DTU Management Engineering
Department of Management Engineering

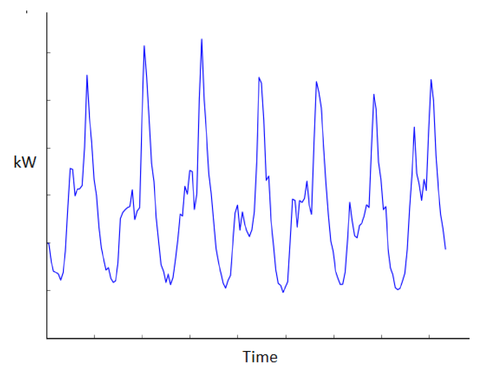
Structured Literature Review of Electricity Consumption Classification Using Smart Meter Data

Alexander Tureczek: PhD Student CITIES WP 1, Smart Meter Analytics

Supervisors: Per Sieverts Nielsen, DTU Management Engineering, Henrik Madsen, DTU Compute

Smart Meters

Over the past decade smart meters for electricity consumption measurement has received large attention. The European Commission has set up goals for roll out of smart meters [1], which are expected to play a significant part in making in reducing the energy consumption and reduce energy waste. To quantify the current state of smart meter analytics this poster describes a structured review [2] of current research methods and learnings using electricity smart meter data for consumption classification.



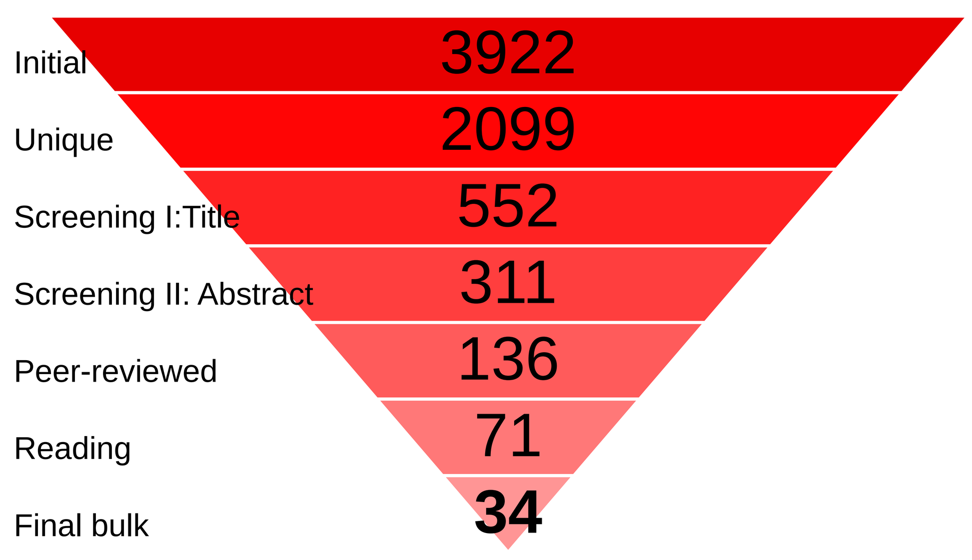
Method

This structured review applies Okolis 8 step method for structured reviews [3] modified to 7 steps here. The method is described for information sciences but is also applicable in an analytical setting. The structure followed is:

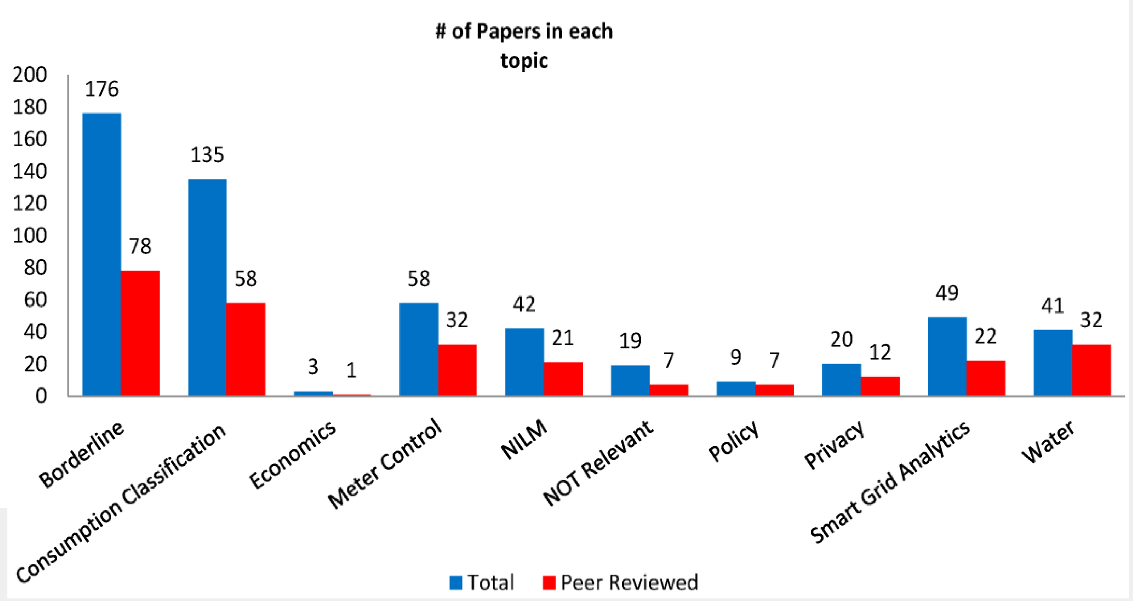
1: **Purpose of the literature review.**
2: **Protocol and training.** Ensure consistency, alignment, and reproducibility.
3: **Searching for literature.** Explicitly describe the search for literature.
4: **Practical screen.** Crude inclusion and exclusion of articles; if in doubt the article should be included.
5: **Quality appraisal.** Screen for exclusion, and explicitly define the criteria for judging articles.
6: **Data extraction and synthesis of studies.** Systematically extract the applicable information of the identified articles and combine the facts.
7: **Writing the review.**

Waterfall stastistic

30 search phrases was applied using Thomson-Reuters Web-of-Science. Resulting in the following paper reduction waterfall.

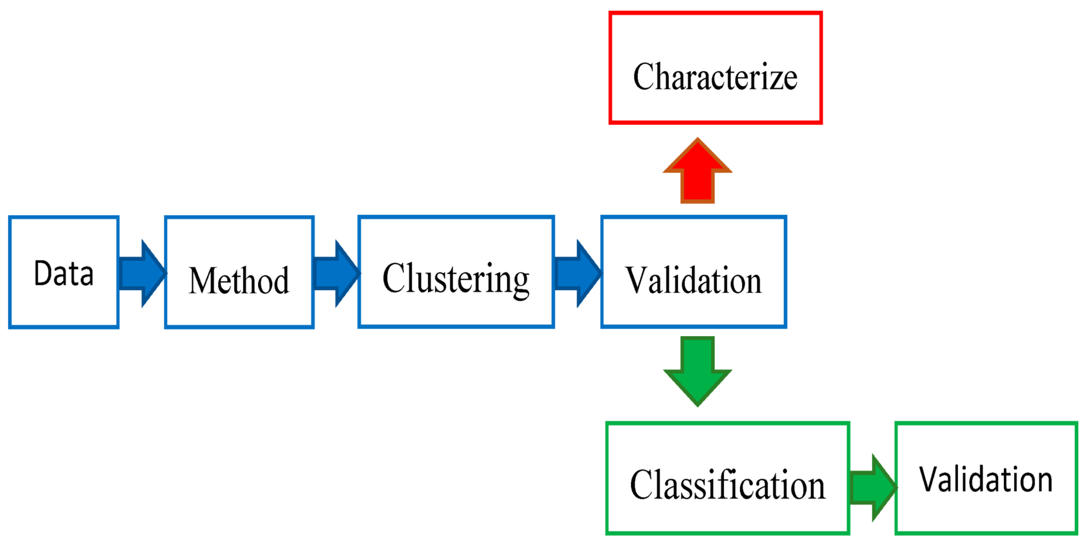


The final number of peer-reviewed papers included in the review is 34. These papers all complied to the quality appraisal criterion: **smart meter data should be applied for end-user consumption classification.** Smart meter consumption data can be binned into 10 distinct categories based on title.



Analysis Workflow

The papers selected for this review follow roughly the same procedure:for analysing smart meter data as outlined below :



BLUE: Outlines the classification procedure followed. By all papers analysed. Smart meter **data** selcting a **method** for **clustering** and finally **validate** the number of clusters

RED: some papers are also investigating the characteristics of the identified clusters.

GREEN: Few papers apply the identified cluster information to classify test set after the training of the clusters algorithms.

Clustering algorithms

Most prevalent clustering techniques

Metode	Total	% of total
K-Means	22	65%
hierarchical	10	45%
fuzzy k-means	4	12%
Follow-The_leader	3	9%
Mixture Model	3	9%
Neural Network	2	6%
K-medoid	2	6%

Cluster Validation

Validation methods applied to help select number of clusters.

Index	Interpretation
DBI (Davies-Bouldin Indictor)	DBI relates the mean diameter of each class with the distance to the closest class. maller values of DBI implies that K-means clustering algorithm separates the data set properly
CDI (Cluster Dispersion Indicator)	CDI prefers Long inter-cluster distance and short intra-cluster distance. Small values indicate good clustering.
Dunn	The ratio between "minimum distance between clusters" and "maximum distance within clusters". When minimum dissimilarity between clusters get large and max cluster diameter gets small the Dunn value gets large and indicates good separation.
Silhouette	Is the average distance between vector x and all other vectors of the cluster c to which x belongs. This is related to the minimum distance between vector x and all other vectors in cluster $\forall\, C' \neq C$. SI is between [–1, 1] higher is better. Negative is miss-clustering.
Entropy	Proportion of correct classified vector i in cluster t. Entropy is a supervised index as the true classes needs to be known.
MIA	Average distance within class to class centroid, summarized across all classes. High MIA indicates large distances within the classes. E.g., large dispersion.

Data Description Score

The papers analysed in this review focus strongly on the description of the methods selected. To such an extend that the description of the data utilized and the data wrangling can be poorly described. We have developed a data description score to help evaluate the data description, and remedy if there are some blank spots.

Category	Attribute	Prevalence	Prevalence %
Geographical information	Country	33	97%
	Region	27	79%
	Origin	27	79%
Data Information	Initial Size	33	97%
	Clear Reduction	32	94%
	Missing Values	17	50%
	Final Size	31	91%
Time Information	Recording Frequency	33	97%
	Start	23	68%
	End	22	65%
	Period	30	88%
Type Information	Industry / Residential	32	94%
Referencing Data	Referral	13	38%

Conclusions:

- **Simple algorithms prevail and perform relatively good**
- **Roughly the same procedure is applied by all papers.**
- **Validation of the clustering is needed in order to select bet clustering**
- **The description of data and the data wrangling is of varying quality. Therefore we have developed a data description score applicable for evaluation and as checklist.**

Acknowledgements

DSF 1305-00027B

Special thanks to

Susanne Balslev for providing the idea and initial literature.

References

1 EU Commission. Smart Grids and Meters. Available online: https://ec.europa.eu/energy/en/topics/markets-and-consumers/smart-grids-and-meters

2 Tureczek. Alexander M, Nielsen. Per S. Structured Literature Review of Electricity Consumption Classification Using Smart Meter Data Alexander. Energies 2017:1–19. doi:10.3390/en10050584.

3 Okoli, C. A guide to conducting a standalone systematic literature review. *Commun. Assoc. Inf. Syst*. **2015**, *37*, 879–910.