

Shapelets to Classify Energy Demand Time Series

Marco G. Pinheiro, Sara C. Madeira, and Alexandre P. Francisco



TÉCNICO
LISBOA

December 15th, 2022



Marco Pinheiro

Instituto Superior Técnico
EDP – Energias de Portugal

Lisboa
marco.pinheiro@edp.com

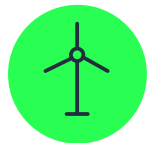
- Master in Computer Science and Eng. (2010)
- Work in energy sector since 2011 as innovation leader in Data, Analytics, Big Data, and AI technology
- Recently as strategist and architecture leader for Data&AI at EDP – Digital Global Unit
- PhD Candidate in Computer Science and Engineering with the subject *Short-Term Electric Load Forecasting* (from national into secondary substations levels)

Examples



Smart grids

increasing power grid sensorization
Increasing data availability



Generation assets

+500 parameters measured in a modern wind turbine
Many energy generation assets are highly monitored



DER, EV, Smart meters, Storage

More time series

Challenges



Data monetization (internally)

Inability to recognize the value of data beyond the siloed application in which data are collected is seen as a barrier.



Complexity

Data interoperability
Data governance complexity
Data streams deluge
Advanced analytics complexity (talent shortage)
Mind Interpretability and Explainability of Machine Learning models

Objectives

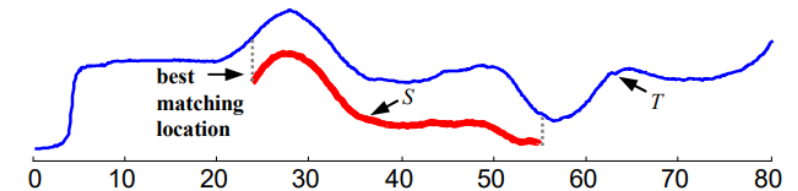
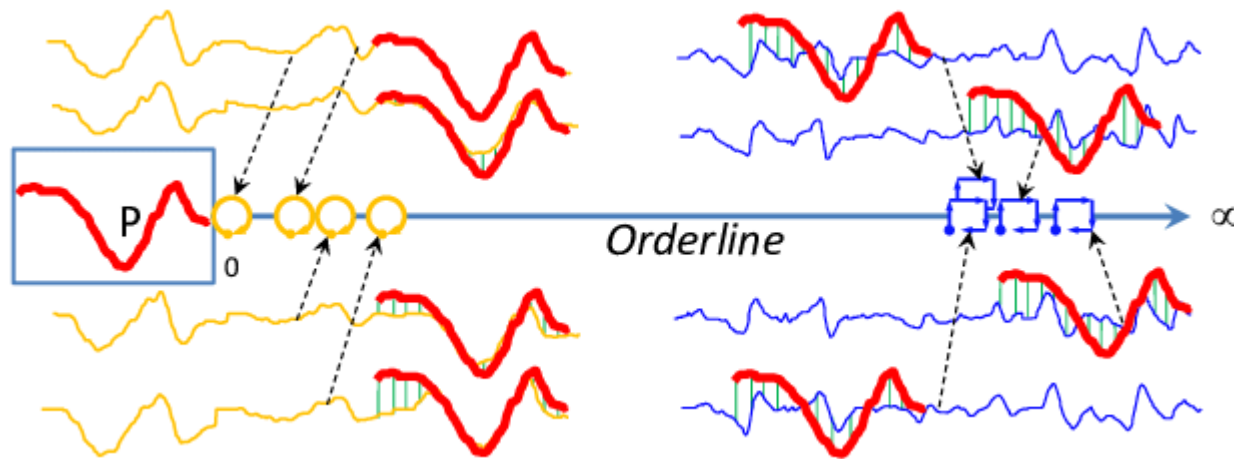
- Demonstrate the value and information that can be extracted as interpretable patterns from time series.
- Reinforces the value of smart grid data in addition to the siloed application in which data are gathered.
- Extend the study of shapelets (and other data mining techniques) to the energy-specific domain time series.

Definition

Shapelets is a primitive for time series data mining.

It represents the discriminative subsequence (in red) of time series which identify classes (yellow and blue)

Shapelet discovery algorithm seek to the subsequence that splits a set of time series into two different classes maximizing information gain (entropy reduction).



National demand

8 year long, 15 min, 1 country (Portugal)

2866 daily time series used as dataset

Dataset	Label A	Label B	Total
Train	40	40	80
Test	2014	772	2786

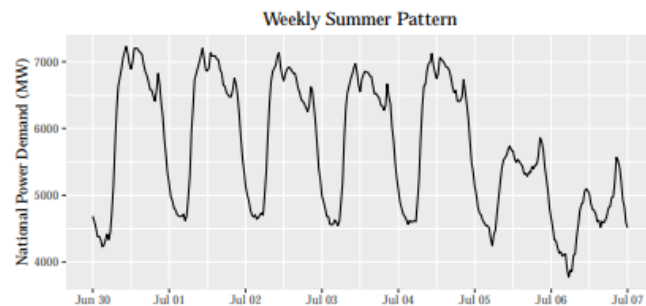
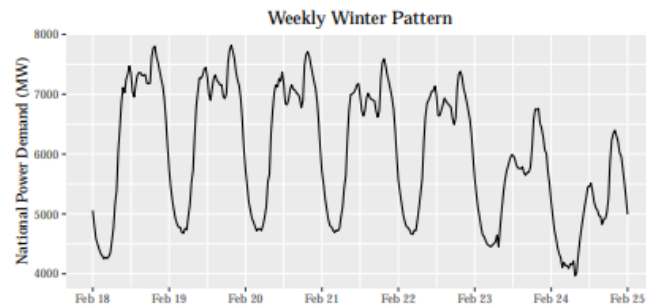


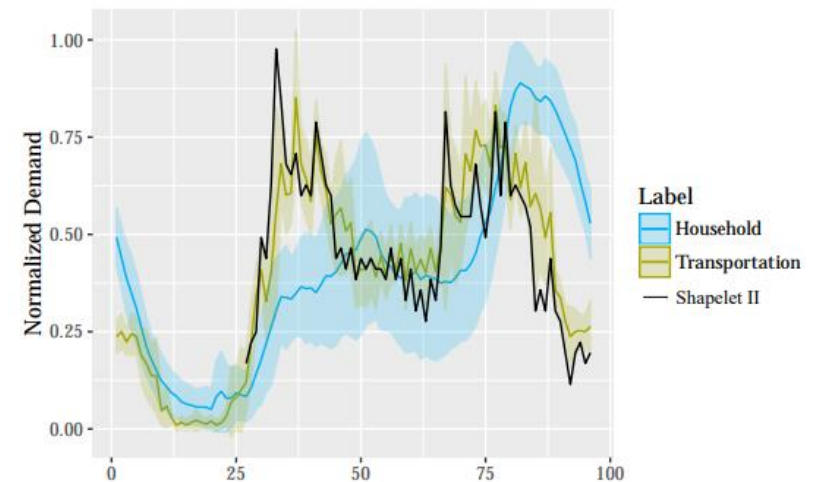
Figure 1. Portuguese demand from Monday to Sunday, in February 2008 and July 2008.

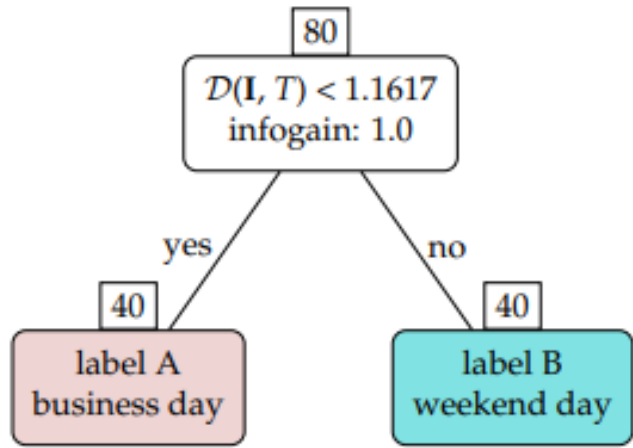
Secondary substations (level) demand

3 year long, 15 min, 28 secondary substations

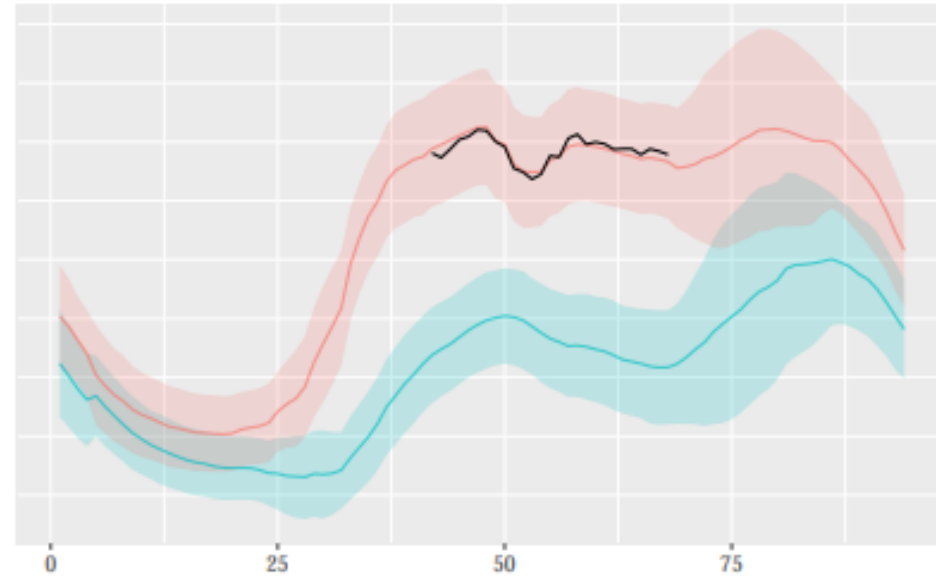
342 daily time series used as dataset

		Household	Industry	Services	Utilities	Transportation	Total
Meters	Train	5	3	4	2	1	15
	Test	4	3	3	2	1	13
Dataset	Train	45	27	36	18	9	135
	Test	37	57	57	38	18	207





(a) decision tree



(b) shapelet I

Figure 4. The decision tree classifier for the weekend problem. During the classification of an arbitrary time series T , its normalized subsequences are compared to the shapelet I (the black line). It is classified as label A if one of its normalized subsequences is *similar* to the shapelet. Note that the shapelet I was positioned according to where it was found, just to match the *similarities*. The shapelet might have a different scale and does not correspond to the coordinate axis.

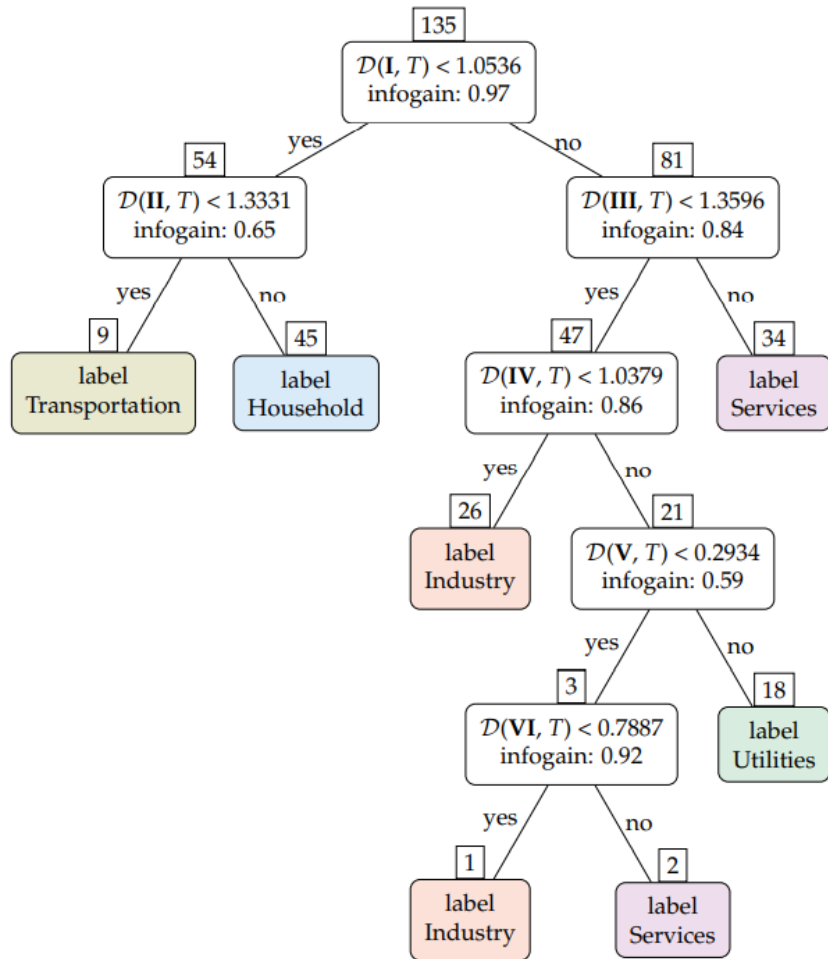


Figure 7. The decision tree classifies the type of power consumption.

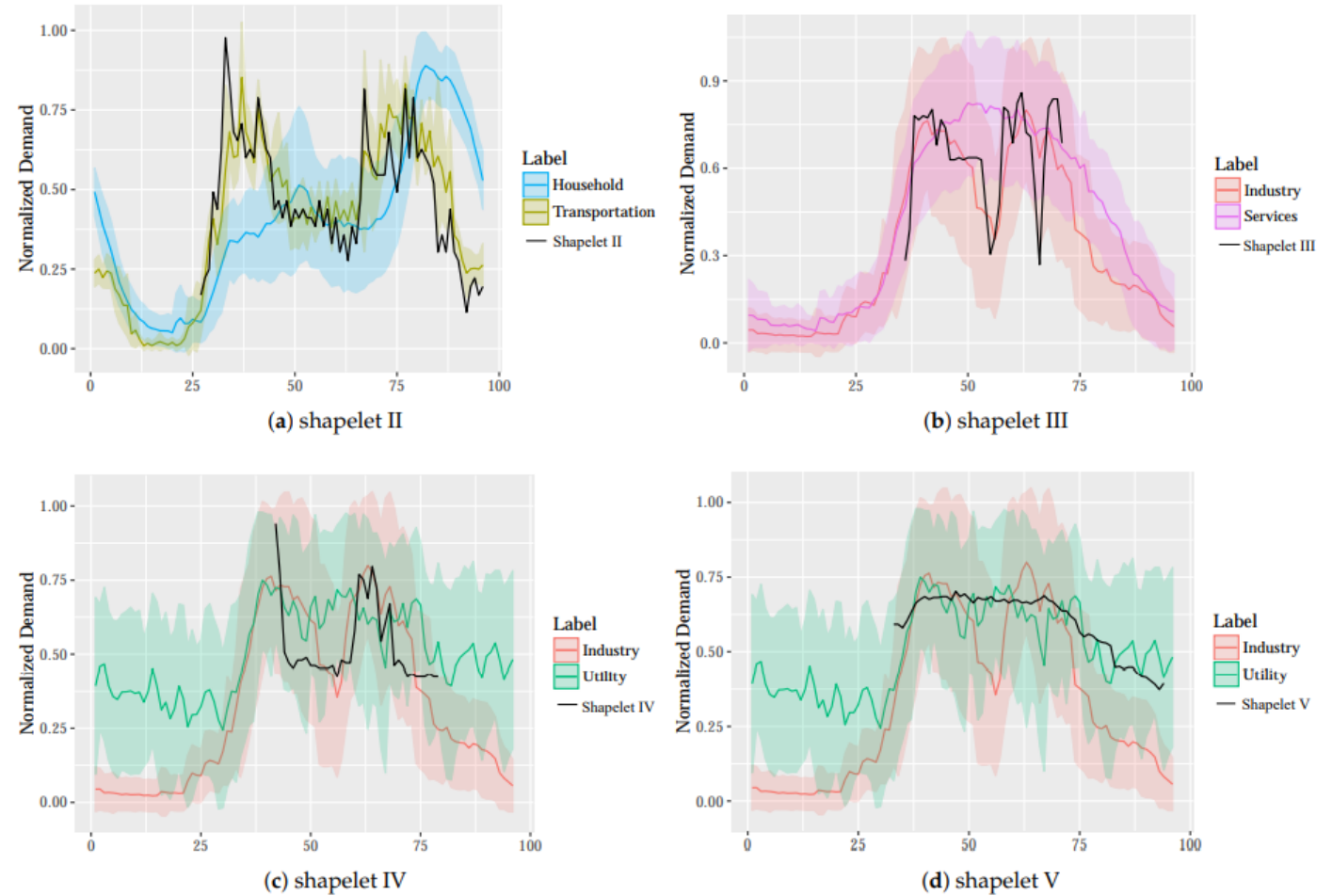


Figure 9. Normalized demand of classes detached by shapelets. The shapelet II detaches “transportation” from “household” through the demand patterns in the morning and evening. The shapelet III separates “industry” from “services” by a demand break around lunch time. The shapelet IV and V splits the time series into two groups: “industry” and “utility”.

Conclusions

- Shapelets demonstrated to be a useful technique to identify shapes in electric load curves in order to split into type of demand.
- Shapelets are mostly interpretable from the domain knowledge point of view.
- Interpretability is one of the crucial characteristics in the context of electricity companies in a high-regulated business in Europe, which seeks to regulate AI across several applications, particularly models that act on or support decisions over critical infrastructure.

The image features a central logo consisting of the lowercase letters 'edp' in a white, sans-serif font. This logo is set within a dark blue circular background. Surrounding this central circle are three overlapping, semi-transparent rings of different colors: a bright cyan ring, a vibrant green ring, and a purple ring. The rings are arranged in a way that they appear to be layered, with the purple ring being the outermost and the cyan ring being the innermost. The overall composition is circular and centered on a dark background.

edp