Classification of EEG During Imagined Mental Tasks by Forecasting with Elman Recurrent Neural Networks

Elliott Forney & Charles Anderson

Colorado State University
Computer Science Department

September 2, 2015
Outline

1. Introduction to EEG & BCI
2. Elman Networks
3. Experiment Setup
4. Forecasting EEG with Elman Networks
5. Classification of EEG with Elman Networks
6. Results
Electroencephalography (EEG) is a technique for measuring brain activity using an array of electrodes placed on the surface of a subject’s scalp.
There are many uses for EEG, both clinical and research.

Today, we are interested in Brain-Computer Interfaces (BCI).

BCI establish a direct channel of communication between brain and machine.

Bypasses ordinary motor-based communication.

BCI can be used in assistive technology:
- operate computers, wheelchairs, telephones, etc.

Reestablish communication for those with Locked-in Syndrome.

May eventually be used in everyday devices.
EEG classification is difficult

- Useful and exciting, but not easy
- Low signal to noise ratio:
  - Microvolt signals
  - Ocular & muscular artifacts
  - Noise from external electronics
- The brain is complex!
  - Billions of neurons
  - Trillions of connections
  - Recurrent & Stateful
- EEG is complex and patterns are both spatial (intra-electrode) and temporal
Current approaches

- Current classification rates are not high enough
- Often rely on power spectrum estimates
- Do not readily capture some patterns
  - Phase differences
  - Short term ordering of events
- We propose using Recurrent Neural Networks
  - Here, we investigate Elman Nets in particular
Elman Networks have two layers

**Hidden layer:**
- fully connected to inputs
- sigmoidal transfer functions
- full recurrent connections

**Visible layer:**
- fully connected to hidden layer and outputs
- linear transfer function
- no recurrent connections

Long history and, relatively, well studied [1]

Universal approximators of Finite State Machines [2]
Training Elman Networks

- Gradient Descent using Scaled Conjugate Gradients (SCG) [3]

- Gradient estimated using Truncated, Batch-Mode, Back-Propagation Through Time (BPTT) [4, 5]
  - Recurrent connections unrolled into feedforward network
  - Unrolling truncated after some timesteps back
  - Gradient accumulated over entire sequence
Classify EEG during imagined mental tasks
- Imagined motor task
- Arithmetic task

Simple dataset for now

Three subjects
- Subject-A and Subject-B: Able bodied, mid-twenties, in laboratory setting
- Subject-C: Quadriplegia, spinal legion at C4, mid-twenties, at home

256 samples per second

Hardware Analog Bandpass filter: 1.5-34Hz

Maximum Noise Fraction Filter [6, 7]

Standardize to zero mean unit variance
First, we consider forecasting EEG a single step ahead

- One input per channel
- One output per channel
- Minimize MSE between current output and next signal value
Parameter Selection

- Three parameters to tune manually
  - Steps unrolled for BPTT
  - Training epochs
  - Hidden units

- Steps unrolled fixed at 20

- Training epochs fixed at 250

- Little improvement with larger values

- Regularization controlled by limiting hidden units
  - Smaller networks are much faster
  - Early stopping doesn’t work quite as well
- Forecasting error vs hidden units
- Six-fold cross validation
- Separates around 15-20
- Levels off around 40
- Only slight over-fitting
Iterated Models

- Placing feedback loop between the output and input
- Forms an autonomous, iterated system
- This gives insight into the temporal information learned
- To the right: 20, 40 and 160 hidden units
Generative approach to classification

- Separate RNN trained to model each class
- Model by forecasting a single step ahead
- Each network is an expert on its class

**Training Data**

- EEG Class 1
- EEG Class 2
- \vdots
- EEG Class K

**Training**

- SCG+BPTT 1
- SCG+BPTT 2
- \vdots
- SCG+BPTT K

**Experts**

- Forecaster/RNN 1
- Forecaster/RNN 2
- \vdots
- Forecaster/RNN K
Novel EEG is labeled by
- apply each forecaster/expert to new EEG
- average error over a short window
- select label associated with model that produced lowest error

Similar to Gupta, Oeda and Coyle [8, 9, 10]
Classification of EEG

- Classification accuracy vs hidden units
- Decisions made every second
- Six-fold cross validation
- Over-fit after 10-20 hidden units
We have a paradox:
- Best modeling error with \( \approx 50 \) hidden units
- Richest iterated dynamics with \( > 150 \) hidden units
- Highest classification accuracy with \( \approx 10 - 20 \) hidden units

Conjecture:
- Simple, short-term patterns seem more discriminative
- Complex, long-term patterns seem to contribute to forecasting error but are not discriminative
Classification Accuracy

<table>
<thead>
<tr>
<th></th>
<th>NH</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-A</td>
<td>18</td>
<td>96.7%</td>
<td>86.7%</td>
<td>80.0%</td>
</tr>
<tr>
<td>Subject-B</td>
<td>10</td>
<td>100.0%</td>
<td>85.0%</td>
<td>57.9%</td>
</tr>
<tr>
<td>Subject-C</td>
<td>16</td>
<td>99.3%</td>
<td>90.0%</td>
<td>94.6%</td>
</tr>
</tbody>
</table>

Table: Average Classification Accuracy

- Subject-A performs well
- Subject-B performs poorly on test partition, possibly lost concentration?
- Subject-C performs very well
- Recall that Subject-C is disabled and data was recorded at home
Bitrates

<table>
<thead>
<tr>
<th></th>
<th>NH</th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject-A</td>
<td>18</td>
<td>47.3bpm</td>
<td>26.0bpm</td>
<td>16.7bpm</td>
</tr>
<tr>
<td>Subject-B</td>
<td>10</td>
<td>60.0bpm</td>
<td>23.4bpm</td>
<td>1.1bpm</td>
</tr>
<tr>
<td>Subject-C</td>
<td>16</td>
<td>56.5bpm</td>
<td>31.9bpm</td>
<td>41.8bpm</td>
</tr>
</tbody>
</table>

**Table: Average Bitrate**

- decisions are made at one second intervals
- bitrates are more comparable, fair and describe user experience better [11, 12]
- state-of-the-art is somewhere around 10-40bpm
Thanks!

Partly funded by National Science Foundation grant #0208958.
Finding structure in time.

On the computational power of elman-style recurrent networks.

A scaled conjugate gradient algorithm for fast supervised learning.

An efficient gradient-based algorithm for on-line training of recurrent network trajectories.

_Neural networks and learning machines._
Signal fraction analysis and artifact removal in eeg.

Geometric subspace methods and time-delay embedding for eeg artifact removal and classification.

[8] Lalit Gupta, Mark McAvoy, and James Phegley.
Classification of temporal sequences via prediction using the simple recurrent neural network.

[9] Ikusaburo Kurimoto Shinichi Oeda and Takumi Ichimura.
Time series data classification using recurrent neural network with ensemble learning.

*Lecture Notes in Computer Science, 4253:742 – 748, 2006.*

A time-series prediction approach for feature extraction in a brain-computer interface.


Eeg-based communication: improved accuracy by response verification.


*An introduction to information theory: symbols, signals & noise.*
Dover Pubns, 1980.