

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

Student Name

Switzerland

**Mourad Khayati**

Senior researcher  
Phone +41 26 300 84 60  
Fax +41 26 300 97 26  
[mkhayati@exascale.info](mailto:mkhayati@exascale.info)

Fribourg, 17. Dezember 2015

## Graphical Interface for Real Time Recovery of Missing Values

MSc Thesis:

Work overview:

The Centroid Decomposition (CD) [1], [2] is a matrix decomposition technique that has been successfully applied for the recovery of blocks of missing values in time series. It takes as input a set of correlated time series and reconstructs the type, the shape and the amplitude of the missing blocks by learning from the history of the time series that contains the missing blocks together with the history of other correlated time series. CD based recovery technique outperforms the state-of-the-art techniques, e.g., REBOM [3], for the recovery of blocks of missing values in shifted time series.

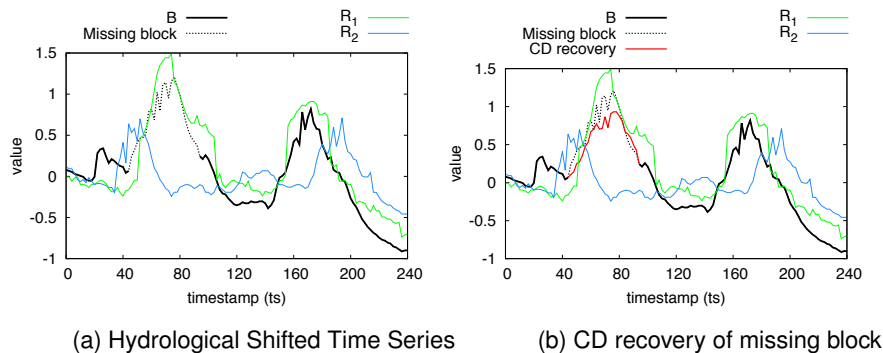


Figure 1: Example of CD based recovery on shifted time series

In Figure 1, CD technique is able to use the trends of both  $R_1$  (aligned with  $B$ ) and  $R_2$  (shifted with respect to  $B$ ) to accurately recover the missing block in  $B$ .

Formally, The Centroid Decomposition technique decomposes an  $n \times m$  matrix  $\mathbf{X} = [X_1 | \dots | X_m]$  into an  $n \times m$  *loading* matrix  $\mathbf{L} = [L_1 | \dots | L_m]$  and an  $m \times m$  *relevance* matrix  $\mathbf{R} = [R_1 | \dots | R_m]$  as follows:

$$CD(\mathbf{X}) = \mathbf{L}, \mathbf{R} \quad (1)$$

$$s.t. \quad \mathbf{X} = \mathbf{L} \times \mathbf{R}^T \quad (2)$$

$$= \sum_{i=1}^d L_i \times R_i^T$$

Where  $d \leq m$  is the number of time series to consider.

The aim of this thesis is to implement a graphical application that visualizes the recovery process of the Centroid Decomposition technique and its result. The tool should also graphically illustrate the steps of the computation of CD technique. Some of the features that the implemented tool should offer are:

- Browse aligned and not aligned time series with different granularities.
- Browse the missing blocks of values throughout the entire history of data.
- Ability to work with raw data, normalized data (Min-Max, Z-Score).
- Display globally and partially the similarities between time series.
- Display the result of the decomposition performed by CD technique.
- Visualize the real time CD based recovery on batch and streams of time series.

#### Work tasks:

1. Familiarize yourself with Centroid Decomposition (CD) algorithm.
2. Implement the scalable version of CD algorithm for the recovery of missing values.
3. Implement a graphical tool to visualize the result of CD based recovery.
4. Write a thesis that describes the algorithm and the tool.
5. Presentation of 20 minutes.

#### Literature:

1. Khayati, M., Böhlen, M.H., and Gamper, J. *Memory-efficient Centroid Decomposition for Long Time Series*, in ICDE, 2014.
2. Chu, M.T., and Funderlic, R.E., : *The Centroid Decomposition: Relationships Between Discrete Variational Decompositions and SVDs*, in SIAM J. Matrix Analysis and Applications, 2002
3. Khayati, M., and Böhlen, M.H.. *REBOM: Recovery of Blocks of Missing Values in Time Series*, in COMAD, 2012.

Starting date of thesis: TBD

Ending date of thesis: TBD

University of Fribourg  
Department of Informatics  
Exascale Infolab

Mourad Khayati  
Senior researcher