MacroPARAFAC for a cellwise and rowwise robust PARAFAC analysis

Mia Hubert[1] and Mehdi Hirari[1]

[1]KU Leuven, Section of Statistics and Data Science, Department of Mathematics, Celestijnenlaan 200B, 3001 Leuven, Belgium

Keywords: Multi-way; cellmap; outliers

Multi-way data extend two-way matrices to a higher dimensional tensor. In many fields, it is relevant to pursue the analysis of such data by keeping it in its initial form without unfolding it into a matrix. Often multi-way data are explored by means of dimensional reduction techniques. Here, we study the Parallel factor analysis (PARAFAC) model, which expresses the multi-way data in a more compact way through a small set of loading matrices and scores.

The most common algorithm to fit this model is by means of an Alternating Least Squares (ALS) algorithm. However, it is well known that ALS is not robust to outliers. Robust alternatives which are resistant towards rowwise outliers have been proposed in the past [1, 2], the latter approach being able to cope with missing values as well.

These methods are however not resistant towards cellwise outliers that might contaminate all observations. A new algorithm is proposed that generalizes the MacroPCA method for two-way data [3] towards multi-way data. We show with simulations and the analysis of real data that our MacroPARAFAC method is robust to rowwise and cellwise outliers while also being able to handle missing elements.

[1] S. Engelen and M. Hubert, "Detecting outlying samples in a parallel factor analysis model," *Analytica Chimica Acta*, vol. 705, pp. 155–165, 2011.

[2] M. Hubert, J. Van Kerckhoven, and T. Verdonck, "Robust PARAFAC for incomplete data," *Journal of Chemometrics*, vol. 26, pp. 290–298, 2012.

[3] M. Hubert, P. Rousseeuw, and W. Van den Bossche, "MacroPCA: An all-in-one PCA method allowing for missing values as well as cellwise and rowwise outliers," *Technometrics*, vol. 61, no. 4, pp. 459–473, 2019.