Shedding Light on Residential Energy Consumer Behaviour in South Africa

Unsupervised Machine Learning for Automated Segmentation and Evaluation of Explainable Household Energy Consumption Patterns

Wiebke Toussaint

Department of Computer Science, University of Cape Town

Research Contribution

1. only data-driven customer archetypes for African residential energy consumers developed in the past 5 years
2. novel framework for evaluating daily load profile clusters in highly dynamic and diverse societies
3. development of pipeline to automate the generation of explainable customer archetypes

Problem Statement

Customer archetypes are frequently used to associate representative daily load profiles (RDLPs) with socio-demographic, geographic, dwelling and economic characteristics of energy users to better understand domestic electricity use. No recent customer archetypes exist for African consumers, and the last customer archetypes developed by domain experts in South Africa in 2013 present several challenges:

- explanations of how the archetypes were developed are not readily available
- only domain experts can generate new customer archetypes
- regenerating the archetypes is time-consuming and costly

Existing approaches for generating customer archetypes in the Global North do not account for the context of highly diverse and highly volatile emerging economies. This motivates the need for automatically generating explainable customer archetypes for highly diverse, dynamic and volatile populations, such as South Africa.

Research Aim and Objectives

This research aims to estimate the daily load profile of a household, given some of its attributes, by establishing the best performing clustering algorithms, normalisation and pre-binning techniques for residential energy consumers in heterogeneous, dynamic populations. Specifically, we aim to:

1. develop a pipeline for automated generation of energy demand clusters and their associated representative daily load profiles
2. develop a framework for evaluating the produced clusters
3. demonstrate how the clusters can be used to establish customer archetypes

Acknowledgements

- Scholarship funding from the Centre of Artificial Intelligence Research and the Department of Science and Technology
- my supervisors Associate Prof Deshen Moodley and Prof Tommie Meyer
- domain experts Prof Trevor Gaunt, Marcus Dekenah and Schalk Heunis

Experimental Design

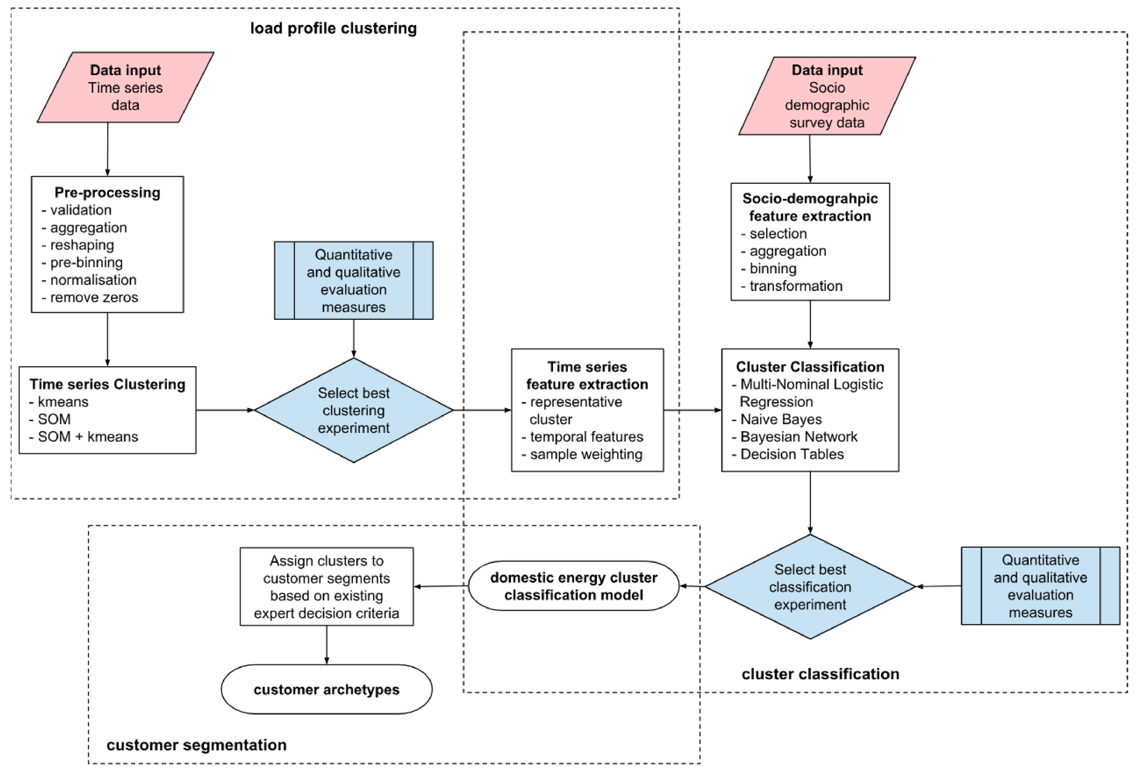


Figure 1: Experimental processing pipeline

Dataset Description

Our dataset contains daily energy demand values of 14 945 households recorded at 5 minute intervals over a one year period between 1994 and 2014. The 5 min observational consumption data of each household was aggregated to hourly intervals and segmented into daily load profiles for the number of days observed.

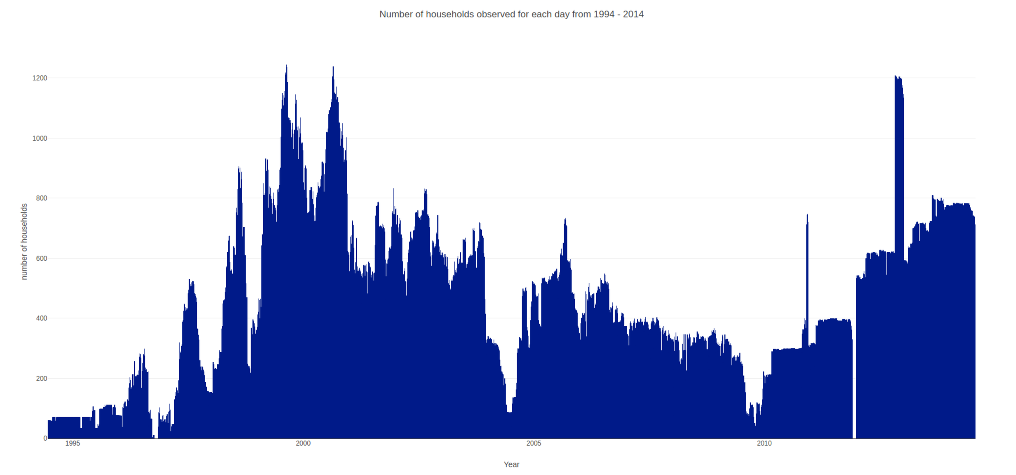


Figure 2: Distribution of households observed per day from 1994 - 2014

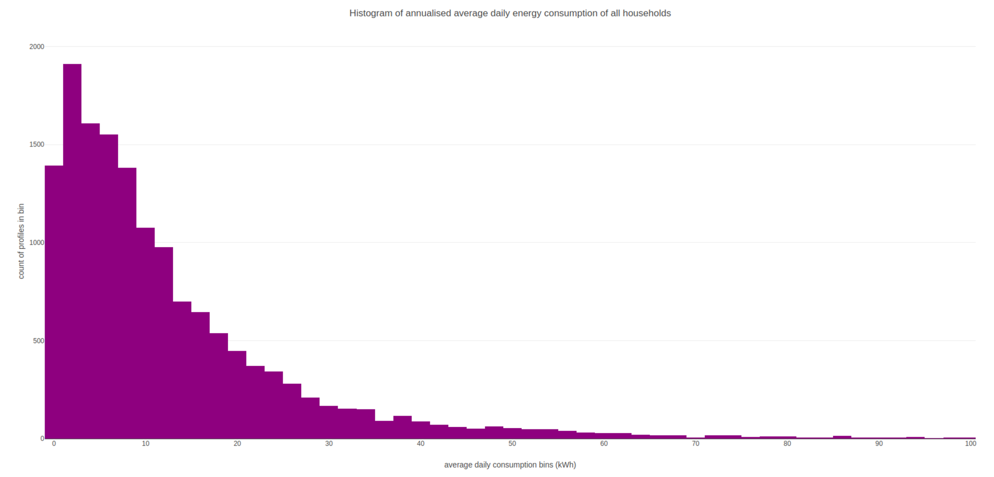


Figure 3: Distribution of mean daily demand of all households

$H(j)$ is the array of all daily 24-hour load profiles of a single household. The daily load profile arrays of all households were joined into an input array X.

$X = H(j) = [24][3295848], where j = 1, 2...14945.$

Clustering

Eight experiments were set up using various combinations of algorithms, normalisation and pre-binning techniques.

algorithms:
kmeans, self-organising maps (SOM) and kmeans+SOM
normalisation:
unit norm, zero-one, deminning, South African norm (used in [1, 2, 3, 4])
pre-binning techniques:
no pre-binning, manually configured annualised average monthly consumption bins, learned bins based on integral of daily load profiles (based on [5])

Experiment	Change from previous experiment
exp 1	test
exp 2	Baseline experiment
exp 3	Removed zero-valued load profiles from exp 2
exp 4	Pre-binned profiles by the household's mean annual monthly consumption with bin sizes manually selected to approximate utility tariffs
exp 5	Increased maximum number of clusters per bin in exp 4 from 9 to 19 clusters
exp 6	Removed zero-valued load profiles from exp 5
exp 7	Pre-binned profiles by daily demand using integral k-means
exp 8	Removed zero-valued load profiles from exp 7

Table 1: Table of clustering experiments

Cluster Evaluation

Quantitative Evaluation

A score was calculated from the products of the Davies-Bouldin Index, the Mean Index Adequacy and the inverse of the silhouette score to measure the **compactness** and **distinctness** of each cluster set.

Qualitative Evaluation

Six competency questions were formulated to frame qualitative evaluation measures that further assess if a cluster set is:

explainable: RDLP makes sense in relation to what we know about local domestic energy use
representative: low error between the total daily consumption and peak demand of a cluster and its member profiles
specific: low cluster entropy implies high information content, which is a prerequisite for useful customer archetypes

These qualitative measures were translated into a weighted scoring matrix, that accounts for the importance of each competency question. The kmeans algorithm with unit norm normalisation and daily demand pre-binning produced the best clusters.

measure	metric	weights	exp2 kmeans unit_norm	exp4 kmeans zero-one	exp5 kmeans unit_norm	exp5 kmeans zero-one	exp6 kmeans unit_norm	exp7 kmeans unit_norm	exp8 kmeans unit_norm
peak_coincR	coincidence_ratio	4	1.00	6.5	5.00	6.50	2.00	3.00	4.0
peak_consE	mean_error	7	6.25	6.0	2.75	5.00	4.75	2.25	1.0
total_consE	mean_error	7	6.75	5.5	3.25	4.25	5.25	2.00	1.0
demand_entropy	peak_entropy	6	7.00	5.5	2.00	5.50	3.00	4.00	1.0
	total_entropy	6	7.00	5.5	2.00	5.50	3.00	1.00	4.0
temporal_entropy	monthly_entropy	5	5.00	6.5	3.00	6.50	4.00	1.00	2.0
	weekday_entropy	5	1.00	6.5	4.00	6.50	3.00	5.00	2.0
		SCORE	209.00	237.5	121.00	221.75	149.00	101.75	80.0

Figure 4: Scored results for best experiments (lowest score = best)

Customer Archetypes

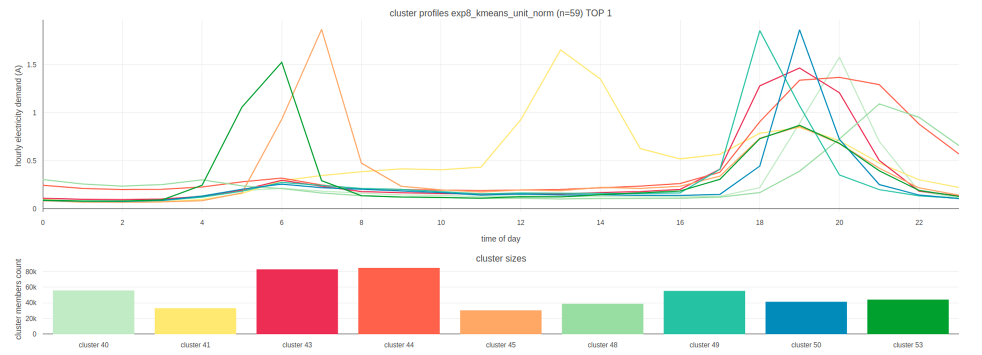
A classification model was built using multinomial logistic regression to quantify the likelihood that a particular feature value characterises a cluster. Customer attributes defined in historic expert archetypes were then used to guide the development of representative daily load profiles for customer archetypes.

archetype	water	roof material	floor area	mean income
rural	river/borehole	thatch	0-50	R0-R3199
informal	tap in yard	corr.iron/zinc	0-50	R0-R3199
township	tap in house	bricks,blocks	50-80	R3.2k-R19.1k
upper middle	tap in house	tiles,blocks	80-150	R19k-R24.5k
urban rich	tap in house	plaster,tiles	250-800	R65.5k+

Table 2: Selection of customer archetypes and some of their characteristics

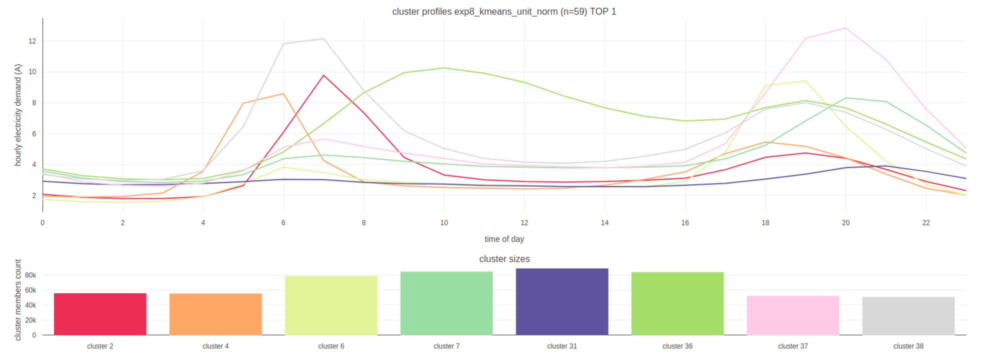
RDLPs for the rural and upper middle class customer archetypes are shown below. The tables describe the season and day type when RDLPs are typically used.

RDLPs for rural consumers in the Free State



cluster	weekday	Friday	Saturday	Sunday
winter	43, 45	43, 45	43, 49	43, 49
summer	40, 45, 50, 53	40, 44, 48, 53	41, 44, 48	40, 41

Upper middle class customers electrified 15+ years



cluster	weekday	Friday	Saturday	Sunday
winter	6, 37, 38	38	36, 37	36, 37
summer	4	2, 4, 31, 34	7, 31, 34	2

References

1. Kwac, Jungsuk, et al. "Utility customer segmentation based on smart meter data: Empirical study." Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on. IEEE, 2013.
2. Dent, Ian, et al. "An approach for assessing clustering of households by electricity usage." (2012).
3. Jin, Ling, et al. "Comparison of clustering techniques for residential energy behavior using smart meter data." AAAI workshopsâĂŤartificial intelligence for smart grids and buildings. 2017.
4. Dekenah, Marcus. "Guide to Regionalised Domestic Subclasses". 2014.
5. Xu, Sharon, Edward Barbour, and Marta C. GonzÃąlez. "Household segmentation by load shape and daily consumption." Proc. 6th ACM SIGKDD Int. Worksh. Urban Computing 2. 2017.

About the Author

I work as a data scientist at UCT's Energy Research Centre and am completing my masters in Computer Science. When I'm not mobilising STEM students and professionals to use their skills for social good, you can find me hiking or rock climbing in the Cape Peninsula.