Overview

- Motivation
- Joint ICA
  - ERP/fMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
  - Epigenetics
- Conclusions

Multimodal Data Collection

Brain Function (spatiotemporal, task-related)
- EEG
- fMRI
- Task 1-N

Brain Structure (spatial)
- T1/T2
- DTI

Genetic Data (spatial/chromosomal)
- SNP
- Gene Expression

Covariates (Age, etc.)

A Role For Data Fusion?:
Possible approaches for joint analyses

- **Voxel-based**
  - Correlation [Worsley 1998]
  - Straightforward, but difficult to visualize
- **Region-based**
  - Interregional correlation [Horwitz, et al. 1984]
  - Structural equation modeling [McIntosh and Gonzalez-Lima 1994; Friston et al., 2003; McIntosh & Gonzalez-Lima, 1991; Buchel & Friston, 1997]
  - Multiple regression and extensions [e.g., Kalman filters, Buchel & Friston, 1998]
  - Useful for model testing, does not take into account all brain regions
- **Transformation-based**
  - A natural set of tools for this problem include those that transform data matrices into a smaller set of modes or components
  - Singular value decomposition [Friston et al., 1993; Friston et al., 1996]
  - Partial Least Squares [McIntosh, Bookstein, et al, 1996]
  - Canonical Variates Analysis [Strother et al, 1995]

Why Joint/Multimodal?

In a non-joint analysis, we maximize the likelihood functions for each modality separately...

Resulting in two unmixing parameters, that then have to somehow be fused together

In contrast: for a joint analysis we maximize the joint likelihood function, resulting in a single fused unmixing parameter

Why Multivariate?

Why Independence?

Uncorrelated: $E\{y_1 y_2\} = E\{y_1\} E\{y_2\}$

Independent: $p(y_1, y_2) = p(y_1) p(y_2)$

$\Rightarrow E\{h(y_1) h(y_2)\} = E\{h(y_1)\} E\{h(y_2)\}$

PCA finds directions of maximal variance (using second order statistics)

ICA finds directions which maximize independence (using higher order statistics)
Why Features?

• What is a feature?
  • Lower dimensional data containing information of interest
  • Examples: An image of activation amplitudes, A gray matter segmentation image, fractional anisotropy image

• Advantages
  • Less-computationally complex/easier to model
  • Takes advantages of existing analytic approaches
  • Can be used to examine inter-relationships between multiple data types at the subject level

fMRI

sMRI

Feature-based ICA

A Family of Multivariate Methods

ICA algorithms are based on cost functions which utilize the higher order statistical information

Explicit computation of higher-order statistics
  • Cumulants — e.g., JADE (Cardoso and Soulloumiac, 1993)
  • Kurtosis — e.g., FastICA (Hyvärinen 1999)

Implicit computation of higher-order statistics
  • Mutual information
  • Maximum likelihood/Information maximization — e.g., Infomax (Bell and Sejnowski, 1995)
  • Nonlinear decorrelations
  • Maximization of non-Gaussianity — e.g., FastICA

and computed using matrix diagonalizations, or iteratively using natural/relative gradient, fixed-point, and Newton variate updates
Joint ICA

\[
\begin{bmatrix}
X_{Task1} \\
X_{Task2}
\end{bmatrix} = A \times \begin{bmatrix}
S_{Task1} \\
S_{Task2}
\end{bmatrix}
\]

Generative Model:
\[
\begin{bmatrix}
\hat{x}^1 \\
\hat{x}^2
\end{bmatrix} = A \begin{bmatrix}
x^{(1)} \\
x^{(2)}
\end{bmatrix}
\]

Update Equation:
\[
\Delta W = \eta \left( 1 - 2y^{(1)}(u^{(1)})^T - 2y^{(2)}(u^{(2)})^T \right) W
\]

Overview

- Motivation
- Joint ICA
  - ERP/fMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fmRI
  - N-way/CASH
- Parallel ICA
  - SNPs
  - Epigenetics
- Conclusions

Joint ERP/fMRI Components

**FMRI Snapshots (movie)**

ERP (temporal) Components: \( T = \begin{bmatrix} t_1 & K & t_N \end{bmatrix} \)

FMRI (spatial) Components: \( S = \begin{bmatrix} s_1 & K & s_N \end{bmatrix} \)

ERP Timecourse Snapshot: \( M_2(v) = T \times S^T(v) \)

---

**Linked EEG/fMRI Results**


---

**Concurrent EEG/fMRI: eyes open vs eyes closed**

These results suggest that changes in neuronal synchronization as indicated by power fluctuations in high-frequency (>1Hz) EEG rhythms such as posterior alpha are partly mediated by widespread changes in inter-regional low-frequency (<.1Hz) functional activities detected in fMRI. They also indicate that generation of local hemodynamic responses is highly sensitive to global state changes that do not involve changes of mental effort or awareness.
Overview

- Motivation
- Joint ICA
  - ERP/fMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
  - Epigenetics
- Conclusions

Preprocessing

Timing/motion correction
Spatial normalization
Spatial smoothing

Feature Extraction

GLM Analysis

AOD
SB-WM

Joint Feature Matrix

Component Selection

Test loading parameters (patients differ from controls)

Joint Histograms: fMRI

Evaluating Multiple Features

- Auditory Oddball fMRI Task
  - Target-related activity
  - Novel-related activity
- Sternberg fMRI Task
  - Recognition-related activity
- MPRAGE
  - Gray matter segmentation


Identifying the “best” combination

- Kullback-Leibler divergence
  \[
  D(s \| u) = \int p_s(\xi) \ln \left( \frac{p_s(\xi)}{p_u(\xi)} \right) d\xi
  \]

s, u ∈ \mathbb{R}^2 ∈ \{F_{SB}, F_{AOD_N}, F_{AOD_T}, S_{GM}\}


Optimal Selection of Discriminative Features

\[
C = -E[\ln f(y)] + \lambda \cdot \sum T_i^2
\]


Joint/Single GLM/ICA Imaging Biomarkers

Overview

- Motivation
- Joint ICA
  - ERP/fMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
  - Epigenetics
- Conclusions

Multi-set Canonical Correlation Analysis

Three-way MCCA Results….

<table>
<thead>
<tr>
<th>Modalities</th>
<th>three modalities</th>
<th>two modalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>fMRI</td>
<td>3.4526, 0.0019</td>
<td>2.1650, 0.0375</td>
</tr>
<tr>
<td>sMRI</td>
<td>2.8604, 0.0011</td>
<td>2.2049, 0.0343</td>
</tr>
<tr>
<td>EEG</td>
<td>3.4079, 0.007</td>
<td>—</td>
</tr>
</tbody>
</table>

Overview

- Motivation
- Joint ICA
  - ERP/fMRI
  - Multitask-fMRI
  - sMRI/fMRI
- CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
  - Epigenetics
- Conclusions
**mCCA and jICA are complementary approaches**

- **Key differences in assumptions:**
  - **mCCA**
    - Modulation profiles: separate uncorrelated
    - Sources: independent
  - **jICA**
    - Modulation profiles: identical
    - Sources: correlated

The connections discovered are based on:
- **mCCA** correlation of modulation profiles
- **jICA** assumption of shared modulation profiles and independence among sources

**By Combining the two approaches we can:**
- Identify both feature-common and feature-distinct connection
- Make full use of the cross-information of two data sets, in order to separate sources more accurately.
- Use mCCA to link features in the two datasets, then use jICA to decomposes the spatial maps
- Addresses limitation of both mCCA & jICA

**Comparison of jICA, mCCA, & mCCA+jICA**

**mCCA+jICA**

**Two-way fusion of fMRI/DTI in BP & SZ**

**Two modalities**
- **FMRI**: Auditory oddball (target) contrast maps (AOD)
- **DTI**: Fractional anisotropy (FA)

**Three groups including participants**
- 54 schizophrenia (SZ, age 37±12, 22 females)
- 48 bipolar disorder (BP, age 37±14, 26 females)
- 62 healthy controls (HC, age 38±17, 30 females)

**mean FA maps**
**mean AOD maps**
Results from Schizophrenia, Bipolar, and Healthy Individuals

If the two ICs have the same frame color in two modalities, they are joint ICs.


Variation of functional spatial maps

Back-reconstruction used to estimate group specific maps.


Extension of mCCA+jICA to N-way fusion

Correlation of mixing matrices after multi-set CCA.

### Group Differences HC vs SZ

<table>
<thead>
<tr>
<th>Group</th>
<th>Number (sex)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>116 (44 female)</td>
</tr>
<tr>
<td>SZ</td>
<td>97 (24 female)</td>
</tr>
</tbody>
</table>

Imaging: fMRI, sMRI, DTI
Site effects regressed out

<table>
<thead>
<tr>
<th>IC No.</th>
<th>fMRI-DTI</th>
<th>fMRI-sMRI</th>
<th>DTI-sMRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.195</td>
<td>0.251</td>
<td>0.407</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.085</td>
<td>0.135</td>
</tr>
<tr>
<td>3</td>
<td>0.069</td>
<td>0.097</td>
<td>0.038</td>
</tr>
<tr>
<td>4</td>
<td>0.023</td>
<td>0.094</td>
<td>0.079</td>
</tr>
<tr>
<td>5</td>
<td>0.025</td>
<td>0.002</td>
<td>0.101</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
<td>0.11</td>
<td>0.247</td>
</tr>
<tr>
<td>7</td>
<td>0.116</td>
<td>0.065</td>
<td>0.188</td>
</tr>
<tr>
<td>8</td>
<td>0.135</td>
<td>0.003</td>
<td>0.093</td>
</tr>
<tr>
<td>9</td>
<td>0.078</td>
<td>0.101</td>
<td>0.494</td>
</tr>
</tbody>
</table>

### Simulation Framework for Data Fusion

- Copulas: a form of bivariate (or multivariate) distributions with uniformly distributed marginals
- Sampling from a bivariate copula leads to two sets of values (each uniformly distributed) which are associated in the way described by the specific copula
- Using the probability integral (or inverse cumulative distribution, icdf) we can assign any desired marginal distribution

### MIALAB

Medical Image Analysis Lab

http://mialab.mrn.org

### Overview

- Motivation
- Joint ICA
  - ERP/fMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI/fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
  - Epigenetics
- Conclusions
Genetic Information

- Genetic: single nucleotide polymorphism (SNP)
- Genetic: copy number variation (CNV)
- Epigenetic: methylation

Simulation

Simulation: Designed to provide a more complete understanding of Parallel ICA while applied to genomic SNP array studies. We specified the parameters for each component and input them into PLINK, an open-source whole genome association analysis toolset [http://pngu.mgh.harvard.edu/purcell/plink/].

Conditions: sample size effect, case to control ratio, SNP array size effect, case-related SNP’s vs. total SNP’s, odds ratio, connection strength between genotype and phenotype effects

Initial Proof of Concept: SNP/fMRI Fusion

Data Description: 20 Sz & 43 Healthy controls
fMRI: one image per subject (Target activation in AOD task)
SNP: one array per subject (384 SNP genotypes -> 367 SNPs)

Control vs Patient
p<0.001

Parallel ICA: Two Goals

- Identify Hidden Features
- Identify Linked Features


Simulation results suggest that parallel ICA, in general, is able to extract more accurately the components and connections than a correlation test, in particular for weak linkages. Results also indicate that the ratio of sample size to SNP size should be at least 0.02. However, when the data have a low odds ratio or cases vs. controls ratio, the correlation test provides results reliably, though with lower accuracy.

Larger scale study of schizophrenia

- Schizophrenia patients and healthy controls
  - MCIC data: Boston, Iowa, Minnesota and New Mexico
  - Genome-wide 1M SNP data - [biallelic coding (AA, AB, or BB)]
  - fMRI sensorimotor task- Block design motor response to auditory stimulation
- SNP data
  - Subject control: heterozygosity, nearest neighbor, 2nd degree or closer relatives, duplication
  - SNP control: missing genotyping ratio, minor allele frequency, Hardy-Weinberg equilibrium, linkage disequilibrium, etc.
  - Population stratification correction: using PCA
- Coding (0 for ‘AA’, 1 for ‘AB’, and 2 for ‘BB’)

fMRI data

- SPM preprocessing (alignment, normalization, filter, GLM) and contrast image
- Outlier subject excluded

Datasets: 208 subjects with SNP (777365 SNPs) and fMRI data (52322 voxels)


Resulting Linked Component

- fMRI component number = 8, SNP component number = 5
- One pair of linked components is identified, with p-value passing Bonferroni correction

<table>
<thead>
<tr>
<th>fMRI component index</th>
<th>SNP component index</th>
<th>r_{fMRI-SNP}</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>-0.065</td>
<td>3.49E-01</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.042</td>
<td>5.48E-01</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.099</td>
<td>1.51E-01</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-0.138</td>
<td>4.54E-02</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.141</td>
<td>4.16E-02</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>-0.128</td>
<td>6.44E-02</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.178</td>
<td>9.95E-03</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0.282</td>
<td>3.39E-05</td>
</tr>
</tbody>
</table>

Bootstrap: Multiple runs of parallel-ICA with 5157 randomly selected SNPs. The median correlation was 0.16

Summary of SNP Results

- Conduct pathway analysis and functional annotation clustering based on identified 94 genes
  - IPA (Ingenuity Pathway Analysis) identifies “Schizophrenia of humans” as one of the top biofunctions, involving 11 genes
  - IPA also identifies a number of significant canonical pathways, four of which are related to neurotransmitter signaling
  - David’s Bioinformatics Resource reports the most significant cluster to be functionally related to synapse. A cluster annotated as “cell projection” is also identified

<table>
<thead>
<tr>
<th>Disease and disorder</th>
<th>Gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia of humans</td>
<td>DRD1, COMT, GRIN2B, DISC1, GAD2, GABA, NR3B2, NR3C1, NR3C2, NR7</td>
<td>6.49E-09</td>
</tr>
<tr>
<td>Neurotransmitter signaling pathway</td>
<td>COMT, DRD1, GRIN2B, DISC1, GAD2</td>
<td>6.49E-09</td>
</tr>
<tr>
<td>GABA receptor signaling</td>
<td>GABA, GABRA1, GAD</td>
<td>2.13E-03</td>
</tr>
<tr>
<td>Dopamine receptor signaling</td>
<td>COMT, DRD1, PPXY</td>
<td>1.15E-02</td>
</tr>
<tr>
<td>Serotonin signaling</td>
<td>NR3B2, NR3C1, ERBB4, AR</td>
<td>1.86E-02</td>
</tr>
<tr>
<td>Glutamate receptor signaling</td>
<td>GRIN2B, GRIK2</td>
<td>3.04E-02</td>
</tr>
<tr>
<td>Functional annotation cluster</td>
<td>DRD1, GABA, GAD, DRD2, DISC1, NR3B1, NR3C1, NR7</td>
<td>3.79E-05</td>
</tr>
<tr>
<td>Synapse</td>
<td>DRD1, GABA, GAD, NR3B1, NR3C1, NR7, PON1, BACE1, NTG</td>
<td>3.04E-02</td>
</tr>
<tr>
<td>Cell projection</td>
<td>DRD1, GAD1, GABA, DRD2, NR3B1, NR3B2, NR3C1, NR7, PON1, BACE1, NTG</td>
<td>3.79E-05</td>
</tr>
</tbody>
</table>

sMRI/SNP

Structural deficits in brain regions consistently implicated in previous schizophrenia reports, including frontal and temporal lobes and thalamus were related to SNPs from 16 genes, several previously associated with schizophrenia risk and/or involved in normal CNS development, including AKT, PI3K, SLC6A4, DRD2, CHRM2 and ADORA2A.

A. sMRI component –A (group difference)

B. sMRI component –B (linked, but no group difference)


Genetics and P3 ERP generation

- Subjects: 41 healthy subjects (24 female, 17 male)
- EEG collected during AOD task, target/novel ERPs extracted
- Blood sample collected, genotyped 384 SNPs from 222 genes 6 physiological systems.

ERP Topography & SNP Associations

Target Stimuli

Novel Stimuli

SNPs | Genes
---|---
rs1800545 | ADRA2A
rs7412 | APOE
rs1128503 | ABCB1
rs6578993 | TH
rs1045642 | ABCB1
rs2278718 | MDH1
rs4784642 | GNAO1
rs521674 | ADRA2A

SNPs | Genes
---|---
rs1800545 | ADRA2A
rs7412 | APOE
rs6578993 | TH
rs1045642 | ABCB1
rs2278718 | MDH1
rs1128503 | PIK3C3
rs284389 | PIK3C3
rs3813065 | PIK3C3
rs4121817 | PIK3C3
rs521674 | ADRA2A

Pathway Analysis


### Classification with SNP & fMRI

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of Training Subjects</th>
<th>Number of Testing Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>4</td>
</tr>
</tbody>
</table>

**SVM Classification**


### Overview

- **Motivation**
- **Joint ICA**
  - ERP/fMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- **mCCA**
  - sMRI/fMRI/ERP
  - mCCA+jICA
    - DTI-fMRI
    - N-way/CASH
- **Parallel ICA**
  - SNPs
  - Epigenetics
- **Conclusions**

### Methylation Sex Correction

**Sex difference:**


### Methylation Sex Correction

- Genomic ~27,000 sites from 23 pairs of chromosomes
- 130 subjects (heavy drinker 33 females, 97 males, age 31.3 ± 9.7)

**Goal:** Association with gender, age, BMI, alcohol use, cigarette use, marijuana use, depression, stress, etc.

**Results:**

Ancestry effect on methylation


CNV and substance abuse

- Effect of double deletion at 22q12.3-13.1 on alcohol use disorder severity and cue-elicited BOLD response in the precuneus


Software

http://mialab.mrn.org/software

5000+ unique downloads
Funded by: 1R01EB006841

1000+ unique downloads
Funded by: R01EB005846

http://mialab.mrn.org/software