TSrepr R package: Time Series Representations

7 March 2018

Summary

`TSrepr` (Laurinec 2018) is an R package for time series representations computing. Time series representations are, in other words, methods for dimensionality reduction, feature extraction or for the preprocessing of time series. They are used for (Esling and Agon 2012):

- Significant reduction of the time series dimensionality,
- Emphasis of fundamental (essential) shape characteristics,
- Implicit noise handling,
- Dimension reduction will reduce the memory requirements and computational complexity of consequent machine learning methods (e.g., classification or clustering).

The `TSrepr` package contains various methods and types of time series representations including the Piecewise Aggregate Approximation (PAA), the Discrete Fourier Transform (DFT), the Perceptually Important Points (PIP), the Symbolic Aggregate approXimation (SAX), the Piecewise Linear Approximation (PLA) and Clipping. Except for these well-known methods, additional methods suitable for seasonal time series are implemented. These methods are based on the model, for example multiple linear regression, robust regression, generalised additive model or triple exponential smoothing (Laurinec and Lucká 2016, Laurinec et al. (2016)). Own developed feature extraction methods from the Clipping representation are also implemented - FeaClip and FeaTrend. In Figure 1, the comparison of all eight available model-based representations in the `TSrepr` on electricity consumption time series from the randomly picked residential consumer is shown.
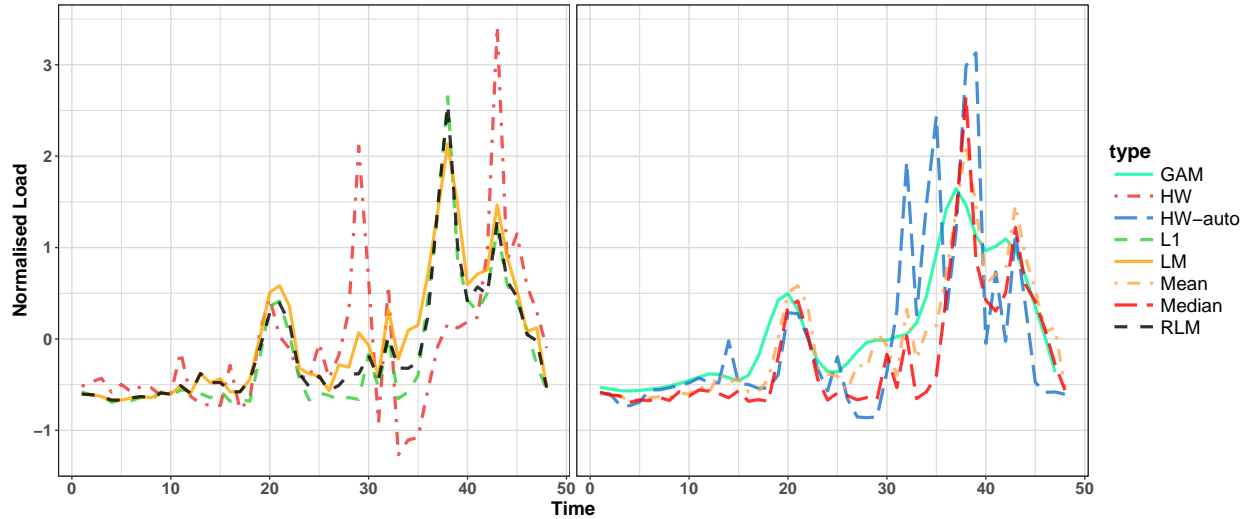
Figure 1. The comparison of model-based time series representations on electricity consumption time series. The length of representations is 48, the same as frequency of the daily season of the used time series.

Additional useful functions and methods related to time series representations were also implemented. The `TSrepr` package includes functions for normalisations and denormalisations of time series - z-score and min-max methods. It supports the simple computation of the windowing method, a matrix of representations and forecasting accuracy measures (MAE, RMSE, MAPE, sMAPE etc.). Methods (functions) were implemented in base R and also in C++ for fast computations. In R, C++ programmes can be written thanks to the package Rcpp (Eddelbuettel and Francois 2011).

The general package for time series representations computations had missed so far. The CRAN's time series task view proves the previous statement. Packages `TSMining` (Fan 2015) and `jmotif` (Senin 2016) both

includes implementations of PAA and SAX time series representations methods. However, these packages are mainly focused on motif discovery in time series.

References

Eddelbuettel, Dirk, and Romain Francois. 2011. "Rcpp: Seamless R and C++ Integration." *Journal of Statistical Software, Articles* 40 (8): 1–18. doi:10.18637/jss.v040.i08.

Esling, Philippe, and Carlos Agon. 2012. "Time-series data mining." *ACM Computing Surveys* 45 (1). ACM: 1–34.

Fan, Cheng. 2015. *TSMining: Mining Univariate and Multivariate Motifs in Time-Series Data.* https://CRAN.R-project.org/package=TSMining.

Laurinec, Peter. 2018. *TSrepr: Time Series Representations.* https://CRAN.R-project.org/package=TSrepr.

Laurinec, Peter, and Mária Lucká. 2016. "Comparison of Representations of Time Series for Clustering Smart Meter Data." In *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering and Computer Science 2016*, 458–63.

Laurinec, Peter, Marek Lóderer, Petra Vrablecová, Mária Lucká, Viera Rozinajová, and Anna Bou Ezzeddine. 2016. "Adaptive Time Series Forecasting of Energy Consumption Using Optimized Cluster Analysis." In *Data Mining Workshops (Icdmw), 2016 Ieee 16th International Conference on*, 398–405. IEEE.

Senin, Pavel. 2016. *Jmotif: Time Series Analysis Toolkit Based on Symbolic Aggregate Dicretization, I.e. Sax.* https://CRAN.R-project.org/package=jmotif.