Book of Abstracts

International Conference on Robust Statistics

ICORS 2022

5 - 8 July 2022

Department of Statistics & Actuarial Science

University of Waterloo, Canada

`https://uwaterloo.ca/international-conference-robust-statistics/`

Sponsor:

Faculty of Mathematics, University of Waterloo

An improved estimation procedure for robust PARAFAC model fitting

Valentin Todorov[1], V. Simonacci[2], M. Gallo[3] and N. Trendafilov[3]

[1] *United Nations Industrial Development Organization (UNIDO), Vienna, Austria; valentin@todorov.at*

[2] *University of Naples Federico II, Italy; violetta.simonacci@unina.it*

[3] *University of Naples-L'Orientale, Naples, 80134, Italy; mgallo@unior.it*

Different techniques exist to analyze multi-way data but PARAFAC is one of the most popular. The usual way of parameter estimation in PARAFAC is an alternating least squares (ALS) procedure which yields least-squares solutions and provides consistent outcomes. Together with these desirable features, the ALS procedure suffers several major flaws which might be particularly problematic for large-scale problems: slow convergence and sensitiveness to degeneracy conditions such as over-factoring, collinearity, bad initialization and local minima. Furthermore, it is well-known that algorithms which rely on least squares easily break down in the presence of outliers. The issue of non-robustness of the ALS procedure was addressed by [1] and software for computing the robust PARAFAC model is available in the R package **rrcov3way** (see [4]). The other issues were addressed in a number of works proposing algorithms more efficient than ALS. However, often these do not provide stable results because the increased speed might come at the expense of accuracy. An integrated algorithm was proposed by [2] and [3] which seems to combine improved speed and stability. The purpose of this work is to elaborate this algorithm by adding capabilities for dealing with outliers which might be present in the data.

The idea of a robust version of ALS is to identify enough "good" observations and to perform the classical ALS on these observations. This is repeated until no significant change is observed. Finally, a reweighting step is carried out to improve the efficiency of the estimators. In order to identify the "good" observations a robust version of principal component analysis on the unfolded array is used. It is obvious that the robust procedure will be much more time consuming than the classical one, repeating many times the ALS optimization. Therefore, any improvement of the performance of the parameter estimation procedure will contribute to the improvement of the performance of the complete robust procedure. We combine the robust PARAFAC procedure proposed by [1] with the highly efficient estimation algorithm INT2 proposed by [2] in order to obtain a fast estimation and robust to outliers PARAFAC modeling technique which we call R-INT2. As before it starts by robust principal components to identify any outlying points and then iterates using the INT2 algorithm until no significant change is observed. After convergence a reweighting step with INT2 is conducted which produces the final solution.

The performance of the newly proposed algorithm R-INT2 for robust estimation of trilinear PARAFAC models is demonstrated in a brief simulation study comparing classical ALS, the robust version based on ALS ([1]) and the new R-INT2. First of, all we want to verify that R-INT2 works well on data sets with and without contamination by identifying the outliers at least as good as the robust ALS algorithm retrieving

solutions with good statistical quality. At the same time we want to verify that the convergence is improved significantly and thus the computational time is reduced.

Future work should bring a thorough investigation of the properties of the algorithm, comparison to the many existing fast alternatives and studying the possibilities for combination with other computational algorithms.

Keywords: PARAFAC; ATLD-ALS; Robustness; Outliers; Computational efficiency.

References

[1] Engelen S, Hubert M (2011) Detecting outlying samples in a parallel factor analysis model. Analytica Chemica Acta 705:155–165

[2] Simonacci V, Gallo M (2020) An ATLD—ALS method for the trilinear decomposition of large third-order tensors. Soft Computing 18

[3] Simonacci V, Gallo M (2019) Improving PARAFAC-ALS estimates with a double optimization procedure. Chemometrics and Intelligent Laboratory Systems 192:103822

[4] Todorov V, Simonacci V, Di Palma M, Gallo M (2022) Robust Tools for Analysis of Three-way Compositional Data: The R package **rrcov3way**, *submitted*