Analysis of Event-Related Potentials in BCIs

EEGLAB Workshop 2013, Track B

Christian A. Kothe
SCCN, UCSD
Outline

1. Task
2. Analysis Approach
3. Review
4. Advanced ERP Topics
1 Task
Experimental Task

- **Flanker Task**: The experiment consists of a sequence of ca. 330 trials with inter-trial interval of 2s +/- 1.5s
- At the beginning of each trial, an arrow is presented centrally (pointing either left or right)
- The arrow is flanked by congruent or incongruent “flanker” arrows:

  ┌───┐
  │   │
  │   │
  └───┘

- The subject is asked to press the left/right button, according to the central arrow, and makes frequent errors (25%)
Consideration

• The peak ERP features discussed so far were chosen for a single channel of EEG

• **Problem:** with multiple channels all channels measure almost the same signal properties, thus little information gain to expect

• **Idea:** Derive a spatial filter and use multiple channels to *computationally focus* on source processes of interest, then extract *source signal features*
Consideration

• How to design an optimal spatial filter for this task?

• **Idea:** Can be done implicitly by a linear classifier when applied to multiple channels

• Works only for source-signal features that are a *linear transform* of channel-signal features

• The classifier must produce the *same solution under rotation and scaling* (not all do, but e.g., LDA does)
2 Analysis Approach
Approach

• Calibration recording is band-pass filtered between 0.5Hz and 15Hz
  – 0.5Hz lower edge removes drifts
  – 15Hz upper edge leaves enough room for sharp ERP features
Approach

- Calibration recording is band-pass filtered between 0.5Hz and 15Hz
  - 0.5Hz lower edge removes drifts
  - 15Hz upper edge leaves enough room for sharp ERP features

- Epochs are extracted for each trial and label is set to A for incorrect trials and B for corrects
Actual Data

- Time courses for all trials super-imposed (color-coded by class) – but here different task
Extracted Epochs

Channel time courses under Condition B

Channel time courses under Condition A

Response (A or B)

Three sample trials (out of 100) shown: mean, -1 std. dev, +1 std. dev
Extracting Linear Features

For each trial segment, calculate signal mean in 3 time sub-windows (→ 3-dim feature vector)
Problem with LDA

- Multi-channel features are too high-dimensional for LDA to handle with few trials!
Fixing LDA

• Given trial segments $x_k$ (in vector form) in $C_1$ and $C_2$,

$$\mu_i = \frac{1}{|C_i|} \sum_{k \in C_i} x_k, \quad \Sigma_i = \sum_{k \in C_i} (x_k - \mu_i)(x_k - \mu_i)^\top$$

$$\theta = (\Sigma_1 + \Sigma_2)^{-1}(\mu_2 - \mu_1), \quad b = -\theta^\top(\mu_1 + \mu_2)/2$$

• $\theta$ often high-dimensional but only few trials available
• Can use a regularized estimator instead, here using shrinkage
  – instead of $\Sigma_i$, we use $\tilde{\Sigma}_i$ above:

$$\tilde{\Sigma}_i = (1 - \lambda)\Sigma_i + \lambda sI$$
Determining $\lambda$

- The regularization parameter is a free “tunable” parameter of the approach, depends on the data.
- Can be found by parameter search (one cross-validation for each possible value) over a value range like [0.0 0.1 0.2 ... 0.9 1.0]
- **Caveat:** Parameter search can be very slow (10 possible values times 5 folds = 50x slower)
- Especially if nested inside an outer cross-validation
Determining $\lambda$

- In the special case of shrinkage LDA, $\lambda$ can be determined analytically or as the result of a convex optimization problem.
- Some further choices exist (e.g., empirical Bayes estimator, information criteria, ...).
3 Review
Resulting Spatial Filters

- Topographically mapped, the following filters emerge:

  - Window 1 (0.25s to 0.3s)
  - Window 2 (0.3s to 0.35s)
  - Window 3 (0.35s to 0.4s)
  - Window 4 (0.4s to 0.45s)
  - Window 5 (0.45s to 0.5s)
  - Window 6 (0.5s to 0.55s)
  - Window 7 (0.55s to 0.6s)
How Good is This Approach?

- Source activation $S$ can be recovered from sensor measurements by a linear mapping if (linear) volume conduction is invertible ($S = WX$)
How Good is This Approach?

• Source activation $S$ can be recovered from sensor measurements by a linear mapping if (linear) volume conduction is invertible ($S = WX$)

• Assuming a jointly Gaussian noise process and a noise distribution that is independent of the condition (A/B), LDA approximates the optimal linear mapping

• Shrinkage LDA on these features yields state-of-the-art ERP performance!
How Good is This Approach?

• Linear classifiers like LDA can operate implicitly on source ERPs, but:
  – EEG variation is often not Gaussian
  – Data variability can depend significantly on condition
  – For limited data samples, LDA is not necessarily optimal
  – Results are only “mildly” interpretable…
4 Advanced ERP Topics
Equivalence under Linear Transforms

- **Note:** LDA on linear features yields the same result (but linearly transformed) with the same performance when applied to any (non-reductive) linear transformation of the data
  - Principle Component Analysis, Independent Component Analysis, Non-adaptive Beamforming

- **But:** These can be used to
  - better interpret or localize underlying sources of a classifier, e.g., artifact/non-artifact components
  - introduce location-dependent constraints or prior knowledge into the classifier
Other Linear Features

- Wavelet transforms of the source time course
- Allow to design features adapted to intricate temporal characteristics of the signal (e.g. ripple, rebound, etc.)
- Can design generic features and employ feature-selection or sparse classification techniques (more later)
Non-Linear Features

- Extracting non-linear source signal features is not easy to get right on channel data.
- In theory, non-linear classifiers could recover such source features, but in practice most fail to capture the necessary structure for the given amount of data.
- Can be handled by a latent-variable model that represents source signals explicitly (more later) such as certain 3+ layer neural networks.
- **Examples:** relative measures (e.g., amplitude ratios), effective connectivity, ...
Signal Detection Aspects

• ERP analysis often amounts to classifying a characteristic ERP vs. a non-ERP / background noise where class ratios are often very imbalanced (e.g., RSVP target detection tasks)
• In such cases other evaluation measures than mis-classification rate rates are needed
• A canonical example are different costs per failure type (e.g., high false negative costs) if such costs are known
Signal Detection Aspects

• A general-purpose measure is Area under Receiver Operator Characteristic (AUC or AUROC) – quantifies performance over all cost choices.

• Can be approximated efficiently for given targets and associated predictions.
Impact on the Classifier Choice

• Most classifiers allow in principle for weighted cost structure, if known (e.g. LDA, logistic regression, Support Vector Machines)

• **Caveat:** Most classifiers assume that the class ratio in the training data equals their prior probability on test data (e.g., logistic regression)

• Some classifiers can be directly trained to optimize the AUC criterion (e.g. boosting, SVMperf) and there are ways to use any binary classifier (active research topic)
Questions?