What can Modern Bayesian Methods Offer Neuroimaging Data Analysis?

Gang Chen Co Organizer
NIH
Bethesda, MD
United States

Amanda Mejia Co Organizer
Indiana University
Department of Statistics
Bloomington, IN
United States

Martin Lindquist Organizer
Johns Hopkins Bloomberg School of Public Health
Baltimore, MD
United States

1812
Symposium

Bayesian methods provide a principled approach for statistical inference that combines prior information about an effect or parameter of interest with information obtained from the sampled data. Bayesian methods have been used to analyze neuroimaging data, but the high computational cost of most models have practically impeded its wider adoption. However, with both modeling and computational advances in recent years, it is time to take stock and ask what Bayesian modeling can offer neuroimaging data analysis. In this symposium, we showcase ongoing work from four experts in the field of Bayesian methods for neuroimaging. These experts are developing complementary approaches that share the unifying theme of utilizing Bayesian multilevel modeling as a regularization mechanism. Gang Chen adopts a global calibration approach using a multilevel model to leverage information across brain regions, with the goal of addressing multiplicity issues that have long plagued the field. Shuo Chen proposes an empirical Bayesian method that sidetracks the arbitrary selection of a primary threshold at the voxel level when performing cluster-extent based thresholding. Amanda Mejia presents a hierarchical Bayesian ICA framework for single-subject brain network organization that utilizes empirical population priors to obtain higher accuracy and computational efficiency. Seyedeh-Rezvan Farahibozorg introduces a stochastic variational Bayes framework to tackle the computational expense of performing resting state network modeling on large datasets. Together, the four talks will illustrate how Bayesian multilevel models can be used to: (i) effectively pool information shared across subjects, space, or time; (ii) address issues of multiplicity that have longed plagued the field; and (iii) efficiently study large datasets such as UK Biobank and Human Connectome Project.

Objective

controls family-wise error rate (FWER) caused by clusters consisting of false positive suprathreshold voxels. It has been well known that the selection of the primary threshold is critical because it determines both statistical power and false discovery rate. However, in most existing statistical packages the primary threshold is selected based on prior knowledge (e.g. p<0.001) without taking into account the information of data. In this current research, we propose a new empirical Bayes method that automatically and objectively determines the primary threshold by fully leveraging the rich information in the multivariate data. We show that our approach can significantly increase statistical power while effectively controlling the false discovery rate based on the results of simulations and data examples. Therefore, it can provide a flexible and objective alternative to select the primary threshold for clusterwise inference.

Presenter

Chen Shou, University of Maryland, Maryland Psychiatric Research Center Catonsville, MD United States

Template ICA: Leveraging big fMRI data through empirical population priors for accurate and fast estimation of subject-level brain networks

Using fMRI data to better understand the organization of the brains of individual subjects is of great scientific interest. Accurate, reliable measures of subject-level brain features would allow researchers to deepen understanding of disease, disorders, development and aging; build imaging-based biomarkers for disease classification; and facilitate translation to clinical care. Unfortunately, the low signal-to-noise ratio of fMRI data makes accurate estimation of brain features using standard statistical approaches difficult, except when very long scans are available. One way of improving estimation when the quantity or quality of data is limited is to "borrow strength" or pool information that is shared across, e.g., subjects, space or time, through Bayesian approaches. While the computational burden posed by some Bayesian models has previously been a barrier to their adoption in practice, the now widespread availability of large fMRI datasets provides an opportunity to establish empirical prior distributions. These empirical priors can be utilized in Bayesian models for computational efficiency. This approach leverages the wealth of information available through existing fMRI studies about the distribution of effects of interest. Here, we propose such an approach in the context of brain network estimation through ICA. Template ICA (tICA) is a hierarchical Bayesian ICA framework for single-subject brain network organization that utilizes empirical population priors or "templates". These empirical priors are spatially varying, providing more "shrinkage" toward the population in areas of the brain where subjects are similar and more flexibility in areas where they tend to differ. Template ICA produces subject-level component estimates that are matched to an existing set of group IC's, and it is applicable to new subjects not included in the original group ICA. In addition to the shared group networks represented in the template, tICA can also estimate subject-specific components representing additional brain networks or sources of noise, thereby avoiding contamination of the components of interest with subject-level noise. The tICA model can be estimated quickly (typically 5-10 minutes) using a two-stage expectation-maximization (EM) algorithm. Through simulation studies and a reliability study based on data from the Human Connectome Project, we find that tICA substantially outperforms dual regression in estimating subject-level RSNs. One limitation of tICA is that it assumes that subject effects are spatially independent. While computationally advantageous, this approach may not have optimal power and efficiency. I will present an extension to the tICA framework, spatial template ICA (stICA), which uses spatial priors to encourage smoothness of the subject effects. Template ICA is implemented in a MATLAB toolbox and an R package.

Presenter

Amanda Mejia, Indiana University
Department of Statistics
Bloomington, IN
United States

Hierarchical Bayesian modelling of individual- and population-level resting state networks from big fMRI data

It is well established that there are structured network fluctuations in fMRI when subjects are at rest. However, characterising this activity, and its variability over subjects, in an interpretable manner from very large population datasets (e.g. up to 100,000 subjects in UK Biobank), is an open problem. Recent advances include modelling subject-specific resting state networks (RSNs) through mapping subject data onto a predefined group model. However, a more accurate estimation of between-subject variability, e.g. due to anatomical misalignments across subjects, requires hierarchical models that provide both bottom-up regularisation of the subject-specific networks, and top-down effect of a flexible group-level model that captures the key aspects of cross-subject variability. These can then yield estimation of functional coupling among RSNs that is less confounded by spatial variations, which is particularly important when studying heterogeneous groups such as clinical populations, developmental or aging cohorts. The framework of PRObabilistic FUnctional MOdes (PROFUMO) accomplishes this goal by hierarchically modelling sets of modes with spatial differences between subjects. However, the computational expense of the algorithm in its current form makes its application to large population datasets impractical. We therefore propose an extension to PROFUMO using stochastic variational Bayes (sPROFUMO) in order to obtain, for the first time, a hierarchical model of RSNs that can readily be applied to datasets like UK Biobank. sPROFUMO utilises an iterative process in order to obtain a noisy update of the group-level model based on a small subsample of subjects in each iteration. The stochasticity of the model can be expected to improve its ability to account for subject-specific deviations from the group, while still yielding a representative group model. We test sPROFUMO's performance in realistic simulation scenarios with one thousand subjects and show its robustness to the choice of specific parameters. We further demonstrate that it yields accurate estimations of functional modes in scenarios that are deemed challenging for spatial ICA (e.g. in the presence of spatial mode overlap and anatomical misalignment across subjects). We finally show the utility of applying sPROFUMO to large-scale datasets, namely the Human Connectome Project and UK Biobank. Applying sPROFUMO to rich epidemiological datasets such as UK Biobank can be expected to improve our understanding of the individualised profiles of the brain function and disorders.

Presenter

Seyedeh-Rezvan Farahibozorg, Oxford University
WIN (FMRIB)
Oxford
United Kingdom