A Clustering Framework for Residential Electric Demand Profiles

M. Jain\textsuperscript{1,2}, T. AlSkaif\textsuperscript{3}, and S. Dev\textsuperscript{1,2}

\textsuperscript{1} UCD School of Computer Science, Dublin, Ireland
\textsuperscript{2} ADAPT SFI Research Centre, Dublin, Ireland
\textsuperscript{3} Wageningen University and Research, Wageningen, The Netherlands

Send correspondence to M. Jain, e-mail: mayank.jain@adaptcentre.ie
Introduction

- Availability of huge amount of electricity consumption data made possible due to large scale adoption of smart-meter systems.
- This data is available with high temporal resolutions, often half-hourly or hourly.
- Crucial task of analyzing energy consumption patterns in the residential areas is now possible with this data.
- Clustering households based on their electricity consumption trends is an important step in this analysis.
- A 2-step clustering framework is defined and an objective validation strategy to validate and compare different frameworks is proposed.
Box-plots of daily consumption pattern for 2 different households, also depicting the median and the scaled up version of normalized median.
Pre-processing the Dataset

**Nomenclature:**

- \( n \) Number of households
- \( r \) Hourly resolution of original data
- \( d' \) Optimal number of reduced dimensions
- \( k \) Optimal number of clusters

**Algorithm 1:** Pre-Processing the Dataset

1. \( n \leftarrow \) number of households
2. \( r \leftarrow \) hourly resolution of original data
3. \( d \leftarrow 24/r \) (dimensionality)
4. \( M_{n \times d} \leftarrow \) median daily consumption of each household stored row-wise
5. \( M'_{n \times d} \leftarrow \ell_2\text{-Normalization}(M, \text{row-wise}) \)
6. return \( M' \)
In this work,

- 2 dimensionality reduction algorithms - *elbow heuristics* at intermediate stage
- 2 clustering algorithms - *gap statistics* at intermediate stage

\[ d_{FA}^' = 7; d_{PCA}^' = 7; \text{ and} \]
\[ k_{FA+SC} = 7; k_{FA+KMC} = 7; k_{PCA+SC} = 7; k_{PCA+KMC} = 7 \]
Algorithm 2: Objective Validation Strategy

1. \( p \) \( \leftarrow \) number of partitions for each household
2. \( \text{dist}(\cdot, \cdot) \) \( \leftarrow \) function to calculate Euclidean distance
3. \( \text{labels} \) \( \leftarrow \) labels assigned to each household
4. \( C_{(k \times d')} \) \( \leftarrow \) cluster centers of each cluster

\textbf{Initialize:}
\begin{align*}
\text{match}, \text{misMatch}, \text{counter} & = 0;
\end{align*}

\textbf{repeat}
\begin{align*}
\textbf{foreach household do}
\end{align*}
\begin{align*}
D & \leftarrow \text{data for each household (days} \times 24/r); \\
D' & \leftarrow \text{randomly shuffled data by rows}; \\
\text{Make } p \text{ equal partitions from rows of } D'; \\
\text{Perform Pre-Processing steps}
\end{align*}
\begin{align*}
M_{p}^{(p \times (24/r))} & \leftarrow \text{new medians of } p \text{ partitions}; \\
M_{p}' & \leftarrow \ell_2\text{-Normalization}(M_{p}, \text{row-wise}); \\
\text{Do Dimensionality Reduction}
\end{align*}
\begin{align*}
N_{p}^{(p \times d')} & \leftarrow \text{dimReduce}(M_{p}', d'); \\
\textbf{foreach partition } \in \{1 \cdots p\} \text{ as part do}
\end{align*}
\begin{align*}
\text{Find Closest Cluster}
\end{align*}
\begin{align*}
CC & \leftarrow \text{argmin}_i(\text{dist}(N_{p}[\text{part}], C[i, \cdot])); \\
\text{if } CC = = \text{labels [household] then}
\end{align*}
\begin{align*}
\text{match} & ++; \\
\text{else}
\end{align*}
\begin{align*}
\text{misMatch} & ++;
\end{align*}
\begin{align*}
\text{counter} & ++;
\end{align*}
\begin{align*}
\textbf{until} \text{ counter} < 100;
\end{align*}
\begin{align*}
\text{avgMatches} & = \text{match}/100; \\
\text{avgMisMatches} & = \text{misMatch}/100;
\end{align*}
\begin{align*}
\textbf{Result:} \text{avgMatches} \& \text{avgMisMatches}
\end{align*}

Results obtained by performing objective validation of the 4 clustering frameworks.

| Clustering Framework | %Matches | %Mis-Matches |
|----------------------|----------|--------------|
| FA + SC \( p = 2 \)  | 22.67    | 77.33        |
| FA + KMC             | 29.07    | 70.93        |
| PCA + SC             | 18.78    | 81.22        |
| PCA + KMC \( p = 2 \) | 76.28    | 23.72        |
| FA + SC \( p = 3 \)  | 21.34    | 78.66        |
| FA + KMC             | 24.98    | 75.02        |
| PCA + SC             | 17.60    | 82.40        |
| PCA + KMC \( p = 3 \) | 67.15    | 32.85        |
Subjective Validation

Two sample clusters as identified by the recommended framework

Sample ill-defined clusters from other frameworks (left: PCA+SC; right: FA+KMC)
Conclusion

Contributions:

- Defined a 2-step generalized clustering framework
- Proposed a novel objective validation strategy to compare results of different frameworks
- Cross-verified the recommendations by subjective validation

Future work:

- Gather data for longer duration to incorporate seasonal variations in consumption behaviour
- Consider more algorithms used in more recent studies
- Compare results of the proposed objective validation strategy with more standard clustering validation indices