



Scalable recovery of missing blocks in time series with high and low cross-correlations

Mourad Khayati¹ · Philippe Cudré-Mauroux¹ · Michael H. Böhlen²

Received: 26 November 2018 / Revised: 14 October 2019 / Accepted: 26 October 2019 /

Published online: 15 November 2019

© Springer-Verlag London Ltd., part of Springer Nature 2019

Abstract

Missing values are very common in real-world data including time-series data. Failures in power, communication or storage can leave occasional blocks of data missing in multiple series, affecting not only real-time monitoring but also compromising the quality of data analysis. Traditional recovery (imputation) techniques often leverage the correlation across time series to recover missing blocks in multiple series. These recovery techniques, however, assume high correlation and fall short in recovering missing blocks when the series exhibit variations in correlation. In this paper, we introduce a novel approach called CDRec to recover large missing blocks in time series with high and low correlations. CDRec relies on the centroid decomposition (CD) technique to recover multiple time series at a time. We also propose and analyze a new algorithm called Incremental Scalable Sign Vector to efficiently compute CD in long time series. We empirically evaluate the accuracy and the efficiency of our recovery technique on several real-world datasets that represent a broad range of applications. The results show that our recovery is orders of magnitude faster than the most accurate algorithm while producing superior results in terms of recovery.

Keywords Recovery of missing blocks · Time series · Centroid decomposition · Correlation

Mourad Khayati received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 732328 (FashionBrain). Philippe Cudré-Mauroux received funding from the European Research Council (ERC) under the European Union Horizon 2020 Research and Innovation Programme (Grant Agreement 683253/Graphint).

✉ Mourad Khayati
mkhayati@exascale.info; mourad.khayati@unifr.ch

Philippe Cudré-Mauroux
philippe.cudre-mauroux@unifr.ch

Michael H. Böhlen
boehlen@ifi.uzh.ch

¹ eXascale Infolab, University of Fribourg, Fribourg, Switzerland

² Department of Informatics, University of Zurich, Zurich, Switzerland