

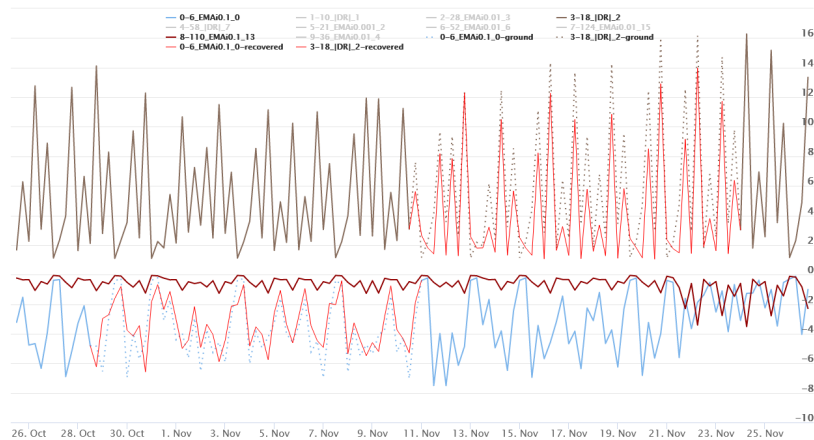
Fribourg, 22. Juli 2022

ImputeVIS: Visualizing Time Series Recovery using ImputeBench

MSc/BSc Thesis:

Work overview:

Real-world time series often contain missing values due to sensor failures, power outages or transmission problems. The recovery of those missing values allows better analysis of time series. Several methods have been proposed to recover missing values in time series, which can be algebraic, statistical, machine learning [1], etc. Selecting the “best” recovery method highly depends on the time series features, the missing rate and type, and error metric. ImputeBench [2] is a popular terminal-based benchmark that compares over 15 missing value imputation techniques algorithms. It relies on a reusable code framework, a large selection of representative time series, and mechanisms to automate benchmarking.



The goal of the thesis is to construct a helper graphical tool aimed at visualizing different aspects of missing value recovery process - parametrization, patterns of missing values, time series features, etc. The tool aims to challenge the existing results of the benchmark through a new dimension of visual analysis of the existing components of ImputeBench. The output of the thesis is a powerful interactive visualization suite that can help do further research in the fields of time series repair, recovery techniques, algorithm parameterization, etc.

Work tasks:

1. Familiarize yourself with the topic of recovery of missing values in time series.
2. Study ImputeBench benchmark, its functionality and contents (algorithms, datasets, missing patterns, etc.).
3. Build a web-based tool to visualize different aspects of missing value recovery process - recovered values, missing patterns, analytical metrics, parameter dependency etc. The tool should have the following features:
 - Visualize real-world time series and various ways of normalizing the data
 - Extract time series features using various extractors (e.g., hctsa [3], KATS [4], TS-Fresh [5], TSFEL, etc.)
 - Display the result of parameterizing each imputation technique
 - Select a set of imputation techniques, execute different recovery scenarios, and visualize the result of the recovery together with some effectiveness/efficiency metrics.
 - Display an summary analysis of the comparison and provide a recommendation based on it.
4. Write a thesis that describes the results.
5. Presentation of 20 minutes.

Literature:

1. Mourad Khayati, Alberto Lerner, Zakhar Tymchenko, and Philippe Cudré-Mauroux. "Mind the Gap: An Experimental Evaluation of Imputation of Missing Values Techniques in Time Series." In Proceedings of the VLDB Endowment, Vol. 13, 2020.
2. ImputeBench: Benchmark of Imputation Techniques in Time Series. <https://github.com/eXascaleInfolab/bench-vldb20>.
3. B. D. Fulcher and N. S. Jones. "hctsa: A computational framework for automated time-series phenotyping using massive feature extraction," Cell Systems, 5, 527, 2017
4. Facebook Infrastructure Data Science. "Kats," <https://facebookresearch.github.io/Kats/>
5. M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, "Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package)," Neurocomputing, 307, pp.72-77, 2018
6. M. Barandas, D. Folgado, Fernandes, L, et al., "TSFEL: Time Series Feature Extraction Library," SoftwareX 11, 2020.
7. Trent Henderson Ben D. Fulcher, "An Empirical Evaluation of Time-Series Feature Sets", IEEE 2021

Starting date of thesis: TBD

Ending date of thesis: TBD

University of Fribourg
Department of Informatics

Mourad Khayati
Senior researcher