Reliable Identification of Mental Tasks Using Time-Embedded EEG and Sequential Evidence Accumulation

Charles Anderson†, Elliott Forney†, Annamalai Natarajan‡
†Department of Computer Science, Colorado State University, Fort Collins, CO
‡Princeton Neuroscience Institute, Princeton University

Abstract. 11 channels of EEG were recorded from a subject performing four mental tasks. A time-embedded representation of the untransformed EEG samples was constructed. Classification of the time-embedded samples was performed by linear and quadratic discriminant analysis and by an artificial neural network. A classifier’s output for consecutive samples is combined to increase reliability. A new performance measure is defined as the number of correct selections that would be made by a BCI user of the system, accounting for the need for an incorrect selection to be followed by a correct one to “delete” the previous selection. A best result of 0.45 correct selections/second (2.2 seconds per BCI decision) was obtained with a neural network using a time-embedding dimension of 50.

1. Introduction

Brain-computer interface (BCI) systems that identify four or more mental tasks provide a larger vocabulary for the user than do the common approaches of detecting a single event-related potential (ERP) or the identification of two or three mental tasks. Reliable classification of EEG during arbitrary mental tasks by current BCI systems is limited by adherence to conventional EEG representations and performance measures. Here experiments are described with an alternative EEG representation consisting of a simple time-embedding of raw EEG values, and an alternative performance measure developed from the BCI user’s viewpoint.

2. EEG Recording and Time-Embedding Representation

A common way to approximately capture the state of a dynamic process is to combine measurements from multiple times. Here an embedding in time of EEG is formed as follows. In the following experiments, 11 channels of EEG are recorded. The EEG sample at time $t$ is the 11-component vector $e_t$. A $d$-dimensional time embedding is formed by concatenating the $d$ samples, $e_t, e_{t-1}, e_{t-2}, \ldots, e_{t-d+1}$. With 11 channels and a time embedding of $d = 10$, each time-embedded EEG sample is a 110-component vector.

This representation makes no assumptions about what aspects of the EEG data are most relevant for mental task classification. Common representations, such as the energy in various frequency bands, assumes the information is contained in certain features of the data and only those features are used for classification. The time-embedding representation with a high-enough dimension includes the information needed to calculate any spatial or temporal feature that helps with classification performance. However, since the relevant features are not already extracted, the responsibility is placed on the classification algorithm to discover the nonlinear combinations of time-embedded samples that provide features which are most useful for classification.

3. Classification

Since untransformed EEG is being classified, it is hypothesized that nonlinear classifiers will perform better than linear ones. To test this, the following classifiers were compared.

**Quadratic Discriminant Analysis (QDA):** Gaussian models are fit to training samples corresponding to each class. For a time embedding of $d = 10$ and $k$ mental tasks, the QDA models are parameterized by $k$ 110-component means and $k$ $110 \times 110$ covariance matrices, for a total of 48,840 parameters.
Linear Discriminant Analysis (LDA): Like QDA, but a common covariance matrix is used for all Gaussian models by averaging the class covariance matrices. For the above example, the LDA models are parameterized by $k$ 110-component means and one $110 \times 110$ covariance matrix for a total of 12,540 parameters.

Artificial Neural Network (NN): A neural network with one hidden layer is used. For $d = 10$, the input is 110-dimensional. The output layer has one output for each class and generates values from a multinomial distribution. The network is trained to maximize the likelihood of the training data using Møller’s Scaled Conjugate Gradient algorithm [4]. For the experiments described here, 20 units are used in the hidden layer. So, for the above example, the neural network is parameterized by $111 \cdot 20 + 21 \cdot 4 = 2,304$ weights.

4. BCI Decision by Evidence Accumulation

Most EEG features used in BCI studies do not appear consistently. Even those features that are most discriminatory between mental tasks come and go over short periods from fractions of seconds to several seconds. Therefore, we used the following procedure to combine evidence from each sample until a class probability exceeds a threshold. We used Baum and Veeravalli’s [1, 2] M-ary Sequential Probability Ratio Test (MSPRT) algorithm that incrementally calculates the joint probability of each class given an increasing window of EEG samples.

Let the classifier’s output for Task $k$ at time $t$ be $p_{t,k} = P(\text{Task} = k|e_t^{(d)})$. MSPRT updates the joint probabilities of $p_{t,k}$ for consecutive $t$ by accumulating the log probabilities, $l_{t,k}$. Using gain parameter, $g$, the joint probability for time span $S$ is calculated.

$$
\begin{align*}
    l_{t+1,k} &= l_{t,k} + \ln p_{t,k} \\
    y_{t,k} &= g \, l_{t,k} \\
    P(\text{Task} = k|e_t^{(d)}, t \in S) &= e^{y_{t,k} - \ln \sum_{k' \in K} e^{y_{t,k'}}}
\end{align*}
$$

If $P(\text{Task} = k|e_t^{(d)}, t \in S)$ exceeds the threshold parameter, $0 < d < 1$, decide Task is $k$ at time $t$ and reset $l_{t,k} = 0$.

4.1. 2-Class Example

MSPRT is illustrated in Figure 1 by an example of EEG samples recorded during performance of two mental tasks. The classifier’s probability of Task 1 is plotted as a thin line. The MSPRT accumulated probability is shown as a thick line. Circles mark points of decision, when accumulated probability exceeds the threshold and when at least one second has elapsed since last decision. This last requirement of at least one second elapsing between decisions was added to restrict the BCI system of making task classification decisions faster than a user can switch between mental tasks.

4.2. Decision Rate

The rate with which a user is able to select an intended symbol must take into account the need for a correct selection to delete an incorrectly selected symbol. Over some period of $s$ seconds, let the number of correct selections be $c$ and the number of incorrect selections be $i$. Then the rate of correct decisions/second is given by $\frac{c}{s}$.

For example, say that over a 10 second period the BCI system correctly identified 9 selections and incorrectly identified 3 selections. The rate of correct decisions/second is $(9 - 3)/10 = 0.6$. If there are 4 possible selections, the “bit rate” would be 1.2 bits/second. However, the decision rate/second give a better impression of the user’s experience.

5. Experiments

This section describes results of experiments with the time-embedding representation and the MSPRT evidence accumulation algorithm.

The mental tasks used in this study are
Identification of Mental Tasks Using Time-Embedded EEG

(i) imagined right hand clenching,
(ii) imagined left foot moving up and down,
(iii) counting backwards from 100 by 3, and
(iv) visualizing the computer screen tumbling in space.

The subject performed 10 repetitions of randomly ordered tasks, each for 5 seconds. This was repeated for 5 sessions, all on the same day.

EEG was recorded from 11 channels (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, O2) at 256 Hz and bandpass filtered to 0.1 - 30 Hz. Recording was done with an Electro-cap and a NeuroPulse Mindset EEG amplifier [5] connected via a SCSI port with a laptop running our open-source Linux software [3].

EEG data was partitioned in the following way. Data from one session was designated as test data, data from another session as validation data, and the remaining three sessions were used as training data. Classification experiments were repeated for all 20 ways of forming this partition into test, validate, and train subsets.

The MSPRT algorithm’s parameters of gain and threshold were approximately optimized by determining which of seven values result in the best correct decision rate for the validation set. Time-embedding dimensions of 1, 3, 10, 20, and 50 were tested.

6. Results

In Figure 2, the number of correct decisions/second averaged over the 20 data partitions are plotted versus the time-embedding dimension. QDA and LDA performance decreases with dimension, but the neural network performance increases. Its best performance is 0.45 correct decisions/second for test data (equivalent to 2.2 seconds per BCI decision.) The average performance on training data is 0.7 and for validation data is 0.48 correct decisions/second. Figure 3 shows the decisions made by the classifier as circles. The actual task performed by the subject is shown by the lines. Decision time was constrained to be one second or greater, assuming a user would not be able to switch mental tasks faster than every second.

7. Conclusions

For the data used in this experiment, a neural network is capable of extracting patterns from time-embedded data that relate to which of four mental tasks the subject is performing. LDA and QDA did not perform as well. More significantly, their performance decreased with embedding dimension while that of the neural network increases with embedding dimension. This shows that the information added as the embedding dimension increases cannot be captured by Gaussian models, but can be extracted with a nonlinear classifier, in this case a neural network.
Identification of Mental Tasks Using Time-Embedded EEG

Figure 2. Performance in correct decisions/second for the three classifiers versus the time-embedding dimension.

Figure 3. Decisions made by BCI system shown as circles. Line shows correct task.
The next step is to extract the patterns learned by the hidden units in the neural network and to investigate if these patterns make sense with what is currently known about how we perform the various mental tasks.

Acknowledgments

This work was partially supported by the National Science Foundation through Grant Number 0208958.

References