Household Energy Consumption Segmentation Using Hourly Data

Jungsuk Kwac, *Student Member, IEEE*, June Flora, and Ram Rajagopal, *Member, IEEE*

***Abstract*—The increasing US deployment of residential advanced metering infrastructure (AMI) has made hourly energy consumption data widely available. Using CA smart meter data, we investigate a household electricity segmentation methodology that uses an encoding system with a pre-processed load shape dictionary. Structured approaches using features derived from the encoded data drive five sample program and policy relevant energy lifestyle segmentation strategies. We also ensure that the methodologies developed scale to large data sets.**

***Index Terms*—Clustering, demand response, segmentation, smart meter data, variability.**

I. Introduction

THE WIDESPREAD deployment of advanced metering infrastructure (AMI) has made available concrete information about user consumption from smart meters. Household load shapes reveal significant differences among large groups of households in the magnitude and timing of their electricity consumption [3]. Hourly smart meter data offers a unique opportunity to understand a household's energy use lifestyle. Further, this consumption lifestyle information has the potential to enhance targeting and tailoring of demand response (DR) and energy efficiency (EE) programs as well as improving energy reduction recommendations. According to the Federal Energy Regulatory Commission, DR is defined as: "Changes in electric use by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized." EE means using less power to perform the same tasks, on a continuous basis or whenever that task is performed.

In this paper, an electricity customer segmentation methodology that uses an encoding system with a pre-processed load shape dictionary is examined. Energy consumers load shape information then is used to classify households according to extracted features such as entropy of shape code which measures the amount of *variability* in consumption. Load shape information enhances our ability to understand individual as well as groups of consumers. For example, time of day building occupancy and energy consuming activities can be interpreted from these shapes.

In the proposed segmentation system, we use a structured approach that uses features derived from the encoded data to drive the segmentation. We also develop segmentation strategies that aligned with specific application purposes such as household targeting for EE programs or recommendations for time of use shifts. In addition, we ensure that methods can readily scale to large data sets. We test our approach in a 220 K household data sample for a large utility.

A. Prior Work

Much of the previous research on audience segmentation takes place in psychology, marketing, and communication. Almost all segmentation in those fields rely on surveys of individuals regarding their self- reported values, attitudes, knowledge, and behaviors [1], [2]. In the last decade, utility companies are increasingly using these psychographic segmentation strategies to support program targeting, recruitment message tailoring and program design for DR and EE programs in [8]. Rarely, however is actual energy use part of the segmentation strategy [6]–[9].

Recently, the wide-spread dissemination of electricity smart meters offers the opportunity to create segmentation strategies based on 15 min, 30 min, or hourly household energy use. Understanding a households time of day energy consumption, daily usage pattern stability over time, as well as actual volume of energy use offers insights into household use of energy [3]. Further, these consumption features can be relevant to marketing and program design tasks. For example, high usage volume consumers or load shapes may signal potential for certain energy efficiency messages, whereas household load shape stability may be more relevant for time of use reduction messages.

Existing literature on analysis of smart meter data focuses on forecasting and load profiling such as [10]–[13], [18]. Some significant contributions in segmentation are [3], [4], [16], [17], [19], [20]. Self-organizing maps (SOM) and K-means are used to find load patterns in [17] and to present a electricity consumer characterization framework in [4]. A two-stage pattern recognition of load curves based on various clustering methods, including K-means, is described in [20]. Various clustering algorithms (hierarchical clustering, K-means, fuzzy K-means, SOM) are used to segment customers with similar consumption behavior in [16]. Similarly, [19] checks the capacity of SOM to filter, classify, and extract load patterns. As an alternative approach to distance-based clustering (K-means, SOM), [18] introduces a class of mixture models, random effects mixture

Manuscript received May 31, 2013; revised July 29, 2013; accepted July 29, 2013. Date of current version December 24, 2013. The work of R. Rajagopal is supported by the Powell Foundation Fellowship. Paper no. TSG-00427-2013.

J. Kwac is with the Department of Electrical Engineering and the Stanford Sustainable Systems Lab, Department of Civil and Environmental Engineering, Stanford University, CA 94305 USA (e-mail: kwjusu1@stanford.edu).

J. Flora is with the Human Sciences & Technologies Advanced Research Institute (HSTAR), Stanford University, CA 94305 USA (e-mail: jflora@stanford.edu).

R. Rajagopal is with the Stanford Sustainable Systems Lab, Department of Civil and Environmental Engineering, Stanford University, CA 94305 USA (e-mail: ramr@stanford.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2013.2278477

models, with its own EM algorithm to fit the mixture models. The current paper proposes a different approach that decomposes the daily usage patterns into daily total usage and a normalized daily load shape. Representative load shapes are found utilizing *adaptive K-means* and summarized utilizing hierarchical clustering, so a stable encoding mechanism can be designed. Various different metrics are computed based on the encoding we propose.

The paper also distinguishes from previous work by analyzing the results of applying the method to more than *66 million* load shapes from a population of 220 K residential consumers. This massive data analytics reveals various important features about the data, including that a consumers' lifestyle is captured by their typical load shapes. We propose five different simple segmentation schemes and illustrate how these segmentation strategies can be selected for certain program development, pricing, and marketing purposes. We also test that the proposed segmentation strategies can be scaled to large data bases by using a load shape dictionary. The robustness of these approaches is verified by contrasting dictionaries learnt from different groups in the population.

The remainder of the paper is organized as follows. Section II describes the proposed methodology and resulting encoding system. Section III evaluates the method on a large consumer dataset. Section IV utilizes the encoding system to design various practical segmentation mechanisms for the dataset. Section VI concludes the paper.

II. Methodology

The proposed methodology for usage based segmentation consists of three stages as shown in Fig. 1. The methodology relies on a simple decomposition of the load profiles. Given a daily consumption profile $l(t)$, we decompose it as $l(t) = a\, s(t)$, where

$$a = \sum_{t=1}^{24} l(t) \text{ and } s(t) = \frac{l(t)}{a}. \tag{1}$$

a is the daily total consumption and $s(t)$ is the normalized load profile, which we denominate *load shape*. The first stage creates a dictionary for representative load shapes by modeling the distribution of a and clustering the load shapes $s(t)$ across the population. The second stage extracts proper dynamic features from the encoded data utilizing the pre-processed dictionary. The last stage performs a second level of clustering depending on segmentation criteria such as a lifestyle or usage variability. The methodology is designed to scale to very large datasets.

A. Daily Total Consumption Characterization

The simplest characterization of daily total usage a is to infer *probability distribution* of values across the population. The empirical distribution exhibits a long tail as shown in Fig. 2 for two climate zones. [10] utilizes a Weibull distribution to model this distribution for a small number of consumers. Instead, we find that a *mixture of log normal distributions* fits best the actual data. The density function for a mixture with M elements is given by

$$f(a) = \sum_{i=1}^{n} \lambda_i g_i(a), g_i(a) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\frac{-(\log(1+a)-\mu_i)^2}{2\sigma_i^2}}, \tag{2}$$

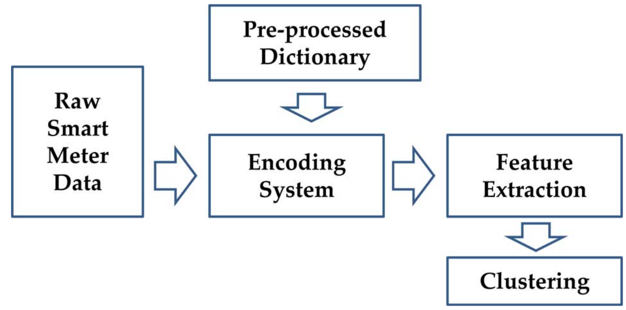


Fig. 1. User segmentation flow.

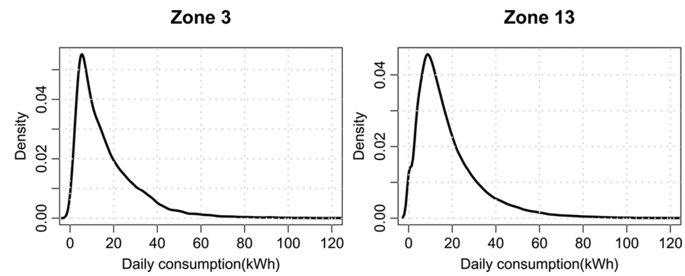


Fig. 2. Daily consumption distribution at Zone 3 and Zone 13.

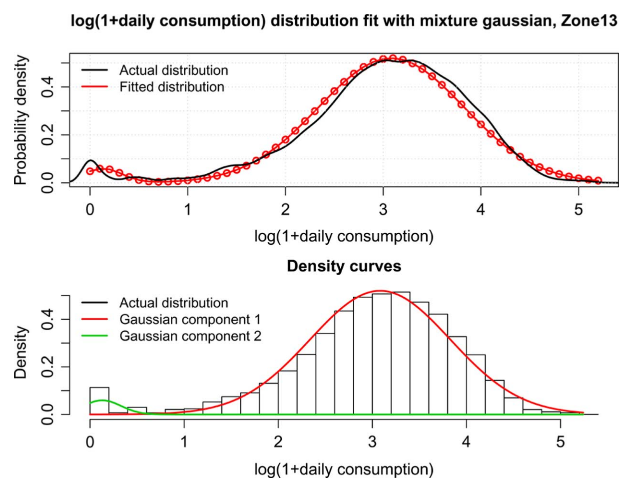


Fig. 3. Mixture of log normal distribution fitting on one zip code area for 2011 June–Aug. data.

where (μ_i, σ_i) are the mean and standard deviation of each mixture element, and λ_i is the proportion of each element in the population. Fig. 3 shows the fit for the population from one zip code during summer. The data fits well with one, two or three element mixture depending on the climate zone. In particular, we find that dry and hot areas require two to three elements, and cooler, coastal areas requires a single element. The parametric model fits well for different zones and seasonality or timing choices (e.g., winter season or a specific day), implying that it is not limited by temporal or spatial locality.

B. Encoding System Based on a Pre-Processed Dictionary

We focus on the normalized load shapes $s(t)$. Many features can be extracted from load shapes. In DR programs, peak usage fraction, peak time and peak duration can be important features to better control the demand at peak time. For EE programs, important information are features which can be used as proxy variables of the existence of specific appliances and their efficiency. For example, load sensitivity to temperature during summer can be a proxy variable of air conditioner existence. Besides, many other features can be extracted from this raw usage data depending on the interests of possible programs.

However, the data generated by sampling a large population hourly is enormous (e.g., for 220 K household data set, we have 66 MM load profiles), creating difficulties for any approach that segments consumers or investigates potential features to be extracted. To address this difficulty, we propose an encoding system using a pre-processed dictionary. The dictionary contains K representative load shapes $C_i(t)$. Every load shape in the data is mapped to the closest shape code. Load shape $s(t)$ is assigned to center $i^*(s)$ that minimizes the squared error:

$$E(s,i) = \sum_{t=1}^{24}(C_i(t) - s(t))^2,$$
$$i^*(s) = \arg\min_i E(s,i). \tag{3}$$

The encoding procedure also records the minimum squared error $E(s, i^*(s))$ for each encoded shape. The total energy a is characterized by its quantile according to the distribution $f(a)$ from the previous subsection. Various properties can be directly computed on the load shape dictionary. Note that given a load shape $s_n^k(t)$ for day n for household k, we can identify a sequence of shape codes $C_{i^*(s_n^k)}$, a sequence of total consumption values a_n^k and the sequence of errors $E((s_n^k, i^*((s_n^k))$. To reduce notation burden, whenever possible we omit the household index k.

Since no dictionary exists a prior, a procedure to uncover it is required. A good dictionary needs to have good coverage meaning every load shape in the data is sufficiently close to some representative shape. A good dictionary is also consistent, meaning that executing the learning procedure in different subsets of the population returns representative load shapes that are not too far from each other. The next subsection addresses this issue.

C. Adaptive K-Means on Normalized Data

Finding K representative shapes that minimize the sum of the mean squared errors $E(s, i^*(s))$ over all shapes s is a standard clustering problem. The K-means algorithm is the most popular statistical clustering approach. To populate a dictionary for representative shapes, the K-means algorithm can be a good starting point as tried in [3], [13]–[16] and [17]. However, the classical K-means algorithm needs to determine the number of clusters before running the algorithm. It is hard to decide on a proper K considering the large number of different load shapes. It is also not appropriate to follow statistical methods that set "K" without proper reasoning that provides a basis for their adoption.

Instead, we propose an adaptive K-means algorithm with a threshold to construct the shape dictionary [5]. The algorithm starts by a set of initialized cluster centers utilizing a standard K-means algorithm, with an initial $K = k_0$. Adaptive K-means then adds additional cluster centers, whenever a load shape $s(t)$ in the dataset violates the mean squared error threshold condition:

$$E(s, i^*(s)) = \sum_{t=1}^{24}\left(s(t) - C_{i^*(s)}(t)\right)^2 \le \theta \sum_{t=1}^{24} C_{i^*(s)}(t)^2, \tag{4}$$

where θ is the threshold choice. The threshold provides flexibility to cope with various practitioners' needs and control of the statistical properties of the load shapes in the same group. Since load shapes are normalized, each cluster center resulting from K-means is also normalized as they are the average of the member shapes. This guarantees that distances on both sides of (4) are bounded, and it is easy to demonstrate the range $0 \le \theta \le 2$ is required for non-trivial solutions. The main differentiation of the proposed algorithm from previous approaches is that the threshold test is utilized to dynamically split clusters that do not satisfy the condition. Together with the normalization utilized in the load shapes, it results in more robust dictionaries and better properties for the algorithm. The detailed algorithm is shown in Algorithm 1.

Algorithm 1 Adaptive K-mean algorithm based on threshold

Require: Daily load shapes for all users $\{s_n(t)\}$, Min and max number of clusters (min.k, max.k)
  Set K $=$ min.k
  while 1 **do**
    Run K-means with the initial centers (if given)
    for all clusters **do**
      Check the threshold condition in (4) for all $s_n(t)$ belonging to it and count the number of clusters violating the condition N_v
      (Meaning: Any data assigned to a cluster is not farther from the cluster center than the given threshold (θ) proportion.)
    end for
  if $N_v = 0$ **then**
      return the clustering results and K
    else if K $+ N_v >$ max.k **then**
      return message: failure to converge.
    end if
    K $=$ K $+ N_v$
    for all clusters violating the threshold condition **do**
      Run K'-means with K$' = 2$
    end for
    Update the set of cluster centers including all split clusters
  end while

Algorithm 2 Hierarchical clustering

Require: Adaptive K-means result
  (C_i: Cluster center (i $= 1, \ldots,$ K)/n_i: Size of i-th cluster)
  Set the target dictionary size, T$(<$ K$)$
  while K $>$ T **do**
    Find the closest two cluster centers, C_i and C_j
    Set $C_i = (n_i\, C_i + n_j\, C_j)/(n_i + n_j)$ and delete C_j
    K $=$ K $- 1$
  end while

D. Hierarchical Clustering

The resulting representative shape dictionary from K-means can be highly correlated as the adaptive K-means algorithm does not guarantee an optimal distance between cluster centers, and instead meets a threshold θ for every cluster. For interpretability and analysis, it is interesting to relax this condition for some clusters. We propose a simple hierarchical clustering algorithm to merge clusters whose centers are too close (Algorithm 2). The algorithm reduces the dictionary to a target size T by merging clusters. The weighted average is exactly the new cluster mean.

It is important to understand the purpose of the two stage clustering for generating the dictionary. If the dictionary size T is set directly, the performance is similar to classical K-means in threshold condition violation perspective which is addressed in Section III-C. However, classical K-means doesn't guarantee that every load shape is within a certain range of the cluster center. Adaptive K-means is needed to find proper K satisfying the desired threshold condition. Except that under this hard constraint, a number of small clusters can arise. Hierarchical clustering is utilized to filter and consolidate these small clusters to result in a small and stable dictionary, that is meaningful in practice.

III. Experiments on Data

A. Description of Smart Meter Data

The data used in this paper is provided by Pacific Gas and Electric Company (PG&E). The data contains the electricity consumption of residential PG&E customers at 1 hour intervals. There are 218 090 smart meters and the total number of 24 hour load profiles is 66 434 179. The data corresponds to 520 different zip codes, and covers climate zones 1, 2, 3, 4, 11, 12, 13, and 16 according to the California Energy Commission (CEC) climate zone definition. The data ranges from April 2008 to October 2011. However, the data range is different depending on smart meters. For example, there are a small number of smart meters having the data before July 2010 and after August 2011. Only data from 123 150 households with data ranging from August 2010 to July 2011 (44 949 750 load profiles) in experiments comparing individual households. Otherwise, the whole data set is utilized.

B. Dictionary Generation on Real Usage Data

It is noteworthy that the daily usage data whose sum is very small are ignored in populating a dictionary. This is because very small usage patterns are usually very irregular, and after scaling up by normalization, they perturb the overall representative shapes in the dictionary generation process. Moreover, typical demand management programs require at least some amount of usage to be applicable. After the dictionary is populated, very small usage data can be assigned to another code or can still be encoded by the dictionary because it is visible in the daily usage sum. Considering that the empirical total daily consumption is 18.68 kWh and the empirical 10% quantile is 4.47 kWh, the load patterns with total energy lower than 3 kWh (6% quantile) are ignored.

We evaluate the adaptive K-means clustering based dictionary construction. Fig. 4 shows the relation between the threshold setting and the size of the dictionary for 1 year of data

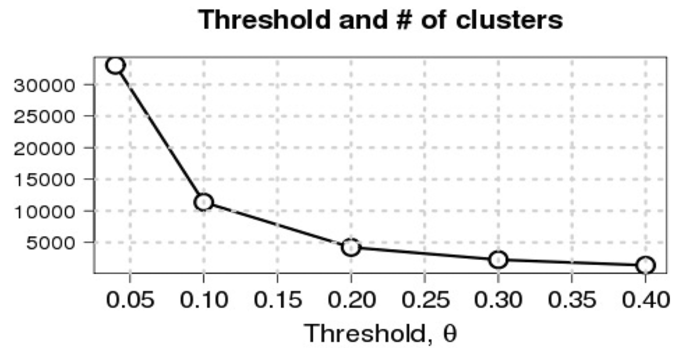


Fig. 4. Relation between threshold choice and number of clusters applying Algorithm 1.

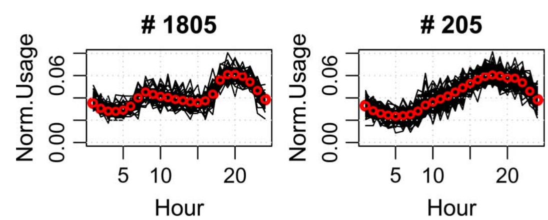


Fig. 5. Example of adaptive K-means result with $\theta = 0.2$: Normalized daily usage patterns (load shapes) and cluster centers.

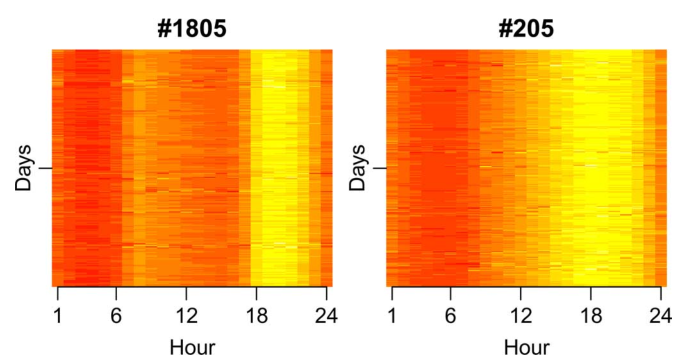


Fig. 6. Example of adaptive K-means result with $\theta = 0.2$: Heat map of normalized data under the same shape code.

(144 147 load patterns) at a chosen zip code. The dictionary size increases with decrease in threshold size as expected. A good choice of threshold is to set it at 0.2 where the number of clusters is not large, yet the marginal gain in error improvement to the explanatory power is small but requires a large number of cluster centers. Large number of cluster centers do not enable a stable dictionary. The resulting typical clusters are shown in Fig. 5, with the cluster center shown as red circles. From Fig. 5, "# number" identifies the load shape code in corresponding dictionary. Fig. 6 summarizes cluster information by plotting all shapes assigned to a cluster. The clusters can be seen to be consistent. We can check the robustness of the dictionary by estimating the coverage of the constructed dictionary in other zip codes and weather zones. Table I shows that the dictionary has good coverage implying that the uncovered clusters possess a stable structure.

C. Dictionary Reduction via Hierarchical Clustering

The dictionary size for $\theta = 0.2$ is on the order of thousands of shapes, making it hard to interpret the shapes and create meaningful metrics. The correlation between cluster centers reveals that some clusters are very strongly correlated, and cluster size

TABLE I
DICTIONARY COVERAGE FROM ZONE 13

Zone 13	Zone 3 coverage	Zone 2 coverage
A dictionary with θ = 0.2 populated	141876 (/143915) load shapes (98.6%)	85393 (/88771) load shapes (96.2%)

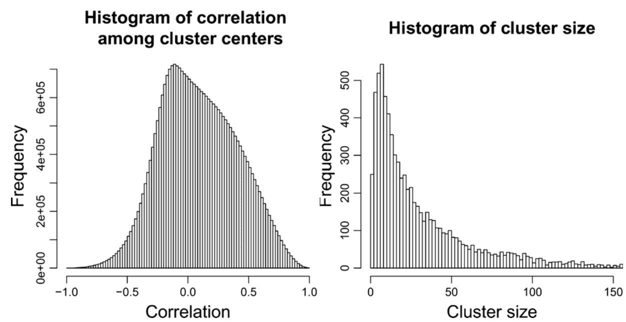


Fig. 7. Clusters correlation and size distribution.

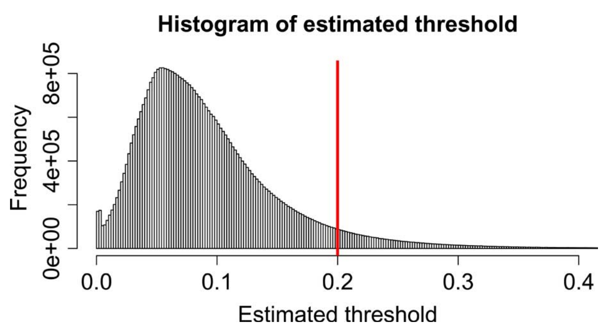


Fig. 8. Estimated threshold distribution.

distribution shows many clusters are of small size (Fig. 7). Hierarchical clustering (Algorithm 2) is applied to merge the clusters by sacrificing quality by the least possible amount. The target number of clusters is set to $T = 1000$ as it is the smallest dictionary size achieving the threshold condition violation less than 5% on sample load profiles.

The quality of the reduced dictionary can be evaluated by encoding all the load shapes in the data set. For each encoded load shape $s(t)$, we can compute the ratio

$$\hat{\theta} = \frac{\sum_{t=1}^{24}(s(t) - C_{i^*(s)}(t))^2}{\sum_{t=1}^{24}\left(C_{i^*(s)}(t)\right)^2}, \tag{5}$$

which would always be smaller than the threshold ($\theta = 0.2$) in the original adaptive K-means dictionary. Fig. 8 shows the distribution of estimated thresholds. Only small portion of load shapes (5.13%) are violating the threshold condition ($\theta < 0.2$). It is worth highlighting that a shape dictionary populated from one area covers 95% of all load shapes over all areas and periods, which means the representative load shapes are consistent regardless of spatial and temporal locality.

Another important question is whether for a given consumer, the deviations of all of his daily consumptions are close to the clustered shapes. Fig. 9 displays the distribution of the standard deviation of the residuals for each household, defined as the deviations from each of their daily shapes. Notice at any given hour, the error in cluster representation is up to 7% of the daily total (assuming 3σ bound).

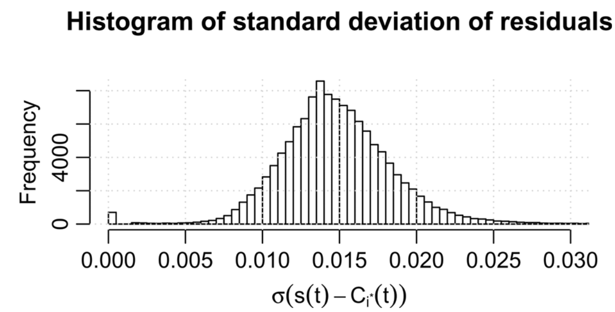


Fig. 9. Distribution of $\sigma(s(t) - C_{i^*(s)}(t))$.

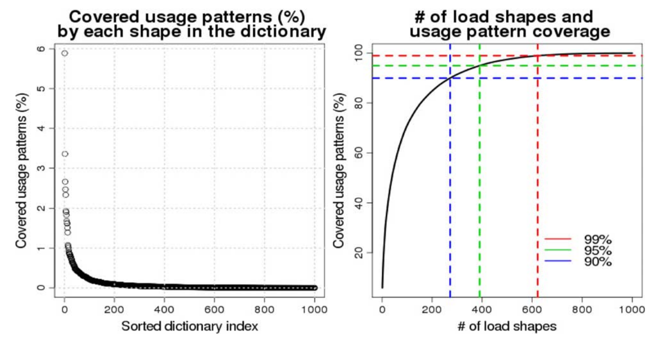


Fig. 10. Covered usage patterns and # of load shapes.

Additional statistics on coverage are provided by empirical distribution and cumulative distribution of cluster size for the whole population shown in Fig. 10. Notice that 90% of the whole data (66 MM load shapes) is covered by 272 representative load shapes (cluster centers). This enormous reduction in representation enables a principled analysis of household lifestyle based on load shapes, which we introduce later.

D. Load Shape Analysis

Basic observations on the encoded load shapes from the population data are described. Fig. 11 shows the most frequent 16 load shapes which account for more than 1% of load shapes in whole encoded data with the final dictionary. It can be seen that most of high usage happens in the late afternoon or evening, which represents the lifestyle of usual households. A sample lifestyle may be that households leave home in the morning, come back home after school or work and consume electricity till they sleep. Notice that many shapes can be differentiated by the timing of peak consumption, indicating this might be a good variable to design programs around. We can compare the number of households that have the top 16 load shapes among their top 5 load patterns to confirm that the top 16 patterns represent a population and not just a small set of consumers. Fig. 12 shows that on average each top 16 load shape appears in the top 5 of 15.3% of households.

IV. Segmentation Analysis

This section develops various analyses based on the load shape encoding developed in the paper.

A. Entropy Analysis

Typical analysis of load shapes for households has focused on average load shapes. Yet, two households with identical average load shapes could have significantly different daily load

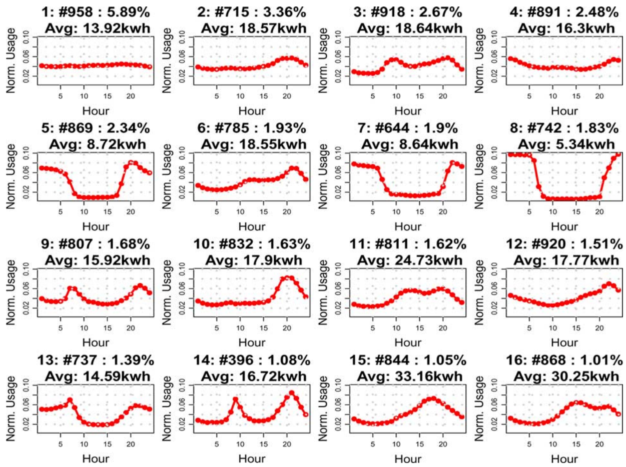


Fig. 11. 16 Most frequent load shapes of whole households.

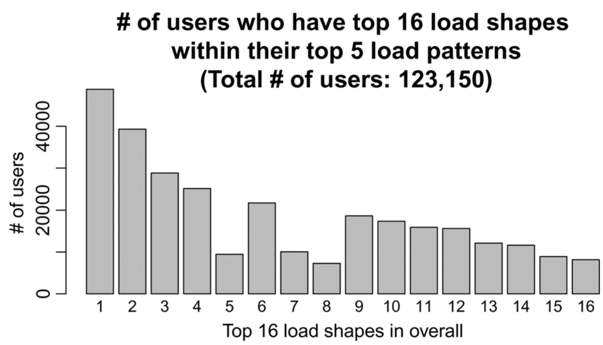


Fig. 12. # of users who have top 16 load shapes within their top 5 load patterns.

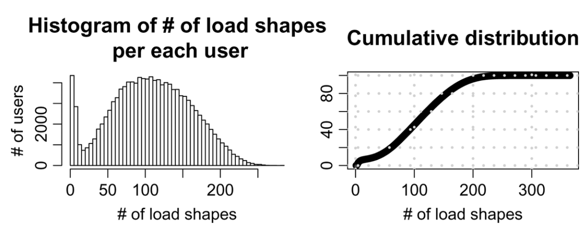


Fig. 13. # of load shapes per each user.

shapes. In fact, this *variability* is a very important factor in program targeting and costumer engagement. For example, it could potentially be easier to target DR to a more stable household that consumes the same load shape everyday, than one that is highly variable. On the other hand, it might be better to target behavioral modification and energy efficiency programs to households with a more diverse set of behaviors.

Fig. 13 displays the histogram and cumulative distribution of the number of (encoded) load shapes observed per household with a full calendar year of data (365 days). It clearly shows that there are households that follow a limited set of load shapes, and a more variable set of households. Since the time horizon in our present analysis is 365 days, the maximum number of load shapes for a customer is 365 (although the dictionary has $T = 1000$ codes). Fig. 13 shows that the most variable household has 285 load shapes, and 45% of households have less than a 100 shapes a year.

We propose utilizing the notion of entropy to create a metric that captures customer variability. For each household n, we record the relative frequency $p_n(C_i)$ of each encoding cluster

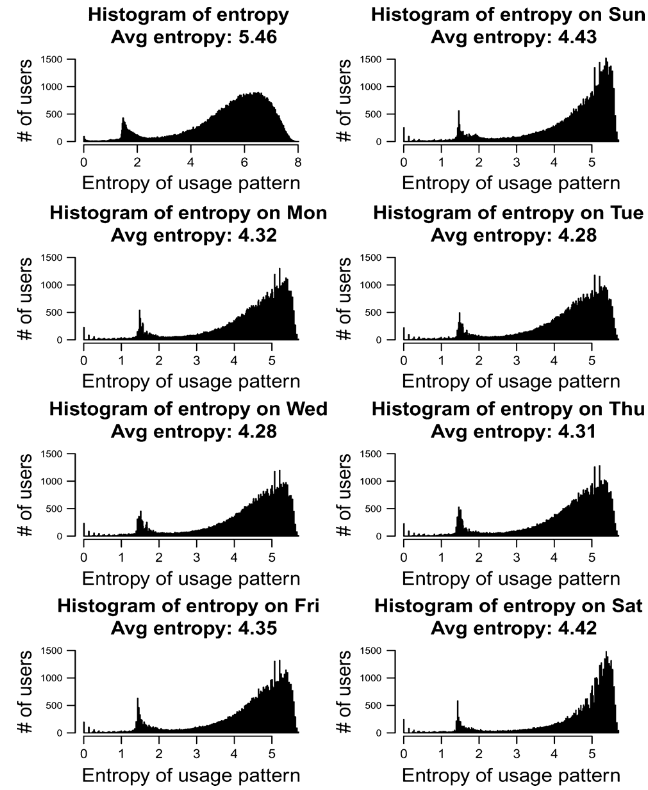


Fig. 14. Load shape entropy distribution.

center C_i in his daily series. Then the *entropy* of household n is given by

$$S_n = -\sum_{i=1}^{K} p(C_i)\, \log\, p(C_i). \tag{6}$$

The entropy is highest if all the cluster centers are equally likely in the data set (i.e., $p_n(C_i) = 1/K$) and lowest (i.e., $S_n = 0$) if the household follows a single cluster center. We can also compute the entropy of households for each day of the week by computing the relative frequency of codes on each day of the week separately. Fig. 14 displays the results. The weekday distribution does not differ much from the overall distribution. The average entropy during weekends is higher than during weekdays, which is reasonable because households have a more regular lifestyle on weekdays.

Households can be segmented by their positions in the distributions in the top left plot of Fig. 14. For example, if a household's entropy has a quantile above 75%, it can be classified as a *variable* household, and if it has a quantile below 25% it can be classified as a *stable* household. Also notice there is a group of load shapes with very low entropy (6480 households of entropy between 1.3 and 2). Fig. 15 shows the 4 most frequent load shapes accounting for 88% of the load shapes in the group. The average daily usage of this group is 8.13 kWh which is much lower than the average daily usage (19 kWh) of the whole data provided. This group has smaller homes that are empty during the day, and can either correspond to a lifestyle or to empty homes.

B. Shape Analysis

[3] suggests 6 representative average load shapes for households: "Evening," "Night," "Afternoon," "Morning," "Day-

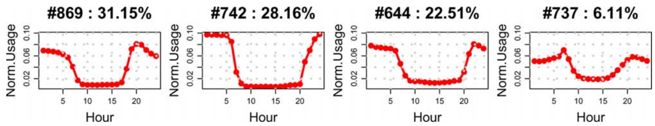


Fig. 15. Most frequent load shapes in the low entropy group.

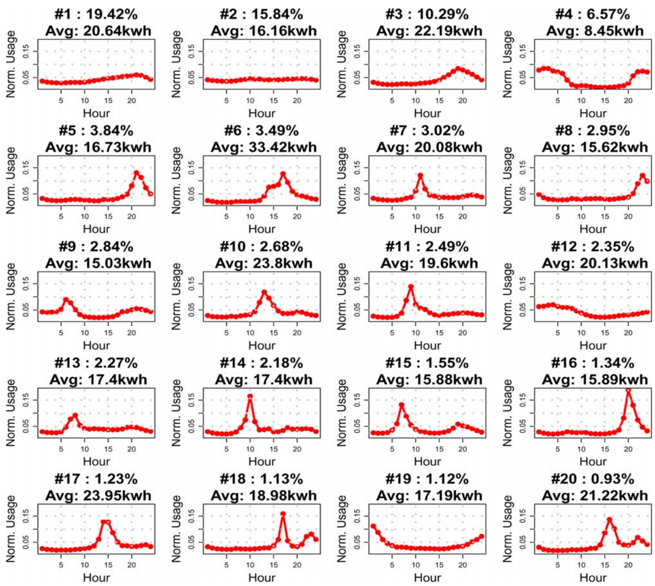


Fig. 16. 20 most frequent load shapes of whole households using the dictionary of size $= 100$.

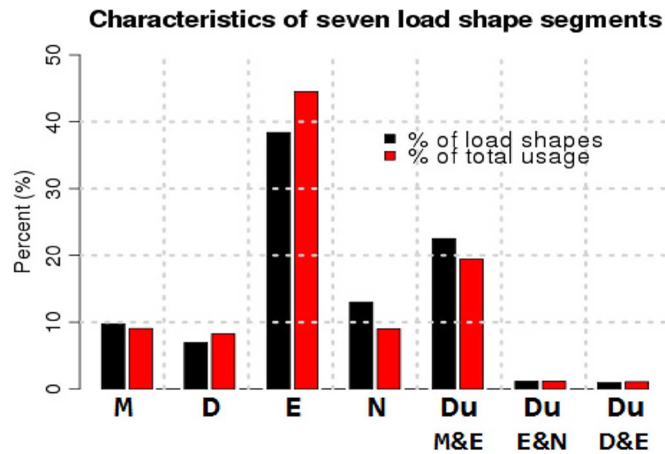


Fig. 17. Characteristics of seven load shape segments.

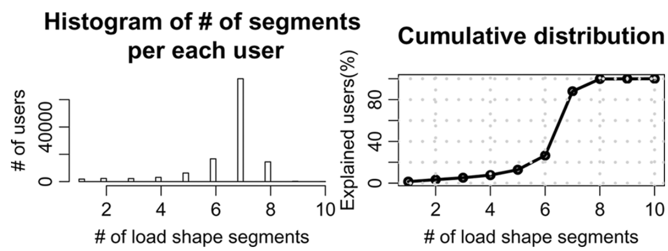


Fig. 18. # of load shape segments per each user.

time," and "Dual peak" segments. We develop the methodology further here. The encoding dictionary is reduced by further hierarchical clustering with $T = 100$. Fig. 16 shows the most frequent 20 load shapes for all households. The top 20 load shapes provide a coverage of 87.5% of all shapes, and each has a frequency higher than 1%. These top 20 shapes can be segmented according to the timing of consumption:

Morning peak (M: 4:00–10:00): Load shapes (#9, 11, 13, 14) belong to this segment. Except #11, the load shapes have relatively low daily average consumption of up to 18.68 kWh, as well as small peak values. The main difference in four load shapes is the peak time (6 A.M., 9 A.M., 8 A.M., 10 A.M. in corresponding order). This segment has low potential for targeting DR programs.

Daytime peak (D: 10:00–16:00): Load shapes (#7, 10, 17) belong to this segment. These load shapes have relatively high average usages (above 20 kWh).

Evening peak (E: 16:00–22:00): Load shapes (#1, 3, 5, 6, 16) are included in this segment. This segment explains the most load shapes (about 40%) among all segments. Load shape #6 has very large average usage (33.42 kwh) with peak time (4 P.M.–6 P.M.). This segment can be a potentially significant target for DR programs.

Night peak (N: 0:00–4:00, 22:00–24:00): Load shapes (#4, 8, 12, 19) belong to this segment. Except load shape #12, these load shapes show low average usage. Especially, #4 has very low average daily total usage (8.45 kWh). This is similar to the low entropy load shapes in Fig. 15. This segment has low potential for targeting of DR programs.

Dual peak Morning & Evening (Du M&E): Load shapes (#2, 15) are in this segment. It may be possible to say #9, 13 have weak dual peaks. The load shapes in this segment have average usage a little below the empirical average and morning peak is the primary peak for these load shapes except #2.

Dual peak Evening & Night (Du E&N): Load shape #18 represents this segment. A sample lifestyle in this segment would be that they cook dinner with electric appliances and take a rest, then they play with computer or some electronics before they sleep.

Dual peak Daytime & Evening (Du D&E): Load shape #20 having two peaks at 4 P.M. & 10 P.M. is at the border line between "Evening and Night peak" and "Daytime and Evening peak. If #20 is 1 hour shifted to right, then it is very similar with #18 which has two peaks at 5 P.M. & 11 P.M.. So, a sample lifestyle can be similar with the previous one with 1 hour shift.

We can define three additional segments in the dual peak shapes: Morning & Daytime (M&D), Morning & Night (M&N), and Daytime & Night (D&N). However, these segments contain less than 0.1% of all load shapes. Fig. 17 plots aggregate statistics of all the segments, including the % of load shapes and the % of the total energy usage. Daytime peak and Evening peak segments can be the high potentials for targeting of DR programs while Morning peak and Night peak segments have low potential.

We can encode the daily household shapes relying on the 10 segments defined above. The distribution of the number of segments each household belongs to is shown in Fig. 18. Notice that most households have load shapes that belong to seven segments. 87.4% of households can be explained by the seven load shape segments in Fig. 17.

We recompute the user entropy, to see the effects of relative rates of each segment, obtaining the result in Fig. 19. The average entropy is 1.76, which means the typical household load

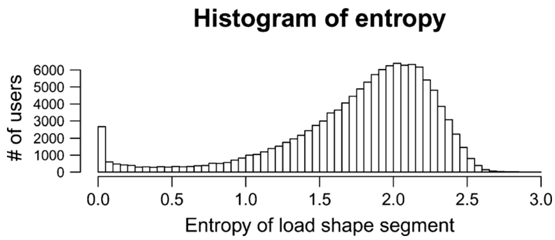


Fig. 19. Load shape segment entropy distribution.

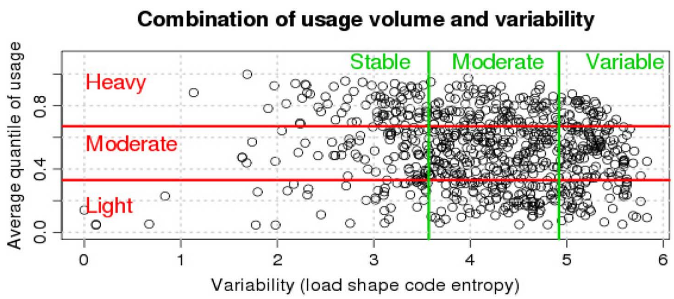


Fig. 20. Combination of usage volume and shape variability.

shapes belong to 3.38 segments throughout the year. There are also significant numbers of households that have low entropy.

C. Multidimensional Segmentation

In this section we show how to segment the households using a combination of multiple clustering criteria.

The segmentation based on consumption timing developed in the previous subsection indicates good subsets of the population of households to apply for different programs, such as DR. We select all users with at least one load profile in the desired segment. Yet, that segmentation does not include two important dimensions that need to be considered: *quantity* and *variability*. Section II shows that a mixture of log-normal distributions fits well the daily consumption distribution. Using the fitted distribution, the quantile of daily consumption can be calculated for each day. Then, using the average quantile, a household can be assigned a group: Heavy, Light, or Moderate. Variability in consumption can be captured by computing the Entropy of each household, and classified into Stable, Moderate, and Variable according to the entropy.

Fig. 20 displays a scatter plot, where each point corresponds to a household's entropy and average usage quantile. Based on this plot, nine classes of households are created. The average quantile consumption is divided into three groups since for this particular subset of households, the mixture model had two distributions, which naturally expresses three classes: light (for those households whose consumption is mostly drawn from the first mixture), moderate (for those households whose consumption is drawn from either mixture), and heavy (for households whose consumption is mostly drawn from the second mixture). For example, to target the households for an automated DR program the focus can be on heavy and stable users, in the appropriate time-based segment (e.g., daytime peak). Fig. 21 shows the four most frequent load shapes among the heavy and stable households corresponding to Fig. 20. Those four load shapes explain about half (47%) of all usage patterns in the filtered users. The first two load shapes could be very good candidates for DR, since they have large relative peaks.

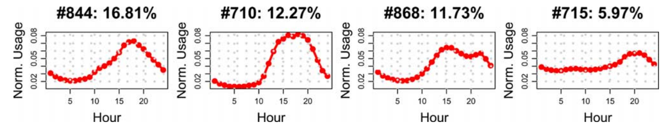


Fig. 21. Most frequent load shapes in the filtered users.

TABLE II
THE NUMBER OF HOUSEHOLDS IN FIG. 20

	Stable	Moderate	Variable	Total
Heavy	79 (10.2%)	103 (13.2%)	38 (4.9%)	220 (28.3%)
Moderate	73 (9.4%)	187 (24.1%)	106 (13.6%)	366 (47.1%)
Light	40 (5.1%)	100 (12.9%)	51 (6.6%)	191 (24.6%)
Total	192 (24.7%)	390 (50.2%)	195 (25.1%)	777 (100%)

TABLE III
COMPARING LOAD SHAPE FREQUENCIES AMONG GROUPS

t-test to check $P(C_i|condition\ A) = P(C_i|condition\ B)$

N_1: sample size satisfying condition A

N_2: sample size satisfying condition B

$\bar{X}_1$: $\frac{\#\ of\ C_i\ among\ N_1}{N_1}$, $\bar{X}_2$: $\frac{\#\ of\ C_i\ among\ N_2}{N_2}$

$S_1^2 = \bar{X}_1(1-\bar{X}_1),\ \ S_2^2 = \bar{X}_2(1-\bar{X}_2)$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}\sqrt{\frac{(N_1-1)S_1^2+(N_2-1)S_2^2}{N_1+N_2-2}}},\quad d.f. = N_1 + N_2 - 2$$

1) $T < t_{0.025}$: $P(C_i|condition\ A) < P(C_i|condition\ B)$
2) $T > t_{0.975}$: $P(C_i|condition\ A) > P(C_i|condition\ B)$
3) Otherwise: $P(C_i|condition\ A) = P(C_i|condition\ B)$

Energy efficiency programs on the other hand would target households that are more variable, and thus, exhibiting behavior choices that could be induced via different forms of interventions. Potentially, the analyst could focus on heavy or moderate users, in the variable class. Table II summarizes the numbers of users on different classes, and shows that filtering can significantly reduce the number of households that require a deeper and potentially much more time intensive analysis.

D. Spatial Locality Analysis

EE and DR programs are managed assuming that household consumption patterns are affected by climate and other factors decided by spatial locality. In fact the usual practice is to focus DR and EE programs in particular zip codes or climate zones. In this subsection we validate whether consumption patterns are indeed influenced by such locality. The total daily consumption is clearly influenced, as coastal and cool climate zones only exhibit a single mixture component with a lower mean during the summer, while inland and hot climate zones have mixtures with at least two components. The higher component corresponds to cooling energy consumption during summer.

It is also interesting to examine whether load shapes exhibit locality effects as well. For example, we compare Zone 3 (an cool coastline area) and Zone 12 (an inland hot area). The frequency of each load shape is compared in two zones. Since the frequency is an estimate of true frequencies of load shapes, we utilize a two sample t-test (Table III). The assumption is that load shapes are drawn independently from a multinomial distribution. Table III shows that the two zones have different frequencies on 80% of load shapes and similar frequencies on the remaining 20%.

More detailed comparisons can be drawn by looking at the frequent load shapes in both zones. Figs. 22 and 23 show Zone 3 has more frequent load shapes with late evening or night peaks

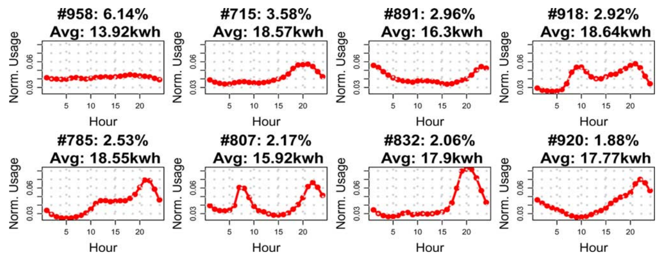


Fig. 22. More frequent load shapes in Zone 3.

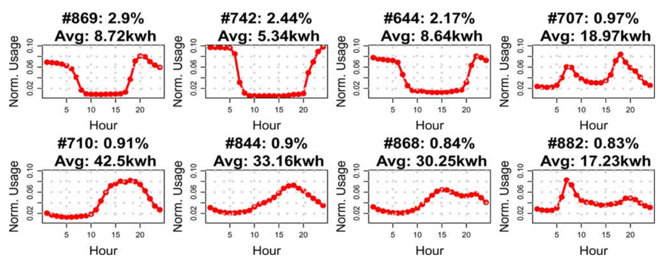


Fig. 23. More frequent load shapes in Zone 12.

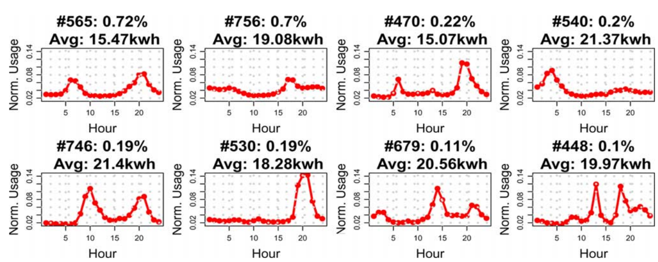


Fig. 24. Common load shapes in both zones.

TABLE IV
T-TEST RESULT

$P(C_i\|Zone3) > P(C_i\|Zone12)$	272
$P(C_i\|Zone3) < P(C_i\|Zone12)$	532
$P(C_i\|Zone3) = P(C_i\|Zone12)$	196
Total	1000

TABLE V
T-TEST RESULT

$P(C_i\|Weekdays) > P(C_i\|Weekends)$	322
$P(C_i\|Weekdays) < P(C_i\|Weekends)$	496
$P(C_i\|Weekdays) = P(C_i\|Weekends)$	182
Total	1000

and moderate consumption, while Zone 12 has afternoon or early evening peak shapes with heavy consumption. Zone 3 has relatively mild climate and use electricity primarily on heating at night. Zone 12 may have many customers using air conditioners during afternoon and early evening due to its hot climate. Fig. 24 shows no common load shape covers more than 0.8% of shapes for both zones, confirming both zones have different load shape distributions.

E. Temporal Locality Analysis

It is interesting to analyze whether there exists temporal locality in load shape choice. For example, we can compare load shape frequency distributions between weekdays and weekends. The two sample t-test is also used. Table V shows that 82% of load shapes are distributed distinctly among both groups.

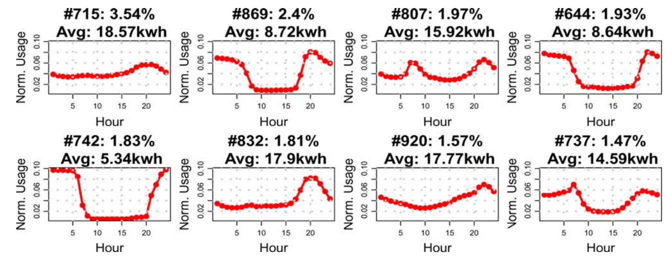


Fig. 25. More frequent load shapes in Weekdays.

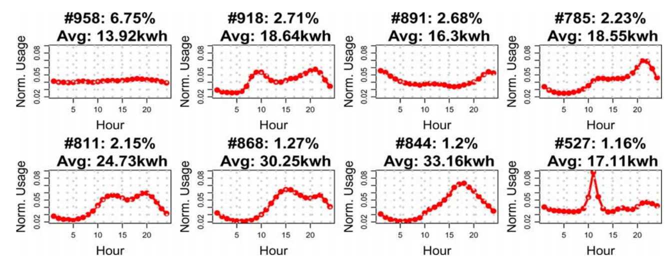


Fig. 26. More frequent load shapes in Weekends.

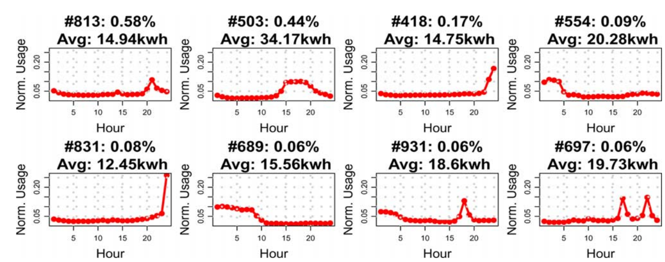


Fig. 27. Common load shapes in Weekdays & Weekends.

Fig. 27 shows no common load shape explains more than 0.6% for weekdays and weekends at the same time. Fig. 25 confirms load shapes corresponding to regular working lifestyles (#644, 742, 869) are frequent during weekdays. Figs. 25 and 26 show that morning peak, night peak, dual peak (morning & evening) load shapes happen more in weekdays with moderate consumption while there are more daytime and afternoon peak shapes in weekends with heavy consumption in three load shapes (#811, 844, 868). Possibly households are empty during daytime and afternoons in weekdays and consume actively during morning or evening. During weekends, households can start a day a bit later than weekdays like #527 or stay home with various activities requiring substantial electricity consumption (e.g., cooking, doing the laundry or cooling) in #811, 844, 868.

Winters and summers have different load shape choices as well, but we omit the analysis which is performed in a similar manner. In case of an actual region based DR program design process, this temporal locality analysis can be more detailed by conditioning specific times and combining with spatial locality investigation.

V. IMPACTS ON LOAD FORECASTING

The methodology can be used to drive improvements in peak load forecasting for a power system zone. The key observation is from the analysis in Fig. 17. If predicting total peak load for a particular hour, only a subset from the set of households that are in a relevant class influence such forecast. Therefore, additional information collected about such households could significantly increase the prediction accuracy.

Moreover the proposed approach can inform load forecasting about individual households. Such forecasting is important for design of microgrids and intelligent distribution systems. The methodology suggests that different consumer classes might require different forecasting approaches. In particular, Figs. 19 and 20 show that households can be classified according to entropy. Low entropy consumers are easier to forecast at an individual level, and high entropy consumers are harder to forecast since they have significantly more variability. Moreover, in analyzing the performance of forecasting, it is important to distinguish the differences for the various classes.

Our method could also drive algorithms for load or load shape forecasting for individuals. After the encoding procedure, each household would have a sequence of load shape code and one of daily consumption. Load shape can be forecasted using various Markov chain type methods or advance classification algorithms after reducing the size of the load shape dictionary. With those results, any daily consumption prediction method can be merged to forecast the load at a specific time.

VI. Conclusions and Future Work

The implications for the methods described here have implications for utility policy and programs such as DR and EE. Using customers load shape profile, we can effectively target residents that have the highest potential for benefiting from DR programs. Load shape based high potential targeting can have significant benefits: increased likelihood of success, energy savings, and public relations benefits from successful engagement in utility programs.

Load shape based energy use profiles that incorporate level of use and entropy offer other potential benefits. For example, recommendations for energy reduction, or critical peak pricing that are "lifestyle" based would be very different from the appliance and device based recommendation currently used by most utilities. Lifestyle recommendations include focusing on shapes such as morning and afternoon or only afternoon peaks and suggesting that they move activities earlier or later in the day. Since it is rare that a single load shape represents a lifestyle, lower energy or off peak load shapes within a household repertoire of shapes also could be recommended as a means of energy reduction and savings.

Beyond load shape segmentation, the extent of entropy within a household could yield further understanding of the potential of success for targeting and recommendation design. For example, high entropy households, indicating variability in occupancy and energy using activity, may have low potential for targeting for DR programs but high potential for energy reduction programs such as appliance rebates.

Future experimental research is needed to validate the targeting and energy saving potential of the segmentation methods.

Acknowledgment

This research was funded in part by the Department of Energy ARPA-E under award number DE-AR0000018, the California Energy Commission under award number PIR-10-054, and Precourt Energy Efficiency Center. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Prof. Rajagopal is funded by TomKat center under award number 1149360-1-WPTAI and by a fellowship from the Powell Foundation. The authors acknowledge the support of PG&E for providing the anonymized data. The authors also acknowledge Brian Smith, Ann George, Arlene Teves, Sylvia Lau, Martin Fischer, Chin-Woo Tan, Steve Eglash, and Carrie Armel's comments and feedback on early versions of the paper.

References

[1] B. Sutterlin, T. A. Brunner, and M. Siegrist, "Who puts the most energy into energy conservation?," *Energy Policy*, vol. 39, pp. 8317–8152, 2011.

[2] T. F. Sanquist *et al.*, "Lifestyle factors in US residential electricity consumption," *Energy Policy*, vol. 42, pp. 354–364, 2012.

[3] B. A. Smith, J. Wong, and R. Rajagopal, "A simple way to use interval data to segment residential customers for energy efficiency and demand response program targeting," in *ACEEE Proc.*, 2012.

[4] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Trans. Power Syst.*, vol. 20, pp. 596–602, May 2005.

[5] S. Bhatia, "Adaptive K-means clustering," in *Proc. Int. Florida Artif. Intell. Res. Soc. Conf.*, 2004.

[6] L. Dethman and D. Thomley, "Comparison of segmentation plans for residential customers," in *Energy Trust*, 2009.

[7] L. Lutzenhiser *et al.*, "Behavioral assumptions underlying California residential energy efficiency programs," in *CIEE Energy & Behavior Program*, Berkeley, CA, 2009.

[8] S. Moss, "Market segmentation and energy efficiency program design," in *CIEE Energy & Behavior Program*, Berkeley, CA, 2008.

[9] Opinion Dynamics Corp., "Final segmentation report—California Public Utilities Commission," 2010.

[10] G. W. Irwin, W. Monteith, and W. C. Beattie, "Statistical electricity demand modelling from consumer billing data," *IEE Proc. C, Gener., Transm., Distrib.*, vol. 133, pp. 328–335, 1986.

[11] C. F. Walker and J. L. Pokoski, "Residential load shape modelling based on customer behavior," *IEEE Trans. Power App. Syst.*, vol. PAS-104, pp. 1703–1711, Jul. 1985.

[12] B. D. Pitt and D. S. Kitschen, "Application of data mining techniques to load profiling," in *Proc. 21st IEEE Int. Conf. Power Ind. Comput. Appl.*, Jul. 1999.

[13] M. Espinoza *et al.*, "Short-term load forecasting, profile identification, and customer segmentation: A methodology based on periodic time series," *IEEE Trans. Power Syst.*, vol. 20, pp. 1622–1630, Aug. 2005.

[14] C. Flath *et al.*, "Cluster analysis of smart metering data—An implementation in practice," *Business Inf. Syst. Eng.*, vol. 4, pp. 31–39, 2012.

[15] T. Rasanen and M. Kolehmainen, "Feature-based clustering for electricity use time series data," in *Proc. 9th Int. Conf. Adaptive Natural Comput. Algorithms*, 2009.

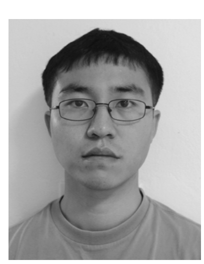
[16] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.*, vol. 21, pp. 933–940, May 2006.

[17] G. Chicco *et al.*, "Load pattern-based classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 19, pp. 1232–1239, May 2004.

[18] G. Coke and M. Tsao, "Random effects mixture models for clustering electrical load series," *J. Time Series Anal.*, vol. 31, no. 6, pp. 451–464, 2010.

[19] S. Verdu *et al.*, "Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps," *IEEE Trans. Power Syst.*, vol. 21, pp. 1672–1682, Nov. 2006.

[20] G. Tsekouras, N. Hatziargyriou, and E. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 22, pp. 1120–1128, Aug. 2007.



Jungsuk Kwac received the B.S. degree from Seoul National University, Korea, and the M.S. degree from Stanford University, Stanford, CA, USA. He is currently a Ph.D. candidate in the Department of Electrical Engineering at Stanford University, and works at Stanford Sustainable Systems Lab. His research focuses on smart grid data analytics, especially segmentation and targeting. He has worked at Oracle (2008 to 2011) as a Senior Application Engineer. He also worked for several years in the IT industry in Korea as a software engineer.

June Flora received the Ph.D. degree in educational psychology from Arizona State University, Tempe, AZ, USA.

She is currently a Senior Research Scientist at the Human Sciences Advanced Technologies Research Institute, Stanford University, Stanford, CA, USA. Her research focuses on the study of energy reduction and conservation via social media and interpersonal communication programs. She also does survey research on the relationship of extreme weather experience, attention to weather news, and climate change communication and the role of family communication on climate friendly behavior.

Ram Rajagopal received the B.S. degree in electrical engineering from the Federal University of Rio de Janeiro, Brazil, the M.S. degree in electrical and computer engineering from University of Texas, Austin, TX, USA, and the M.A. degree in statistics and the Ph.D. degree in electrical engineering and computer sciences from the University of California Berkeley, CA, USA.

He is an Assistant Professor of Civil and Environmental Engineering at Stanford University, Stanford, CA, USA, where he directs the Stanford Sustainable Systems Lab (S3L), focused on large scale monitoring, data analytics and stochastic control for infrastructure networks, in particular energy and transportation. His current research interests in power systems are in integration of renewables, smart distribution systems, and demand-side data analytics.