UCLA Math REU 2013: Epilepsy and EEG/fMRI

Heal, K., Navarro, K., Wollner, M., Yan, E., Douglas, P., Gilles, J., Kerr, W., and Meyer, T.

August 7, 2013
Part One

Epilepsy Classification Using Positron Emission Tomography (PET) Data
Temporal Lobe Epilepsy (TLE)

- Most common of localized epilepsies
- Patients with Non-Epileptic Seizures (NES), Left Temporal Lobe Epilepsy (LTLE), Right Temporal Lobe Epilepsy (RTLE), and Bilateral Temporal Lobe Epilepsy (BTLE)
- Early detection and treatment are important
Positron Emission Tomography (PET)

- 3D image of metabolic processes
- Manual analysis used to detect abnormalities
- Atrophy is associated with hypometabolism
Regions of Interest (ROI)

- Combine pixels into biologically relevant features
- Focus on analyzing these 47 ROIs
Project Goal

- Can computers be used to detect abnormalities in a different way?
- At base we have a classifier for NES vs LTLE vs RTLE
- We want to incorporate Bilateral patients
- Adding in these patients increases complexity
Cyclical Leave-One-Out Cross Validation (CL1OCV) was used to evaluate classifier performance

- Leave one instance of data out (leave-one-out)
- Train on the rest
- Try to classify the data left out
- Repeat until all data has been left out once (cyclical)
Spectral clustering & \( k \)-means

Data not easily separable by common similarity metrics such as Gaussian similarity

Unsupervised clustering techniques failed to beat the naïve classifier

- Naive classifier: classify everything as the most frequent class e.g. NES RTLE LTLE BTLE unspec 32 34 39 14 5
- Naive performance: 34.7\% \pm 8.4\%
Feed data to an input layer of nodes, data is processed through \( n \) hidden layers with \( m_i, i = 1, 2, \ldots, n \) nodes per layer

Number of nodes per layer and number of layers to use is a difficult problem

Grid Search!

Beat naïve classifier case... barely

\[^1\text{http://www.texample.net/tikz/examples/neural-network/}\]
Neural Networks Heatmap

Neural Network Layers, Nodes Grid Search (OVA)
Neural Network Perf. cont.

Mean Neural Network Accuracy: 1 Layer, 15 Nodes, 5 iterations

Accuracy

50%
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%

[Bar chart showing comparison between Naive and Neural Network accuracy.]

- Naive
- Neural Network
- Naive 95% Confidence Interval
Support Vector Machines

classification = \text{sign} (w_0 \cdot x_i + b_0)

https://upload.wikimedia.org/wikipedia/commons/2/2a/Svm_max_sep_hyperplane_with_margin.png
Many datasets, however are not *linearly* separable

In 1995, Cortes et al. introduced the **soft margin** hyperplane described by the objective function \( \frac{1}{2}w^2 + CF \left( \sum_{i=1}^{l} \xi_i \right) \) subject to the constraints \( y_i (w \cdot x_i + b) \geq 1 - \xi_i \), \( \xi_i \geq 0 \), where \( \sum_{i=1}^{l} \xi_i \) is the sum of training errors, \( F \) is a monotonic convex function, and \( C \) is a constant

Allows classification of non-linearly separable data, but introduces additional parameters to the problem \((C, F)\)
Feature Selection

- We wish to avoid the “curse of dimensionality”
- Two main classes of feature selection:
  - “Unsupervised”: Principal Component Analysis (PCA), Independent Component Analysis (ICA), Non-negative Matrix Factorization (NMF)
  - “Supervised”: Sequential Forward Selection (SFS), Sequential Backward Selection (SBS), Sequential Floating Forward Selection (SFFS), Sequential Floating Backward Selection (SFBS)
- “Unsupervised” techniques rely on some objective parameters—which is problematic e.g. if the axis with the highest variance is not a good feature for classification
Sequential Feature Selection

- Sequential feature selection describe a class of feature selection algorithms that sequentially add (bottom-up) or remove (top-down) features based on some objective criteria.
- We can define this objective function to be the training accuracy ("supervised" case).
SFFS Flowchart

Start

Are there unselected features?

No

Yes

Select most significant feature

Does selecting a feature improve criteria?

Yes

Select most significant feature

No

Deselect least significant feature

Does deselecting least significant feature improve criteria?

Yes

Deselect least significant feature

No

Can any more least significant features be deselected?

Yes

Deselect least significant feature

No

Stop
Pairwise Classification Results

![Pairwise vs. BTLE Accuracy](image)
The intersection of features selected by SFFS and SFBS were all regions of the temporal lobe.
Part Two

Electroencephalography (EEG) Empirical Wavelet Analysis and EEG-Functional Magnetic Resonance Imaging (fMRI) Fusion
Neuroscience Background

Figure: Current EEG Experimentation and Analysis Processes
Neuroscience Background

EEG records brain waves. Brain waves travel at different frequencies. Group brain wave types into “spectral bands”.

<table>
<thead>
<tr>
<th>Band Name</th>
<th>Frequency Range (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta</td>
<td>approx. 0.1 – 4</td>
</tr>
<tr>
<td>Theta</td>
<td>4 – 8</td>
</tr>
<tr>
<td>Alpha</td>
<td>8 – 13</td>
</tr>
<tr>
<td>Beta</td>
<td>13 – 30</td>
</tr>
<tr>
<td>Gamma</td>
<td>30 – variable</td>
</tr>
</tbody>
</table>

**Figure: Traditional (Fixed) Spectral Bands**

However, these classifications are very old; they predate EEG and the field of signal processing. Are they still relevant? Can their definitions be mathematically justified?
Currently Used Visualization Methods

**Figure:** CWT With Morlet Wavelets
Empirical Wavelets

Goal: Define a new set of spectral bands that is adaptive to an individual’s brain.

How to do this? Define a set of wavelets as a band pass filter bank:

\[
\pi \omega_1 \omega_2 \omega_3 \ldots \omega_n \omega_{n+1} \]

\[
2\tau_1 2\tau_2 2\tau_3 \ldots 2\tau_n 2\tau_{n+1} \tau_N
\]

Figure: An Empirical Wavelet, Each \(\omega_i\) Is a Boundary

Use an Empirical Wavelet Transform on the signal in the time domain, then decompose the signal into spectral components:

\[
f(t) = \sum_j a_j(t)\cos(\phi_j(t))
\]
Wavelet Construction

**Figure:** Desirable Boundary Choices
Wavelet Construction

Figure: Undesirable Boundary Choices
Adaptive Boundary Search Methods

Figure: “Epsilon” Boundary Search
Adaptive Boundary Search Methods

Figure: “Closure” Boundary Search
EEG Signal Processing

Amplitude vs. Frequency (Hertz)

Amplitude vs. Time Since Flash (Seconds)

Frequency (Hertz) vs. Time Since Flash (Seconds)
Time-Frequency Plane

Figure: Hilbert Transform With Traditional Bands
Time-Frequency Plane

Figure: Hilbert Transform With Traditional Bands
Time-Frequency Plane

Amplitude

Frequency (Hertz)

Time Since Flash (Seconds)

Figure: Hilbert Transform With “Closure” Bands
Time-Frequency Plane

Figure: Hilbert Transform With “Closure” Bands
fMRI and EEG Colocalization

- fMRI: dense spatial information, sparse temporal information
- EEG: sparse spatial information, dense temporal information

Goal: dense spatial, dense temporal
Standardized Low Resolution Brain Electromagnetic Tomography

Solution to the inverse problem for EEG: localizing the exact sources of the neural activity measured as scalp electric potentials

Standardized Low Resolution Brain Electromagnetic Tomography (sLORETA):
  ▶ Fairly good accuracy for deep sources (other methods misplace these on outer cortex)
  ▶ Gives a smoothed result
0.28 seconds before stimulus

Time of stimulus

0.5 seconds after stimulus
Temporal Kernel Canonical Correlation Analysis

- **Goal:** find the maximal correlation between simultaneous EEG-fMRI, both spatially and temporally
- **Method:** Temporal Kernel Canonical Correlation Analysis (tkCCA)
  - Canonical Correlation Analysis (CCA): multivariate correlation between two data sets \((X, Y)\)
    - finds the maximally correlated features of \(X\) and \(Y\)
    - CCA uses covariance matrices (covariance of \(X\) and \(Y\), variance \(X\), variance of \(Y\))
  - Kernel Canonical Correlation Analysis (kCCA): reduces the dimensions of CCA by using linear kernel matrices in place of covariance matrices
  - tkCCA: allows non-instantaneous coupling
    - Use shifted \(\tilde{X}\) = multiple stacked copies of \(X\) with incremental time-shifts of size \(\tau\)
- **Idea:** tkCCA on voxel-space EEG (after transformation from time-frequency) and voxel-space fMRI
Future Research

- Use of In-Scanner data
- Use of spectral bounds found by Empirical Wavelets
- Spatial shift instead of time shift, allowing for even higher temporal accuracy


References II


Thank you

Any questions?