, in this study, we have implemented the legacy DTW distance.

\begin{align}
D(i, j) &= \delta(p_i, q_j) + \min \left\{ D(i - 1, j - 1), D(i - 1, j), D(i, j - 1) \right\} \\
\delta(p_i, q_j) &= \text{distance between the } i-th \text{ sample of } P \text{ and the } j-th \text{ sample of } Q.
\end{align}

in which $\delta(p_i, q_j)$ is the distance between the $i-th$ sample of $P$ and the $j-th$ sample of $Q$. The above equation can be solved by dynamic programming. Upon measuring the DTW distance, CI-DTW is defined as:

$$
\text{CIDTW} = DTW(P, Q) \cdot CF(P, Q)
$$

in which $CF(P, Q)$ is a correction factor as follows:

$$
CF(P, Q) = \frac{\max(CE(P), CE(Q))}{\min(CE(P), CE(Q))}
$$

and $CE(A)$ is a complexity estimate as:

$$
CE(A) = \sqrt{\frac{1}{N} \sum_{i=1}^{N-1} (a_i - a_{i+1})^2}
$$

2) ITERATIVE MERGING

Using the CI-DTW distance measurement between cluster centroids, we merge the closest cluster pair $(i, j)$ in each iteration, and update the content and centroid of the remaining ones. A control parameter ($\tau$) is considered as the maximum cluster density upon merging $i, j$ such that:

$$
\| C_i \| + \| C_j \| \leq \tau \cdot N
$$

where $\| C_i \|$ and $\| C_j \|$ are the numbers of profiles associated with clusters $i$ and $j$, respectively, and $N$ is the total # of profiles in the dataset. Therefore, $\tau$ controls post-merging cluster size to avoid the formation of highly dense clusters. In the presented results, $\tau = 0.2$ (i.e., 20% of the dataset) was considered, allowing for both the formation of dense clusters and capturing a minimum of basic energy lifestyle patterns, scattered at different times of a day (peak time of morning, noon, afternoon, late evening, and midnight). The other configuration parameter ($K'$) that sets the initial number of clusters has been evaluated in a sensitivity analysis in Section V-B. Algorithm 1 shows the pseudocode for the cluster merging process in the second stage.

V. RESULTS AND DISCUSSION

A. VISUALIZATION AND EMPIRICAL INVESTIGATION

We applied the two-stage clustering approach described in section IV using SOM and K-means as the first stage baseline because they outperform fuzzy c-means and hierarchical clustering, which tend to generate outliers. To estimate the initial number of clusters ($K'$), the elbow curve for WCSS in Figure 3 was used. To overpopulate the initial clusters, a set of $K' = \{50,70,90\}$ spanning the range with low error declines between consecutive $K$ values was considered. For the second stage, the final library size of $K = \{10,20,30,40\}$ was used to study the impact of cluster merging at different levels.

**FIGURE 7.** Load shapes with similar behaviors and temporal shift.

1) CI-DTW

CI-DTW [36] is a variation of DTW distance measure. In DTW [39], the optimal alignment of the two time-series $(P = \{p_1, p_2, \ldots, p_T\}$ and $Q = \{q_1, q_2, \ldots, q_T\}$) is recursively found by calculating the cost defined by:

$$
D(i, j) = \delta(p_i, q_j) + \min \left\{ D(i - 1, j - 1), D(i - 1, j), D(i, j - 1) \right\}
$$
Figure 8 shows the pairs of merged clusters at different iterations for dataset 1, i.e. subsequent merging trials ($K' = 90; K = 40; K' - K = 50$). In this figure, SOM was applied for creating the initial cluster library. The value above each subplot is the iteration number. The results show that the selected clusters are subjectively close in both temporal shape patterns and peak magnitudes. Figure 9 shows the final clusters for this example after merging. Visual evaluations show that clusters are well-separated and distinct in their temporal shapes and power magnitude. Furthermore, they reveal the useful features in load shapes such as peak magnitude, peak timing, peak duration, and energy consumption (area under load shape curve), which could be of interest to utilities for planning customized energy programs.

**B. QUANTIFIED INVESTIGATION**

1) **METRICS**

We used the weighted average correlation (WAC) coefficient and WCSS as two quantified metrics for measuring the extent to which the cluster centroids represent their associated profiles. Both of these metrics reflect the compactness of clusters. WAC first measures the correlation coefficient between each cluster centroid and its associated profiles and uses the average value as the correlation indicator of each cluster.
Two-stage clustering outcome ($K' = 90$, $K = 40$). Subplots’ horizontal axis: time of day (hr); vertical axis: power (kW).

Algorithm 1 Pseudocode for Cluster Merging

**Input:** Overpopulated clustering results, cluster centroids with DBA, initial cluster number ($K^1$)

1: Set the target cluster number $K$

2: While $K' > K$:

3: Find the closest cluster centroids $\mu_i$ and $\mu_j$ based on CI-DTW metric.

4: While $||C_i|| + ||C_j|| > \tau * N$:

5: Find the next set of closest $\mu_i$ and $\mu_j$.

6: Set $\mu_i = (||C_i|| * \mu_i + ||C_j|| * \mu_j)/(||C_i|| + ||C_j||)$

7: Delete $\mu_j$

8: Update cluster index from cluster $j+1$ to the last one.

9: $K' = K' - 1$

Thereafter, the frequency of each cluster is utilized to have the weighted average as a single correlation score. More specifically:

$$WAC = \frac{\sum_{i=1}^{K} ||C_i|| \cdot corr_i}{N}, \quad WAC \in [0, 1]$$  

In which $WAC$ is the weighted average correlation and $corr_i$ is the average correlation coefficient for cluster $i$ as follows:

$$corr_i = \frac{\sum_{x \in C_i} corr(x, \mu_i)}{||C_i||}$$  

WCSS error is as defined in section III-C. A lower value of WCSS indicates a higher compactness.

Figure 10 presents the $WAC$ for different cases of analysis. Each row represents one dataset, and each column is one clustering method (SOM on the left and K-means on the right). The two-stage bars show the final results after overpopulating the clusters (with $K'$ values shown in the subplots) and then merging them up to $K$ clusters (shown on the horizontal axis). The benchmark bar represents the conventional clustering by directly selecting $K$ clusters. As the results show, using the investigated two-stage approach improves the correlation in most cases, specifically, for datasets that include more observations and higher diversity in energy use patterns. Specifically, it improves the average correlation by 8.2%, 8.9%, and 2.6% for dataset 1, dataset 2, and dataset 3, respectively.
Figure 11 shows the WCSS metric results. The two-stage approach results in lower error values in most cases, indicating the increased compactness in clusters. Specifically, the average of WCSS is reduced by 9.3%, 9.5%, for datasets 1 and 2, respectively. However, for dataset 3, an average of 3.4% increase was observed. A possible interpretation for the error increase is the smaller size of dataset 3 (Table 1). Therefore, the initial $K'$ values used in Figure 3 may not be appropriate for this dataset. Other solutions such as measuring the eigenvalues of the correlation matrix [40] for...
VI. DISCUSSION FOR CUSTOMIZED ENERGY PROGRAMS

Based on the findings in section V, the investigated clustering approach based on cluster merging could help reveal useful load shape features (beneficial in planning and decision-making) such as peak magnitude, peak timing, peak duration, and energy consumption, which are important to utilities’ policies and program planning. In other words, by increasing the quality of clusters, the cluster centroids will better represent all their members. Therefore, targeting those members for different energy management programs would be more productive as their energy use profiles fit the characteristics of the clusters. To provide some context on the implications of findings, we interpreted results of Figure 9 from the application standpoint.

DR events are typically scheduled for the evening timeframe, in which the net demand of the network is excessively high. The DR implementation could be achieved by partial load shedding or load shifting during peak times. Based on Figure 9, clusters 4, 5, 6, 7, 11, 13, 17, 19, 22, 33 have sharp peaks, which makes them potential candidates for DR events during 5pm-7pm timeframe. Therefore, customers with load profiles belonging to those clusters are potential targets for DR marketing and engagement. Similarly, clusters 2, 8, 21, 30, 34, and 37 also have a sharp peak at a later timeframe, which make them potential candidates for DR for timeframes after 8 pm. Especially, clusters 30, 32, 33, 34, and 37 have considerably high usage (peak demand of more than 10kW), whose peak consumption could be tied with high power appliances such as pool pump, multiple AC units, and simultaneous usage of wet appliance or electric vehicle (EV) charging [41]. Appliance-level data from a subset of buildings tied with the smart meter data have also been implemented to infer the possible drivers of consumption for daily load shapes (e.g., [42]).

In another class of distributed energy management applications, load profiling could facilitate the integration of renewable resources such as solar generation. Accordingly, the increased integration of renewables is a key element to move towards decarbonization and distributed energy management in the smart grid [43], [44]. In particular, in Figure 9, clusters 9, 16, 26, and 35, whose demand patterns in the noon timeframe coincide with the solar generation, could benefit more from photovoltaic (PV) integration. Therefore, customers with this load pattern are potential targets for PV adoption due to the higher chance of maximizing the self-consumption of solar energy for their units. Additionally, clusters 20 and 29 have almost zero demand at noon, implying their storage capabilities if they have PV-battery systems, and could self-consume the solar energy at later times when the electricity price rate is higher. Alternatively, with the rise of new technologies such as peer-to-peer energy trading between prosumers and consumers [44], these clusters, for PV-equipped houses, could also offer their high amount of on-site generation to their consumer neighbors.

Finally, clusters 14, 15, 26, and 27 have sharp peaks after midnight, presumably due to EV charging and wet appliances (e.g., dishwasher) operation. Since the heavy usage is during the off-peak demand time, the energy consumption pattern is already compatible with DR plans, assuming the typical evening/night peak demand for the network.

VII. CONCLUSION

The ubiquity of smart meter infrastructure and energy time-series data provides opportunities for customers’ energy behavior analytics. Clustering energy load shapes into a few representative patterns helps utilities in resource allocation and energy-efficiency program design to help consumers engage in energy management programs. In this paper, we evaluated a cluster merging technique through a two-stage approach to preserve temporal patterns and peak magnitude of load shapes for segmentation and to increase the quality of clusters in representing their corresponding members. In doing so, We first provided a comparative assessment of conventional clustering techniques and cluster validation indices (CVIs). We then showed that using common CVIs could underestimate the number of clusters and results in reduced representativeness of the clusters. The proposed approach utilizes overpopulation and merging as a method to improve cluster quality, measured by the correlation between cluster centroids and individual members. The merging is carried out using a DTW-based distance measure to account for the similarity of temporal variations. Empirical investigation and quantified assessment compared to the benchmark solutions were presented to investigate the impact of the framework. The approach was tested on three datasets from the Pecan Street Project Dataport with varied load shapes in each dataset. The results showed that compared to classical clustering techniques (i.e., using SOM or K-means clustering along with optimizing the number of clusters using an error function), the proposed investigated merging-based clustering resulted in an increase in correlation by 8.2%, 8.9%, and 2.6% for datasets 1 to 3, respectively.

In the future, we plan to 1) investigate this approach for other use-cases and generic datasets rather than the building science domain, 2) develop pre-processed cluster library for classifying energy use behavior of new customers, and 3) investigate the correlation of time-of-use of appliances with clusters of electricity load shapes obtained by the investigated cluster merging approach.

REFERENCES


MILAD AFZALAN received the B.Sc. and M.Sc. degrees from the University of Tehran, Iran, in 2013 and 2015, respectively, and the M.Sc. degree in computer science and the Ph.D. degree in civil engineering from Virginia Tech, USA, in 2020. His research interests include intelligent cyber-physical systems through data sensing and machine learning, control of building systems, and infrastructure operational analytics.

FARROKH JAZIZADEH received the M.Sc. degrees in computer science and civil engineering and the Ph.D. degree in civil engineering with a focus on informatics for the intelligent built environment from the University of Southern California (USC), Los Angeles, CA, USA, in 2013 and 2015, respectively. He is currently an Associate Professor with the Via Department of Civil and Environmental Engineering, Virginia Tech, where he is also the Founding Director of the INFORM Laboratory. His research interests include the intersection of cyber-physical systems, data-driven informatics, infrastructure operational analytics, and built environment sustainability and adaptability. Among his recent leadership efforts, he has Co-Chaired the Technical Program Committee of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys 2019) and organized and ran the first workshop on Building Blocks for Smarter Cities workshop, Arlington, VA, USA, in 2019.

HODA ELDARDIRY (Member, IEEE) received the B.E. degree in computer and systems engineering from Alexandria University, Egypt, in 2003, and the M.Sc. and Ph.D. degrees in computer science from Purdue University, USA, in 2006 and 2012, respectively. She is currently an Associate Professor of computer science at Virginia Tech, USA, where she directs the Machine Learning Laboratory. She previously worked at the Palo Alto Research Center (aka Xerox PARC), where she led the Machine Learning Research Group, managed technical commercial portfolios, and spearheaded machine learning for sensor research. Her research interests include artificial intelligence and machine learning with a focus on building human–machine collaborative AI systems that can learn context-aware and explainable models from multisource and interconnected data.

***