Decoding EEG Data Using Machine Learning

Dr. Firat Soylu
The University of Alabama
http://fsoylu.people.ua.edu

Sabancı University, CS Seminar

June 16, 2023
Motivation

• Improve analysis of EEG data using machine learning (ML) methods.
• Develop heuristics for analyzing neuroscience data with ML methods.
Outline

- EEG (Electroencephalogram)
- Use of EEG in cognitive neuroscience
- Traditional EEG data analysis methods
- Machine learning with EEG (and other neuro) data
- A path for future
Invention of EEG

- Hans Berger (1929)
  - Also discovered Alpha waves
- Initially met with skepticism
Electroencephalogram

- Used for studying sensory, attentional & cognitive processes
- Cheap
- Non-invasive
- Easy to use with children

Figure taken from Nagel, S. (2019).
What does EEG measure?

- Voltage in each electrode relative to a reference electrode
- Signals are weak and have to be amplified (x10K-50K)
- Sampling rate ranges from 500 – 2000 Hz
- Low/high pass filters to filter noise
Principles of EEG

- Measures electrical activity in the (mostly) cortex
- Activity oscillates at different frequencies
- Measured in hertz (Hz) cycles/sec
- Most activity is between 0-100 Hz
Delta (1 – 3 Hz)

Theta (4 – 7 Hz)

Alpha (8 – 12 Hz)

Beta (13 – 35 Hz)

Gamma (36 – 100 Hz)
• Event-related potentials

• Frequency analysis

• Time-frequency analysis
ERPs
Example: N170

Rossion & Jacques (2009)
N400

• Violation of a semantic expectation

I take my coffee with cream and sugar

I take my coffee with cream and dog

Figure: Courtesy of Steve Luck, ERP Bootcamp
Traditional EEG/ERP Analysis

- Determine relevant electrode sites
- Determine relevant time intervals
- Extract measurements
- Conduct inferential stats
Enter Machine Learning ...

- Fewer a priori assumptions about the spatial location and time intervals for effects
- Allows testing effects distributed across different electrode sites and time intervals
- Good for conducting research in new domains
Case Study

- 38 adult participants (13 female)


Research Question:

Do previous experiences with monitoring and counting lead to higher automaticity?
Results

• Higher behavioral performance for montring
• Higher P1/N1 positivity for montring
  – Feature-based attention system, focusing on features matching with a template.
• Similar P3 for montring and counting
  – Memory allocation as opposed to use of counting or subitizing
“The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.”

-Marcel Proust
Decoding with SVM

What is decoding?

- Scalp distributions (SD) are varied across tasks & subjects
- Decoding with SD can allow prediction of task-related processes

Grootswagers, Wardle, and Carlson (2017)
Decoding in Neuroscience

- Lower accuracy than engineering applications (e.g., brain-machine interfaces)
- The goal is to test research hypotheses
- Data SNR is crucial
- Preprocessing steps might differ from traditional analysis
Approach

● Decode SD at each time point to predict numerical magnitudes, for each configuration.

● Compare temporal aspects of decoding accuracy to compare finger configurations

● Compare results with traditional analysis.
Mean Accuracy of Alpha-Based Decoding and ERP-Based Decoding Averaged Across All 12 FNCs

26.7% (at 490 ms)

12.3 (at 220 ms) %

Note. The black horizontal line indicates the chance-level performance (1/12 ≈ 0.083 ≈ 8.3%). The shaded areas indicate ± 1 SEM.
Comparison of configuration types:
- Rerun the procedure for each conf.
- Compare the temporal aspects of avg. accuracy

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Accuracy</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>36.5%</td>
<td>472 ms</td>
</tr>
<tr>
<td>C</td>
<td>57.7%</td>
<td>506 ms</td>
</tr>
<tr>
<td>NC</td>
<td>60.4%</td>
<td>604 ms</td>
</tr>
</tbody>
</table>

Note: The black horizontal line indicates the chance-level performance (0.25 = 1/4).
The Percentage of Accurate Classifications for Each of the 4 FNCs
<table>
<thead>
<tr>
<th>FNC</th>
<th>Average accuracy (%)</th>
<th>Peak accuracy (%)</th>
<th>Peak accuracy time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
</tr>
<tr>
<td>Montring (M)</td>
<td>36.5 (6.6)</td>
<td>47.2 (8.9)</td>
<td>445.789 (186.629)</td>
</tr>
<tr>
<td>Counting (C)</td>
<td>44.3 (9.6)</td>
<td>57.7 (11.4)</td>
<td>506.316 (184.192)</td>
</tr>
<tr>
<td>Noncanonical (NC)</td>
<td>46.3 (10.9)</td>
<td>60.4 (12.6)</td>
<td>515.263 (194.377)</td>
</tr>
</tbody>
</table>

![Montring (M)](image1)

![Counting (C)](image2)

![Non-canonical (NC)](image3)
Results & Comparisons

• ERP-based more accurate than Alpha-based

• Early peak for alpha-decoding might be associated with the previous P1/N1 effect (caution due to low acc. level)
Results & Comparisons

- Decoding better informs when the three conditions diverge in processing (post-500ms).
Results & Comparisons

- Decoding allows a more detailed inspection of differences across conditions.
What about other ML algorithms?
Decoding Performance of 6 ML Algorithms Compared

<table>
<thead>
<tr>
<th>Classifier</th>
<th>MATLAB function</th>
<th>Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>fitcecoc()</td>
<td>One. vs all, linear kernel</td>
</tr>
<tr>
<td>LDA</td>
<td>fitcecoc()</td>
<td>One. vs all, linear discriminant, Kernel: normal</td>
</tr>
<tr>
<td>NB</td>
<td>fitcecoc()</td>
<td>One. vs all</td>
</tr>
<tr>
<td>KNN</td>
<td>fitcecoc()</td>
<td>One. vs all, k=12 for decoding all FNCs and K=4 for category-specific decoding</td>
</tr>
<tr>
<td>DT</td>
<td>fitcecoc()</td>
<td>One. vs all</td>
</tr>
<tr>
<td>NN</td>
<td>fitcnet()</td>
<td>Input: [1,32], FC1: [10,32], FC2: [4,10] or [12, 10], output: [1,1]</td>
</tr>
</tbody>
</table>

4 Comparison Metrics

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \\
\text{Precision} = \frac{TP}{TP + FP} \\
\text{Recall} = \frac{TP}{TP + FN} \\
F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

*TP: true positives, TN: true negatives*  
*FP: false positives, FN: false negatives*
<table>
<thead>
<tr>
<th>Classifier</th>
<th>Avg. Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>15.46</td>
<td>15.57</td>
<td>15.46</td>
<td>15.51</td>
</tr>
<tr>
<td>LDA</td>
<td>13.45</td>
<td>13.53</td>
<td>13.45</td>
<td>13.49</td>
</tr>
<tr>
<td>NB</td>
<td>9.46</td>
<td>10.96</td>
<td>9.46</td>
<td>10.15</td>
</tr>
<tr>
<td>DT</td>
<td>10.98</td>
<td>11.13</td>
<td>10.98</td>
<td>11.06</td>
</tr>
<tr>
<td>NN</td>
<td>13.59</td>
<td>13.62</td>
<td>13.59</td>
<td>13.60</td>
</tr>
</tbody>
</table>
Category specific decoding
Confusion Matrices

(SVM)

Monring (M)

(NN)
Results

• SVM scores highest across all measures -> most reliable. NN & KNN perform well as well.

• DT close to chance-level in category specific, but better in all-categories → DT better suited to decoding with high num. of classes.

• NC & C had higher accuracy than M across all
  → automaticity leads to more similar neural patterns

• Highest true positive rates for Non-Canonical, numbers 2 and 4 → most distinguishable
Future Work

- Explore adjustments in preprocessing
- Test different parameter values with each algorithm.
- Test different forms of perceptual, cognitive, and socio-emotional processing
Current Study