Composite Kernel SVM in Conjunction with Spatial Filter for Brain Computer Interface

Jyoti Singh Kirar, R. K. Agrawal

Biomedical Signal Processing and Control, 33 (2017) 151-160

School of Computer & Systems Sciences, JNU
Overview

- Brain Computer Interface (BCI)
- Types of BCI
- Components of a BCI
- Feature extraction Techniques
- Feature Selection Techniques
- Classification
- Proposed Method
- Results
- Conclusion
BCI: Motivation

• There exist diseases of the nervous system that gradually cause the body’s motor neurons to degenerate
  – Example: Amyotrophic Lateral Sclerosis (ALS)
• Eventually causes total paralysis
• The affected individual becomes trapped in his own body, unable to communicate
**BCI: Motivation**

- “Locked-in” patients are not able to produce any voluntary muscle movement.
- Fortunately, their brains are healthy and active.

- [Assistive BCI](#)
BCI: Motivation

- In USA, more than 200,000 patients live with the motor sequelae (consequences) of serious injury. There are two ways to help them restore some motor function:
  - Repair the damaged nerve axons
  - Build neuroprosthetic device
(a) In healthy subjects, primary motor area sends movement commands to muscles via spinal cord.

(b) But in paralyzed people this pathway is interrupted.

(c) A Computer based decoder is used, which translates this activity into commands for muscle control.
Brain-Computer Interface (BCI) addresses this concern by translating brain signals into control signals without using muscles or peripheral nerves.

BCI allows you to control a computing device just with your thoughts!!
BCI: Definition

“BCI is a system which takes a biosignal measured from a person and predicts some abstract aspect of the person’s cognitive state.”
Types of BCI

- **Dependent**
  - Uses the activity in the brain’s normal output pathways

- **Independent**
  - Not based on the brain’s normal output pathways
  - Based on cognitive brain activity
### Types of BCI

- **Invasive BCI**
  - Electrodes implanted inside the brain tissue
  - High quality signals
  - Prone to scar tissue build-up

- **Semi invasive BCI**
  - Electrodes implanted inside the skull but outside the brain cortex
  - Fine spatial resolution but at the expense of surgical incision into the skull. E.g. ECoG

- **Non invasive BCI**
  - Electrodes placed on the scalp
  - Fine temporal resolution but low spatial resolution
  - Majority employ EEG recordings
EEG: Electrodes Position

- **Letter represents region**
  - F - Frontal Lobe
  - T - Temporal Lobe
  - C - Center
  - O - Occipital Lobe
  - P - Parietal Lobe

- **Number represents position**
  - Odd number - left
  - Even number - right

Standard “10-20” system for electrodes placement
Application Areas & Users

- Spelling
- Controlling a prosthesis
- Communication by thoughts
- Interacting within a virtual reality environment
- Gaming
- Controlling a mechatronic device
Communication and Control

Spelling Device

Surfing Internet
Movement control

Assistive Mobility

Grasping
Environmental Control & Neurorehabilitation
Locomotion
Lie detection
BCI COMPONENTS

Operating Protocol

Signal acquisition

Signal processing
  Feature extraction
  Feature translation

Output device

Digitized signals

Commands
The Feature Extraction techniques are broadly categorized into four types (Wolpaw, 2012):

- Temporal
- Spatial
- Spectral
- Temporal-spectral
These are applied in the time domain:

- Peak picking
- Integration
- Moving average
Spectral Techniques

Fourier transform: Let \( f(t) \) be a signal. The Fourier transform of \( f(t) \) is

\[
F(u) = \int_{-\infty}^{\infty} f(t) \exp[-j2\pi ut] \, dt
\]

• Signal with 2Hz + 10Hz + 20 Hz

Stationary

![Graph showing time and frequency analysis](image-url)
Spectral Techniques

• However, FT has problem with non-stationary Signals.

Different in Time Domain
Spatial Techniques

- Common spatial Patterns
- Independent component analysis
COMMON SPATIAL PATTERN

- CSP finds spatial filters that maximizes variance for one class and minimizes variance for the second class simultaneously.
- Mathematically, it does simultaneous diagonalization of two covariance matrices.
- $X^i_1$ and $X^i_2$ are raw EEG data of trial $i$ having dimension $N \times T$ with $N$ the no of channels and $T$ the no of samples per channel for class 1 and 2 respectively.
- The normalized covariance matrices for class 1 and 2 are given as
  \[ \Sigma^i_1 = \frac{x^i_1 x^i_1'}{\text{trace}(x^i_1 x^i_1')} \quad \text{and} \quad \Sigma^i_2 = \frac{x^i_2 x^i_2'}{\text{trace}(x^i_2 x^i_2')} \]
- The spatial covariances $\Sigma_1$ and $\Sigma_2$ for the two classes are obtained by averaging over the trials of each group.
- The composite covariance matrix is $\Sigma = \Sigma_1 + \Sigma_2$
  which can be factored into its eigenvectors by $\Sigma = U \Lambda U'$
  where $U$ is the matrix of normalized eigenvectors and $\Lambda$ is the diagonal matrix of eigenvalues which we assume to be sorted in descending order.
CSP (contd.)

- The whitening transformation $\mathbf{W} = \sqrt{\mathbf{U}^{-1}} \mathbf{U}'$ equalizes the variance in the space spanned by $\mathbf{U}$, i.e., all eigenvalues of $\mathbf{W} \Sigma \mathbf{W}'$ are equal to 1.
- After whitening transformation, the covariances of each class data would be $\mathbf{s}_1 = \mathbf{W} \Sigma_1 \mathbf{W}'$ and $\mathbf{s}_2 = \mathbf{W} \Sigma_2 \mathbf{W}'$.
- $\mathbf{s}_1$ and $\mathbf{s}_2$ share the same eigenvectors, since $\mathbf{s}_1 + \mathbf{s}_2 = \mathbf{W} \Sigma \mathbf{W}' = \mathbf{I}$ and the corresponding eigenvalues for the two matrices sum up to 1, i.e.,

  $$\text{if} \quad \mathbf{s}_1 = \mathbf{V} \lambda_1 \mathbf{V}' \quad \text{then} \quad \mathbf{s}_2 = \mathbf{V} \lambda_2 \mathbf{V}' \quad \text{and} \quad \lambda_1 + \lambda_2 = 1.$$

- Consequently, the eigenvector with largest eigenvalue for $\mathbf{s}_1$ has the smallest eigenvalue for $\mathbf{s}_2$ and vice-versa.
- The projection of whitened EEG onto the first and last eigenvectors in $\mathbf{V}$ will give feature vectors that are optimal for discriminating two populations of EEG.
- The projection matrix is $\mathbf{P} = \mathbf{V}' \mathbf{W}$, with which a signal $\mathbf{X}$ would be mapped as $\mathbf{Z} = \mathbf{PX}$. 
CSP Limitations

- CSP limitations
  - Sensitive to artifacts in the EEG
  - Covariance matrix is of the order of the square of the number of channels but we have comparatively small number of trials
  - Does not use temporal information
  - Accurate classification depends upon the choice of optimal frequency bands – highly user specific
Temporal–Spectral

• **Wavelet transform**
  • It represents the non stationary signal in time as well as frequency domain.
  
  • It employs a ‘mother wavelet’ having specific time-frequency characteristics and mathematical properties.
    
    $$\psi_{s,t}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-t}{s}\right)$$
    
  • The signals obtained from **scaled** and **translated** versions of the mother wavelet are correlated with the segments of the original signal.
  
  • The average of all such correlation terms is used as a ‘feature’.
Feature Selection Techniques

- Feature Selection is used to retrieve the most relevant and non-redundant features from a large set of features.

- Features Selection is done prior to the application of classification algorithm (i.e. SVM, kNN, LDA etc.) to an unlabeled problem set.

- In BCI, Feature selection leads to reduction in the high dimension feature space, resulting in better classifier performance in terms of accuracy without causing overfitting of that classifier (Guyen, 2000).
Feature Selection Techniques

Three main types of Feature selection techniques exist:

- **Wrapper**: Wrappers evaluate and rank feature subsets using a specific classifier. They do well as the FS is tuned for the particular classifier but are also very slow as a classifier needs to be built for every subset and evaluated using cross validation.

- **Embedded**: In this, Feature Selection procedure is embedded into the learning algorithm itself, using the ability to ignore a subset of features.

- **Filter**: Filters evaluate and rank features or feature subsets prior to learning and independent of the classification algorithm. So they act faster. E.g. Information Gain, FCBF (Fast Correlation based Filter), ReliefF. Filters are most commonly used for BCI data.
Feature Selection Techniques

Correlation coefficient (CC):

\[ \rho(X_i, X_j) = \frac{1}{n} \sum_{k=1}^{n} (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) \sqrt{\text{var}(X_i)\text{var}(X_j)} \]

- \[ \rho(X_i, X_j) = 0 \]
- \[ \rho(X_i, X_j) = 1 \]

- For uncorrelated variables , whereas for linearly dependent variables we have .
Classification

Training examples

Linear classifier:

\[ q(x) = \begin{cases} 
H & \text{if } (\mathbf{w} \cdot \mathbf{x}) + b \geq 0 \\
J & \text{if } (\mathbf{w} \cdot \mathbf{x}) + b < 0 
\end{cases} \]
Linear Classifier

- denotes +1
- denotes -1

\[ f(x, w, b) = \text{sign}(w x + b) \]

Any of these would be fine..

..but which is best?
Support Vector Machine

What we know:

- $\mathbf{w} \cdot \mathbf{x}^+ + b = +1$
- $\mathbf{w} \cdot \mathbf{x}^- + b = -1$
- $\mathbf{w} \cdot (\mathbf{x}^+ - \mathbf{x}^-) = 2$

$M = \frac{(\mathbf{x}^+ - \mathbf{x}^-) \cdot \mathbf{w}}{||\mathbf{w}||} = \frac{2}{||\mathbf{w}||}$

Support Vectors are those datapoints that the margin pushes up against.

$\mathbf{M}$ = Margin Width
Support Vector Machine

- **Goal:**
  
  1. **Correctly classify all training data**
     
     \[
     \mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \quad \text{if } y_i = +1
     \]
     
     \[
     \mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \quad \text{if } y_i = -1
     \]
     
     \[y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \text{for all } i\]
  
  2. **Maximize the Margin**
     
     \[M = \frac{2}{|\mathbf{w}|}\]
     
     same as minimize
     
     \[
     \frac{1}{2} \mathbf{w} \cdot \mathbf{w}
     \]

- **We can formulate a Quadratic Optimization Problem and solve for \(w\) and \(b\)**

<table>
<thead>
<tr>
<th>Minimize (\Phi(w) = \frac{1}{2} \mathbf{w} \cdot \mathbf{w})</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject to (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \quad \forall i)</td>
</tr>
</tbody>
</table>
Support Vector Machine

\[
L = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} - \sum_{i=1}^{m} \alpha_i (y^{(i)}(\mathbf{w} \cdot \mathbf{x}^{(i)} + b) - 1)
\]

- Requiring the derivatives with respect to \( w, b \) to vanish yields:

- KKT conditions yield:

\[
\mathbf{w} = \sum_{i=1}^{m} \alpha_i y^{(i)} \mathbf{x}^{(i)}
\]

for any \( \alpha_i \neq 0, \ b = y^{(i)} - \mathbf{w} \cdot \mathbf{x}^{(i)} \)
Support Vector Machine

\[ w = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} \]

for any \( \alpha_i \neq 0 \), \( b = y^{(i)} - w \cdot x^{(i)} \)

maximize \( \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} x^{(i)} \cdot x^{(j)} \)

Subject to : \( \sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \)

\( \alpha_i \geq 0 \quad \forall i \)
Support Vector Machine

• The resulting separating function is:

\[ f(x) = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} \cdot x + b \]

• If \( f(x) > 0 \), \( x \) will be assigned class label +1 else -1
Support Vector Machine

• The resulting separating function is:

\[ f(x) = \sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} \cdot x + b \]

• Notes:
  – The points with \( \alpha = 0 \) do not affect the solution.
  – The points with \( \alpha \neq 0 \) are called support vectors.
  – The equality conditions hold true only for the Support Vectors.
maximize \[ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} K(x^{(i)}, x^{(j)}) \]

Subject to: \[ \sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \]
\[ \alpha_i \geq 0 \quad \forall i \]

\[ f(\Theta(x)) = \sum_{i=1}^{m} \alpha_i y^{(i)} K(x^{(i)}, x) + b \]
Composite Kernel SVM in Conjunction with Spatial Filter for Brain Computer Interface

Jyoti Singh Kirar, R. K. Agrawal

*Biomedical Signal Processing and Control, 33* (2017) 151-160
The Proposed Method

First Phase

Anatomical Division of Brain EEG electrodes

• In order to determine relevant electrodes which are actually participating to discriminate two different motor imagery tasks, electrodes placed on brain scalp are divided according to their neurological anatomy. There are five major lobes of brain cortex viz (frontal, central, temporal, parietal and occipital)
Stationary common spatial pattern (SCSP)

Max $R(w) = \frac{w^T \Sigma_1 w}{w^T \{\Sigma_1 + \Sigma_2\} w + \mu P(w)}$

Where $P(w) = w^T (\bar{\Delta}_1 + \bar{\Delta}_2) w$

$$\bar{\Delta}_c = \frac{1}{K} \sum_{k=1}^{K} \Delta_c^k$$

$$\Delta_c^{(k)} = s(\Sigma_c^k - \Sigma_c), \ c=1,2$$
Composite Kernel Support Vector Machine

\[
\begin{aligned}
\max_{\alpha} & -\frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \sum_{l=1}^{L} k_l(f_{i,l}, f_{j,l}) + \sum_{i} \alpha_i \\
\text{s.t} & \sum_{i} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq C \\
& 1 \leq i \leq N \\
& 1 \leq l \leq L \\
y & = \text{sgn}(\sum_{i=1}^{N} \alpha_i \sum_{l=1}^{L} k_l(f_{i,l}, f_{\text{test},l}) + b) \\
\end{aligned}
\]

\[
||h_l||^2 = \alpha^T K_l \alpha
\]
Experimental Results

<table>
<thead>
<tr>
<th>Subject</th>
<th>CSP</th>
<th>SCSP</th>
<th>CKSCSP at $\mu = 0$</th>
<th>CKSCSP at $\mu \neq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>75.37</td>
<td>80.45</td>
<td>80.01</td>
<td>81.14</td>
</tr>
<tr>
<td>al</td>
<td>97.73</td>
<td>94.38</td>
<td>98.04</td>
<td>98.34</td>
</tr>
<tr>
<td>av</td>
<td>69.14</td>
<td>69.82</td>
<td>70.25</td>
<td>77.56</td>
</tr>
<tr>
<td>aw</td>
<td>82.27</td>
<td>82.43</td>
<td>86.63</td>
<td>86.64</td>
</tr>
<tr>
<td>ay</td>
<td>82.17</td>
<td>89.33</td>
<td>87.83</td>
<td>88.17</td>
</tr>
<tr>
<td>Mean</td>
<td>81.33</td>
<td>83.28</td>
<td>84.55</td>
<td>86.37</td>
</tr>
</tbody>
</table>

Observations:

- Maximum average classification accuracy is achieved with the proposed method CKSCSP ($\mu \neq 0$).
- The proposed method CKSCSP ($\mu \neq 0$) outperformed standard CSP in terms of classification accuracy for all subjects.
- The proposed method CKSCSP ($\mu \neq 0$) outperformed SCSP in terms of classification accuracy for all subjects except subject ay.
## Experimental Results

<table>
<thead>
<tr>
<th>Subject</th>
<th>CSP</th>
<th>SCSP</th>
<th>CKSCSP at $\mu_R = 0$</th>
<th>CKSCSP at $\mu_R \neq 0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ds1a</td>
<td>73.10</td>
<td>81.75</td>
<td>62.60</td>
<td>67.05</td>
</tr>
<tr>
<td>ds1b</td>
<td>65.40</td>
<td>59.95</td>
<td>65.35</td>
<td>71.45</td>
</tr>
<tr>
<td>ds1c</td>
<td>70.50</td>
<td>75.1</td>
<td>73.05</td>
<td>75.35</td>
</tr>
<tr>
<td>ds1d</td>
<td>76.80</td>
<td>89.2</td>
<td>87.30</td>
<td>90.3</td>
</tr>
<tr>
<td>ds1e</td>
<td>83.30</td>
<td>90.35</td>
<td>89.70</td>
<td>89.55</td>
</tr>
<tr>
<td>ds1f</td>
<td>82.80</td>
<td>86.9</td>
<td>84.40</td>
<td>88.3</td>
</tr>
<tr>
<td>ds1g</td>
<td>79.00</td>
<td>91.45</td>
<td>92.45</td>
<td>94.75</td>
</tr>
<tr>
<td>Mean</td>
<td>75.84</td>
<td>82.1</td>
<td>79.26</td>
<td>82.39</td>
</tr>
</tbody>
</table>

### Observations:

- Maximum average classification accuracy is obtained with the proposed method CKSCSP ($\mu \neq 0$).
- The proposed method CKSCSP ($\mu \neq 0$) outperformed standard CSP in terms of classification accuracy for all subjects except for subject ds1a.
- The proposed method CKSCSP ($\mu \neq 0$) outperformed SCSP in terms of classification accuracy for all subjects except for subject ds1a and ds1e.
Experimental Results

Average Classification Accuracy using SVM classifier for BCI competition III dataset IVa

Frequency of each brain region for BCI competition III dataset IVa using CKSCSP

Average Classification Accuracy using SVM classifier for BCI competition IV dataset Ia

Frequency of Groups selected in 10 fold 10 times cross validation for dataset 1 at using CKSCSP

Average Classification Accuracy using SVM classifier for BCI competition IV dataset Ia

Frequency of Groups selected in 10 fold 10 times cross validation for dataset 2 using CKSCSP
Conclusions

- Exploits neurological information to divide the electrodes among five brain regions.
- Alleviates problem of small sample size and non-stationarity of existing CSP method.
- Employs Recursive Feature Elimination and composite kernel SVM to rank brain regions.
- Determines a reduced set of electrodes to discriminate motor imagery tasks.
- Outperforms existing methods on publicly available datasets.
<table>
<thead>
<tr>
<th>References</th>
</tr>
</thead>
</table>


References
