Time Series Data Mining - representation and clustering

Peter Laurinec

May 14, 2019

POWEREX

Highlights

Time Series Data Mining - representation and clustering

Highlights

Time Series Data Mining - representation and clustering

- About me,

Highlights

Time Series Data Mining - representation and clustering

- About me,
- What is time series,

Highlights

Time Series Data Mining - representation and clustering

- About me,
- What is time series,
- Time series data mining methods,

Highlights

Time Series Data Mining - representation and clustering

- About me,
- What is time series,
- Time series data mining methods,
- Time series clustering - problems,

Highlights

Time Series Data Mining - representation and clustering

- About me,
- What is time series,
- Time series data mining methods,
- Time series clustering - problems,
- Time series representations,

Highlights

Time Series Data Mining - representation and clustering

- About me,
- What is time series,
- Time series data mining methods,
- Time series clustering - problems,
- Time series representations,
- TSrepr R package,

Highlights

Time Series Data Mining - representation and clustering

- About me,
- What is time series,
- Time series data mining methods,
- Time series clustering - problems,
- Time series representations,
- TSrepr R package,
- Data (time series) streams,

About me

About me

- Master's degree from FMFI UK - statistics. Thesis about Model-based cluster analysis, supervisor: doc. **Radoslav Harman**.

About me

- Master's degree from FMFI UK - statistics. Thesis about Model-based cluster analysis, supervisor: doc. **Radoslav Harman**.
- **PhD.** from FIIT STU - intelligent information systems. Thesis: Improving Forecasting Accuracy through the Influence of Time Series Representations and Clustering, supervisor: prof. Mária Lucká.

About me

- Master's degree from FMFI UK - statistics. Thesis about Model-based cluster analysis, supervisor: doc. **Radoslav Harman**.
- **PhD.** from FIIT STU - intelligent information systems. Thesis: Improving Forecasting Accuracy through the Influence of Time Series Representations and Clustering, supervisor: prof. Mária Lucká.
- Now: Data Scientist at **PowereX**. **P2P** energy sharing marketplace. Forecasting and analysing large amount of time series from smart meters.

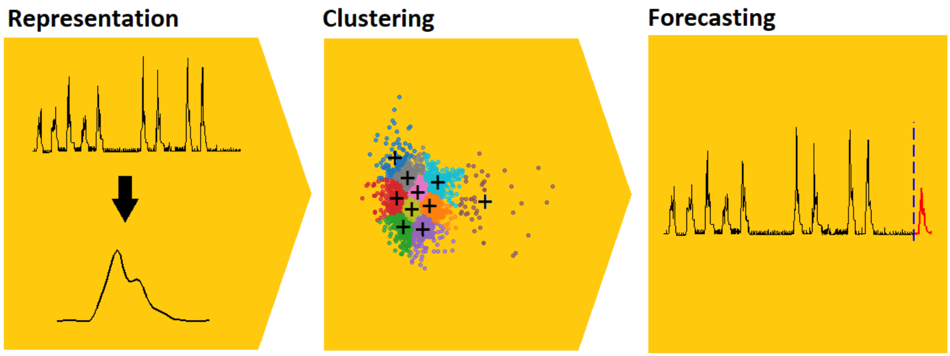
PhD. Thesis Goals

- The thesis had the goal to investigate, in the broader context, the **usage of time series data mining (analysis) methods** in order to **improve the predictive performance of machine learning methods** and its combinations.

PhD. Thesis Goals

- The thesis had the goal to investigate, in the broader context, the **usage of time series data mining (analysis) methods** in order to **improve the predictive performance of machine learning methods** and its combinations.
- In more detail, the goal was to investigate the usage of various **time series representations** for seasonal time series, **clustering**, and **forecasting methods** for electricity consumption forecasting accuracy improvement.

Approach Overview



But...start from the beginning

What is a time series?

Definition
A time series $\mathbf{x}$ is an ordered sequence of n real-valued variables

$$\mathbf{x} = (x_1, x_2, \ldots, x_n), x_i \in \mathbb{R}.$$

But...start from the beginning

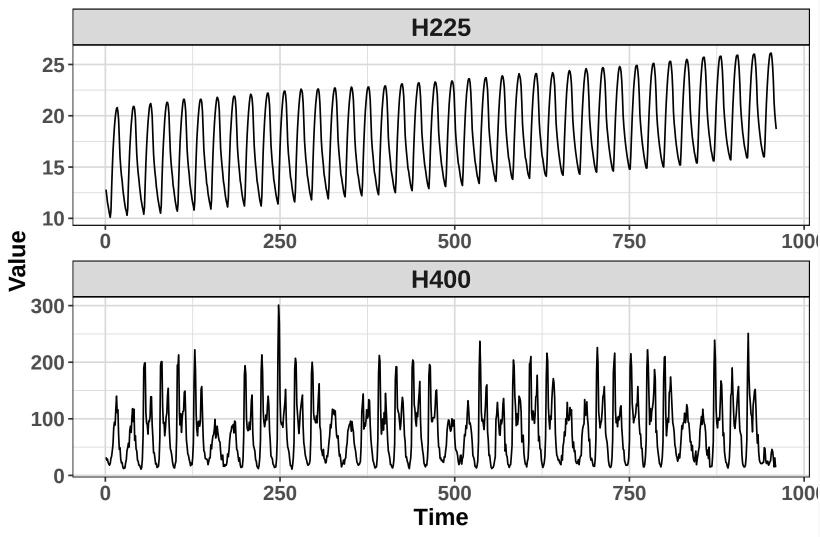
What is a time series?

Definition
A time series $\mathbf{x}$ is an ordered sequence of n real-valued variables

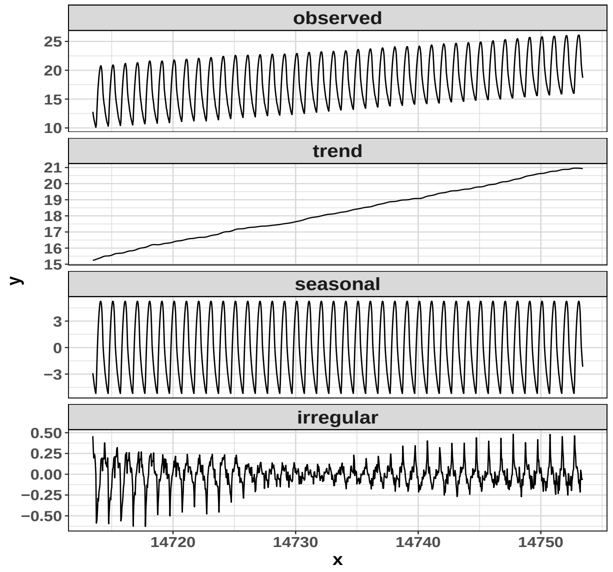
$$\mathbf{x} = (x_1, x_2, \ldots, x_n), x_i \in \mathbb{R}.$$

Domains, where TS can occur:

- Economy, stock exchange, demography, energetics, weather, web traffic, insurance, IoT sensors and many more.


H225
25
20
15
10
0
250
500
750
Value
H400
300
200
100
0
Time

Parts of time series



TS Data Mining Methods

- Methods for working with TS:

TS Data Mining Methods

- Methods for working with TS:
    - TS representations,

TS Data Mining Methods

- Methods for working with TS:
  - TS representations,
  - TS distance measures,

TS Data Mining Methods

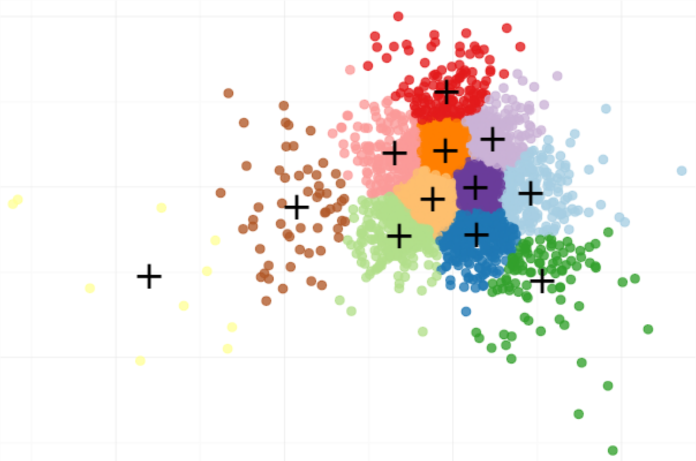
- Methods for working with TS:
  - TS representations,
  - TS distance measures,
- Tasks:

TS Data Mining Methods

- Methods for working with TS:
  - TS representations,
  - TS distance measures,
- Tasks:
  - TS classification,
  - TS clustering,
  - TS forecasting,
  - TS anomaly detection,
  - TS motif discovery,
  - TS indexing.

Clustering

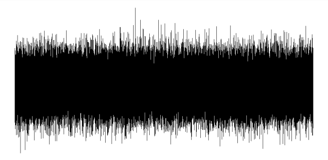
You should know...



Clustering TS

What is the difference?
(problems)

Clustering TS



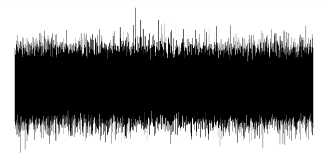
What is the difference? (problems)

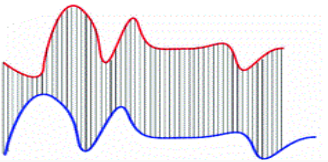
- TS can be very long - high dimensionality,

Clustering TS

What is the difference? (problems)

- TS can be very long - high dimensionality,
- TS can be lagged or moved to some direction,

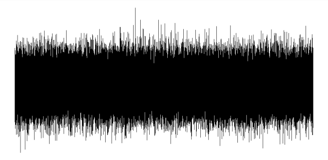


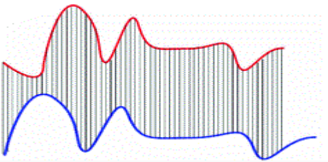


Clustering TS

What is the difference? (problems)

- TS can be very long - high dimensionality,
- TS can be lagged or moved to some direction,
- TS can constantly grow - new values coming at time.





N -> ∞

Time Series Representations

What can we do for solving problems with high-dimensional TS?

Time Series Representations

What can we do for solving problems with high-dimensional TS?

- Use time series representations!

Time Series Representations

What can we do for solving problems with high-dimensional TS?

- **Use time series representations!**

They are excellent to:

- Reduce memory load.
- Accelerate subsequent machine learning algorithms.
- Implicitly remove noise from the data.
- Emphasize the essential characteristics of the data.
- Help to find patterns in data (or motifs).

TS representation methods

Definition
Let $\mathbf{x}$ be a time series of length n, representation of $\mathbf{x}$ is a model $\hat{\mathbf{x}}$ of reduced dimensionality d ($d \ll n$) such that $\hat{\mathbf{x}}$ closely approximates $\mathbf{x}$.

TS representation methods

Definition
Let $\mathbf{x}$ be a time series of length n, representation of $\mathbf{x}$ is a model $\hat{\mathbf{x}}$ of reduced dimensionality d ($d \ll n$) such that $\hat{\mathbf{x}}$ closely approximates $\mathbf{x}$.

Four types of time series representation methods:

1. Nondata adaptive,
2. Data adaptive,
3. Model based,
4. Data dictated.

Nondata adaptive repr.

In nondata adaptive representations, the parameters of transformation remain the same for all time series, irrespective of their nature.

Nondata adaptive repr.

In nondata adaptive representations, the parameters of transformation remain the same for all time series, irrespective of their nature.

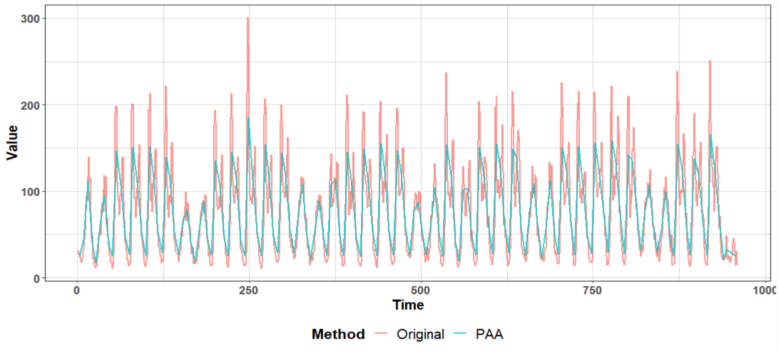
Methods:

- Piecewise Aggregate Approximation (PAA),
- Discrete Wavelet Transform (DWT),
- Discrete Fourier Transform (DFT),
- Discrete Cosine Transform (DCT),
- Perceptually Important Points (PIP).

PAA

PAA - Piecewise Aggregate Approximation.

$$\hat{x}_i = \frac{d}{n} \sum_{j=(n/d)(i-1)+1}^{(n/d)i} x_j.$$



We can also extract: median, standard deviation, maximum ...

Data adaptive

In data adaptive representations, the parameters of transformation vary depending on the available data.

Data adaptive

In data adaptive representations, the parameters of transformation vary depending on the available data.

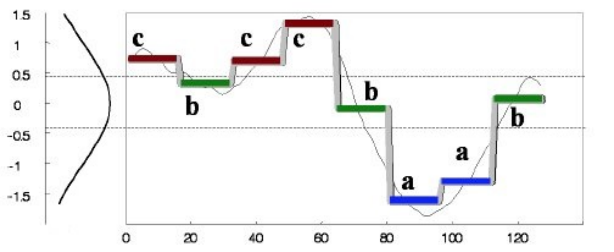
Methods:

- Symbolic Aggregate approXimation (SAX),
- Adaptive Piecewise Constant Approximation (APCA),
- Piecewise Linear Approximation (PLA),
- Singular Value Decomposition (SVD),
- Principal Component Analysis (PCA).

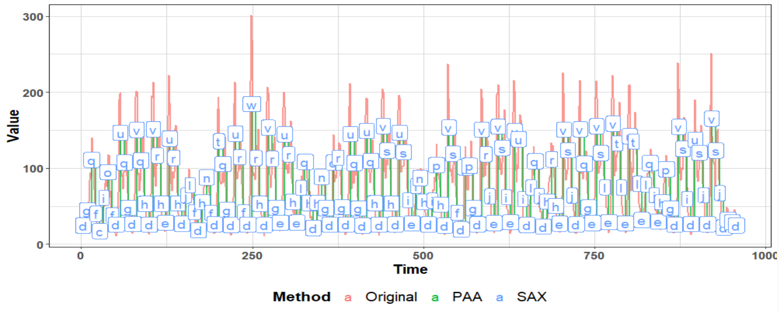
SAX

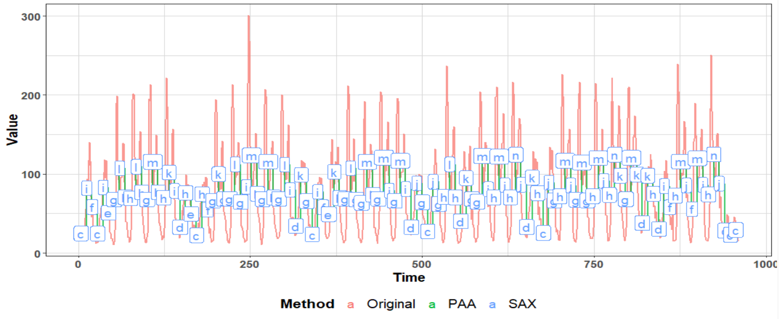
SAX - Symbolic Aggregate approXimation.

Firstly transforms a time series by PAA and then averages are transformed to symbols according to normal distribution quantiles.



We can change length of a "piece" and length of an alphabet.

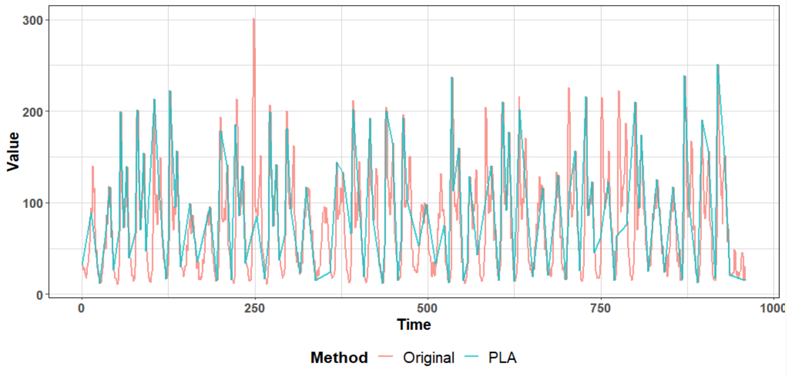




PLA

PLA - Piecewise Linear Approximation.

It begins by creating a simple approximation of the time series, i.e., $n/2$ segments are used and then iteratively connects pairs of segments with the least losses, until it reaches to the given number of segments.



Model based

The aim is to find the parameters of a model as a representation. Two time series are then considered as similar if they were created by the same set of parameters of a basic model.

Model based

The aim is to find the parameters of a model as a representation. Two time series are then considered as similar if they were created by the same set of parameters of a basic model.

Methods:

- ARIMA,
- Hidden Markov Chains,
- Seasonal models:
  - Seasonal averages,
  - Regression coefficients (MLR, RLM, GAM),
  - Holt-Winters exponential smoothing seasonal coefficients,

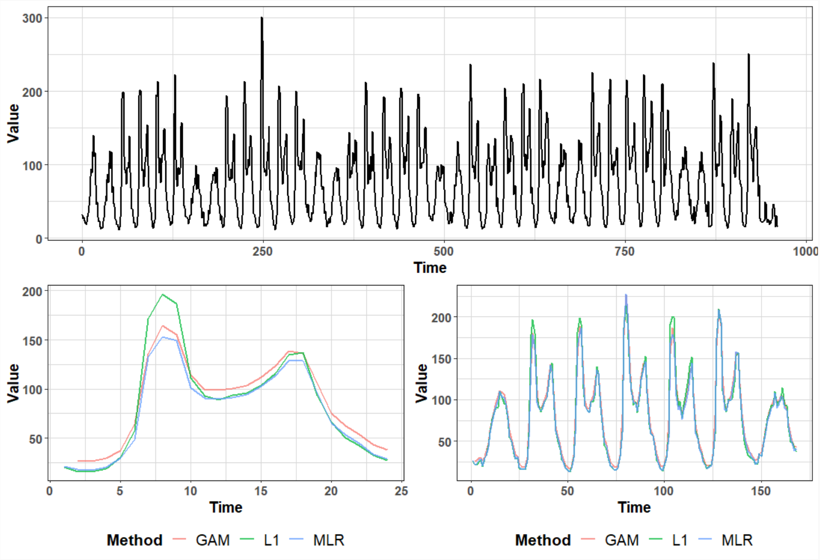
Seasonal model-based repr.

Creation of a representation which is long as a frequency of a time series.

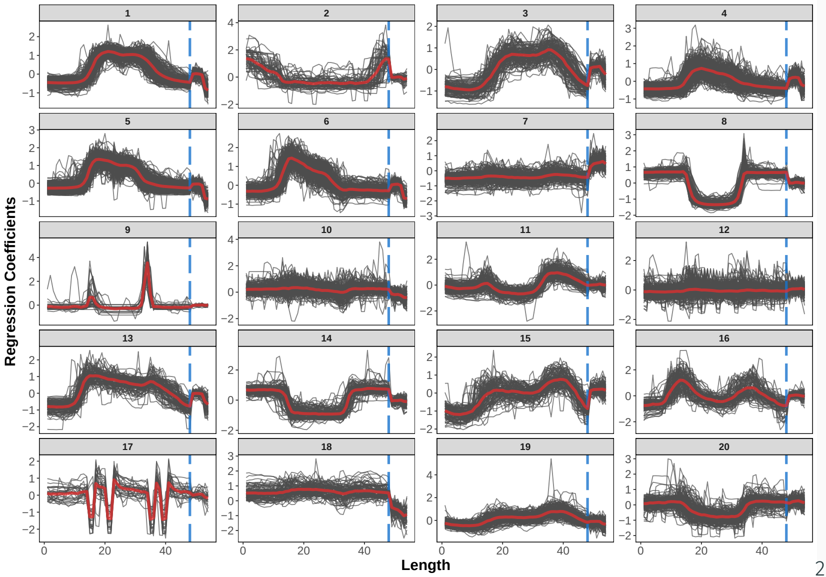
$$x_i = \beta_1 u_{i1} + \beta_2 u_{i2} + \cdots + \beta_{seas} u_{iseas} + \varepsilon_i, \; \textit{where} \; i = 1, \ldots, n$$

New representation: $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_{seas})$.

Applied methods:
Multiple Linear Regression. Robust Linear Model. Quantile Regression. Generalized Additive Model.


Value
Time
Value
Time
Method GAM L1 MLR
Value
Time
Method GAM L1 MLR

Clustered TS Representations



Data dictated

In data dictated approaches, the compression ratio is defined automatically based on a raw time series such as a clipped representation.

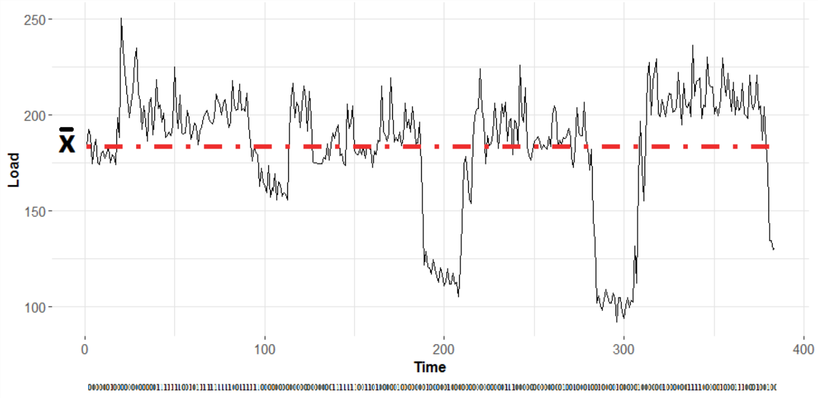
Data dictated

In data dictated approaches, the compression ratio is defined automatically based on a raw time series such as a clipped representation.

Clipping:

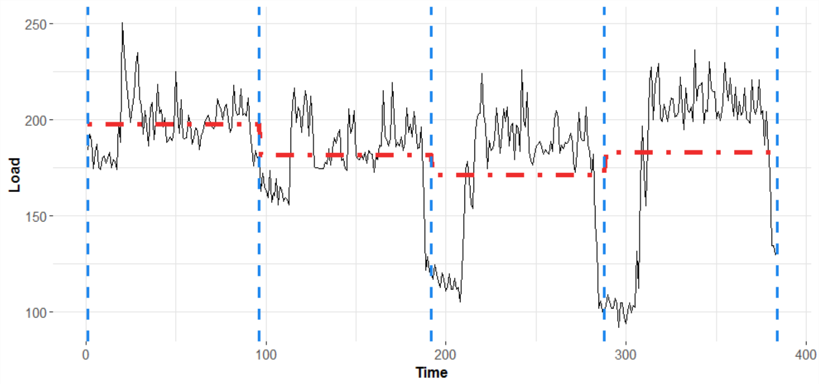
$$\hat{x}_t = \begin{cases} 1 & \text{if } x_t > \mu \\ 0 & \text{otherwise} \end{cases}$$

Clipped - bit level representation



Clipping - RLE

RLE - Run Length Encoding. Windowing - one day.



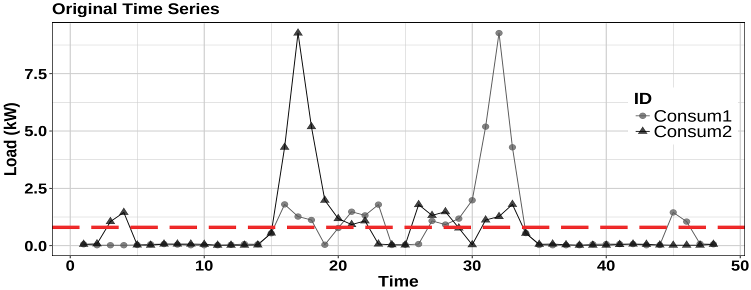
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
6	1	12	7	2	1	1	10	2	6	20	7	2	2	1	1	5	1	7	1	1	4	1	15	3	13	1	2	1	3	1	2	1	1	3	1	5	1	6	1	6	4	5	1	3	3	3	1	2	1	2

FeaClip

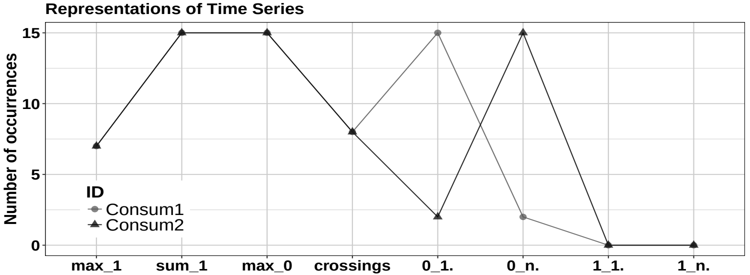
Feature extraction from the clipped representation.

$$
\begin{aligned}
\hat{x} = \{ & max_1 = \text{max. from run lengths of ones}, \\
& sum_1 = \text{sum of run lengths of ones}, \\
& max_0 = \text{max. from run lengths of zeros}, \\
& crossings = \text{length of RLE encoding} - 1, \\
& f_0 = \text{number of first zeros}, \\
& l_0 = \text{number of last zeros}, \\
& f_1 = \text{number of first ones}, \\
& l_1 = \text{number of last ones}, \}.
\end{aligned}
$$

FeaClip

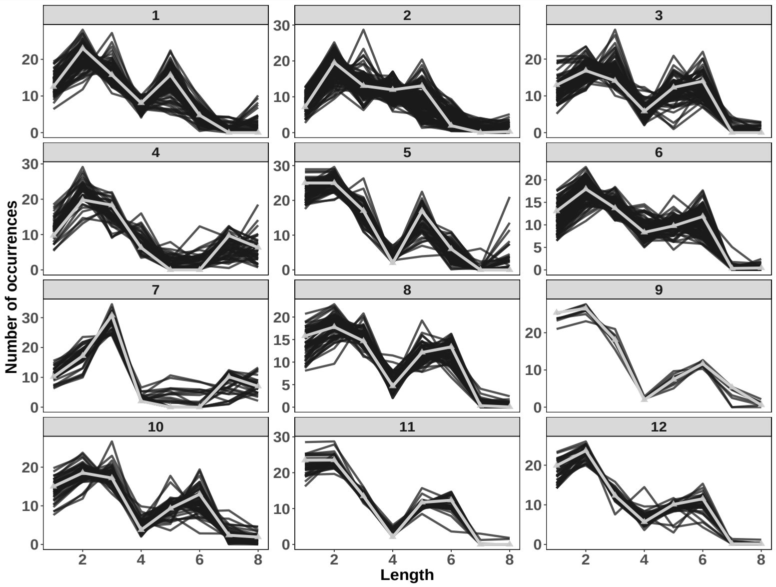


Original Time Series
Load (kW)
7.5
5.0
2.5
0.0
0
10
20
30
40
50
Time
ID
Consum1
Consum2



Representations of Time Series
Number of occurrences
15
10
5
0
max_1
sum_1
max_0
crossings
0_1.
0_n.
1_1.
1_n.
ID
Consum1
Consum2

Clustering FeaClip



Summarize TS representations

Summarize TS representations

- Various interesting methods,

Summarize TS representations

- Various interesting methods,
- Nondata adaptive methods have limits,

Summarize TS representations

- Various interesting methods,
- Nondata adaptive methods have limits,
- Data adaptive methods have own distance metrics (can be limiting),

Summarize TS representations

- Various interesting methods,
- Nondata adaptive methods have limits,
- Data adaptive methods have own distance metrics (can be limiting),
- For seasonal TS, the model-based and data dictated methods are best to use.

TSrepr

TSrepr - CRAN[1], GitHub[2]

- R package for time series representations computing
- Large amount of various methods are implemented
- Several useful support functions are also included
- Easy to extend and to use

[1] https://CRAN.R-project.org/package=TSrepr
[2] https://github.com/PetoLau/TSrepr/

TSrepr

TSrepr - CRAN[1], GitHub[2]

- R package for time series representations computing
- Large amount of various methods are implemented
- Several useful support functions are also included
- Easy to extend and to use

```
data <- rnorm(1000)
repr_paa(data, func = median, q = 10)
```

[1] https://CRAN.R-project.org/package=TSrepr
[2] https://github.com/PetoLau/TSrepr/

All type of time series representations methods are implemented, so far these:

- PAA - Piecewise Aggregate Approximation (`repr_paa`)
- DWT - Discrete Wavelet Transform (`repr_dwt`)
- DFT - Discrete Fourier Transform (`repr_dft`)
- DCT - Discrete Cosine Transform (`repr_dct`)
- PIP - Perceptually Important Points (`repr_pip`)
- SAX - Symbolic Aggregate Approximation (`repr_sax`)
- PLA - Piecewise Linear Approximation (`repr_pla`)
- Mean seasonal profile (`repr_seas_profile`)
- Model-based seasonal representations based on linear model (`repr_lm`)
- FeaClip - Feature extraction from clipping representation (`repr_feaclip`)

Additional useful functions are implemented as:

- Windowing (`repr_windowing`)
- Matrix of representations (`repr_matrix`)
- Normalisation functions - z-score (`norm_z`), min-max (`norm_min_max`)

TS can be lagged, with different speed. . .

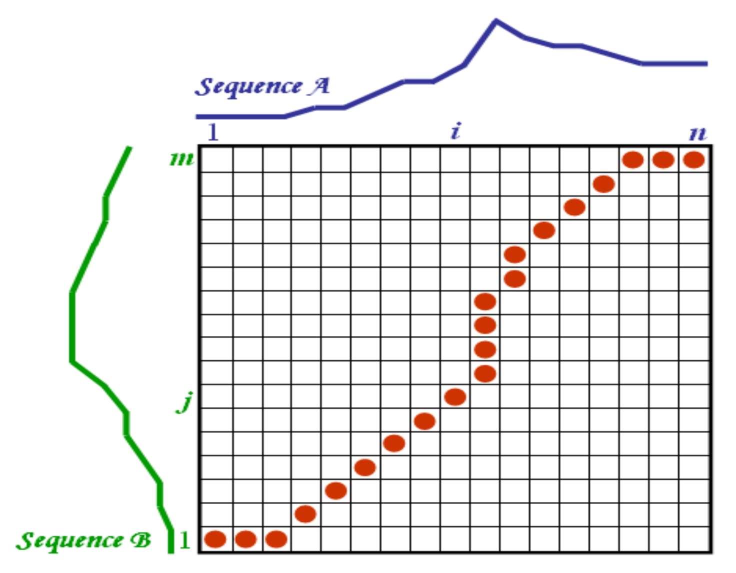
Dynamic Time Warping (DTW) distance -

Suppose we have two time series, a sequence Q of length n, and a sequence C of length m, where

$$Q = q_1, q_2, \ldots, q_i, \ldots, q_n$$
$$C = c_1, c_2, \ldots, c_j, \ldots, c_m.$$

To align these two sequences using DTW, we first construct an n-by-m matrix where the (i^{th}, j^{th}) element of the matrix corresponds to the squared distance, $d(q_i, c_j) = (q_i c_j)^2$, which is the alignment between points q_i and c_j. To find the best match between these two sequences, we retrieve a path through the matrix that minimizes the total cumulative distance between them.

DTW



Clustering Methods

Clustering Methods

- K-means with the most presented TS representations,

Clustering Methods

- K-means with the most presented TS representations,
- K-medoids (PAM) with custom distance function,

Clustering Methods

- K-means with the most presented TS representations,
- K-medoids (PAM) with custom distance function,
- Hierarchical clustering with custom distance function,

Clustering Methods

- K-means with the most presented TS representations,
- K-medoids (PAM) with custom distance function,
- Hierarchical clustering with custom distance function,
- K-shape and other PAM-like algorithms adapted based on own distance function.

Clustering Methods

- K-means with the most presented TS representations,
- K-medoids (PAM) with custom distance function,
- Hierarchical clustering with custom distance function,
- K-shape and other PAM-like algorithms adapted based on own distance function.

In R packages `dtwclust`, `TSclust`, `TSdist`.

Data (Time Series) Streams

Time series can constantly grow...

Definition
Data stream $\mathbf{s}$ is a sequence of objects $\mathbf{s} = x_1, x_2, \ldots, x_n$, or $\mathbf{s} = \{x_t\}_{t=1}^{n}$, which is potentially unbounded ($n \to \infty$).

Problems:

Data (Time Series) Streams

Time series can constantly grow...

Definition
Data stream $\mathbf{s}$ is a sequence of objects $\mathbf{s} = x_1, x_2, \ldots, x_n$, or $\mathbf{s} = \{x_t\}_{t=1}^{n}$, which is potentially unbounded ($n \to \infty$).

Problems:

- Real-time processing,

Data (Time Series) Streams

Time series can constantly grow...

Definition
Data stream $\mathbf{s}$ is a sequence of objects $\mathbf{s} = x_1, x_2, \ldots, x_n$, or $\mathbf{s} = \{x_t\}_{t=1}^{n}$, which is potentially unbounded ($n \to \infty$).

Problems:

- Real-time processing,
- Limited memory storage,

Data (Time Series) Streams

Time series can constantly grow...

Definition
Data stream $\mathbf{s}$ is a sequence of objects $\mathbf{s} = x_1, x_2, \ldots, x_n$, or $\mathbf{s} = \{x_t\}_{t=1}^{n}$, which is potentially unbounded ($n \to \infty$).

Problems:

- Real-time processing,
- Limited memory storage,
- Noise and anomalies in DS,

Data (Time Series) Streams

Time series can constantly grow...

Definition
Data stream $\mathbf{s}$ is a sequence of objects $\mathbf{s} = x_1, x_2, \ldots, x_n$, or $\mathbf{s} = \{x_t\}_{t=1}^{n}$, which is potentially unbounded ($n \rightarrow \infty$).

Problems:

- Real-time processing,
- Limited memory storage,
- Noise and anomalies in DS,
- Evolving nature of DS,

Data (Time Series) Streams

Time series can constantly grow...

Definition
Data stream $\mathbf{s}$ is a sequence of objects $\mathbf{s} = x_1, x_2, \ldots, x_n$, or $\mathbf{s} = \{x_t\}_{t=1}^{n}$, which is potentially unbounded ($n \to \infty$).

Problems:

- Real-time processing,
- Limited memory storage,
- Noise and anomalies in DS,
- Evolving nature of DS,
- High-dimensionality of DS.

Data (Time Series) Streams

In clustering:

Data (Time Series) Streams

In clustering:

- Windows -
  - Sliding,
  - Damped,
  - Landmark,

Data (Time Series) Streams

In clustering:

- Windows -
  - Sliding,
  - Damped,
  - Landmark,
- Online-Offline style -
  - Online - synopsis - representation,
  - Offline - clustering and other discoveries,

Data (Time Series) Streams

In clustering:

- Windows -
    - Sliding,
    - Damped,
    - Landmark,
- Online-Offline style -
    - Online - synopsis - representation,
    - Offline - clustering and other discoveries,
- Incrementality,

Data (Time Series) Streams

In clustering:

- Windows -
  - Sliding,
  - Damped,
  - Landmark,
- Online-Offline style -
  - Online - synopsis - representation,
  - Offline - clustering and other discoveries,
- Incrementality,
- Number of clusters and theirs character can vary,

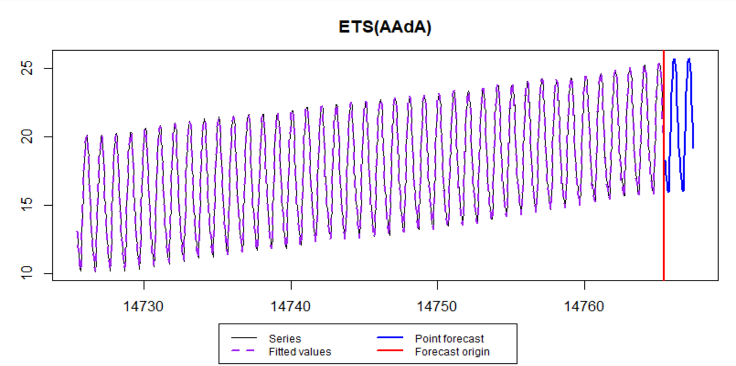
Data (Time Series) Streams

In clustering:

- Windows -
  - Sliding,
  - Damped,
  - Landmark,
- Online-Offline style -
  - Online - synopsis - representation,
  - Offline - clustering and other discoveries,
- Incrementality,
- Number of clusters and theirs character can vary,
- Automatic outlier detection,
- Automatic change detection.

Forecasting

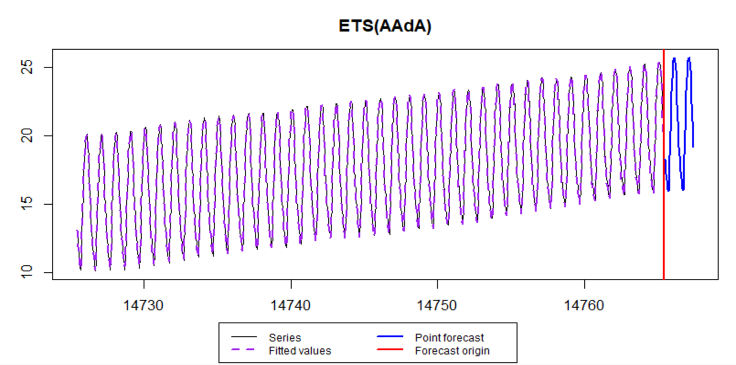
Methods:



Forecasting

Methods:

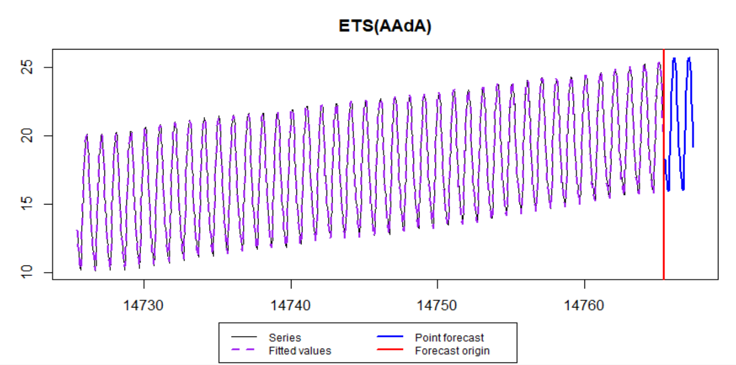
- Time series analysis -
  - ARIMA,
  - Exponential smoothing,
  - Theta ...



Forecasting

Methods:

- Time series analysis -
  - ARIMA,
  - Exponential smoothing,
  - Theta . . .
- Regression -
  - Multiple Linear Regression (LASSO),
  - Trees, Forests, Boosting,
  - Support Vector Regression,
  - ANN - RNN.



Forecasting - Feature Engineering

What to care about?

Forecasting - Feature Engineering

What to care about?

- Trend,
- Seasonalities (possible multiple),
- Holidays,

Forecasting - Feature Engineering

What to care about?

- Trend,
- Seasonalities (possible multiple),
- Holidays,
- Lag features of dependent and also independent variables,

Forecasting - Feature Engineering

What to care about?

- Trend,
- Seasonalities (possible multiple),
- Holidays,
- Lag features of dependent and also independent variables,
- Moving averages (and other statistics),
- Decompositions,

Forecasting - Feature Engineering

What to care about?

- Trend,
- Seasonalities (possible multiple),
- Holidays,
- Lag features of dependent and also independent variables,
- Moving averages (and other statistics),
- Decompositions,
- Interactions.

Conclusions

TS data mining:

Conclusions

TS data mining:

- TS representations are our fiends in clustering, forecasting, classification etc.,

Conclusions

TS data mining:

- TS representations are our fiends in clustering, forecasting, classification etc.,
- Implemented in TSrepr package,

Conclusions

TS data mining:

- TS representations are our fiends in clustering, forecasting, classification etc.,
- Implemented in TSrepr package,
- DTW for lagged TS,

Conclusions

TS data mining:

- TS representations are our fiends in clustering, forecasting, classification etc.,
- Implemented in **TSrepr** package,
- DTW for lagged TS,
- TS streams how to clustering them,

Conclusions

TS data mining:

- TS representations are our fiends in clustering, forecasting, classification etc.,
- Implemented in **TSrepr** package,
- DTW for lagged TS,
- TS streams how to clustering them,
- Intro to forecasting.

Conclusions

TS data mining:

- TS representations are our fiends in clustering, forecasting, classification etc.,
- Implemented in **TSrepr** package,
- DTW for lagged TS,
- TS streams how to clustering them,
- Intro to forecasting.

Questions: `laurinec.peter@gmail.com`
Blog: `https://petolau.github.io`
Work: `https://powerex.io`
Code: `https://github.com/PetoLau/`