Appliance based model for energy consumption segmentation

Zi Yin, Thanchanok Teeraratkul, Nutthavuth Tamang

December 11, 2014

Abstract

We propose two steps energy segmentation at the appliances level. The first step is to model a schedule of appliances usage of each household, using convex optimization and backward search features selection. The second step is to use hierarchical clustering to group households appliance usage patterns into a small number of classes. Using Dynamic Time Warping (DTW) as a clustering metric results in a smaller number of classes and higher clustering efficiency measure, comparing to L_1 metric and K-means.

1 Introduction

Problem statement and Contribution The previous works on classification of consumer load shapes focus on clustering aggregated energy profile using L_2 based mechanism such as K-means [KFR14], [KTS+13], [THD07], [YV93]. While this is a reasonable thing to do, two main concerns arise. First, the aggregated energy profile does not clearly represent the underlying behavior of household energy usage. The second concern is that L_2 metric used in K-means penalizes the difference uniformly across time, and might result in a poor representation of consumer class. Our contribution in this paper is as follows :

1. Given a set of hourly electricity consumption, $x \in \mathbf{R}^{24}$, and power levels of appliances, $p \in \mathbf{R}^{20}$, we model **Appliance Usage Matrix** A for each user, where $A_{ij} \in \{0, 1\}$ indicates whether an appliance with power level j is used at time interval i. This step is a supervised learning as we find a minimum norm A which satisfies $Ap = x$, and round out A_{ij} to 0 or 1 . Furthermore, for each user, we use backward search feature selection, with leave-one-out cross validation (LOOCV) to choose a subset of features $f \subseteq \{p_j\}_{j=1}^{20}$.

2. We present a new metric for clustering **Appliance Usage Matrix** A based on Dynamic Time Warping (DTW) [Mul07], which is a technique to find an optimal alignment between two given time dependent sequences. Our contribution is the design of DTW based metric that accounts for elastic changes across each column of A, which reflects human randomness in usage schedule of each appliance. Using hierarchical clustering based on DTW metric results in a relatively smaller number of classes and higher clustering efficiency measure, comparing to L_1 metric and K-means.

2 Data Description

The data used in this paper is provided by Pecan Street. The data contains the daily electricity consumption of residential Pecan Street customers at 1 hour intervals. The data used in the analysis comes from 223 smart meters. The total number of 24 hours load profiles is 6690. The data ranges from March 1, 2013 to March 30, 2013.

3 Methodology

3.1 Modeling Household Appliance Schedule

For each household l, we model an **Appliance Usage Matrix** $A^{(l)}$, where $A_{ij}^{(l)} \in \{0,1\}$ indicates whether an appliance with power level j is used during hour i. Let $\bar{x} \in \mathbf{R}^{24}$ be an hourly mean energy usage of a household over D samples (D days of data), and p be a vector of power levels. Suppose that p is known, we define

$$\tilde{A} = \underset{Ap=\bar{x}}{\operatorname{argmin}} \sum_{i,j} |A_{ij}|, \tag{1}$$

then, we approximate $A^{(l)}$ by setting $A_{ij}^{(l)} = \mathbb{1}\{\tilde{A}_{ij} > \gamma\}$ where γ is a median of $\{\tilde{A}_{ij}\}$. Initially, p is the same for all households, *i.e.*, 20 power levels of appliances. However, since each house might not have appliances at all power levels, we use backward search feature selection with leave-one-out cross validation (LOOCV) to choose a subset of $\{p_j\}$ for each house. The LOOCV and backward search algorithms are listed in Appendix 1.

After feature selection, we solve (1) for A using p^*, and round out entries of A to 0 and 1. Now, the number of columns in A will be dimension of p^*. We expand the matrix A to have number of columns equal to dimension of the full feature vector, p, by padding zeros to the column of A corresponding to left out features.

3.2 Non-probabilistic Approach to Clustering

Hierarchical clustering To obtain a smaller number of load shape representations, we cluster the **Appliance Usage Matrix** $A^{(l)}$, $l = 1 \dots L$ (L households) using agglomerative hierarchical clustering and divisive hierarchical clustering. The detailed clustering methodology is included in Appendix 2.

Drawback of euclidean metric Our model of energy usage assumes that each appliance usage schedule has human randomness, which introduces time shift, stretch, or contraction across time within a reasonable bound. An appropriate clustering metric should not penalize elastic changes within a reasonable bound. However, Euclidean metric does not possess this property, and may mistakenly combine pairs generated from different routines.

3.3 Dynamic Time Warping (DTW)

To account for the drawback of euclidean metric, we explore an alternative dissimilarity measure, Dynamic Time Warping (DTW). As described in [Mul07], DTW is a technique to find an optimal alignment between two given time dependent sequences based on dymamic programming. Figure 1 shows the possible comparison path DTW can produce.

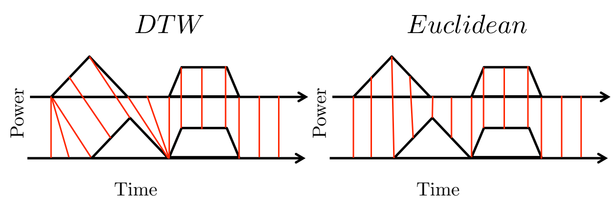
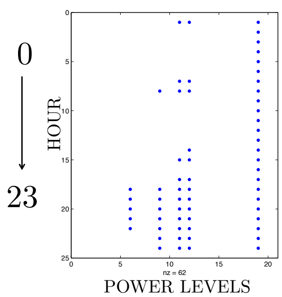


Figure 1: DTW alignment gives zero dissimilarity between two load curves while Euclidean penalizes mismatch across time.

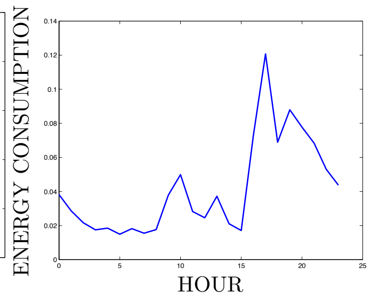
4 Results and Analysis

4.1 Result from Appliance Usage Matrix Modeling

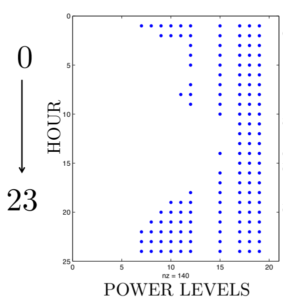
For each household, sparsity pattern of the output A agrees with the majority of its load curves x, *i.e.*, non-zeros elements concentrate in those rows where hourly consumption peaks. An example sparsity pattern of **Appliance Usage Matrix** for **Evening Peak** user and **Dual Morning & Evening Peaks** user is presented (Figure 2a and 3a). Table **??** shows test error from LOOCV for userID 26 after each feature has been removed.



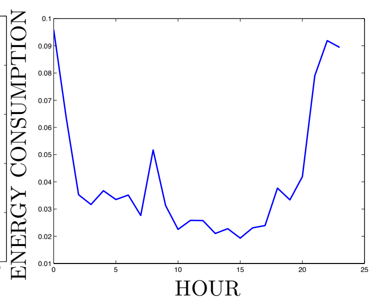
(a) A for userID 26 (Evening Peak)



(b) Load curve for userID 26 on January 1 2012

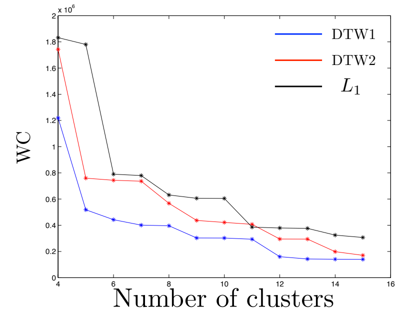


(a) A for userID 94 (Dual Morning & Evening Peak)

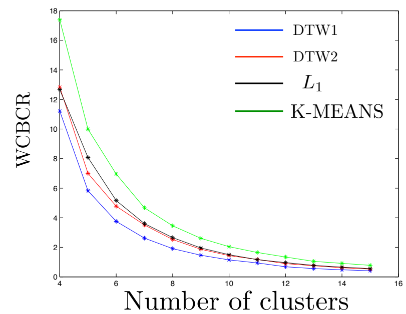


(b) Load curve for userID 94 on January 1 2012

Figure 3: Examples of Load curve and its corresponding sparsity pattern for matrix A.



(a) Total within-cluster distance



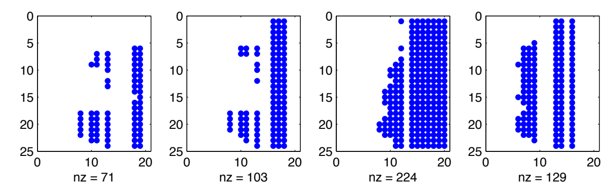
(b) Ratio of intra-cluster sum to inter-cluster variation

Figure 4: Efficiency measures of clusters produced from Agglomerative hierarchical clustering

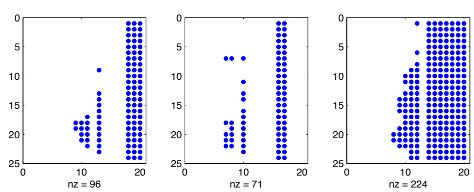
4.2 Result from Hierarchical Clustering

We apply two patterns of DTW and L_1 metrics for clustering. To determine the number of clusters, we set a limit on L_1 dissimilarity between each **Appliance Usage Matrix** A and its cluster center. DTW achieves less clusters than L_1. Figure 5 shows cluster centers using Agglomerative hierarchical clustering.

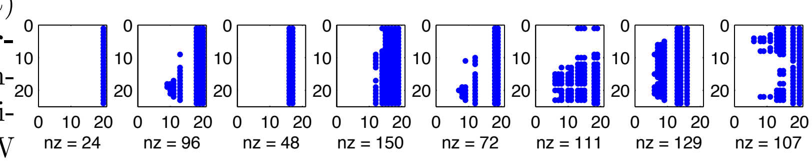
Figure 4 plots **Within Cluster Sum** (WC) and **Ratio of within cluster sum to inter-cluster variation** (WCBCR) against the number of cluster K. For Agglomerative and Divisive clustering, WC and WCBCR using DTW is lower than those using L_1 and K-means. It suggests that besides achieving smaller cluster number, DTW also yields more compact clusters and larger inter-cluster variation, which is encouraging.



(a) Cluster centers with members from DTW1 metric



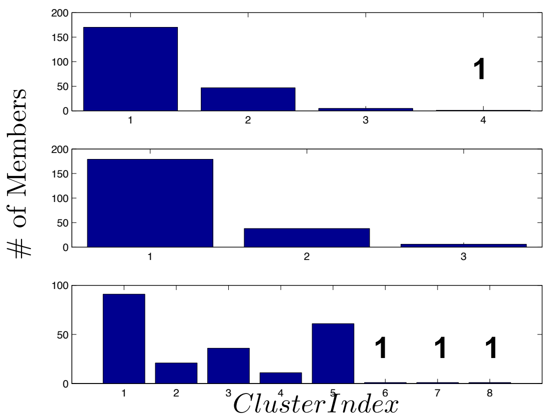
(b) Cluster centers with members from DTW2 metric



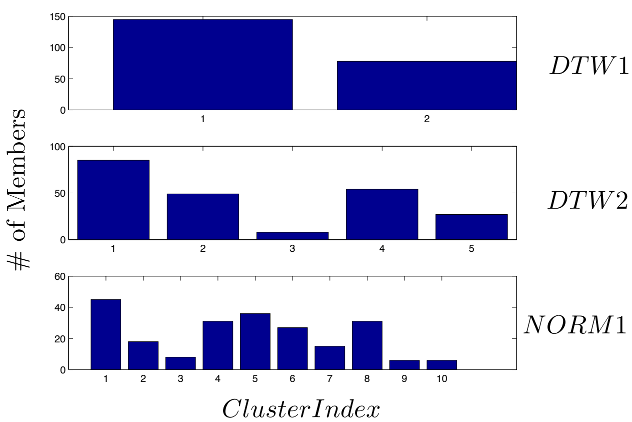
(c) Cluster centers with members from L_1 metric

Figure 5: Clusters produced from Agglomerative hierarchical clustering

Comparison between Agglomerative and Divisive Clustering While the Agglomerative clustering yields lower WCBCR (Figure 7b), the produced clusters are less compact (Figure 7a). Furthermore, the Agglomerative clustering tends to leave out clusters with a single member or a few members (Figure 6a), which is undesirable since those clusters poorly represent an entire population.

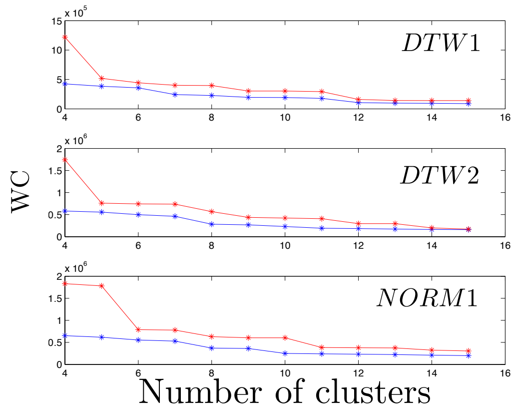


(a) Number of members within each clusters from Agglomerative hierarchical clustering

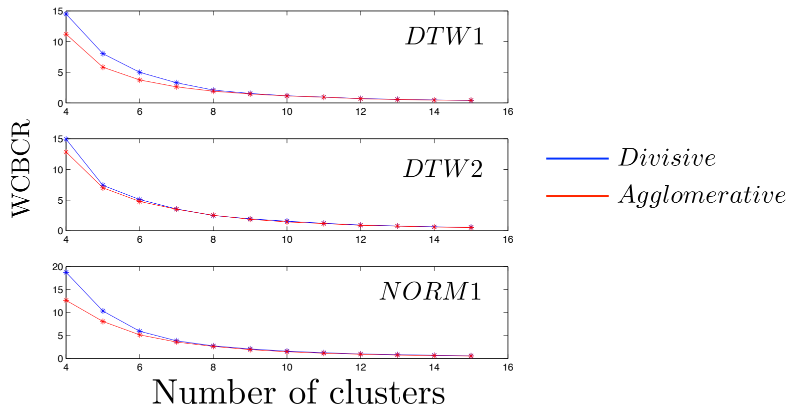


(b) Number of members within each clusters from Divisive hierarchical clustering

Figure 6: Comparison of chaining effect of clusters produced from Agglomerative hierarchical clustering and Divisive hierarchical clustering



(a) Total within-cluster distance



(b) Ratio of within cluster sum to between cluster variation

Figure 7: Comparison of Efficiency measures of clusters produced from Agglomerative hierarchical clustering and Divisive hierarchical clustering

5 Conclusion and Future Work

1. We model **Appliance Usage Matrix** A for each household, where $A_{ij} \in \{0, 1\}$ indicates whether an appliance with power level j is used during hour i.

$$\tilde{A} = \underset{Ap=\bar{x}}{\operatorname{argmin}} \sum_{i,j} |A_{ij}| \qquad (2)$$

where $\bar{x} \in \mathbf{R}^{24}$ is an hourly mean power consumption of a household, and p is a vector of power level of appliances selected by backward search with LOOCV. For each household, sparsity pattern of the output A agrees with the majority of its load curves x, *i.e.*, non-zeros elements concentrate in those rows where hourly consumption peaks. Hence, our model of A suffices to produce a reasonable pattern and time efficient since it can be solved by minimum norm least square.

2. We present a new metric for clustering **Appliance Usage Matrix** A based on Dynamic Time Warping (DTW). Our contribution is the design of DTW based metric that accounts for a small shift, contraction or expansion of each column of A, which reflects human randomness in a usage schedule of each appliance. Using hierarchical clustering based on DTW metric results in a relatively smaller number of classes and higher clustering efficiency measure, comparing to L_1 metric and K-means. DTW yields more compact clusters, with larger between cluster variation. The fact that DTW achieves lower Ratio of within cluster sum to between cluster variation (WCBCR) is not trivial, since DTW gives the optimal alignment between two patterns, which would give lower within cluster sum, but does nothing to maximize between cluster distance. The smaller and higher quality sets of clusters by DTW is preferable to the DR program.

In the future, in supervised learning part, we would like to find a formulation to fit A and p simultaneously. One possible way is to use convex optimization. In unsupervised learning part, we would visualize the cluster centers using PCA, to gain more insight into appliances usage pattern.

References

[AR13] Adrian Albert and Ram Rajagopal. Smart meter driven segmentation: What your consumption says about you. *IEEE Transactions on Power Systems*, 28(4), 2013.

[Has] *The Elements of Statistical Learning.* Springer.

[KFR14] Jungsuk Kwac, June Flora, and Ram Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1), 2014.

[KTS+13] Jungsuk Kwac, Chin-Woo Tan, Nicole Sintov, June Flora, and Ram Rajagopal. Utility customer segmentation based on smart meter data: Empirical study. In *IEEE SmartGridComm 2013 Symposium - Support for Storage, Renewable Resources and Micro-grids*, 2013.

[Mul07] M. Muller. *Information Retrieval for Music and Motion.* Springer, 2007.

[THD07] George J. Tsekouras, Nikis D. Hatziargyriou, and Evangelos N. Dialynas. Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Transaction on Power Systems*, 22(3):1120 – 1128, August 2007.

[YL11] Jaewon Yang and Jure Leskovec. Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011.

[YV93] Fung Y.H. and Rao Tummala V.M. Forecasting on electricity consumption: A comparative analysis of regression and artificial neural network models. In *IEEE 2nd International Conferennce on Advances in Power System Control, Operation and Management*, 1993.

[YV05] Fung Y.H. and Rao Tummala V.M. An electric energy consumer characterization framework based on data mining techniques. In *IEEE Transactions on Power Systems* [YV93], pages 596–602.