Clustering of ICA components

Arnaud Delorme

(with Julie Onton, Romain Grandchamp, Nima Bigdely Shamlo, Scott Makeig)
ICA and PCA

ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix,

While PCA simply decorrelates the outputs (using an orthogonal mixing matrix), ICA attempts to make the outputs statistically independent, while placing no constraints on the mixing matrix.
Central limit theorem

Scalp channels = linear mixture of A and B (more gaussian)
ICA Training Process

• Remove the mean
  \[ x = x - \langle x \rangle \]

• ‘Sphere’ the data by diagonalizing its covariance matrix,
  \[ x = \langle xx^T \rangle^{-1/2}(x-\langle x \rangle). \]

• Update \( W \) according to
  \[ \Delta W \propto \frac{\partial H(y)}{\partial W} W^T w. \]
Entropy

\[ H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b p(x). \]

Dice: \(1/6\)

\( H = 6 \left( -\frac{1}{6} \log_2 \left( \frac{1}{6} \right) \right) = 2.58 \)

Fake dice (make a 6 half of the time): entropy 2.16 (base 2)

\( H = 5 \left( -\frac{1}{10} \log_2 \left( \frac{1}{10} \right) \right) - \frac{1}{2} \log_2 \left( \frac{1}{2} \right) = 2.16 \)
Entropy

\[ H(X) = - \sum_{x \in X} p(x) \log_b p(x). \]

Joint entropy

\[ H(X, Y) = - \sum_{(x, y) \in X \times Y} p(x, y) \log_b p(x, y). \]

Mutual Information

\[ H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2). \]

Shannon in his landmark 1948 paper `A Mathematical Theory of Communication.'

From http://planetmath.org/encyclopedia/ShannonsTheoremEntropy.html
Contingency table for stress and emotionality

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMOT= 1</td>
<td>19</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>63</td>
<td>64</td>
<td>3</td>
<td>1</td>
<td></td>
<td>142</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>16</td>
<td>18</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>33</td>
<td>87</td>
<td>84</td>
<td>35</td>
<td>13</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

From http://tecfa.unige.ch/~lemay/thesis/THX-Doctorat/node149.html
Contingency frequencies for stress and emotionality

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMOT=1</td>
<td>0.07</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.24</td>
<td>0.25</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02</td>
<td></td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Joint entropy 3.46; exercise: compute mutual information

\[
H(X, Y) = - \sum_{(x,y) \in X \times Y} p(x, y) \log_b p(x, y)
\]
ICA learning rule

How to make the outputs statistical independent?
Minimize their redundancy or mutual information.

Consider the joint entropy of two components,

\[ H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2). \]

Maximizing \( H(y_1, y_2) \) \( \Leftrightarrow \) minimizing \( I(y_1, y_2) \).

The learning rule:

\[ \Delta W \propto \frac{\partial H(y)}{\partial W} W^T W \]

=0 if the two variables are independent

Entropy extremum

Natural gradient (Amari)
Sub-gaussian

Super-gaussian

Sphering

ICA
Steps of clustering

• Select ICA components for clustering

• Precompute measures of interest

• Cluster measures

• Plot clusters and edit them if necessary
Edit dataset info
ICs to cluster
Precompute data measures
Precompute data measures

TIP: Compute all measures so you can test different combinations for clustering

Time-frequency options
Cluster components
Precluster: Use singular values from PCA

Mean ERSPs

ICs (all subj)

ERSP (time/freq)

PC templates

# PCs

Normalized singular values

~ relative variance of principal components

10% of max singular value
Precluster schematic

Each component is a dot. Clustering will group these dots.

ICs (all subj)
- ERSP
- Spectrum
- Dipoles

OR

ICs (all subj)
1. k initial "means" (in this case k=3, shown in color) are randomly selected from the data set (shown in grey).

2. k clusters are created by associating every observation with the nearest mean.

3. The centroid of each of the k clusters becomes the new means.

4. Steps 2 and 3 are repeated until convergence has been reached.

Classical KMean
Cluster components
Choosing data measures

What measure(s) should you use?

It depends on your final cluster criteria…
- If for example, your priority is dipole location, then cluster only based on dipole location…

But consider:
- What is the difference between these two components?
Choosing data measures

Similar dipole location, very different orientation.

Obvious dramatic effect on scalp map topography:

But, do they perform the same functions?
Choosing data measures

ERPs seem different...
Subject differences?

Significant ITC differences (by bootstrap) between the LOC and fLOC clusters immediately follow Probe presentation (5-11 Hz).
Subject differences?

**Cluster ERSPs show significant activity determined by bootstrap statistics within subject and binomial probability between subjects (p < 0.01)**

***Difference ERSP shows significant differences between the two clusters by bootstrap statistics (p < 0.001)***
Subject differences?
Plot/edit clusters
Plot cluster data

Plot mean scalp maps for easy reference.
Plot cluster data

Choose which cluster

Choose which components
Plot cluster data
Plot cluster ERP
Reassigning components
Issue with standard clustering

Large parameter space problem: many different clustering solutions can be produced by changing parameters and measure subsets. Which one should we choose?

EEGLAB clustering has ~12 parameters
Measure projection

(EEGLAB extension by Nima Bigdely Shamlo) only has one pre-clustering parameter.

(Affinity clustering by Pernet, Martinez, Delorme)
Exercise

• Load the STUDY
• Precluster and cluster components using spectrum and dipoles location
• Look at your cluster. Identify frontal midline theta cluster and occipital alpha cluster
• Plot significant difference for one component cluster spectrum between the two conditions in the default design