



UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zurich

Michal Koltonik
09-762-253
Onnhaldenstrasse 4
8302, Kloten
Switzerland

Prof. Dr. Michael Böhlen
Professor

Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, 7. April 2011

Msc project: Engine Tool for Recovering Missing Values

Description: The goal of this project is to develop a new online search tool that efficiently supports the recovery and the visualization of missing values in databases that store hydrological measurements. The recovery of missing values is based on using some replacement techniques for the approximated missing values. In order to assist the development of single missing values imputation strategies, basic techniques with their corresponding visualizations have to be implemented. The basic imputation techniques are mainly the Minimum imputation, the Mean imputation, the Nearest Neighbor imputation, the Linear Interpolation imputation and the Pairwise Deletion. We propose to compare the results of the single missing value imputation techniques and implement techniques able to approximate blocks of missing values. We propose to use two data decomposition methods which are the Singular Value Decomposition (SVD) and the Centroid Decomposition (CD) for the imputation of blocks of missing values. A prototype of the tool has to be implemented. This prototype will be implemented in Java and will offer the following features:

- Browse aligned and not aligned time series
- Browse the missing blocks of values throughout the entire history of data
- Ability to work with raw data, normalized data (Min-Max, Z-Score)
- Align time series having different time granularity
- Rank the alignments performed between time series with respect to different missing values imputation methods
- Alignment of subsequences of time series
- Compute the precision error of each implemented technique
- Native support for different operating systems via SWT libraries

- Build-in user guide

The first online version of the tool is implemented and described in http://www.ifi.uzh.ch/dbtg/teachingall/master_project/. A second version of the tool can be found in http://www.ifi.uzh.ch/dbtg/software/rmv_tp/

Tasks Description:

1. Implementation of the Singular Value Decomposition for the imputation of missing values
2. Implementation of the Centroid Decomposition which approximates SVD with low cost computation.
3. Empirical precision comparison between the two techniques
4. Write a report of max 10 pages in which you describe formally the implemented techniques.
5. Presentation of results 15 min

Task assignment and supervisor:

- Prof. Michael Böhlen
- Mourad Khayati

Starting date of project: 15/06/2010

Ending date of project: 11/09/2011

Report deadline: 18/09/2011

Presentation deadline: 27/09/2011.

University of Zurich
Department of Informatics

Prof. Dr. Michael Böhlen
Professor