GIFT Course
Lecture 1: overview: independent component analysis, ICA of fMRI, SPM/Matlab preliminaries,

Vince D. Calhoun, Ph.D.
Director, Medical Image Analysis Laboratory
Otto Neuropsychiatry Research Center,
Institute of Living
Assistant Clinical Professor, Psychiatry
Yale University School of Medicine

Overview

- CDROM
  - Lot’s of ICA of FMRI and ICA articles
  - GIFT Software
- Binder
  - CD Contents
  - Course notes
  - GIFT Walkthrough
- ICA books
  - Hyvarinen ICA Book
  - Stone ICA Book
Lecture 1

- What is ICA
- A few algorithms (we’ll do the equations early!)
- ICA of fMRI overview

Intro to ICA

- The long standing problem: How to find a suitable representation of multivariate data?
- The general formulation: What could be a function from an $k$-dimensional space to an $n$-dimensional space such that the transformed variables give information about the data that is otherwise hidden in the large data set?
- Let’s consider linear functions only (to simplify things), so we want to find coefficients $w$ such that:

$$u_i(v) = \sum_j w_{ij} x_j(v), \text{ for } i = 1, \ldots, N, j = 1, \ldots, K$$
Intro to ICA

- In matrix notation:

\[
\begin{pmatrix}
  u_1(v) \\
  \vdots \\
  u_N(v)
\end{pmatrix} = W
\begin{pmatrix}
  x_1(v) \\
  \vdots \\
  x_K(v)
\end{pmatrix}
\]

\[ u = Wx \]

- What principle do we use to find $W$?
  - Dimension reduction (PCA)
  - Independence
    - Factor Analysis (assumes data is gaussian...in this case uncorrelatedness→independence)
    - Independent Component Analysis (does not assume data is gaussian and directly attempts to determine independence of data)→model can be uniquely estimated using supplemental information not contained in the covariance matrix

---

Intro to ICA

- What about non-gaussianity assumption?
  - Many real-world data sets have \textit{supergaussian} distributions
    - The random variables take relatively more often values that are very close to zero or are large
      - Sound recordings
      - fMRI activations
ICA vs. PCA

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

ICA vs. PCA

ICA finds directions which maximize independence
(using higher order statistics)
ICA vs. PCA

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

ICA finds directions which maximize independence
(using higher order statistics)
Blind Source Separation

- We observe the signals:
  \[ x_1(v) = a_{11}s_1(v) + a_{12}s_2(v) + a_{13}s_3(v) \]
  \[ x_2(v) = a_{21}s_1(v) + a_{22}s_2(v) + a_{23}s_3(v) \]
  \[ x_3(v) = a_{31}s_1(v) + a_{32}s_2(v) + a_{33}s_3(v) \]

- With the goal of finding estimates of the original sources:
  \[ u_1(v) = w_{11}x_1(v) + w_{12}x_2(v) + w_{13}x_3(v) \]
  \[ u_2(v) = w_{21}x_1(v) + w_{22}x_2(v) + w_{23}x_3(v) \]
  \[ u_3(v) = w_{31}x_1(v) + w_{32}x_2(v) + w_{33}x_3(v) \]

- We can, in fact, do this just using the assumption of non-gaussianity of the signals.

Independent Component Analysis

- Given signals \( x \) which are linear mixtures of independent sources \( s \):
  \[ x = As \]

- We want to find \( W \) such that \( y \) is a good estimate of the original sources:
  \[ u = Wx \]

- But what about algorithms?
Maximum Likelihood

- Very flexible approach
- Rich theoretical foundation
- More sensitive to incorrect estimation of pdf than other approaches
- May be sensitive to outliers if pdfs have certain shapes
**Maximum Likelihood**

- Given mixtures $x = As$
- We formulate the density of the mixed signals $x$ as:
  $$p_x(x) = |\text{det } W| p_s(s) = |\text{det } W| \prod_i p_i(s_i)$$
- Given $V$ observations of $x$ the likelihood is:
  $$L(W) = \prod_{v=1}^{V} \prod_i p_i(w_i^T x(v)) |\text{det } W|$$
- And the solution is obtained by finding:
  $$W_{ML} = \arg \max_W \ln L(W)$$

**Infomax**

- Background
  - Proposed by Linsker in 1992
  - Popularized by Bell & Sejnowski in 1995
  - First applied to fMRI data in 1998
  - Promising performance on a number of BSS
- Properties
  - Intuitively meaningful contrast function (mutual information)
  - Typically provides a simple learning rule
  - Choice of nonlinearity is required
  - Used on majority of fMRI applications
**Information Maximization (Infomax)**

- The goal is to maximize the mutual information that the output $z$ of a neural network contains about its input $x$. 

**Real-valued case:**

$$\mathcal{MI}(z, x) = H(z) - H(z | x)$$

\[\begin{array}{c}
\text{X} & \xrightarrow{W, w_o} & \text{z} = g(Wx) \\
\end{array}\]

**InfoMax**

$x =$ input, $y = g(w, x) =$ output, $u = wx$

$$I(x, y) = H_s(y) - H_s(y | x)$$

$$p(y) = \frac{p(x)}{J}, J = \left| \frac{\partial y}{\partial x} \right|$$

$$H_s(y) = -\int p(y) \ln \left( \frac{p(x)}{J} \right) dy = E[\ln J] - E[\ln p(x)]$$

Key: for deterministic mappings, $H_s(y | x)$ is not a function of the weights, the goal is thus to maximize $H_s(y)$.
InfoMax

\[ H_y(y) = E[\ln J] - E[\ln p(x)] \]

\[ \Delta w \propto \frac{\partial H(y)}{\partial w} = \frac{\partial}{\partial w} \ln |\frac{\partial y'}{\partial x'}| = \left( \frac{\partial y'}{\partial x'} \right)^{-1} \frac{\partial}{\partial w} \left( \frac{\partial y'}{\partial x} \right) \]

if \( y = g(w, x) = \frac{1}{1 + e^{-wx-w_0}} \) then \( \Delta w \propto \frac{1}{w} + x(1-2y) \)

Information Maximization (Infomax)

- For N sources:

update equation: \( \Delta W = \eta \{I + \varphi(u)u^T\} W \)

where \( \varphi(u) \equiv \frac{g''(u)}{g'(u)} \) and \( u = Wx \)

if \( g(\cdot) = \tanh(\cdot) \) then

\[ \Delta W = \eta \{I - 2yu^T\} W \]

Note: the update equation is identical to the ML approach provided the nonlinearity chosen has the same form as the cdf of the densities used for the ML approach.
Information Maximization (Real Case)

Output pdf given a Gaussian input

\[ y = \tanh(wx) \quad \text{and} \quad y = \sin(wx) \]

\[ KL_{\mu_y} = 0.101 \quad \text{and} \quad KL_{\mu_y} = 0.168 \]

\[ J(\hat{s}) = H(\hat{s}_G) - H(\hat{s}) \]

Negentropy/Kurtosis

- Assumes non-Gaussian sources
- Motivated by central limit theorem, the mixtures will be (more) Gaussian
- The independent components are calculated by maximizing the KL divergence between the output pdfs and a Gaussian pdf.

- Negentropy is always non-negative and is zero only if the variable is Gaussian.
Negentropy/Kurtosis

- Approximations of negentropy are used in practice. One such approximation uses kurtosis as a measure of non-Gaussianity

\[ kurt(x) = E\{x^4\} - 3\left(E\{x^2\}\right)^2 \]

- We can derive a learning rule and use instantaneous approximations for the expectation

\[ w \propto x(t)\left(w(t)^T x(t)\right)^3 - 3\|w(t)\|^2 w(t) \]
\[ + f\left(\|w(t)\|^2\right)w(t) \]

ICA of FMRI
How ICA Fits In Brain Imaging

Brain Imaging Techniques
- EEG
- DTI
- fMRI
- sMRI

Collection
- Data Driven Techniques
  - Principal Component Analysis
  - Independent Component Analysis
- Group ICA Of fMRI Toolbox
- fMRI Lab

Analysis
- Model Based Techniques
  - Statistical Parametric Mapping
- Preprocessing Techniques
  - Motion Correction/Realign
  - Co register/Normalize

Preprocessing Techniques
- fMRI process chain
- Functional Images
  - Time (secs): 1, 2, 3, ..., 750
- Phase Fix
- Registration
- Threshold/Overlay
  - Detection/Estimation
    - $y = X\beta + e$
- Normalization
### General Linear Model

1. **Model**
   - (1 or more Regressors)
   - Regressors: $x_i(j)$

2. **Data**
   - $y(j)$

3. **Fitting the Model to the Data at each voxel**
   - $y(j) = \hat{\beta}_0 + \sum_{i=1}^{M} \hat{\beta}_i x_i(j) + e(j)$

### General Linear Model (GLM)

- **Data(X)**
  - $\text{Voxels} \rightarrow$ Time

- **Design matrix**
  - $G$

- **Time courses**
  - $\hat{\beta}$

- **“Activation maps” corresponding to columns of G**

### Independent Component Analysis (ICA)

- **Data(X)**
  - $\text{Voxels} \rightarrow$ Time

- **Mixing matrix**
  - $\hat{W}^{-1}$

- **Components (C)**

### Spatially Independent Components

- In spatial ICA, there is no model for the fMRI time course, this is estimated along with the hemodynamic source locations.
A little more detail

\[ \text{Data}(X) = R \uparrow \hat{W}^{-1} \times \text{Components (C)} \]

Voxels \rightarrow \text{Data}(X) \leftarrow \text{Spatially Independent Components}

ICA of fMRI Illustration

The ICA model assumes the fMRI data, \( x \), is a linear mixture of statistically independent sources, \( s \).

\[ x = As \]

\[ p(s_1, s_2) = p(s_1) p(s_2) \]

The goal of ICA is to separate the sources given the mixed data and thus determine the \( s \) and \( A \) matrices.

movie of fMRI data from one slice over time, \( x \)

Source 1 \times \text{Time course 1}

Source 2 \times \text{Time course 2}
ICA Example

ICA Halloween (Un)Mixer!

\[ X = A \times S \]

- background
- Time
- candle 1
- candle 2
- candle 3

Candle out
Model for Applying ICA to fMRI

Data Generation (synthesis)

Data Processing (analysis)

Brain in Magnet
MR Scanner

(a) Preprocessing, Normalization
(b) Data Reduction
(c) ICA

Spatial versus Temporal ICA

- Does it matter?
- Why is spatial ICA more common?
- Some examples:
Temporally and Spatially low-correlated Components

SICA  SPM  TICA

Spatially Dependent Components

SICA  SPM  TICA
Impact of preprocessing/algorithms/etc

Criterion: Kullback-Leibler (KL) divergence

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

"True" source distributions

\[ p_s(\xi) = \prod_{i=1}^{N} p_i(\xi_i) \]

Estimated source distributions

\[ p_u(\xi) = \prod_{i=1}^{N} p_{u_i}(\xi_i) \]

- Define sources
- Generate sources
- For all:
  - Add noise
  - Smooth
  - Reduce (PCA, cluster, etc.)
  - Unmix (Info., fastICA, jade, etc.)
  - Evaluate (KL)

\[ \text{min}(KL) \text{ is winner} \]

Note: Other criteria for evaluation such as correlation of the true and estimated sources are possible. We have performed optimization using correlation, mean, and bias calculations and found similar results to the KL criterion above.


Simulated Sources

We generated four simulated mixtures, and applied an optimization algorithm for two types of ICA, two types of clustering, and different levels of smoothing.

Source 1
Source 2
Source 3
Source 4

**Temporal Pattern**
- Periodic, slowly varying
- Periodic, transient
- Random fluctuations
- Very slowly varying with occasional large transients (discontinuities)

**Signal Types, \( x(t) \)**
- Task-related
- Transiently task-related
- Function-related
- Motion-related
In general, the infomax algorithm performed slightly better than the fastICA algorithm (with the performance difference increasing at higher noise levels) and the clustering performance was better than PCA at lower noise levels. Different types of smoothing did not change the results much and thus only the results for no smoothing are shown. PCA had the least amount of variability in performance while clustering exhibited some variability in performance. The best combination for the higher SNR was infomax with clustering as the data reduction approach. However as the SNR decreased, PCA began to outperform clustering.

“We hybrid” fMRI Experiment

We generated a “hybrid” fMRI data by adding a known source to an fMRI data set. We performed ICA estimation, extracted the known source, and applied our optimization algorithm for two types of ICA, two types of clustering, and different levels of smoothing.
The results from the fMRI experiment are presented for CNR=0.41 and CNR=2.07. In general, the infomax approach outperformed the fastICA approach, and PCA outperformed clustering. Results from two smoothing kernels are presented, but did not have a significant effect on the outcome. The best overall combination for this case appears to be infomax and PCA, as in the low SNR case for the simulation results. Note that the variability of the results is also quite low when PCA is combined with infomax.
Signal Types

- Task related
- Cardiac
- Motion
- Vasomotor oscillation/
  High order visual

Motion Artifact

Motion-related signal due to mouth movement from inferior temporal and orbitofrontal regions
Hemodynamic Model
Artifact Detection and Reduction

Note: PREPROCESSING MAY DIFFER FOR Art. Hunting Approach

Eye movements

N/2 Nyquist Ghost

Source: Christian Beckmann’s “Little Shop of fMRI Horrors”:
http://www.fmrib.ox.ac.uk/~beckmann/homepage/academis/littleshop/

A Couple Recent Review Articles


How? A Few Software Packages

- The ICA-DTU toolbox
  (http://mole.imm.dtu.dk/toolbox/ica/index.html)
  - matlab
  - three different ICA algorithms
  - fMRI specific with demo data

- FMRIB Software Library, which includes the ICA tool MELODIC
  (http://www.fmrib.ox.ac.uk/analysis/research/melodic/):
  - C
  - FastICA
  - Complete Package

- AnalyzefMRI
  (http://www.stats.ox.ac.uk/~marchini/software.html)
  - R
  - FastICA

- BrainVoyager (http://www.brainvoyager.com/)
  - Commercial
  - FastICA
  - Complete Package

- FMRLAB (http://www.sccn.ucsd.edu/fmrlab)
  - matlab
  - infomax algorithm
  - fMRI specific with additional tools

- ICALAB
  - matlab
  - many ICA algorithms
  - not fMRI specific although one fMRI example included

- GIFT (http://icatb.sourceforge.net)
  - matlab
  - 8 ICA algorithms (more coming) including infomax and fastICA
  - Visualization tools and sorting options.
  - Sample data and a step-by-step walk through