Introduction to Common Spatial Pattern Filters for EEG Motor Imagery Classification

Tatsuya Yokota

Tokyo Institute of Technology

July 17, 2012

Outline

Motor Imagery [Hema et al., 2009] I

Motor Imagery is a mental process of a motor action. It includes preparation for movement, passive observations of action and mental operations of motor representations implicitly or explicitly. The ability of an individual to control his EEG through imaginary mental tasks enables him to control devices through **a brain machine interface (BMI)** or **a brain computer interface (BCI)**.
The **BMI/BCI** is a direct communication pathway between the brain and an external device. BMI/BCIs are often aimed at assisting, augmenting, or repairing human cognitive or sensory-motor functions. BMI/BCIs are used to rehabilitate people suffering from neuromuscular disorders as a means of communication or control.
The methodologies of BMI/BCIs can be separated into two approaches:

- Invasive (or partially-invasive) BMI/BCIs (ECoG),
- Non-invasive BMI/BCIs (EEG, MEG, MRI, fMRI etc).

Invasive and partially-invasive BCIs are accurate. However there are risks of the infection and the damage. Furthermore, it requires the operation to set the electrodes in the head.
On the other hand, non-invasive BCIs are inferior than invasive BCIs in accuracy, but costs and risks are very low. Especially, EEG approach is the most studied potential non-invasive interface, mainly due to its fine temporal resolution, ease of use, portability and low set-up cost.

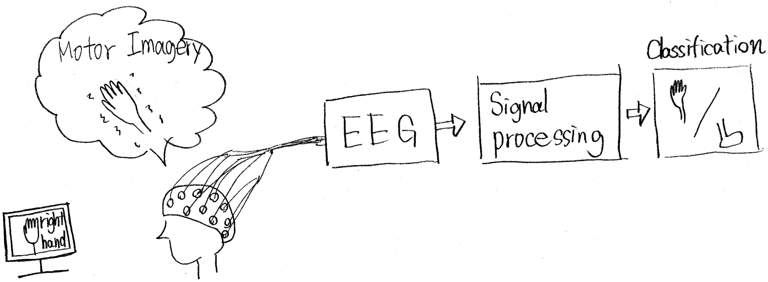
EEG is the recording of electrical activity along the scalp. EEG measures voltage fluctuations resulting from ionic current flows within the neurons of the brain.

(a) Electrodes(32 channels)

(b) EEG data

In this research, we analyze the EEG signals to extract the important features of them.

EEG Motor Imagery Classification



- Here, we consider the EEG motor imagery classification problem by BCI competition III [Blankertz et al., 2006].
- Number of subjects is **5 (aa,al,av,aw,ay)**.
- **118 channels** of electrodes were used.
- The problem is to classify the given EEG signal to **right hand** or **right foot**.
- They recorded their EEG signals for **3.5 seconds** with **100 Hz** sampling rates for each trial.
- For each subject they conducted **280 (hand 140 / foot 140)** trials.

Problem Formalization

EEG signals are formalized as

$$\{\boldsymbol{E}_n\}_{n=1}^{N} \in \mathbb{R}^{ch \times time}, \tag{1}$$

where

- $N = 280$ is the number of trials,
- $ch = 118$ is the number of channels,
- $time = 300$ is the range of time domain.

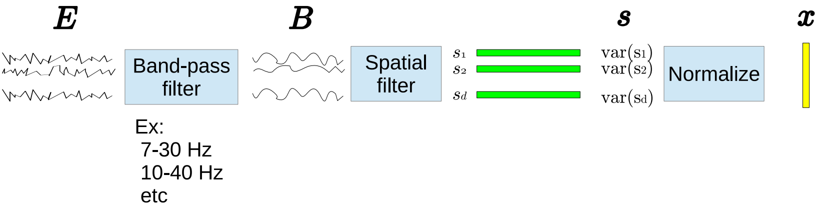
We used 0.5 to 3.5 seconds from visual cue for each trial, then time range is 3.0 seconds.
To apply the EEG signals to classification, we have to transform the each EEG signal to feature vector. Thus, some transformation

$$\boldsymbol{E}_n \in \mathbb{R}^{ch \times time} \mapsto \boldsymbol{x}_n \in \mathbb{R}^{d}, \tag{2}$$

is necessary.
The problem is how to extract the important features for best classification.

Feature Extraction Scheme

We explain about a basic feature extraction scheme at first



The key-points here are

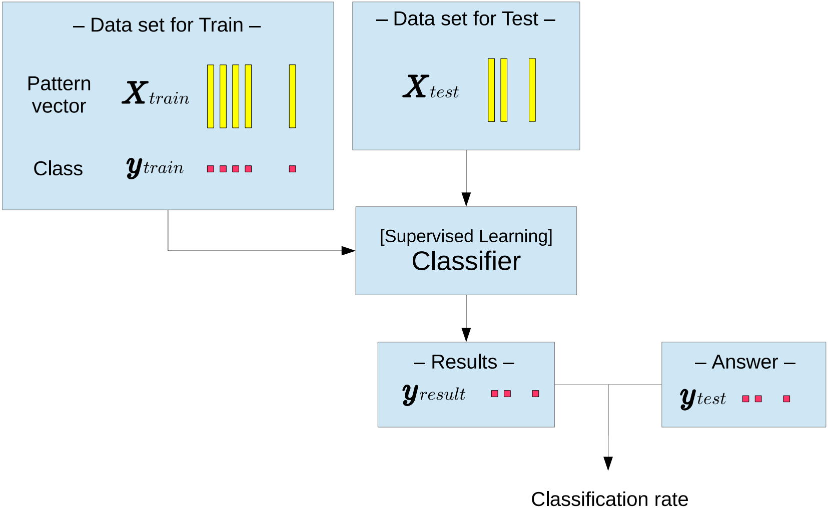
- EEG signals are very noisy, i.e., **noise reduction**,
- There will exist the best frequency band for classification, i.e., **frequency band selection**,
- There will exist important channels and un-important channels for classification, i.e., **channel selection (spatial filter)**.

We can obtain a set of feature vector as a matrix

$$\boldsymbol{X} \in \mathbb{R}^{d \times N}. \tag{3}$$

Classification Scheme

The feature extraction is done, then we can apply the data set to the classification.

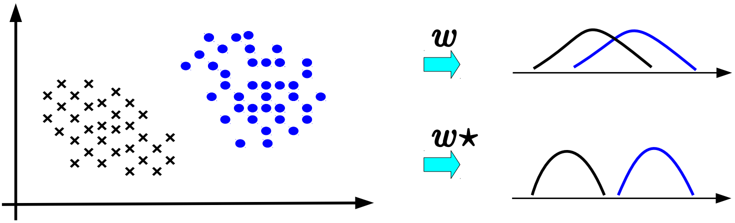


Fisher's Linear Discriminant Analysis/Classifier I

We introduce the Fisher's linear discriminant analysis (LDA). The LDA is a very famous binary classification method. It is based on mean vectors and covariance matrices of patterns for individual classes.
we consider to transform a d-dimensional vector $\boldsymbol{x}$ to a scalar z as

$$z = \boldsymbol{w}^T \boldsymbol{x}. \tag{4}$$

The LDA give an optimal projection $\boldsymbol{w}$ so that the distribution of z is easy to discriminate. See the following figure. $\boldsymbol{w}^\star$ is clearly easier to discriminate than $\boldsymbol{w}$.



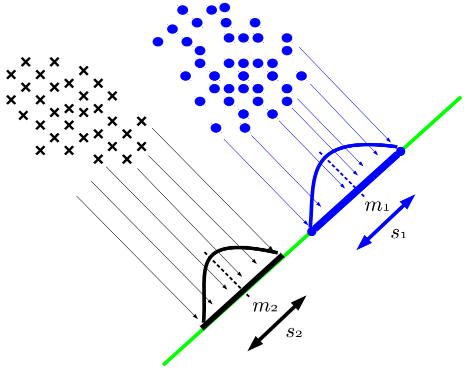
Fisher's Linear Discriminant Analysis/Classifier II

The criterion of LDA is given by

$$\text{maximize} \quad J(\boldsymbol{w}) = \frac{(m_1 - m_2)^2}{s_1 + s_2}, \tag{5}$$

where

- m_1 and m_2 denote averages for $z_n \in \{$ class 1 $\}$ and $z_n \in \{$ class 2 $\}$, respectively,
- s_1 and s_2 denote variances for $z_n \in \{$ class 1 $\}$ and $z_n \in \{$ class 2 $\}$, respectively.



Fisher's Linear Discriminant Analysis/Classifier III

We have

$$
\begin{aligned}
(m_1 - m_2)^2 &= \left(\boldsymbol{w}^T \boldsymbol{\mu}_1 - \boldsymbol{w}^T \boldsymbol{\mu}_2\right) \left(\boldsymbol{w}^T \boldsymbol{\mu}_1 - \boldsymbol{w}^T \boldsymbol{\mu}_2\right)^T \\
&= \boldsymbol{w}^T \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right) \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^T \boldsymbol{w} = \boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w},
\end{aligned} \tag{6}
$$

and

$$
\begin{aligned}
s_1 + s_2 &= \boldsymbol{w}^T \boldsymbol{\Sigma}_1 \boldsymbol{w} + \boldsymbol{w}^T \boldsymbol{\Sigma}_2 \boldsymbol{w} \\
&= \boldsymbol{w}^T \left(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right) \boldsymbol{w} = \boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w},
\end{aligned} \tag{7}
$$

where

- $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ denote mean vectors for $\boldsymbol{x}_n \in \{$ class 1 $\}$ and $\boldsymbol{x}_n \in \{$class 2 $\}$, respectively,
- $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ denote covariance matrices for $\boldsymbol{x}_n \in \{$ class 1 $\}$ and $\boldsymbol{x}_n \in \{$ class 2 $\}$, respectively.

Fisher's Linear Discriminant Analysis/Classifier IV

The cost function $J(\boldsymbol{w})$ can be transformed as

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \boldsymbol{S}_B \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{S}_W \boldsymbol{w}}, \tag{8}$$

and then this solution is given by

$$\widehat{\boldsymbol{w}} \propto \boldsymbol{S}_W^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2). \tag{9}$$

Finally, we choose an optimal threshold z_0, then we can classify any $\boldsymbol{x}$ by

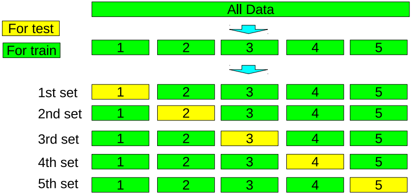
$$\widehat{\boldsymbol{w}}^T \boldsymbol{x} \geq z_0 \rightarrow \boldsymbol{x} \in \{\text{class 1}\}, \tag{10}$$

$$\widehat{\boldsymbol{w}}^T \boldsymbol{x} < z_0 \rightarrow \boldsymbol{x} \in \{\text{class 2}\}. \tag{11}$$

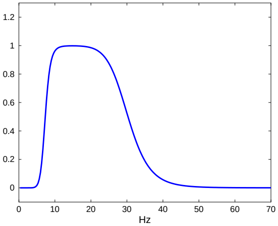
For example, $z_0 = (m_1 + m_2)/2$ could be usable.

An Example

- 7-30 Hz band-pass filter (butterworth filter)
- Use variances of individual channel's signals (i.e., $\boldsymbol{x}$ is a 118-dimensional feature vector)
- Normalize
- Apply the LDA into the data set
- Calculate classification rate by 5-fold cross validation for each subject



- Separate a data set into 5 parts
- Use one for test, the others for train one by one
- Calculate classification rates for individual CV data set
- Summarize the 5 results



aa	68.9 ± 7.3
al	83.9 ± 6.4
av	58.6 ± 2.6
aw	82.9 ± 2.7
ay	76.8 ± 5.8
ave	74.2

Common Spatial Pattern Filter I

In previous example, we do not use a spatial filter. The proper spatial filter would provide signals so that easy to classify. The goal of this study is to design spatial filters that lead to optimal variances for the discrimination of two populations of EEG related to right hand and right foot motor imagery. We call this method the "Common Spatial Pattern" (CSP) algorithm [Muller-Gerking et al., 1999].
We denote the CSP filter by

$$\boldsymbol{S} = \boldsymbol{W}^T \boldsymbol{E} \quad \text{or} \quad \boldsymbol{s}(t) = \boldsymbol{W}^T \boldsymbol{e}(t), \tag{12}$$

where $\boldsymbol{W} \in \mathbb{R}^{d \times ch}$ is spatial filter matrix, $\boldsymbol{S} \in \mathbb{R}^{d \times time}$ is filtered signal matrix.
The criterion of CSP is given by

$$\text{maximize} \quad \text{tr}\boldsymbol{W}^T \boldsymbol{\Sigma}_1 \boldsymbol{W}, \tag{13}$$

$$\text{subject to} \quad \boldsymbol{W}^T (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2) \boldsymbol{W} = \boldsymbol{I}, \tag{14}$$

where

$$\boldsymbol{\Sigma}_1 = \underset{\boldsymbol{E}_n \in \{\text{class } 1\}}{\text{Exp}} \frac{\boldsymbol{E}_n \boldsymbol{E}_n^T}{\text{tr}\boldsymbol{E}_n \boldsymbol{E}_n^T}, \tag{15}$$

$$\boldsymbol{\Sigma}_2 = \underset{\boldsymbol{E}_n \in \{\text{class } 2\}}{\text{Exp}} \frac{\boldsymbol{E}_n \boldsymbol{E}_n^T}{\text{tr}\boldsymbol{E}_n \boldsymbol{E}_n^T}. \tag{16}$$

This problem can be solved by **generalized eigen value problem**. However, we can also solve it by **two times of standard eigen value problem**.

Common Spatial Pattern Filter II

First we decompose as

$$\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2 = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T, \tag{17}$$

where $\boldsymbol{U}$ is a set of eigenvectors, and $\boldsymbol{D}$ is a diagonal matrix of eigenvalues.
Next, compute $\boldsymbol{P} := \sqrt{\boldsymbol{D}^{-1}}\boldsymbol{U}^T$, and

$$\widehat{\boldsymbol{\Sigma}}_1 = \boldsymbol{P}\boldsymbol{\Sigma}_1\boldsymbol{P}^T, \tag{18}$$

$$\widehat{\boldsymbol{\Sigma}}_2 = \boldsymbol{P}\boldsymbol{\Sigma}_2\boldsymbol{P}^T. \tag{19}$$

Please note that we have $\widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_2 = \boldsymbol{I}$, here. Thus, any orthonormal matrices $\boldsymbol{V}$ satisfy $\boldsymbol{V}^T(\widehat{\boldsymbol{\Sigma}}_1 + \widehat{\boldsymbol{\Sigma}}_2)\boldsymbol{V} = \boldsymbol{I}$.
Finally, we decompose as

$$\widehat{\boldsymbol{\Sigma}}_1 = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T, \tag{20}$$

where $\boldsymbol{V}$ is a set of eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues.
A set of CSP filters is obtained as

$$\boldsymbol{W} = \boldsymbol{P}^T\boldsymbol{V}. \tag{21}$$

Common Spatial Pattern Filter III

We have

$$\boldsymbol{W}^T \boldsymbol{\Sigma}_1 \boldsymbol{W} = \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{ch} \end{pmatrix}, \tag{22}$$

$$\boldsymbol{W}^T \boldsymbol{\Sigma}_2 \boldsymbol{W} = \boldsymbol{I} - \boldsymbol{\Lambda} = \begin{pmatrix} 1-\lambda_1 & & \\ & \ddots & \\ & & 1-\lambda_{ch} \end{pmatrix}, \tag{23}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{ch}$. Therefore, first CSP filter $\boldsymbol{w}_1$ provides maximum variance of class 1, and last CSP filter $\boldsymbol{w}_{ch}$ provides maximum variance of class 2.

We select first and last m filters to use as

$$\boldsymbol{W}_{csp} = \begin{pmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_m & \boldsymbol{w}_{ch-m+1} & \cdots & \boldsymbol{w}_{ch} \end{pmatrix} \in \mathbb{R}^{2m \times ch}, \tag{24}$$

and filtered signal matrix is given by

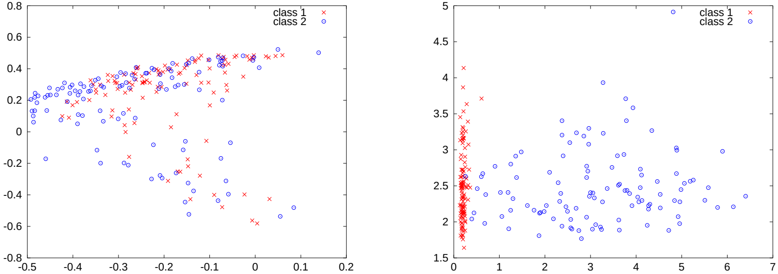
$$\boldsymbol{s}(t) = \boldsymbol{W}_{csp}^T \boldsymbol{e}(t) = \begin{pmatrix} s_1(t) & \cdots & s_d(t) \end{pmatrix}^T, \tag{25}$$

i.e., $d = 2m$.

Feature vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^T$ is calculated by

$$x_i = \log \left(\frac{\mathrm{var}[s_i(t)]}{\sum_{i=1}^{d} \mathrm{var}[s_i(t)]} \right). \tag{26}$$

Common Spatial Pattern Filter IV

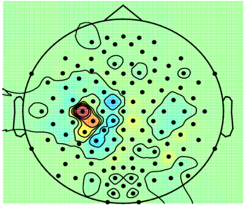


(c) PCA feature

(d) CSP feature

Figure 1: 2D-plot of CSP feature

Common Spatial Pattern Filter V


(a) Right Hand (al)

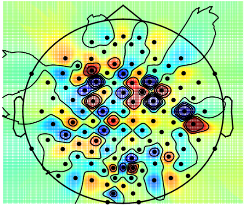

(b) Right Foot (al)

Figure 2: Topographies of Common Spatial Patterns

Common Spatial Pattern Filter VI

- We extracted their features by the CSP filter and classified them by the LDA.
- Changing the frequency band of butterworth filter, and number of m.

band	m	aa	al	av	aw	ay	ave
7-30 Hz	non-CSP	68.9 ± 7.3	83.9 ± 6.4	58.6 ± 2.6	82.9 ± 2.7	76.8 ± 5.8	74.2
5-40 Hz	1	76.1 ± 5.4	98.2 ± 1.8	52.1 ± 7.9	87.1 ± 1.5	**90.4 ± 3.2**	80.8
5-40 Hz	10	73.9 ± 7.2	98.6 ± 1.5	69.3 ± 2.3	88.6 ± 3.9	86.1 ± 4.1	83.3
7-30 Hz	1	78.6 ± 4.2	**98.9 ± 1.6**	53.2 ± 8.7	92.1 ± 2.0	**90.4 ± 3.7**	82.6
7-30 Hz	10	72.5 ± 7.8	**98.9 ± 1.6**	**72.9 ± 2.3**	96.1 ± 2.9	87.5 ± 3.3	**85.6**
10-30 Hz	1	80.4 ± 5.9	98.6 ± 1.5	51.1 ± 7.4	95.4 ± 1.0	84.6 ± 5.4	82.0
10-30 Hz	10	**81.4 ± 3.5**	**98.9 ± 1.6**	72.1 ± 2.4	98.2 ± 2.2	73.9 ± 5.7	84.9
10-35 Hz	1	80.7 ± 5.4	98.6 ± 1.5	50.7 ± 8.2	96.1 ± 2.3	87.1 ± 3.7	82.6
10-35 Hz	10	78.9 ± 4.6	98.6 ± 1.5	72.1 ± 5.1	**97.5 ± 2.4**	75.0 ± 4.2	84.4

- CSP methods are basically superior to non-CSP method
- In ‘aa’, and ‘ay’, $m = 1$ is matched
- In ‘av’ and ‘aw’, $m = 10$ is matched
- Frequency band is also important factor

Common Spatio-Spectral Pattern Filter I

The Common Spatio-Spectral Pattern (CSSP) filter is an extension of the CSP filter [Lemm et al., 2005]. The CSSP can be regarded as a CSP method with the **time delay embedding**.

The algorithm is not so different from the standard CSP. In CSP, we consider the following transformation

$$\boldsymbol{S} = \boldsymbol{W}^T \boldsymbol{E} \quad \text{or} \quad \boldsymbol{s}(t) = \boldsymbol{W}^T \boldsymbol{e}(t). \tag{27}$$

But the CSSP's transform is given by

$$\boldsymbol{S} = \boldsymbol{W}^T \boldsymbol{E} + \boldsymbol{W}_\tau^T \boldsymbol{E}_\tau = \widehat{\boldsymbol{W}}^T \begin{pmatrix} \boldsymbol{E} \\ \boldsymbol{E}_\tau \end{pmatrix}, \tag{28}$$

$$\text{or} \quad \boldsymbol{s}(t) = \boldsymbol{W}^T \boldsymbol{e}(t) + \boldsymbol{W}_\tau^T \boldsymbol{e}(t+\tau) = \widehat{\boldsymbol{W}}^T \begin{pmatrix} \boldsymbol{e}(t) \\ \boldsymbol{e}(t+\tau) \end{pmatrix}, \tag{29}$$

where $\boldsymbol{E}_\tau$ is a τ-time delayed signal matrix of $\boldsymbol{E}$, and $\widehat{\boldsymbol{W}}^T = [\boldsymbol{W}^T,\ \boldsymbol{W}_\tau^T]$ is a CSSP matrix. It can be also regarded that the number of channels increase to double. The deference between CSP and CSSP is only this point, and then we can apply this method easily in a same way to the CSP algorithm. However, τ is a hyper-parameter.

Common Spatio-Spectral Pattern Filter II

The key-point here is this method can be interpreted into a spatial and a spectral filter. Therefore let $\widehat{\boldsymbol{w}}$ denote the i-th column of the CSSP matrix $\widehat{\boldsymbol{W}}$, then the projected signal $s(t)$ can be expressed as

$$\begin{aligned} s(t) &= \boldsymbol{w}^T \boldsymbol{e}(t) + \boldsymbol{w}_\tau^T \boldsymbol{e}(t+\tau) && (30) \\ &= \sum_{j=1}^{ch} w_j e_j(t) + (w_\tau)_j e_j(t+\tau) && (31) \\ &= \sum_{j=1}^{ch} \gamma_j \left(\frac{w_j}{\gamma_j} e_j(t) + \frac{(w_\tau)_j}{\gamma_j} e_j(t+\tau) \right), && (32) \end{aligned}$$

where

- γ_j can be regarded as **a pure spatial filter**,
- $\left[\frac{w_j}{\gamma_j},\ 0, \ldots, 0,\ \frac{(w_\tau)_j}{\gamma_j} \right]$ can be regarded as a **finite impulse response (FIR) filter** for each channel.

Common Spatio-Spectral Pattern Filter III

Table 1: Experimental results

band	m	τ	aa	al	av	aw	ay	ave
7-30 Hz	non-CSP		68.9 ± 7.3	83.9 ± 6.4	58.6 ± 2.6	82.9 ± 2.7	76.8 ± 5.8	74.2
7-30 Hz	1		78.6 ± 4.2	**98.9 ± 1.6**	53.2 ± 8.7	92.1 ± 2.0	90.4 ± 3.7	82.6
7-30 Hz	10		72.5 ± 7.8	**98.9 ± 1.6**	**72.9 ± 2.3**	**96.1 ± 2.9**	87.5 ± 3.3	**85.6**
7-30 Hz	1	5	80.0 ± 5.1	98.6 ± 2.0	52.5 ± 5.7	92.5 ± 2.3	89.3 ± 2.8	82.6
7-30 Hz	10	5	72.1 ± 6.8	98.2 ± 1.3	65.0 ± 7.5	91.8 ± 2.7	88.6 ± 2.0	83.1
7-30 Hz	1	10	82.9 ± 5.7	98.6 ± 1.5	53.2 ± 5.1	94.3 ± 2.0	90.4 ± 2.4	83.9
7-30 Hz	10	10	80.7 ± 5.6	97.9 ± 2.0	63.6 ± 5.1	93.2 ± 2.0	87.5 ± 4.0	84.6
7-30 Hz	1	15	84.3 ± 7.1	98.6 ± 1.5	46.8 ± 8.0	93.2 ± 2.9	**91.4 ± 3.2**	82.9
7-30 Hz	10	15	76.4 ± 6.6	**98.9 ± 1.6**	66.4 ± 6.1	94.3 ± 3.9	85.4 ± 2.6	84.3
5-35 Hz	1		75.7 ± 7.2	98.2 ± 1.8	51.8 ± 7.9	86.4 ± 2.0	**91.4 ± 3.4**	80.7
5-35 Hz	10		70.7 ± 6.5	98.2 ± 1.8	**72.9 ± 4.3**	87.1 ± 5.6	87.1 ± 2.9	83.2
5-35 Hz	1	5	78.6 ± 2.8	98.2 ± 1.8	53.9 ± 7.9	91.1 ± 1.8	87.5 ± 4.6	81.9
5-35 Hz	10	5	67.9 ± 5.2	98.6 ± 1.5	67.9 ± 4.9	85.7 ± 4.2	85.0 ± 5.0	81.0
5-35 Hz	1	10	**85.7 ± 3.8**	98.2 ± 1.8	52.9 ± 6.5	91.1 ± 3.6	90.4 ± 3.0	83.6
5-35 Hz	10	10	77.9 ± 3.2	97.1 ± 2.7	60.4 ± 5.0	89.3 ± 2.2	85.4 ± 3.4	82.0
5-35 Hz	1	15	82.1 ± 5.6	98.6 ± 1.5	49.3 ± 5.9	88.9 ± 3.9	90.0 ± 2.0	81.8
5-35 Hz	10	15	76.1 ± 6.5	97.1 ± 3.0	61.8 ± 6.5	89.3 ± 3.6	87.1 ± 3.2	82.3

- 'aa' is obviously improved
- τ selection is necessary

Common Sparse Spatio Spectral Pattern Filter I

The Common Sparse Spatio Spectral Pattern (CSSSP) filter is a further extension of the CSSP [Dornhege et al., 2005]. In CSSP's FIR filter consists of only one time delay. The CSSSP's FIR filter is given by

$$\boldsymbol{f}(t|\boldsymbol{b}) = b_0\boldsymbol{e}(t) + b_1\boldsymbol{e}(t+\tau) + b_2 + \boldsymbol{e}(t+2\tau) + \cdots + b_T\boldsymbol{e}(t+T\tau) \tag{33}$$

so that $\boldsymbol{b}$ is sparse spectral filter. Therefore, final signals are given by

$$\boldsymbol{s}(t) = \boldsymbol{W}^T\boldsymbol{f}(t|\boldsymbol{b}) = \sum_{k=0}^{T} b_k\boldsymbol{W}^T\boldsymbol{e}(t+k\tau). \tag{34}$$

The criterion of CSSSP is given by

$$\max_{\boldsymbol{b}}\max_{\boldsymbol{W}} \boldsymbol{W}^T\left[\mathsf{Exp}_1\left\{\boldsymbol{f}(t|\boldsymbol{b})\boldsymbol{f}(t|\boldsymbol{b})^T\right\}\right]\boldsymbol{W} - \frac{C}{T}||\boldsymbol{b}||_1, \tag{35}$$

$$\text{subject to} \quad \boldsymbol{W}^T\left[\mathsf{Exp}_1\left\{\boldsymbol{f}(t|\boldsymbol{b})\boldsymbol{f}(t|\boldsymbol{b})^T\right\} + \mathsf{Exp}_2\left\{\boldsymbol{f}(t|\boldsymbol{b})\boldsymbol{f}(t|\boldsymbol{b})^T\right\}\right]\boldsymbol{W} = \boldsymbol{I} \tag{36}$$

where

- $\mathsf{Exp}_1\{\cdot\}$ and $\mathsf{Exp}_2\{\cdot\}$ denote the expectations for samples of class 1 and class 2, respectively.

Common Sparse Spatio Spectral Pattern Filter II

In this paper, the authors say that since the optimal $\boldsymbol{W}$ can be calculated by the usual CSP techniques for each $\boldsymbol{b}$ the problem for $\boldsymbol{b}$ remains which we can solve with usual line-search optimization techniques if T is not too large. In other words, we define the cost function as

$$\begin{aligned} J(\boldsymbol{b}) &:= \max_{\boldsymbol{W}} \boldsymbol{W}^T \left[\mathsf{Exp}_1 \left\{ \boldsymbol{f}(t|\boldsymbol{b}) \boldsymbol{f}(t|\boldsymbol{b})^T \right\} \right] \boldsymbol{W} - \frac{C}{T} ||\boldsymbol{b}||_1, \\ &\text{subject to} \quad \boldsymbol{W}^T \left[\mathsf{Exp}_1 \left\{ \boldsymbol{f}(t|\boldsymbol{b}) \boldsymbol{f}(t|\boldsymbol{b})^T \right\} + \mathsf{Exp}_2 \left\{ \boldsymbol{f}(t|\boldsymbol{b}) \boldsymbol{f}(t|\boldsymbol{b})^T \right\} \right] \boldsymbol{W} = \boldsymbol{I}, \end{aligned} \tag{37}$$

and maximize the $J(\boldsymbol{b})$ by line-search algorithms (e.g., the gradient method and the Newton's method).

Common Sparse Spatio Spectral Pattern Filter III

Table 2: Experimental results

band	m	τ	T	C	aa	al	av	aw	ay	ave
7-30 Hz	non-CSP				68.9 ± 7.3	83.9 ± 6.4	58.6 ± 2.6	82.9 ± 2.7	76.8 ± 5.8	74.2
7-30 Hz	1				78.6 ± 4.2	$\mathbf{98.9 \pm 1.6}$	53.2 ± 8.7	92.1 ± 2.0	90.4 ± 3.7	82.6
7-30 Hz	10				72.5 ± 7.8	$\mathbf{98.9 \pm 1.6}$	$\mathbf{72.9 \pm 2.3}$	$\mathbf{96.1 \pm 2.9}$	87.5 ± 3.3	**85.6**
7-30 Hz	1	15			84.3 ± 7.1	98.6 ± 1.5	46.8 ± 8.0	93.2 ± 2.9	$\mathbf{91.4 \pm 3.2}$	82.9
7-30 Hz	10	15			76.4 ± 6.6	$\mathbf{98.9 \pm 1.6}$	66.4 ± 6.1	94.3 ± 3.9	85.4 ± 2.6	84.3
5-35 Hz	1	10			$\mathbf{85.7 \pm 3.8}$	98.2 ± 1.8	52.9 ± 6.5	91.1 ± 3.6	90.4 ± 3.0	83.6
5-35 Hz	10	10			77.9 ± 3.2	97.1 ± 2.7	60.4 ± 5.0	89.3 ± 2.2	85.4 ± 3.4	82.0
5-35 Hz	1	5	3	0.1	61.1 ± 16.5	96.1 ± 5.0	53.6 ± 4.2	88.9 ± 4.3	55.4 ± 20.2	71.0
5-35 Hz	10	5	3	0.1	61.4 ± 10.4	94.6 ± 2.8	51.8 ± 8.5	91.4 ± 3.2	79.6 ± 10.6	75.8
5-35 Hz	1	10	3	0.1	48.9 ± 3.7	79.6 ± 23.1	52.5 ± 7.5	53.2 ± 4.3	53.9 ± 6.4	57.6
5-35 Hz	10	10	3	0.1	67.9 ± 8.4	95.4 ± 4.3	53.2 ± 10.4	77.9 ± 14.5	55.7 ± 13.3	70.0
5-35 Hz	1	5	10	0.1	50.0 ± 7.5	51.4 ± 5.0	50.4 ± 4.3	53.6 ± 9.4	51.4 ± 10.0	51.4
5-35 Hz	10	5	10	0.1	54.6 ± 6.3	55.4 ± 8.7	55.7 ± 6.6	52.9 ± 9.5	54.6 ± 7.8	54.6

- CSSSP results were not good
- The computational cost of CSSSP is very expensive compared with the CSP and the CSSP
- Many hyper-parameters; it is difficult to adjust all the parameters

Sub-band Common Spatial Pattern I

- The CSP, the CSSP need the selection of frequency band to use band-pass filter at first
- The CSSSP solutions are depend greatly on the initial points

Here, we introduce an alternative method based on Sub-band CSP (SBCSP) method and score fusion [Novi et al., 2007].

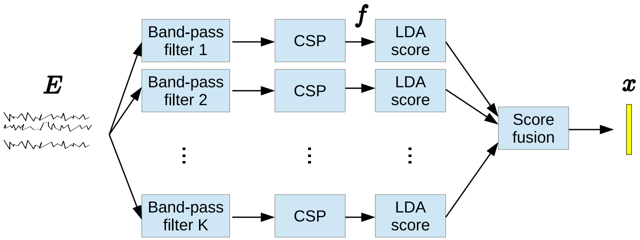


Figure 3: System Flowchart of SBCSP: for example frequency bands of individual filters are 4-8Hz, 8-12Hz, ... , 36-40Hz.

The processes until CSP filters are the same as it is. And then we obtain individual CSP feature data matrices $\{\boldsymbol{F}^{(1)}, \boldsymbol{F}^{(2)}, \ldots, \boldsymbol{F}^{(K)}\}$. The process of LDA score is given by follow:

Sub-band Common Spatial Pattern II

- Calculate the LDA projection vector $\widehat{\boldsymbol{w}}^{(k)}$ for each data matrices $\boldsymbol{F}^{(k)}$.
- LDA score vector is given by

$$\boldsymbol{s}_n = \begin{pmatrix} \widehat{\boldsymbol{w}}^{(1)T} \boldsymbol{f}_n^{(1)} \\ \widehat{\boldsymbol{w}}^{(2)T} \boldsymbol{f}_n^{(2)} \\ \vdots \\ \widehat{\boldsymbol{w}}^{(K)T} \boldsymbol{f}_n^{(K)} \end{pmatrix}. \tag{38}$$

There are two approaches of score fusion method.

- Recursive Band Elimination (RBE)
- Meta-Classifier (MC)

In RBE, first we set a RBE-order that is a integer number $r \in [1, K]$. The first data set is $\boldsymbol{X} = [\boldsymbol{s}_1, ..., \boldsymbol{s}_N] \in \mathbb{R}^{K \times N}$. In this method, we remove $K - r$ rows from $\boldsymbol{X}$ by SVM feature selection. The algorithm is follow

1. Train $\boldsymbol{w}$ by SVM
2. Remove the row of $\boldsymbol{X}$ with the smallest w_k^2
3. If number of rows of $\boldsymbol{X}$ is r algorithm is finished, else back to 1.

Sub-band Common Spatial Pattern III

The survival data $\boldsymbol{X}$ can be used as a training data matrix.
In MC, we use a Bayesian classifier for score fusion. We assume that each band score $s_k \in \{\text{class } 1\}$ and $s_k \in \{\text{class } 2\}$ are distributed normally. And we can estimate individual parameters of normal distributions (i.e., $\{\mu_1^{(k)}, \sigma_1^{(k)}\}$ and $\{\mu_2^{(k)}, \sigma_2^{(k)}\}$). We calculate the training data matrix by

$$x_{kn} = \log \left\{ \frac{p(s_{kn} | \mu_1^{(k)}, \sigma_1^{(k)})}{p(s_{kn} | \mu_2^{(k)}, \sigma_2^{(k)})} \right\}, \tag{39}$$

where x_{kn} is a (k, n)-element of $\boldsymbol{X}$.

Sub-band Common Spatial Pattern IV

Table 3: Experimental results

band	m	τ	aa	al	av	aw	ay	ave
7-30	non		68.9 ± 7.3	83.9 ± 6.4	58.6 ± 2.6	82.9 ± 2.7	76.8 ± 5.8	74.2
7-30	1		78.6 ± 4.2	98.9 ± 1.6	53.2 ± 8.7	92.1 ± 2.0	90.4 ± 3.7	82.6
7-30	10		72.5 ± 7.8	98.9 ± 1.6	$\mathbf{72.9} \pm \mathbf{2.3}$	96.1 ± 2.9	87.5 ± 3.3	85.6
7-30	1	15	84.3 ± 7.1	98.6 ± 1.5	46.8 ± 8.0	93.2 ± 2.9	$\mathbf{91.4} \pm \mathbf{3.2}$	82.9
7-30	10	15	76.4 ± 6.6	98.9 ± 1.6	66.4 ± 6.1	94.3 ± 3.9	85.4 ± 2.6	84.3
5-35	1	10	$\mathbf{85.7} \pm \mathbf{3.8}$	98.2 ± 1.8	52.9 ± 6.5	91.1 ± 3.6	90.4 ± 3.0	83.6
5-35	10	10	77.9 ± 3.2	97.1 ± 2.7	60.4 ± 5.0	89.3 ± 2.2	85.4 ± 3.4	82.0
SBCSP	m	r						
RBE	1	6	$\mathbf{85.7} \pm \mathbf{2.8}$	99.3 ± 1.0	48.9 ± 3.2	93.6 ± 2.4	89.3 ± 3.3	83.4
RBE	5	6	78.6 ± 6.7	98.9 ± 1.6	72.5 ± 4.7	95.4 ± 2.7	90.4 ± 3.9	87.1
RBE	10	6	79.3 ± 7.2	99.6 ± 0.8	62.9 ± 6.7	96.8 ± 2.3	85.7 ± 4.6	84.9
RBE	1	9	83.9 ± 4.7	99.3 ± 1.0	50.4 ± 5.8	94.3 ± 2.0	88.9 ± 3.9	83.4
RBE	5	9	80.0 ± 7.2	98.9 ± 1.6	71.4 ± 4.0	96.8 ± 2.9	87.9 ± 6.0	87.0
RBE	10	9	78.9 ± 7.6	$\mathbf{99.6} \pm \mathbf{0.8}$	64.3 ± 7.0	97.1 ± 2.0	86.1 ± 5.6	85.2
MC	1		85.0 ± 4.8	97.9 ± 0.8	47.9 ± 12.5	93.9 ± 2.4	90.7 ± 4.8	83.1
MC	5		81.1 ± 7.7	98.6 ± 1.5	71.1 ± 4.6	96.4 ± 3.6	90.0 ± 3.2	**87.4**
MC	10		80.0 ± 6.5	98.9 ± 1.0	62.9 ± 8.1	$\mathbf{98.6} \pm \mathbf{1.5}$	88.2 ± 5.6	85.7

- SBCSP provides stably and good results without frequency band selection
- Especially, there is only one hyper-parameter in MC approach

Summary

Summary

- The CSP filters is an effective method of feature extraction for EEG motor imagery classification.
- Frequency band selection is very important to obtain good classification in CSP.
- To overcome it, the CSSP, the CSSSP, and the SBCSP methods are proposed.
- In CSSP, there is a little improvement, but we still need to select a frequency band and τ.
- In CSSSP, hyper-parameters are too much, and it is difficult to adjust, and computational cost is also too much.
- In SBCSP, it provides a successful result that we do not need to select a frequency band, and hyper-parameter is only one or two.

Bibliography I

[Blankertz et al., 2006] Blankertz, B., Muller, K.-R., Krusienski, D., Schalk, G., Wolpaw, J. R., Schlogl, A., Pfurtscheller, G., del R. Millan, J., Schröder, M., and Birbaumer, N. (2006).
The bci competition iii: Validating alternative approaches to actual bci problems.
IEEE Trans. Neural Systems and Rihabilitation Engineering, 14:153–159.

[Dornhege et al., 2005] Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G., and robert Muller, K. (2005).
Optimizing spatio-temporal filters for improving brain-computer interfacing.
In *in Advances in Neural Inf. Proc. Systems (NIPS 05*, pages 315–322. MIT Press.

[Hema et al., 2009] Hema, C., Paulraj, M., Yaacob, S., Adom, A., and Nagarajan, R. (2009).
Eeg motor imagery classification of hand movements for a brain machine interface.
Biomedical Soft Computing and Human Sciences, 14(2):49–56.

[Lemm et al., 2005] Lemm, S., Blankertz, B., Curio, G., and Muller, K.-R. (2005).
Spatio-spectral filters for improving the classification of single trial eeg.
Biomedical Engineering, IEEE Transactions on, 52(9):1541 –1548.

[Minka et al., 1999] Minka, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K. (1999).
Fisher discriminant analysis with kernels.
In *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop*, pages 41 –48.

Bibliography II

[Muller-Gerking et al., 1999] Muller-Gerking, J., Pfurtscheller, G., and Flyvbjerg, H. (1999).
Designing optimal spatial filters for single-trial eeg classification in a movement task.
Clinical Neurophysiology, 110(5):787 – 798.

[Novi et al., 2007] Novi, Q., Guan, C., Dat, T. H., and Xue, P. (2007).
Sub-band common spatial pattern (sbcsp) for brain-computer interface.
In *Neural Engineering, 2007. CNE '07. 3rd International IEEE/EMBS Conference on*, pages 204 –207.

Thank you for listening