One Step Feature Extraction and Classification with Tikhonov Regularization for BCI

Arun K Bharathan

NSS College of Engineering, Palakkad

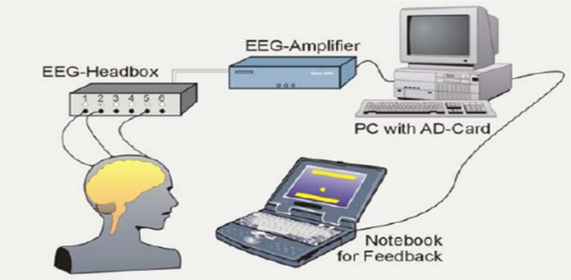
arunkbharathan@gmail.com

December 13, 2013

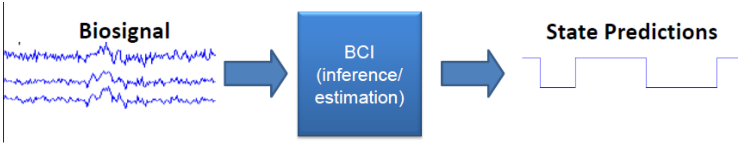
Overview

Brain Computer Interface

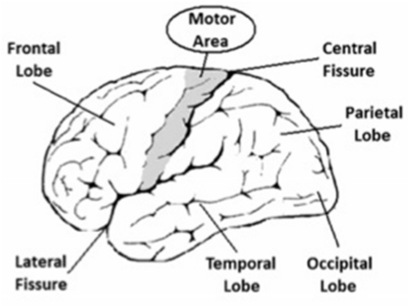
- Brain Computer Interface (BCI) is a hardware and software communications system that permits cerebral activity alone to control computers or external devices.
- The immediate goal of BCI research is to provide communications capabilities to severely disabled people who are totally paralysed or 'locked in' by neurological neuromuscular disorders.



As a Signal Processing Scheme



Brain Rhythms



- Delta (less than 4 Hz)
- Theta (4 to 7 Hz)
- Alpha (8 to 12 Hz) and Mu (7 to 13 Hz)
- Beta (12 to 30 Hz)
- Gamma (30 to 100 Hz)

Events

- Different parts of brain generate oscillations when idle.
- The amplitude of oscillations starts to decrease when we think of moving a limb.
- It reaches a minimum just after the onset of motion called Event Related De-synchronization (ERD), then it reverts back called Event Related Synchronization (ERS)

Steps that form a standard BCI

1. **Signal Acquisition**
2. **Pre-processing**
3. **Feature Extraction**
4. **Classification**
5. **Control Interface**

Signal Acquisition

Dataset Description

BCI Competition IV Dataset 4a

- 4 class EEG dataset with 22 electrodes and 3 EOG electrodes.
- Left hand [Class 1], Right hand [Class 2], Feet [Class 3], Tongue [Class 4].
- 9 subjects, each with 72 testing and 72 training trials.
- Sampled at 250Hz, BPF at 0.5Hz and 100Hz and a notch filter at 50Hz.

Cont.

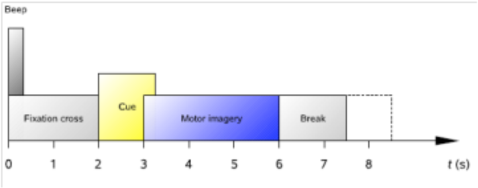


Figure : Timing Scheme

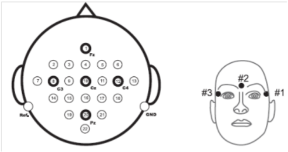


Figure : Electrode Positions

Pre-processing

EEG

- Only left hand ($Y = -1$) and right hand ($Y = +1$) trials of 9 subjects were used.
- RLS filtered EEG using EOG signals and BPF at 7-30 Hz.
- 0.5 to 3.5 sec epoch was cut out after visual cue.
- Down-sampled to 100 Hz and whitened the spatial covariance matrices.

Feature Extraction

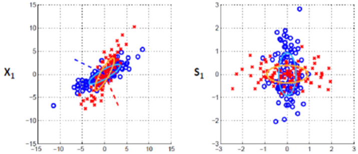
Common Spatial Patterns

- A data dependent spatial filtering.
- Can exploit ERS and ERD localized in the motor cortex.
- Simple, fast and relatively robust.
- Projects multichannel EEG signals into a subspace, where differences are maximized and similarities are minimized.
- Works on normalized spatial covariance matrix.

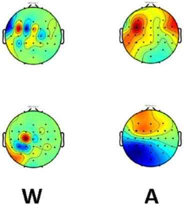
$$C = \frac{X^T X}{trace\left(X^T X\right)}$$

- $X \in \Re^{T \times C}$ $C \in \Re^{C \times C}$.
- $$f_\theta\left(X\right) = \sum_{j=1}^{J} \beta_j \log\left(w_j^T X^T B_j B_j^T X w_j\right) + \beta_0$$

Common Spatial Patterns



- $X = AS, \quad X, S \in \Re^{T \times C}, \;\; A \in \Re^{C \times C}$
- $S = WX, \quad X, S \in \Re^{T \times C}, \;\; W \in \Re^{C \times C}$
  - A is forward mapping matrix.
  - W is inverse mapping matrix.



Common Spatial Patterns

CSP spatial filters

- W is the generalized eigenvectors of covariance matrices C_1 and $(C_1 + C_2)$
- By optimization approach
  - $$w_c = \max_{w} \frac{w^T C_1 w}{w^T C_2 w} \;\; s.t. \; w^T C_2 w = 1$$
- But we take only 1-3 pair of filters and patterns from either side of W and A, respectively.
- If we take only 1 pair
  - $$X_{T \times C} \times W_{C \times 2} = S_{T \times 2}$$
- Taking log variance of S, LDA has to learn only 3 parameters (including bias term).

CSP with Tikhonov Regularization

$$w_c = \max_w \frac{w^T C_1 w}{w^T C_2 w + \alpha w^T I w} \;\; s.t. \; w^T C_2 w = 1$$

- 'C_x' is the average of covariance matrices of class 1 and 2.
- Where 'I' is identity matrix or any diagonal matrix that encode channel prior.
- The above regularization is equal to minimizing the squared euclidean norm of each channel.
- α is the regularization parameter, that is to be fine tuned by cross validation on training data.
- α fixes how much it should believe between identity matrix and covariance matrix.

Cont.

- To prevent over-fitting in CSP due to short data set or noisy data set.
- Restricts w to have small norms.
- Performs better than CSP for noisy and short data set.
- The weight matrix W is can be learned by taking eigenvector of:
- $$(C_2 + \alpha I)^{-1} C_1 \; and \; (C_1 + \alpha I)^{-1} C_2$$

One Step Feature Extraction and Classification Framework

- The method learns in a single step both the spatial filters and the relative weights.
- A unified, globally optimal solution to spatial filter estimation (an alternative to CSP+LDA).
- This is an optimization-based approach.
- Can impose various regularizers and loss terms while optimizing.
- The loss function is defined as the negative logarithmic pay off.
- Using various regularizers different assumptions on the structure of data can be made.

The framework consists of three components.

1. Probabilistic predictor model.
2. Detector function.
3. Regularization.

Detector Function

- Uses a CSP based detector function.
  - $$f_\theta(X) = \sum_{j=1}^{J} \beta_j \log\left(w_j^T X^T B_j B_j^T X w_j\right) + \beta_0$$
  - $$f_\theta(X) = \langle w^T w, X^T X \rangle + b$$
  - $$f_\theta(X) = \langle W, C \rangle + b$$
- The detector function outputs a positive number if $y = +1$ is more likely than $y = -1$ and vice versa.

Probabilistic Model

- Logistic model generates probabilistic output from detector function output.
- The logistic model converts it into the probability of $y = +1$ given X and vice versa.
  - $$q_\theta\left(Y = y \mid X\right) = \frac{1}{1 + \exp\left(-y f_\theta\left(X\right)\right)} \quad \left(y \in \{+1, -1\}\right)$$
- The $-\log\left(q_\theta\left(Y = y \mid X\right)\right)$ is set as loss function.
- Loss is smaller if the predictor predicts the actual intention of the user with high confidence.

Regularization

- Can impose various prior assumption on learned weight matrix W.
- W is forced to be a diagonal matrix and $trace(W)$ is added as regularization.
- since W is ww^T, the ($loss\ function + regularization$) minimizes the ℓ_2 distance.

Results (Cont.)

Results for EEG data set in percentage.

Table : Performance Comparison of TRCSP Using One Step Process and by eigen decomposition

	One Step TRCSP	TRCSP
A1	85.68	88.89
A2	60.42	54.17
A3	90.72	96.53
A4	73.61	70.83
A5	58.33	62.50
A6	70.56	67.36
A7	81.94	81.25
A8	89.56	95.87
A9	90.00	91.67

Results (Cont.)

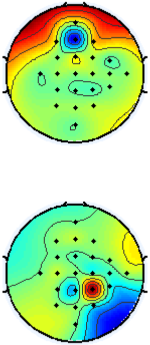


Figure : a

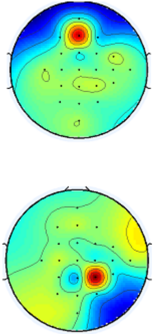


Figure : b

- (a) Spatial filter learned by generalised eigenvalue method.
- (b) Spatial filter learned by one step method.

Conclusion

The tikhonov regularized common spatial pattern filters were learned using one step feature extraction and classification framework with on an average similar performance.

References

Nicolas-Alonso, Luis Fernando, and Jaime Gomez-Gil (2012)
Brain computer interfaces, a review
Sensors 12.2, 1211-1279.

Wang, D., Miao, D., and Blohm, G. (2012)
Multi-class motor imagery EEG decoding for brain-computer interfaces
Frontiers in neuroscience 6.

Lotte, Fabien, and Cuntai Guan (2011)
Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms.
Biomedical Engineering, IEEE Transactions on 58, no. 2: 355-362.

Lotte, Fabien, and Cuntai Guan (2010)
Spatially regularized common spatial patterns for EEG classification.
International Conference on Pattern Recognition pp. 3712-3715.

Tomioka, Ryota, and Kazuyuki Aihara. (2007)
Classifying matrices with a spectral regularization
In Proceedings of the 24th international conference on Machine learning pp. 895-902.

Tomioka, R., and Mller, K. R. (2010)
A regularized discriminative framework for EEG analysis with application to brain-computer interface
Neuroimage pp 49(1), 415-432.

Fukunaga, Keinosuke (1990)
Introduction to statistical pattern recognition
Introduction to statistical pattern recognition

Ramoser, H., Muller-Gerking, J., and Pfurtscheller, G. (2000).
Optimal spatial filtering of single trial EEG during imagined hand movement
Rehabilitation Engineering, IEEE Transactions on, 8(4) pp.441-446.

References

Lewis, Adrian S., and Michael L. Overton. (1996)
Eigenvalue optimization.
*Acta numerica 5 no.1*pp. 149-190.

Vandenberghe, Lieven, and Stephen Boyd. (1996)
Semidefinite programming
SIAM review 38, no. 1 pp. 49-95.

Stephen Boyd, and Lieven Vandenberghe (2004)
Convex optimization
Convex optimization

Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Muller, K. R. (2008).
Optimizing spatial filters for robust EEG single-trial analysis
Signal Processing Magazine IEEE, 25(1), 41-56.

Tomioka, Ryota, Kazuyuki Aihara, and Klaus-Robert Mller. (2007)
Logistic regression for single trial EEG classification.
Advances in neural information processing systems vol 19 pp. 1377-1384.

He, P., G. Wilson, and C. Russell(2004)
Removal of ocular artifacts from electro-encephalogram by adaptive filtering.
*Medical and biological engineering and computing*vol 48 no. 3: 407-412.

Grant, Michael, Stephen Boyd, and Yinyu Ye (2008)
CVX: Matlab software for disciplined convex programming
CVX: Matlab software for disciplined convex programming

Delorme, Arnaud, and Scott Makeig (2004)
EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis.
Journal of neuroscience methods 134.1 (2004): 9-21.

The End