Clustering and Temporal Evaluation of Energy Demand

Akshay Uttama Nambi S. N., Evangelos Pournaras and Venkatesha Prasad R

I. Introduction

In this article we describe the cluster and temporal evaluation of energy demand. Cluster evaluation validates the number of clusters derived using unsupervised EM clustering for CER data set [2]. Further, the temporal analysis presented in [1] is evaluated for REFIT data set to show the applicability of the temporal analysis across different data sets.

II. Cluster evaluation

The number of clusters formed are validated using two well-known cluster evaluation metrics: DBI and Silhouette [3]. Both the metrics measure the quality of clustering. The cluster evaluation is performed on the CER data set [2].

Davies-Bouldin Index (DBI)

DBI is the ratio of within-cluster and between-cluster distances. DBI [3] measures the intra-cluster distances (intra-cluster similarity) and inter-cluster distances (inter-cluster similarity) to determine optimal cluster sizes. DBI is defined as,

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{j\neq i}\left(\frac{\delta(o_i)+\delta(o_j)}{\delta_{o_i,o_j}}\right), \quad (1)$$

where, $\delta(o_i)$, $\delta(o_j)$ are the intra-cluster distance of cluster o_i and o_j, which measures the average distance of the cluster members to its centroid. δ_{o_i,o_j} is the inter-cluster distance between clusters o_i and o_j, which measures the Euclidean distance between the centroids of the two clusters. k is the number of clusters. Thus, the optimal clustering solution has a minimum DBI value with maximum intra-cluster similarity and minimum inter-cluster similarity. Fig. 1 shows the DBI values for different number of clusters with daily and weekly granularity. The optimal clusters identified by DBI for AVG, RSD, LF features with daily demand properties are 7, 5, and 5 respectively. These are the same number of clusters identified by the clustering approach employed in the paper, thus validating the clustering results. Similarly, the number of clusters computed matches the optimal clusters identified by DBI for weekly demand properties, thus validating the clustering results.

Akshay Uttama Nambi S. N and Venkatesha Prasad are with the Embedded Software Group, EEMCS, Delft University of Technology, The Netherlands (email: Akshay.Narashiman@tudelft.nl, R.R.VenkateshaPrasad@tudelft.nl).

Evangelos Pournaras is with Professorship of Computational Social Science, ETH Zurich, Zurich, Switzerland (email: epournaras@ethz.ch).

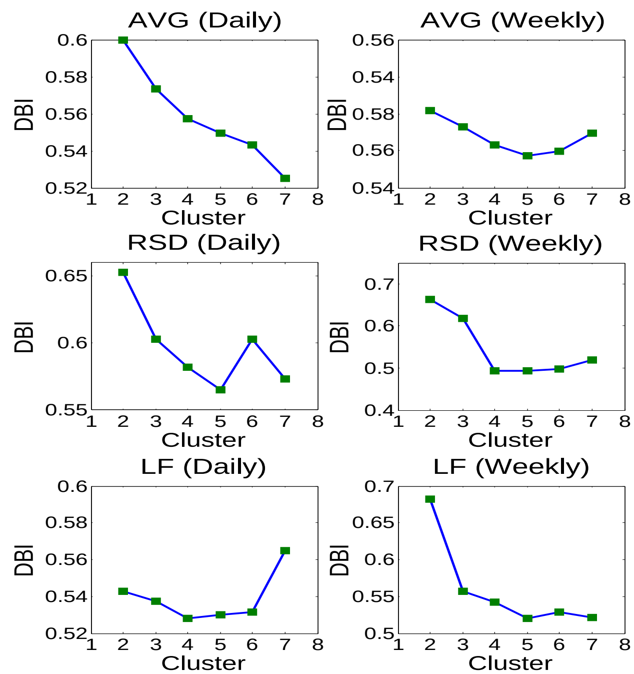

Fig. 1: DBI for each cluster.

Silhouette Value

Silhouette value is based on the average distance to members in the same cluster and the average distance to members in other clusters. Silhouette values [3] represent how well each member is associated in a cluster. The silhouette value for member x is defined as,

$$Sil_x = \frac{\delta_2(x)-\delta_1(x)}{max\{\delta_1(x),\delta_2(x)\}}, \quad (2)$$

where, $\delta_1(x)$ is the average distance from member x to all other members in the same cluster, and $\delta_2(x)$ is the average distance of member x to any other cluster centroid of which x is not a member. Euclidean distance is used to measure the average distance between member x to other members and/or other cluster centroids. The silhouette value ranges from -1 to $+1$. A silhouette value close to $+1$ means the member is appropriately clustered and a value close to -1 indicates that the member should have been clustered in its neighboring cluster. A value close to 0 indicates the member is on the border of two clusters. The average silhouette value over all members of a cluster is a measure of how tightly members are grouped in the cluster. Similarly, the mean silhouette value over all data of the entire data set is a measure of how

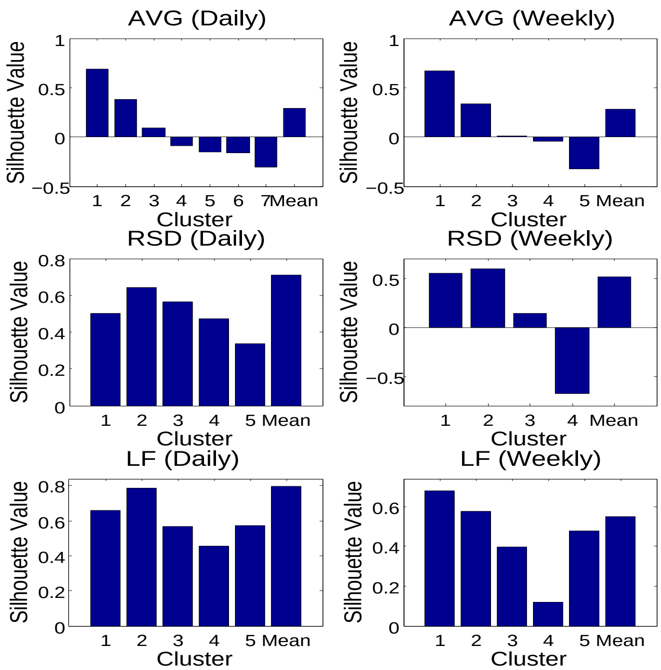

Fig. 2: Mean silhouette values for each cluster.

appropriately the data are clustered in the data set. The higher the silhouette value the better is the clustering. Fig. 2 shows the average silhouette value for each cluster and mean silhouette value for the entire data set. Most of the silhouette values are on the positive range indicating that all its members are clustered to the correct cluster. The silhouette values for the AVG quality feature have few negative values. This is due to the fact that AVG demand of the households is sparse and scattered in these clusters. This could be due to occasional high average consumption by the households. Moreover, the mean silhouette value for each quality feature is close to 1, indicating correct number of clusters is computed in the data set. Thus, the above metrics validate the clustering results.

III. Temporal evaluation for REFIT data set

To evaluate the methodology presented in [1] across other data sets, we employed a publicly available data set from UK. The data set includes 20 households energy consumption data for a year. The REFIT data set [4] was released as part of the Smart Home and Energy Demand Reduction project at the University of Strathclyde. The data set contains active power measurements from 20 homes in the Loughborough area of the UK, at a resolution of 1 sample every 8 seconds.

Key results obtained from this data set using the temporal analysis are:

(a) With the help of unsupervised clustering mechanism proposed, the number of demand states obtained for AVG daily and weekly feature across 20 households are 7 and 4 respectively.

(b) Fig. 3 shows the temporal membership of the households across various demand states for daily and weekly AVG features. It can be seen that 50% of the households belong to intermediate demand states (States 3,4,5) for daily

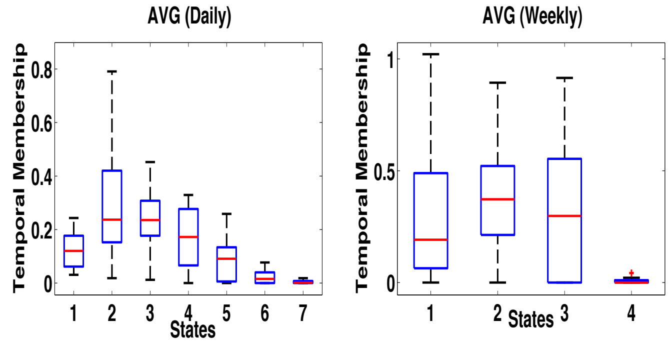

Fig. 3: Temporal membership for AVG feature daily and weekly.

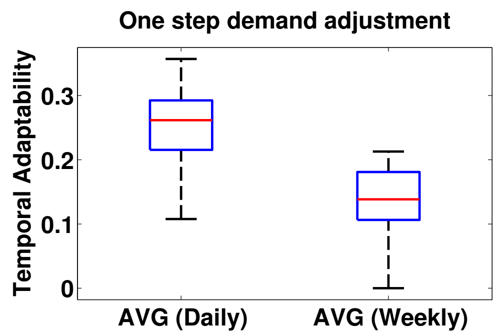

Fig. 4: Temporal adaptability with one step demand adjustment for AVG daily and weekly features.

AVG properties. Whereas, around 70% of the households has membership to low energy consumption states (States 1,2) for weekly properties indicating low energy consumers.

(c) Fig. 4 shows the one step demand adjustments for daily and weekly AVG properties. The analysis on REFIT dataset identifies 25% of the households can reduce average daily energy demands. Similarly around 14% of the households can participate in AVG weekly demand regulation.

The above evaluation allows researchers to employ the temporal analysis presented in [1] across various data sets to analyze temporal dynamics of consumers.

References

[1] S. N. A. U. Nambi, E. Pournaras, and R. V. Prasad, "Temporal Self-regulation of Energy Demand," *in IEEE Transactions on Industrial Informatics,* 2016.

[2] Electricity customer behaviour trial, The commission for Energy Regulation (CER), 2012.

[3] E. Rendon, I. Abundez, A. Arizmendi and E. M. Quiroz, "Internal versus external cluster validation indexes," *International Journal on Computer Communication*, vol. 5, no. 1, pp. 27-34, 2011.

[4] D. Murray, J. Liao, L. Stankovic, "A data management platform for personalised real-time energy feedback," *in Int Conf Energy Efficiency in Domestic Appliances and Lighting, EEDAL,* 2015.