MYDATAMODELS
The Automated Machine Learning Company

Case Studies – Master M1 – Université Cote d'Azur

PROPOSED STUDY

Time series as signals: depth of training and forecasting horizon

Aim:

1. Given time varying input data, determine the optimal depth for the back-test method (known also as *time-lag*) in order to perform predictions with the TADA algorithm.
2. Likewise, evaluate the forecast obtained by varying the *prediction horizon* – namely the time interval ahead in the future and judge on the obtained accuracy.
3. Envisage potential extensions to multi-variate time series.

The algorithm in TADA was originally designed to perform Machine Learning predictions in Classification and Regression tasks – thus *not* forecasting. Recently though, it has been preliminarily tested on real world uniformly spaced time series with satisfactory performances.

One method to address the necessary data transformation is the back-test method (also known as *time-lag*). However, besides the difficulties of the traditional train-test split, not applicable 'as-is' in this instance - parametrization remains undefined, especially for the choice of the time intervals – both in the past (how many past observations should we use?) – and in the future (how far can we predict?). Hence existing methods to determine these time lags need to be tested and novel ones need finding. The variability of these time intervals permits exploring scenarios which are interesting for real-world IoT applications. Benchmarks of the algorithm as a function of these 'lags' should be performed against traditional techniques like ARIMA – where applicable - but also to more contemporary implementations.

Additionally, real world time series are 'signals,' whose pre-processing can be done with techniques that come from signal processing (from filtering to smoothing, down to Fourier Transform and more advanced techniques). Thus, also the influence of the chosen pre-processing techniques on the overall performance needs to be evaluated.

Datasets will be provided from publicly available repositories and from real world situation of commercial interest.

Contact:
Carlo Fanara: cf@mydatamodels.com