Appliance-level Residential Consumer Segmentation from Smart Meter Data

Lily Buechler and Zonghe Chua

{ebuech, chuazh}@Stanford.edu

Department of Mechanical Engineering, Stanford University

Introduction

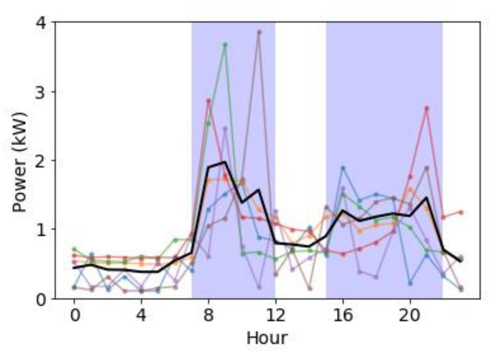
Background:

- Deployment of distributed energy resources (e.g. solar, electric vehicles) is driving greater uncertainty in power demand, requiring expansion of residential demand response programs (curtailment/shift in power consumption in response to prices or incentives)

Objective:

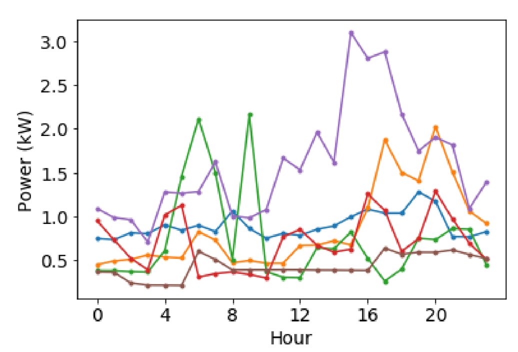
- Perform consumer segmentation for DR program targeting based on appliance-level power consumption
- Segment consumers based on three characteristics:

Availability



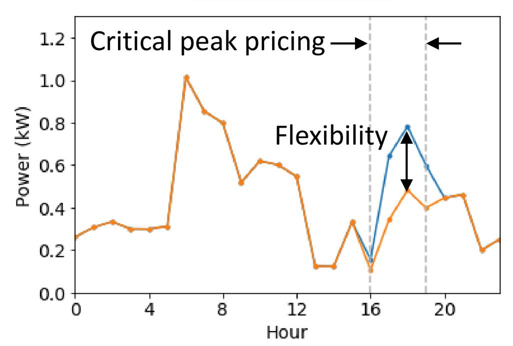
Identify consumer clusters with similar temporal use patterns

Variability



Identify consumer clusters with consistent use patterns

Flexibility



Predict % reduction in consumption during critical peak pricing periods

Dataset and Features

Dataset:

- Pecan Street database [1]
- 3 years with 1 minute resolution: training (2014), dev (2015), test (2016)
- Critical peak pricing trial [2] during 12 days in summer 2013 (32 homes)

Features:

- Availability: Empirical hourly start-time distribution ($\mathbb{R}^{24}$)
- Variability: Hourly load profile normalized by daily energy consumption ($\mathbb{R}^{24}$)

Appliance	# of homes	Analysis
Air conditioner	129	V
Refrigerator	103	V
Dishwasher	82	V, A
Clothes washer	80	V, A
Dryer	62	V, A
Electric vehicle	29	V, A
Water heater	19	V, A
Clothes washer + dryer	13	V, A
Pool pump	9	V
Total home demand	132	V, F

V = variability, A = availability, F = flexibility

Flexibility Features (per home)	Units
Home size	ft^2
Year built	year
# of stories	
Mean 6-hr energy consumption (24:00-6:00, 6:00-12:00, 12:00-18:00, 18:00-24:00)	kWh
Mean, 10%ile, 90%ile of daily energy consumption above baseload	kWh
Mean and variance of hourly power	kW
Mean, 10%ile, 90%ile maximum and minimum daily energy consumption	kWh
Entropy of load profile (from variability analysis)	

Methods

Availability:

Methods:

- K-Means clustering - Euclidean distance measure
- Hierarchical clustering – Ward's linkage with symmetrized KL divergence distance measure:

$$D_{KL}(P(h)|Q(h)) = \sum_i P(i)\log\left(\frac{P(i)}{Q(i)}\right) + \sum_i Q(i)\log\left(\frac{Q(i)}{P(i)}\right)$$

- Gaussian Mixture Model (GMM)
- Latent Dirichlet Allocation (LDA)

Evaluation metrics:

- Increase in availability: $A_\tau^{(c,l)} = \frac{\max_c\left[\frac{1}{|c||\tau|}\sum_{t\in\tau}\sum_{j\in c} p_t^{(j,l)}\right]}{\frac{1}{|H||\tau|}\sum_{t\in\tau}\sum_{j\in H} p_t^{(j,l)}}$
- Completeness score:

$$c = 1 - \frac{H(K|C)}{H(K)} \qquad H(K|C) = -\sum_{c=1}^{|C|}\sum_{k=1}^{|K|}\frac{n_{c,k}}{n}\log\left(\frac{n_{c,k}}{n_c}\right) \qquad H(K) = -\sum_{c=1}^{|C|}\frac{n_k}{n}\log\left(\frac{n_k}{n}\right)$$

Variability:

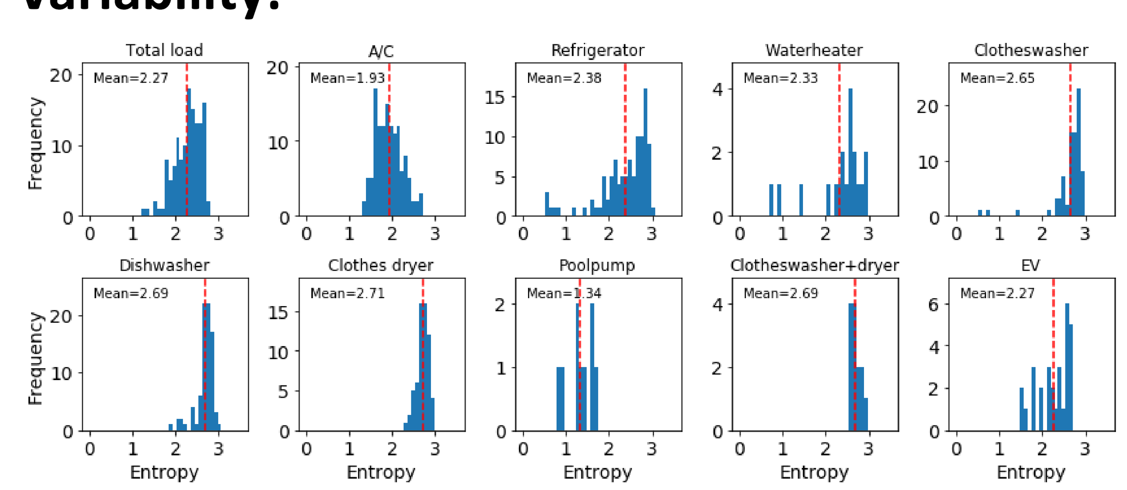
- K-Means clustering – assign load profiles to clusters
- Entropy of load profile assignment distribution: $S = -\sum_i P(i)\log(P(i))$
- Hierarchical clustering – segment homes by load profile distribution

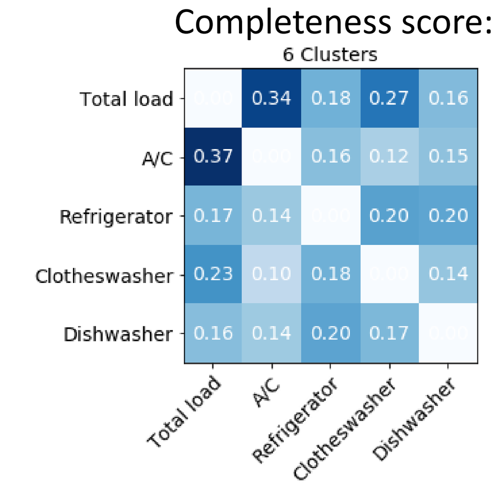
Flexibility:

- Linear regression (LR) with recursive feature selection
- KNN regression
- Random forest (RF) with recursive feature selection

Results

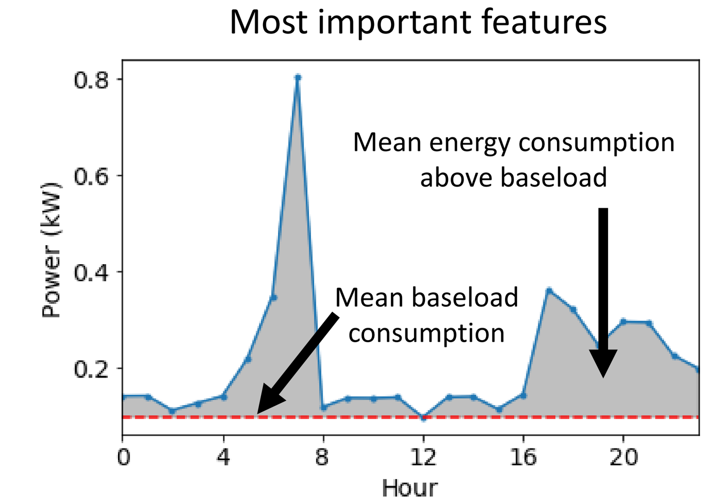
Variability:





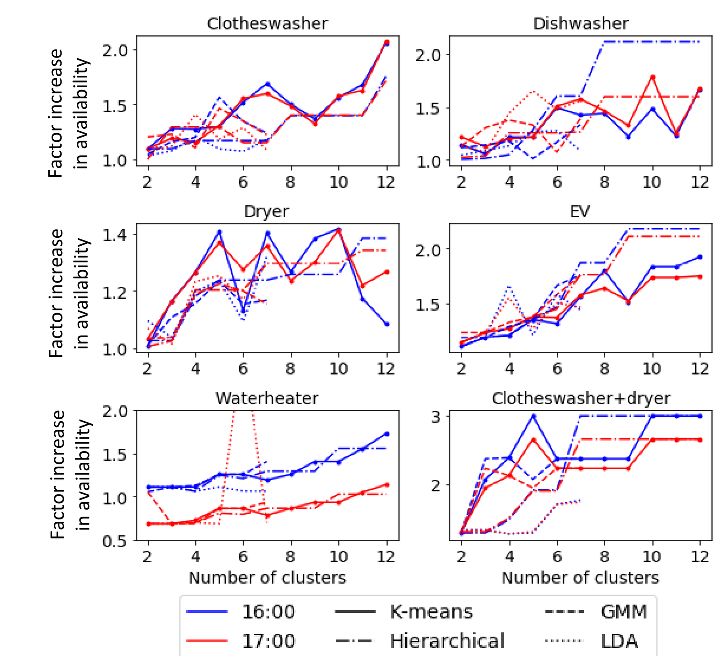
Flexibility:

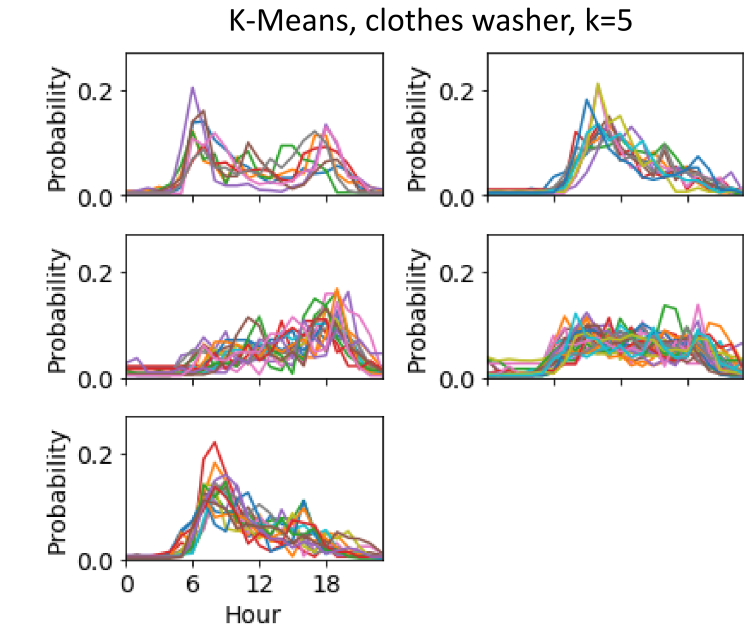
Method	Train MSE (n=20)	Dev MSE (n=6)	Test MSE (n=6)
LR	0.0346	0.0709	0.0307
KNN	0.0148	0.0589	0.0608
RF	1.680	0.7855	0.4466

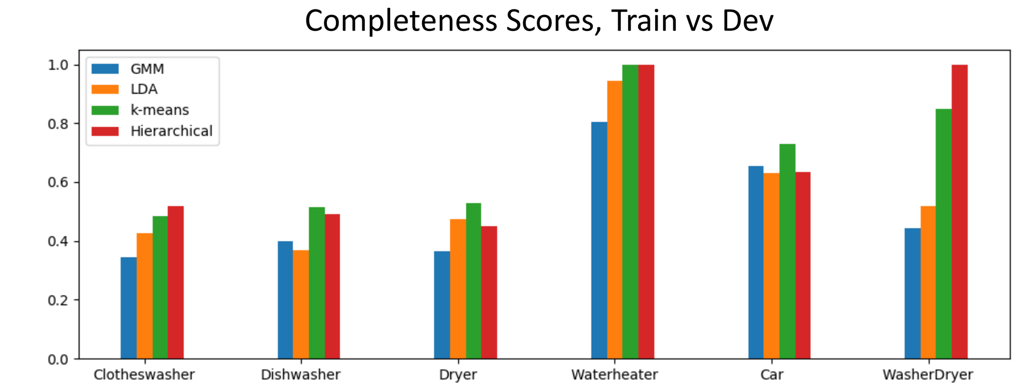


Results

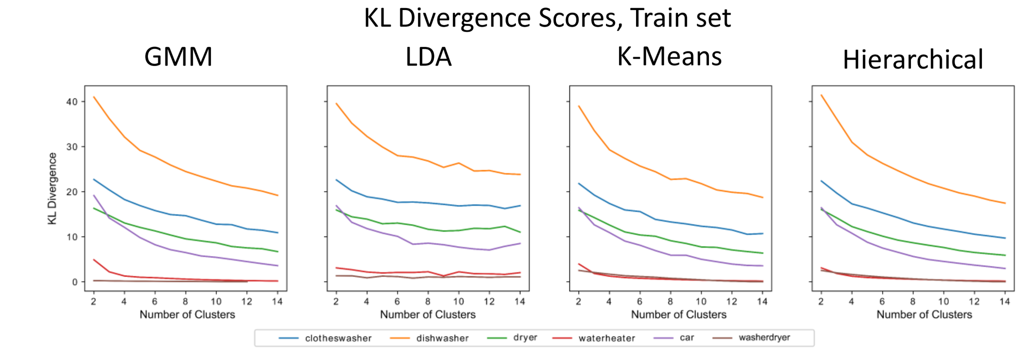
Availability:







Appliance	Optimal # of clusters
Dishwasher, clothes washer, dryer, EV	5
Clothes washer + dryer, water heater	3



Conclusions and Future Work

Conclusions:

- Consumer segmentation can yield up to a x2 increase in load availability during specific hours
- Appliance-level cluster assignments can differ significantly from aggregate load cluster assignments, requiring separate segmentation analysis
- Linear regression with feature selection yields highest accuracy for predicting consumer responsiveness to price

Future work:

- Incorporate additional factors (day of week, season) into availability analysis
- Perform analysis on larger dataset

References:

[1] Pecan Street Inc. (2017) Dataport from pecan street. [Online]. Available: https://dataport.cloud/

[2] *Technology solutions for wind integration in ERCOT*. Center For The Commercialization Of Electric Technology, Austin, TX (United States), 2015.