Multi-view learning with shared but delayed sources

Supervision:

- Alexandre Gramfort, Parietal, Inria (alexandre.gramfort@inria.fr)
- Pierre Ablin, CNRS, Lamsade, Univ. Paris-Dauphine (pierreablin@gmail.com)
- Bertrand Thirion, Parietal, Inria (bertrand.thirion@inria.fr)

Context Group studies involving large cohorts of subjects are important to draw general conclusions about brain functional organization. However, the aggregation of data coming from multiple subjects is challenging, since it requires accounting for large variability in anatomy, functional topography and stimulus response across individuals. Data modeling is especially hard for ecologically relevant conditions such as movie watching, where the experimental setup does not imply well-defined cognitive operations.

To tackle this statistical machine learning problem, different models have been proposed in the literature. In the Parietal team, we have recently developed a new model called MultiView Independent Component Analysis (ICA), where each subjects' data are modeled as a linear combination of common, shared independent sources plus noise [3, 2]. Contrary to most group-ICA procedures, the likelihood of the model is available in closed form and excellent mathematical guarantees can be obtained, including identifiability of the model [2]. So far, this model has been successfully applied on functional MRI and magnetoencephalography (MEG) data with hundreds of patients (see Figure 1). However, one limitation of such models is that the sources are supposed to be perfectly identical between subjects, hence so far ignoring the temporal variability in the neural responses of each subjects. The aim of this internship is to develop algorithms that take such temporal variabilities into account, so that they are no longer detrimental to the quality of the decomposition of the algorithm.

This work is part of a large effort in the machine learning community to be able to learn from data observed from multiple views [1].

Methods We have data coming from m subjects. The samples coming from subject i are denoted by $x_i \in \mathbb{R}^k$, and they are modeled as

$$x_i(t) = A_i(s(t - \tau_i) + b_i), i = 1, \dots, m \tag{1}$$

where t is the time. The matrices $A_i \in \mathbb{R}^{k \times k}$ are called mixing matrices, the vectors $s \in \mathbb{R}^k$ are the sources, τ_i is a time that corresponds to patient i response lag, and b_i is some residual noise. Here we only have access to the x_i, all other variables (A_i, s, τ_i and b_i) are to be estimated, using only an independence assumption on the sources s.

The aim of the internship will be to develop a fast algorithm for this inference problem, *e.g.,* using techniques likes alternate quasi-Newton method [3] or joint-diagonalization [2]. The theoretical identifiability of such model will also be studied.

The validation will be done on simulations and large public databases of MEG and electroencephalography (EEG) signals.

All code will be written in Python based on existing code developed in the team [1] [2] .

[1] https://github.com/hugorichard/multiviewica

[2] https://github.com/hugorichard/ShICA

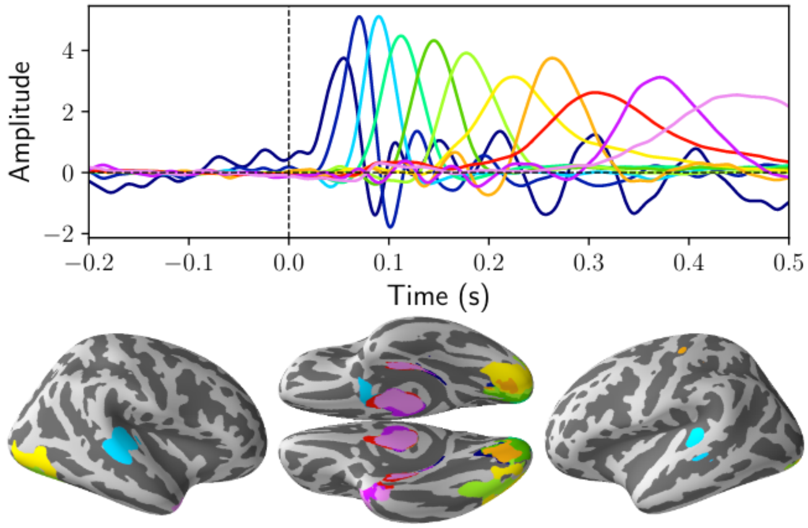

Figure 1: Multiview ICA applied on MEG data from the Cam-CAN public biobank.

Environment The internship will take place in Inria Saclay, in the Parietal team. This is a large team working focused on mathematical methods for statistical modeling of brain function using neuroimaging data (fMRI, MEG, EEG). Particular topics of interest include machine learning techniques, numerical and parallel optimization, applications to human cognitive neuroscience, and scientific software development.

Requirements

- Training in statistical machine learning at the master level.
- Undergraduate training in applied mathematics and/or computer science.
- Good programming skills in Python.
- Interest of machine learning for healthcare applications will be appreciated.

References

[1] Ronan Perry, Gavin Mischler, Richard Guo, Theodore Lee, Alexander Chang, Arman Koul, Cameron Franz, Hugo Richard, Iain Carmichael, Pierre Ablin, Alexandre Gramfort, and Joshua T. Vogelstein. mvlearn: Multiview machine learning in python. *Journal of Machine Learning Research*, 22(109):1–7, 2021.

[2] Hugo Richard, Pierre Ablin, Bertrand Thirion, Alexandre Gramfort, and Aapo Hyvärinen. Shared independent component analysis for multi-subject neuroimaging, 2021.

[3] Hugo Richard, Luigi Gresele, Aapo Hyvärinen, Bertrand Thirion, Alexandre Gramfort, and Pierre Ablin. Modeling shared responses in neuroimaging studies through multiview ica, 2020.