Scalable group level probabilistic sparse factor analysis

Many data-driven approaches exist to extract neural representations of functional magnetic resonance imaging (fMRI) data, but most of them lack a proper probabilistic formulation. We propose a scalable group level probabilistic sparse factor analysis (psFA) allowing spatially sparse maps, component pruning using automatic relevance determination (ARD) and subject specific heteroscedastic spatial noise modeling. For task-based and resting state fMRI, we show that the sparsity constraint gives rise to components similar to those obtained by group independent component analysis. The noise modeling shows that noise is reduced in areas typically associated with activation by the experimental design. The psFA model identifies sparse components and the probabilistic setting provides a natural way to handle parameter uncertainties. The variational Bayesian framework easily extends to more complex noise models than the presently considered.

Sparse Probabilistic Parallel Factor Analysis for the Modeling of PET and Task-fMRI Data

Modern datasets are often multiway in nature and can contain patterns common to a mode of the data (e.g. space, time, and subjects). Multiway decomposition such as parallel factor analysis (PARAFAC) take into account the intrinsic structure of the data, and sparse versions of these methods improve interpretability of the results. Here we propose a variational Bayesian parallel factor analysis (VB-PARAFAC) model and an extension with sparse priors (SP-PARAFAC). Notably, our
Archetypal Analysis for Modeling Multisubject fMRI Data

Functional magnetic resonance imaging (fMRI) is widely used to measure brain function during various cognitive states. However, it remains a challenge to obtain low-rank models of functional networks in fMRI that have interpretable latent features and generalize across groups of subjects, due to significant intersubject variability in the signal structure and noise. Group-level modeling is typically performed using component decompositions such as independent component analysis (ICA), which represent data as a linear combination of latent brain patterns, or using clustering models, where data are assumed to be generated by a set of ‘prototype’ time series. Archetypal analysis (AA) provides a promising alternative, combining the advantages of component-model flexibility with highly interpretable latent ‘archetypes’ (similar to cluster-model prototypes). To date, AA has not been applied to group-level fMRI; a major limitation is that it does not generalize to multi-subject datasets, which may have significant variations in blood oxygenation-level-dependent signal and heteroscedastic noise. We develop multi-subject AA (MS-AA), which accounts for group-level data by assuming that archetypal temporal profiles have a common latent generator across subjects, ensuring that the temporal components are derived from a consistent set of brain regions. In addition, the model accounts for noise heteroscedasticity by modeling subject-and voxel-specific noise variance. This provides a novel approach to group-level modeling and an alternative to preexisting methods that account for inter-subject variability by extracting individual maps as a postprocessing step (e.g., dual-regression ICA), or assuming spatial dependency of maps across subjects (e.g., independent vector analysis). MS-AA shows robust performance when modelling archetypes for a motor task experiment. The procedure extracts a ‘seed map’ across subjects, used to provide brain parcellations with subject-specific temporal profiles. Our approach thus decomposes multisubject fMRI data into distinct interpretable component archetypes that may help to model both consistent group-level measures of fMRI data and individual variability.
Variational group-PCA for intrinsic dimensionality determination in fMRI data

Functional Magnetic Resonance Imaging (fMRI) is widely used to gain a better understanding of the human brain's functional organization. As fMRI data are high dimensional it is challenging to analyze using conventional methods thus low-rank approximations such as principal component analysis (PCA), and independent component analysis (ICA) is often applied as a preprocessing step before any additional analysis. Low-rank methods generally require that the rank or latent dimensionality is known beforehand. When this is not the case a range of plausible dimensionalities have to be tested and compared. Furthermore, in an fMRI-context it is not fully understood how information from multiple subjects should best be incorporated when applying dimensionality reduction. We propose a Bayesian group principal component analysis (Group-BPCA) model with an automatic relevance determination (ARD) prior to determine the number of active components supported by the data. All subjects share the same spatial maps (components), but the uncertainties on these maps as well as the noise is subject specific. We find an approximate solution using the mature variational Bayesian framework and develop a fast and scalable implementation using a graphical processing unit (GPU). We test the model on fMRI data from 29 healthy subjects performing a block-design fingertapping experiment. The model identified 10 active components. Neither variational Bayesian PCA on temporally concatenated data nor Group-BPCA, where uncertainties on the spatial maps are shared, leads to pruning components, but provide better generalization in two of three scenarios. We show that the right level of subject variability is highly dependent on the chosen validation scheme.