

CLUE — Clustering-Based Load Understanding and Exploration: Summarizing High-Dimensional Electricity Grid Data for Scenario Analysis

Context

Energy systems are undergoing a major transformation:

- Digitization: smart meters, sensors, and IoT devices
- Integration of Renewables
- New Demands: electrification (EVs, heating), efficiency

The need: analyzing large volumes of time-series data from smart electricity meters.

Application: support grid management, forecasting & customer segmentation.

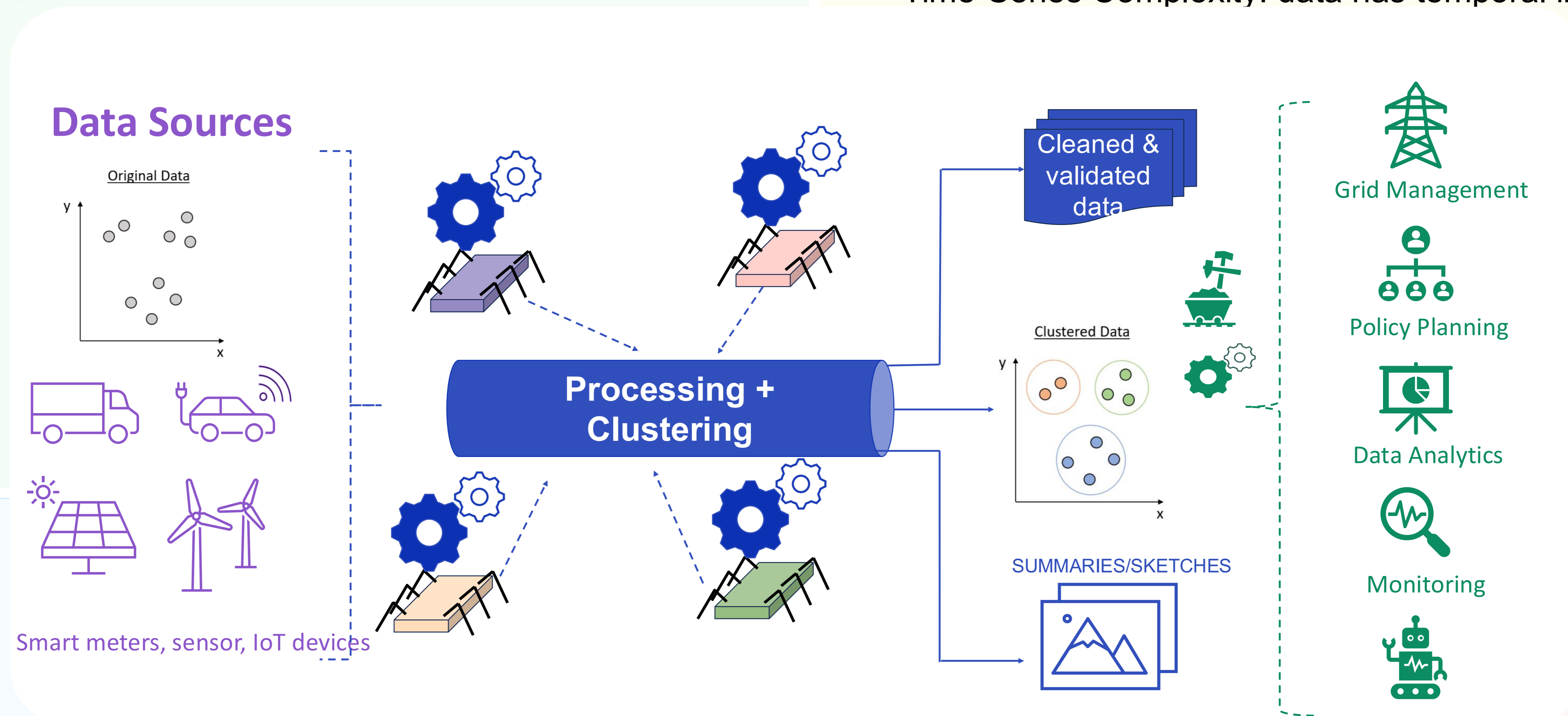
Clustering as a backbone

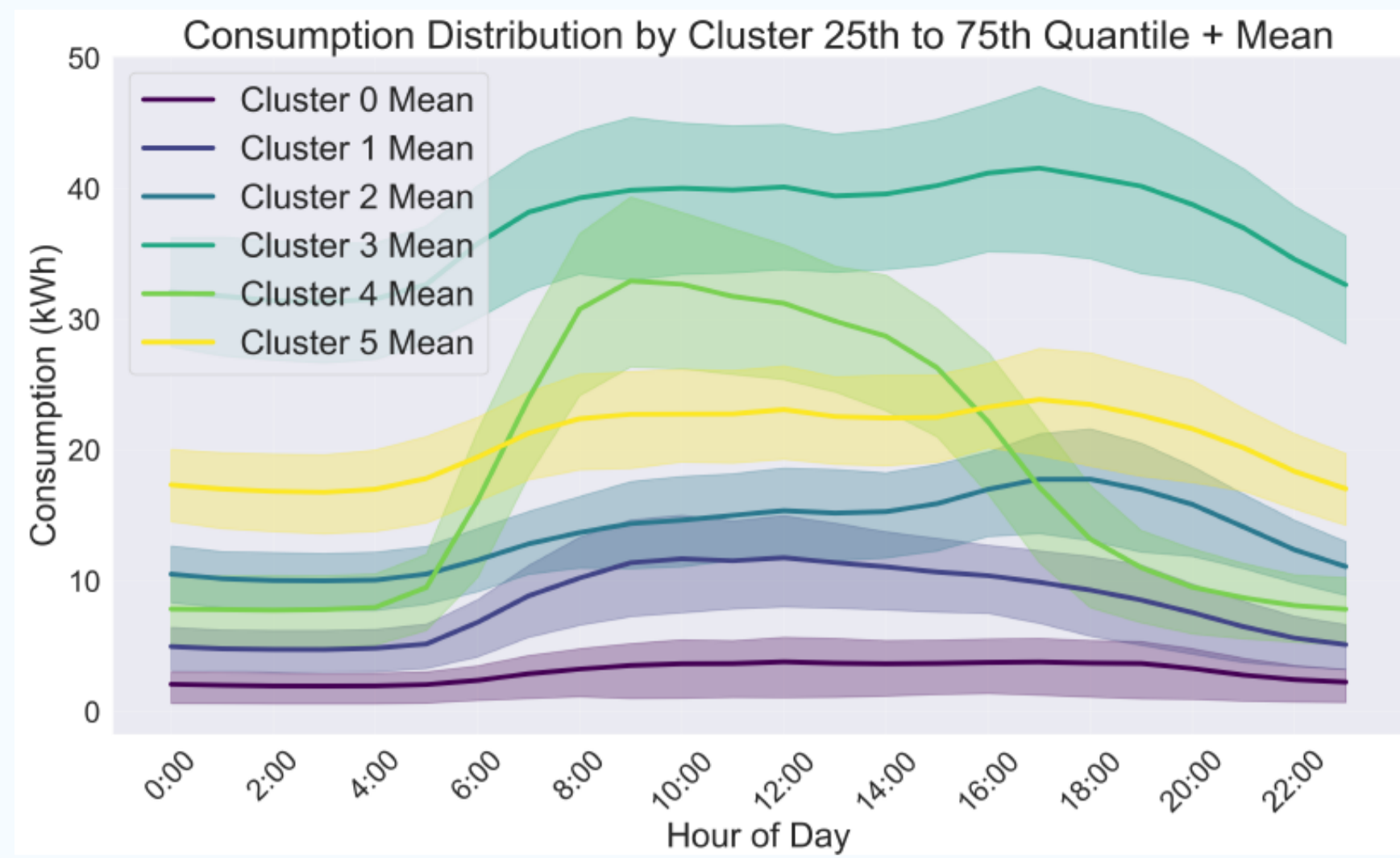
What: clustering is a powerful data analysis technique that groups similar data-points based on shared characteristics.

Why: clustering can **serve as the backbone** to analyze the data and develop applications to support the need.

Challenges:

- Massive data: from smart meters and sensors
- Streaming & Timeliness: data arrives continuously
- Time-Series Complexity: data has temporal information

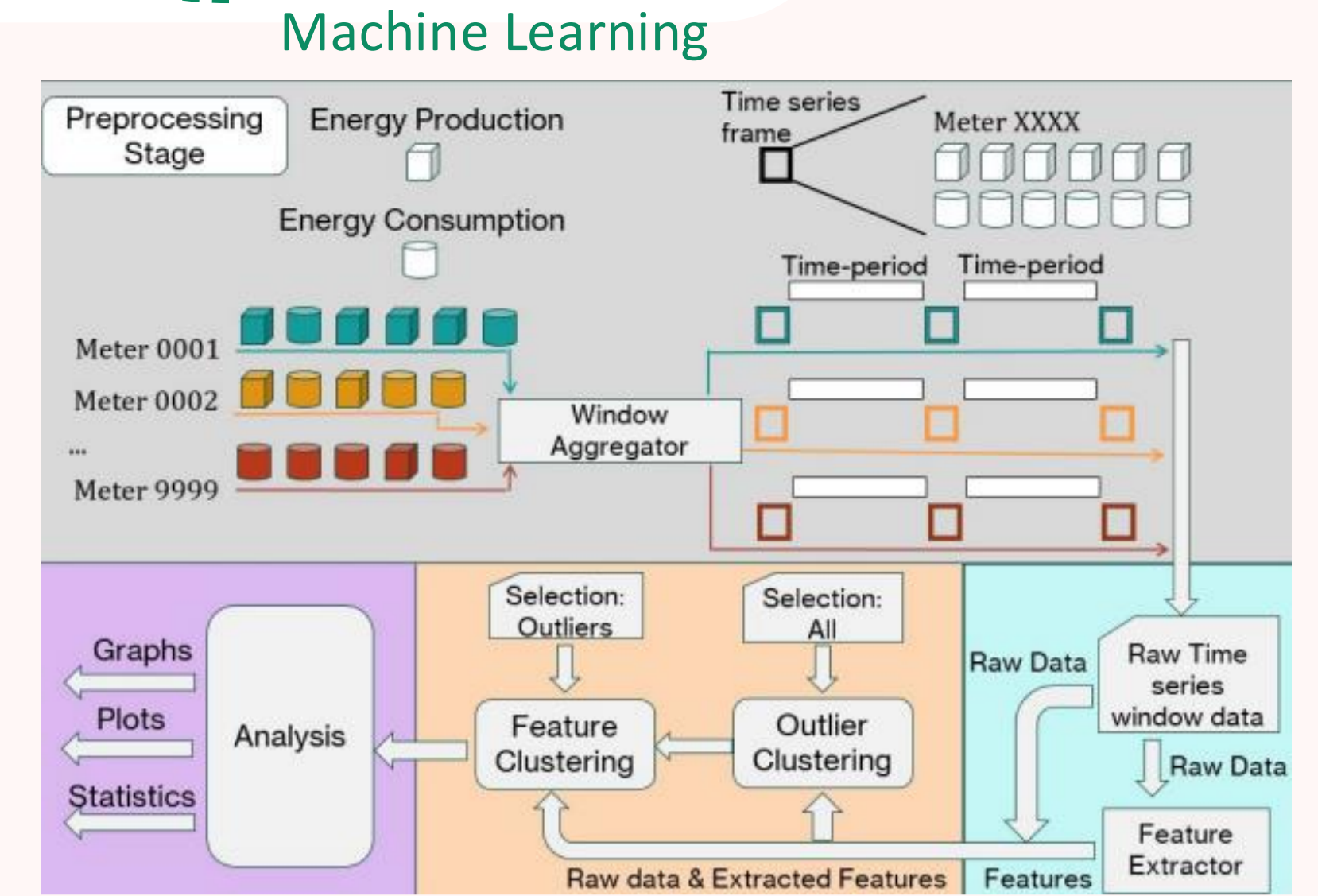




Macrobenchmark - Execution Time Ratio with different LSH parameters

Algorithm (Configuration)	Dataset 1	Dataset 2	Dataset 3
DBSCAN	1.0×	1.0×	1.0×
L=15, M=5	11.92×	22.22×	200.9×
L=50, M=10	11.22×	18.01×	186.0×
L=100, M=10	9.57×	16.01×	135.67×
L=100, M=50	7.32×	12.64×	81.43×

Table 5.6: Impact of LSH parameters on complete pipeline performance when combining IP.LSH.DBSCAN with K-means. Higher L and M values trade speed for accuracy. Dataset 1 (7,500 points, 24 dimensions), Dataset 2 (7,500 points, 96 dimensions), Dataset 3 (50,000 points, 96 dimensions).



This work's questions

Explore the power of data summarization with **clustering** for time-series electricity data.

Data summaries, transformations & data structures: what to summarize and how? Impact on downstream analysis. Facilitate new use-cases.

Efficiency & performance: how to summarize well by efficiently utilizing hardware features (e.g. parallelism)?

Balancing tradeoffs: lean processing at acceptable cost and timeliness.

CLUE & examples

Clustering + Summarization enables:

- Scalability + efficiency: orders of magnitude faster analysis
- Hence, larger exploration space

CLUE: toolchain⁽¹⁾ integrating:

- Efficient exploration & configurable feature engineering; round-based clustering
- Computationally efficient high-dimensional data processing, maintaining the analytical depth necessary for operational decision-making

(1) <https://github.com/rasmusthorsson/CLUE>

References

A. A. Keramatian, V. Gulisano, M. Papatriantafilou, P. Tsigas: IP.LSH.DBSCAN: Integrated Parallel Density-Based Clustering Through Locality-Sensitive Hashing. Parallel Processing Euro-Par 2022, Springer, https://doi.org/10.1007/978-3-031-12597-3_17

B. Z. Fu, M. Almgren, O. Landsiedel, M. Papatriantafilou. Online temporal-spatial analysis for detection of critical events in Cyber-Physical Systems. 2014 IEEE Int'l Conf. on Big Data (Big Data). 129–134. <https://doi.org/10.1109/BigData.2014.7004221>

C. B. Havers, R. Duvignau, H. Najdataei, V. Gulisano, M. Papatriantafilou, A. Koppisetty: DRIVEN: A framework for efficient Data Retrieval and clustering in Vehicular Networks. Future Gener. Comput. Syst. 2020 <https://doi.org/10.1016/j.future.2020.01.050>

D. L. Magnusson, M.Thorsson. 2025. CLUE – Cluster-based Load Understanding and Exploration: Summarizing High-Dimensional Electricity Grid Data for Scenario Analysis. Master's thesis. Gothenburg University, <https://gupea.ub.gu.se/handle/2077/89783>

E. V. Botev, M. Almgren, V. Gulisano, O. Landsiedel, M. Papatriantafilou, J. van Rooij: Detecting non-technical energy losses through structural periodic patterns in AMI data. IEEE BigData 2016 <https://doi.org/10.1109/BigData.2016.7840967>

Linus Magnusson, Rasmus Thorsson, Vinh Quang Ngo, Marina Papatriantafilou, Joris van Rooij, Mihail Chigrichenko

Access the thesis work digitally:



Distributed Computing and Systems
Chalmers university of technology



EPITOME - Summarization and structuring of continuous data in concurrent



MSCA Doctoral Network RELAX-DN



EU Horizon 2021-27 - Relaxed Semantics



Swedish Electricity Storage and Balancing Centre

