Multi-view methods for imaging genetics

Jonas Richiardi Co Organizer
Lausanne University Hospital
Lausanne, Vaud
Switzerland

Andre Altmann Organizer
University College London
London, NA
United Kingdom

2132
Symposium

Large-scale imaging genetics data is becoming increasingly available freely, both for clinical studies and large-scale national cohorts. With this, there has been growing interest in techniques beyond traditional mass-univariate analysis (i.e., genome-wide association studies; GWAS) to uncover gene-function and gene-diseases associations using the brain as a complex trait. Thus far, striking results have been achieved concerning brain morphology, resting-state networks, and neuropsychiatric disorders. In this symposium, the goal will be to provide a self-contained overview of the techniques necessary to perform such analyses, and the results that they can yield.

The symposium will start with an introduction to multi-view methods including PLS, CCA, and RRR, together with a new toolbox. The second talk will introduce classical imaging genetics approaches and recent results in the field. The third and fourth talk will focus on modern multi-view methods for imaging genetics together with experimental results. The format is designed to enable non-methods specialists and/or non-genetics specialists to follow along, and to extend the learnings of this session to other types of multimodal data, including imaging + neuropsychological assessment scores or other clinical scores.

Objective

Illustrate multi-view methods for analyzing imaging data + X, including available toolboxes

Outline imaging genetics and explain classic imaging genetics methods

Demonstrate the application of multi-view methods in imaging genetics

Target Audience

This symposium has two key target audiences:
1) Firstly, we target non-genetics specialists that are interested in starting to work on imaging genetics.
2) Secondly, we target non-methods specialists that are interested in learning about Multiview methods to analyze their data - as we will show in the symposium these techniques have broad applicability).

Presentations

Introduction to regularized canonical correlation analysis (RCCA): a unifying approach to perform CCA/PLS analysis

Recently, there has been a growing interest in using multivariate approaches such as Canonical Correlation Analysis (CCA) and Partial Least Squares (PLS) to investigate associations between multiple views/sources of data e.g. brain imaging and genetics. To control the complexity of the CCA model and the risk of overfitting, a regularized version of CCA (RCCA) was proposed by Hardoon et al. (2004). Interestingly, PLS and CCA can be also seen as specific cases of RCCA. In this talk, I will present RCCA and its dual formulation, kernel CCA (KCCA) which improves the computational efficiency of the algorithm. Moreover, I will demonstrate that the iterative solution of these models enables to select the optimal regularization parameter for each associative effect, and to specify orthogonality constraints for the subsequent effects e.g. resulting in different PLS variants (Mihalik et al. 2019). Finally, I will present a recently developed toolkit which includes CCA, PLS, KCCA as well as other algorithms (e.g., sparse PLS) that can be easily run using MATLAB and automatically selects the regularization parameters as well as evaluates the statistical significance of the results. References: Hardoon, D.R., Szedmak, S. & Shawe-Taylor, J. (2004) Canonical correlation analysis: An overview with application to learning methods. Neural Comput. 16, 2639–2664 Mihalik, A., ..., Shawe-Taylor, J., Dolan, R., Mourao-Miranda J. (2019) Multiple holdouts with stability: improving the generalizability of machine learning analyses of brain-behaviour relationships. Biol. Psychiatry, DOI: 10.1016/j.biopsych.2019.12.001

Presenter

Agoston Mihalik, University College London London, NA
United Kingdom

Functional insights from univariate genetic associations with brain structure -- a benchmark for advanced multivariate analyses

Recent large-scale efforts from efforts including the UK Biobank (Elliot et al, 2018), the CHARGE consortium (Satizabal et al, 2019) and the ENIGMA Consortium (Grasby et al biorxiv 2019), have pooled tens of thousands of datasets with brain imaging and genome-wide genetic information to test the association between individual nucleotide variations in the genome and variation in brain structure. A surge in data collection and availability has opened the doors for more extensive investigations into the functional significance of these findings, in relation to human behavior and disease. We describe efforts to 1) more precisely map individual genomic associations, including harmonizing voxelwise efforts, and parallel efforts to create more detailed maps of localized functional effects; 2) aggregate genetic association information across genomic loci using polygenic risk assessments, to predict brain structure in independent individuals, and aggregate information across brain regions to identify overlapping patterns of genetic risk with disease; 3) identify genetic connectivity patterns in the brain related to functionally annotated genetic pathways. Furthermore, we showcase how such univariate findings can be used as benchmarks to test the validity of multivariate methods aimed at boosting power to discover genetic loci associated with brain structure. References: Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature 562, 210–216 (2018). Grasby, K.L. et al.. The genetic architecture of the human

cerebral cortex. BioRxiv (2019) Satizabal, C. L. et al. Genetic architecture of subcortical brain structures in 38,851 individuals. Nat Genet 51, 1624–1636 (2019).

Presenter

Neda Jahanshad, University of Southern California Marina del Rey, CA
United States

ICA-based multimodal data mining and its application on imaging genetics

Independent component analyses (ICA) is a data-driven technique to extract hidden factors within a dataset. Based on this technique, hidden relations among multiple datasets from diverse modalities can be uncovered and explored. As example, parallel ICA is designed to simultaneously extract data-modality specific hidden factors from two datasets and inter-modality correlations among the independent factors. Specific features for each data modality, such as genetic regions, brain ROIs, or super sparsity, can be incorporated into ICA decomposition, in order to prioritize and refine the hidden factors, as implemented in the parallel ICA with reference and sparse parallel ICA algorithms. For many studies, more than two modalities are of great interest, such as brain structure, brain function, genetic, epigenetic, and cognition. Methods have been developed as a N-way parallel ICA to accommodate such applications. Examples of data mining on gray matter networks, fMRI network, EEG components, SNP factors, DNA methylation, etc. will be presented to illustrate the functions of ICA based multimodal methods.

Presenter

Jingyu Liu, Georgia State University Atlanta, GA
United States

Latent variable models in imaging-genetics for multi-variate, multi-view and multi-centric analyses

During this talk, we will introduce multi-variate latent variable models for jointly analyzing the relationship between high-dimensional genetic information and brain imaging data (Lorenzi et al., 2018). We will extend classical multivariate analysis approaches to jointly account for data sources represented by multiple information channels (a.k.a. multi-view datasets). In particular, we will introduce the "multi-channel variational auto-encoder" (McVAE), a scalable (deep) Bayesian association method for inference of latent relationships across multiple data channels, such as clinical, genetic, and multi-modal imaging information (Antelmi et al., 2019). Finally, we will extend current approaches to cope with the constraint of data privacy and security typical of multi-centric studies. This problem is here tackled by extending current latent variable models (as well as their Bayesian counterparts) to the federated learning setting (Silva et al., 2019). References: Luigi Antelmi, Nicholas Ayache, Philippe Robert and Marco Lorenzi. Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data. Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, 2019. Marco Lorenzi, Andre Altmann, Boris Gutman, Selina Wray, Charles Arber, Derrek P. Hibar, Neda Jahanshad, Jonathan M. Schott, Daniel C. Alexander, Paul M. Thompson and Sebastien Ourselin. Susceptibility of brain atrophy to TRIB3 in Alzheimer's disease: Evidence from functional prioritization in imaging genetics. Proceedings of the National Academy of Sciences of the United States of America (PNAS). March 20, 2018. 115 (12) 3162-3167. Santiago Silva, Boris Gutman, Barbara Bardoni, Paul M Thompson, Andre Altmann, Marco Lorenzi. Multivariate Learning in Distributed Biomedical Databases: Meta-analysis of Large-scale Brain Imaging Data. IEEE International

Symposium on Biomedical Imaging (ISBI), Venice, 2019.

Presenter

Marco Lorenzi, Université Côte d'Azur Sophia Antipolis, NA France