Online Adaptive Clustering Algorithm for Load Profiling

Guillaume Le Ray, Pierre Pinson

Technical University of Denmark

Introduction

The increasing share of renewable energy sources (RES) engenders an increase intermittency of the production. Consumers' behaviors are also becoming more complex and dynamic. In the same time, the production should meet the demand at all time.

However typical load profiles, which are the main tool to evaluate the demand side, were historically estimated from yearly consumption due to a lack of data.

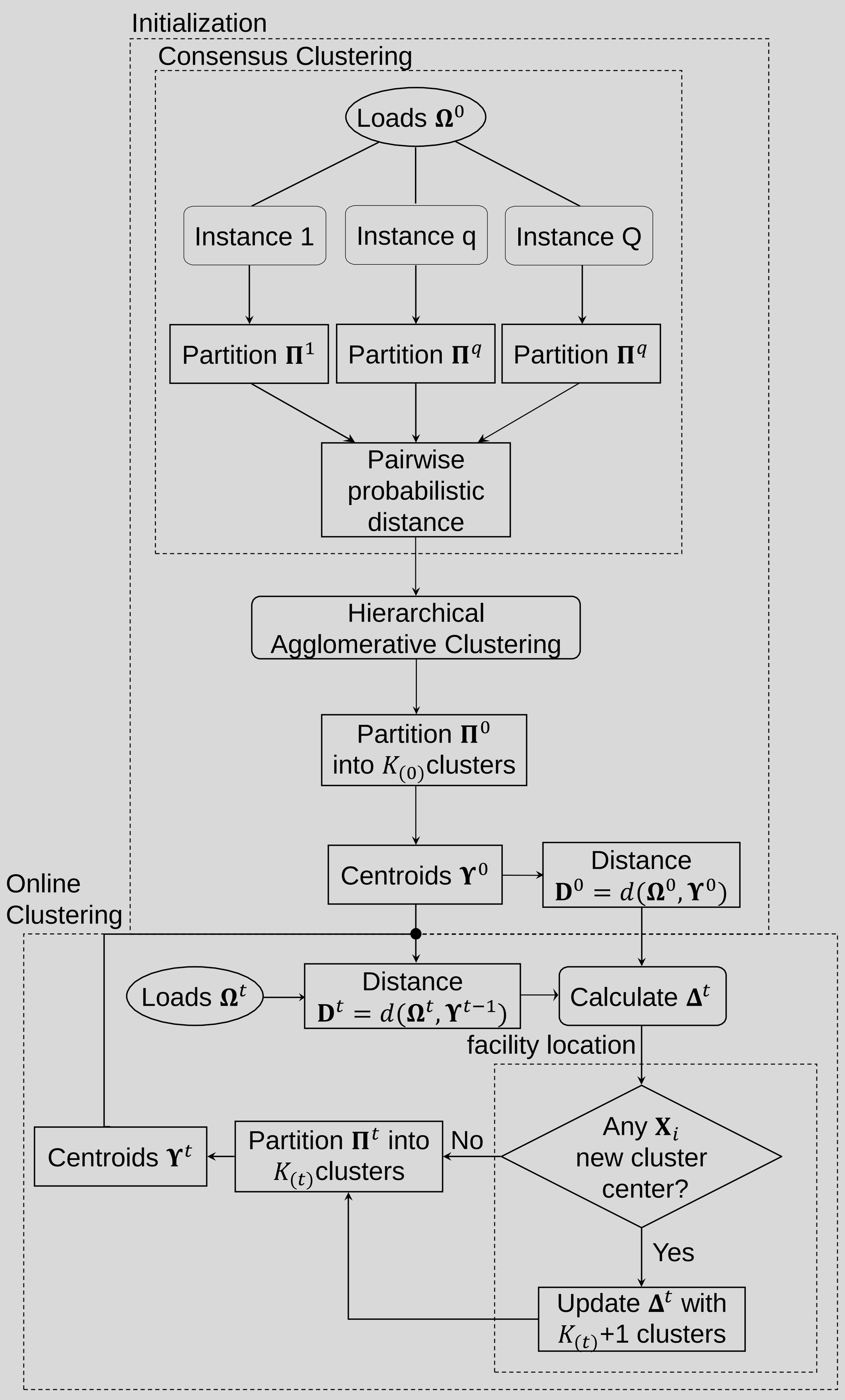
In this context the development of Advanced Metering Infrastructures (AMI) is a game changer. Indeed, we change from scarcity (1point/year) to profusion of data (1 point/15 min.) which constitutes streams of data.

Research on energy consumption data has focused on developing cluster-based load profiling using well-known clustering algorithm. Despite being data-driven they are not adapted to dynamic data as they use batch of historical data.

We propose a new methodology to cluster online loads and generate dynamic load profiles.

In addition to existing applications (e.g. long and short time planning), new applications can emerge to dynamically manage the demand flexibility to balance the system.

Methodology



Initialization

Consensus Clustering:

- run several instances with different number of clusters $K_{(q)}$ on a initial block of loads $\mathbf{\Omega}^0$.
- Generate a probability distance matrix of 2 points ending in the same cluster.
- Use the HAC to choose $K_{(0)}$.
- Set the initial online clustering parameters: partition $\mathbf{\Pi}^0$, centroids $\mathbf{\Upsilon}^0$, distance between loads and centroids $\mathbf{D}^0 = d(\mathbf{\Omega}^0, \mathbf{\Upsilon}^0)$.

Online clustering

- Based on K-means algorithm (assign step, update centroids step).
- The distance used is an adaptive dissimilarity index combining a Dynamic Time Warping (DTW) and a first order temporal correlation using adaptive tuning function.

In each time step:

- calculate the distance $\mathbf{D}^t$ between centroids $\mathbf{\Upsilon}^{t-1}$ at the previous step and the newly generated loads $\mathbf{\Omega}^t$.
- The assumption is made that loads are relatively stable and it uses exponential smoothing on the distance calculation with, $\mathbf{\Delta}^t = \sum_{\tau=0}^{t} \lambda^{t-\tau}\mathbf{D}^\tau$ to propagate in time structural information from previous steps.
- Evaluate if $K_{(t)}$ is accurate with facility location (see below).
- Generate partition $\mathbf{\Pi}^t$ into $K_{(t)}$ clusters.
- Calculate centroids $\mathbf{\Upsilon}^t$.

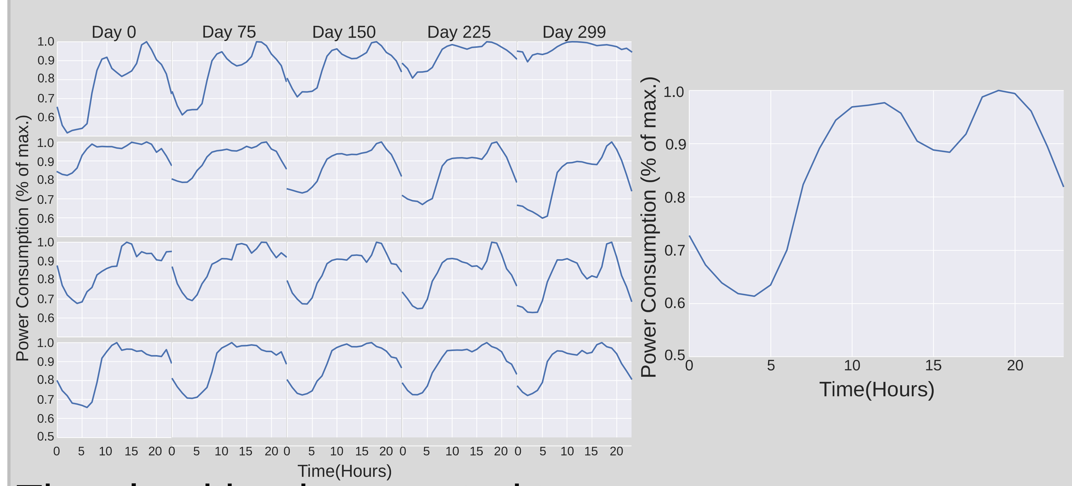
Facility location:

- Calculate an assignment cost C_a.
- Set cost C_f, C_f>C_a of an extra cluster center.
- Calculate the probability that a load become a new cluster center, $p(\mathbf{Y}_{K+1} = \mathbf{X}_i) = \min(\frac{d_i}{C_f}, 1)$ where d_i is the smallest distance of load $\mathbf{X}_i$ to an existing cluster center.
- A threshold of how many loads should be picked is used to define a new load behavior.
- Update $\mathbf{\Delta}^t$ with $K_{(t)} = K+1$ clusters.

Online Clustering Evaluation

Synthetic data presentation

1000 customers sampled from four typical load profiles (left figure) at t=0 with additional white noise. At t=100, 250 customers exhibit an extra typical load profile (right figure) and at t=200 they come back to one of the four typical load profiles assigned at t=0.

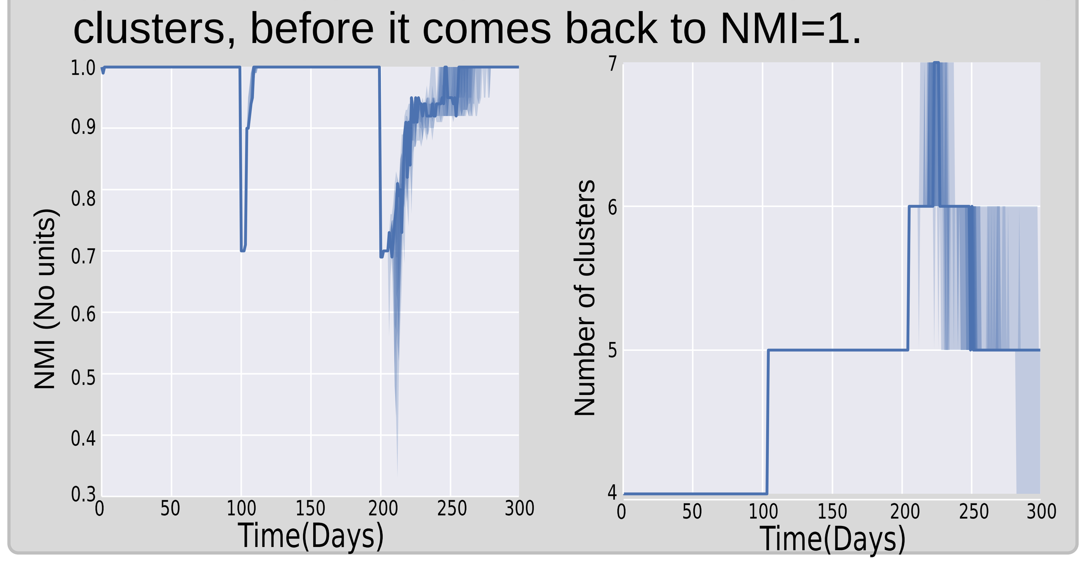


The algorithm is expected to:

- Change from 4 to 5 at t=100.
- Classify correctly the customers according to the typical load profiles they are sampled from.

Results

- classification accuracy → NMI (Normalized Mutual Information) between real and algorithm assignments. It takes values in [0,1] where 1 is a perfect match.
- Algorithm catches the change from 4 to 5 clusters accurately, drop to NMI=0.7 and back to 1.
- Transition from 5 to 4 (day 200) generates more error (NMI drops for 50 days) as the algorithm generates (up to 7) and deletes (down to 5-4) clusters, before it comes back to NMI=1.



Conclusion and Future Work

The results on the synthetic data have proven that the online adaptive clustering algorithm is well suited for profiling streams of load data. Moreover, the synthetic data is simulating an extreme case as loads usually display seasonality (slow changing shape) which is handled by the exponential smoothing and not often brutal changes. Tests are being run on real data and an extension to multi-energy profiling is plan as future work.

Acknowledgements

This research was funded by the EnergyLab Nordhavn project, Grant No. EUDP 64015-0055.

Contact: **Guillaume Le Ray, PhD student, Email: gleray@elektro.dtu.dk**