An Approach for Paralleling PLS Mode B Algorithm

Alba Martínez Ruiz[1], Cristina Montañola Sales[2]

[1]Universidad Católica de la Santísima Concepción, Chile

[2] Universitat Politècnica de Catalunya - Barcelona Supercomputing Center, España

Abstract

In this work, we first present the theoretical background of PLS Mode B path modelling, a multi-block method for estimating the relationships among several block of variables. This method is used for estimating structural equation models using a component-based perspective. Second, we introduce a set of guidelines to approach the parallelization process of the PLS algorithm in order to analyzed large data sets. Third, some key aspects of parallel computing to tackle a project of this nature are presented along with the R-project library pbdR that allows us to analyze large distributed data and implement methods for statistical analysis. Finally, we report some results of our research in progress related to the algorithm and its computational performance in analyzing large data sets. This work contributes to the understanding of methodological aspects of PLS methods for big data analysis.