UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

Department of Informatics

University of Fribourg
Department of Informatics
Bd. de Pérolles 90
CH-1700 Fribourg
Phone. +41 26 300 84 65
Fax +41 26 300 97 26
http://exascale.info/

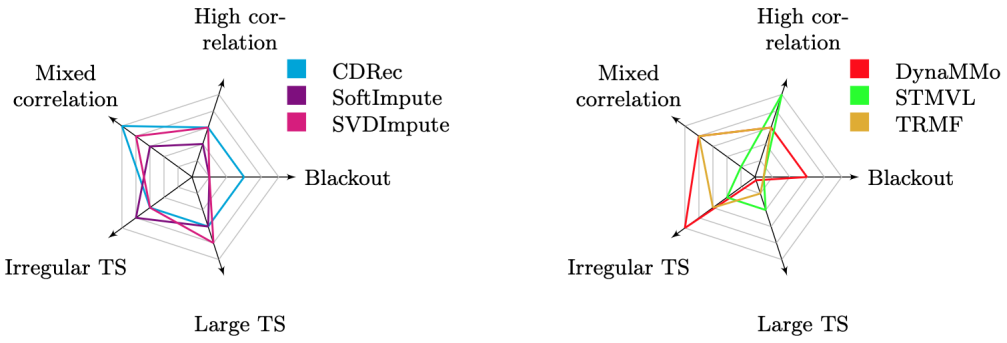
University of Fribourg, Dept. of Informatics, CH-1700 Fribourg

Dr. Mourad Khayati, Ines Arous
Senior researcher
Phone +41 26 300 84 60
Fax +41 26 300 97 26
mkhayati@exascale.info

Fribourg, Nov 20, 2020

A Configuration-Free Repair of Time Series.

Work overview: Real-world time series often contain missing values due to sensor failures, power outages and transmission problems. The recovery of these missing values allows better analysis of time series. Several methods have been proposed to recover missing values in time series, which can be matrix-based, pattern-based or machine learning-based technique [2]. Selecting "the best" recovery method highly depends on the dataset properties and often requires users to run multiple approaches with a different set of configuration parameters [3]. To perform these tasks, the user needs to acquire knowledge about the dataset domain and the recovery algorithms (cf. Figure 1).



(a) Efficient algorithms.

(b) Time consuming algorithms.

Figure 1. Algorithm selection based on time series features.

The aim of this thesis is to study and compare different ways to perform a recommendation of recovery techniques. The recommendation will relieve the user from the task of selecting and configuring recovery algorithms The thesis will focus on two classes: feature-based approach and parameter-based approach. The output of the thesis will be a solution that allows to select the most appropriate recovery technique in [3] in a systematic way.

Work tasks:

1. Familiarize yourself with recovery of missing values [1] and *ImputeBench* Benchmark [2,3]
2. Create clusters of time series having similar properties [4]
3. Feature-based recommendation
   - Extract the features of each cluster as a vector. The features include for example the correlation, the regularity and the length of the time series within a cluster of time series.
   - Apply a classifier that uses the features of the time-series to select the recovery algorithm based on Figure 1.
4. Parameter-based recommendation [5]
   - Generate a set of strategies from [3] i.e., all combinations of algorithms from *ImputeBench* with their corresponding set of parameters.
   - Sample from each cluster a sequence of the time series and evaluate the performance of all strategies on the sampled data. Assign the evaluation of a strategy on a cluster as its label and propagate this labeling through clusters.
   - Train a classifier that takes as input the feature vector of different clusters and selects the best strategy for recovery.
5. Evaluate the performance of the implemented techniques.
6. Write a thesis that describes the implemented technique and the result of the experiments.
7. Make a presentation of 20 minutes.

References:

1. P. García-Laencina, J. Sancho-Gómez and Aníbal Figueiras-Vidal *Pattern classification with missing data: a review*, **Neural Computing and Applications volume**, 2010.
2. Khayati, M., Lerner, A., Tymchenko, Z. and Cudré-Mauroux, P. *Mind the gap: an experimental evaluation of imputation of missing values techniques in time series.* **Proceedings of the VLDB Endowment 2020**
3. ImputeBench: Benchmark of Imputation Techniques in Time Series. Source code: `https://github.com/eXascaleInfolab/bench-vldb20.git`
4. Paparrizos, J. and Gravano, L., 2015, May. *k-shape: Efficient and accurate clustering of time series.* **In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data**.
5. Mahdavi, M., Abedjan, Z., Castro Fernandez, R., Madden, S., Ouzzani, M., Stonebraker, M. and Tang, N., 2019, June. *Raha: A configuration-free error detection system.* **In Proceedings of the 2019 International Conference on Management of Data**

Starting date of thesis: TBD

Ending date of thesis: TBD

University of Fribourg
Department of Informatics
eXascale Infolab

Mourad Khayati, Ines Arous