DSSV 2023 Data Science, Statistics & Visualisation
ECDA 2023 European Conference on Data Analysis

Clusterwise Joint Independent Component Analysis: discovering disease related subtypes using multi-modal neuroimaging data

J. Durieux[1,2,3], S. A. R. B. Rombouts[2,3,4], and T. F. Wilderjans[2,4]

[1]Erasmus University, Erasmus School of Economics, Department of Econometrics, Rotterdam, The Netherlands
[2]Leiden University, Psychology Institute, Methodology and Statistics Unit, Leiden, The Netherlands
[3]Leiden Institute for Brain and Cognition, Leiden University, Leiden The Netherlands
[4]Leiden University Medical Center, Leiden, The Netherlands

In this digital era, more and more interesting research questions in different fields of science (e.g., psychology, neuroscience, genetics) call for the collection and joint analysis of big data from different sources or modalities. In neuroscience, for example, brain scans yield simultaneous information on both brain functioning (i.e., connectivity as measured by fMRI) and structure (e.g., cortical thickness) for a set of subjects, implying a two-modal data set.

To extract the mechanisms underlying each modality and investigate the covariation of mechanisms across modalities, Joint-Independent Component Analysis (Joint-ICA) was proposed [1]. Often, however, sample heterogeneity in these underlying mechanisms exists. For example, brain diseases, like dementia and depression, are known to be very heterogeneous across patients (groups). To capture this heterogeneity, subjects should be clustered based on similarities and differences in the multi-modal components underlying their coupled data. These estimated cluster-specific brain-based multi-modal components may function as a novel biomarker. This allows for a categorization of diseases that may complement (or challenge) existing diagnostic criteria that are based on clinical symptoms solely.

To obtain this, in this presentation, Clusterwise Joint-ICA, which combines (K-means type of) clustering with Joint-ICA, is proposed. In this model, subjects are clustered and the multi-modal components characterizing each cluster are estimated simultaneously, using an Alternating Least Squares (ALS) type of algorithm. The performance is evaluated by means of an extensive simulation study and by an illustrative application to coupled brain data from a group of Alzheimer's disease patients and elderly control subjects.

Keywords: Cluster analysis, ICA, neuroimaging data, multi-modal data

[1] V. D. Calhoun, T. Adali, N. Giuliani, J. Pekar, K. Kiehl, and G. Pearlson, "Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data," *Human Brain Mapping*, vol. 27, no. 1, pp. 47–62, 2006.