STATIONARY COMMON SPATIAL PATTERNS: TOWARDS ROBUST CLASSIFICATION OF NON-STATIONARY EEG SIGNALS

Wojciech Wojcikiewicz[†‡], Carmen Vidaurre[†], Motoaki Kawanabe[‡†]

[†]Machine Learning Group
Technical University Berlin

[‡]Intelligent Data Analysis Group
Fraunhofer Institute FIRST

May 27, 2011

Contents

1. Introduction
2. Methods
   - Common Spatial Patterns
   - Stationary Common Spatial Patterns
3. Experimental Results
   - Dataset and Setup
   - Classification Performance
   - Interpretation
4. Summary

Introduction

Brain-Computer Interfacing

- Brain-Computer Interface systems aim to translate the intent of a subject measured from brain signals e.g. EEG into control commands for a computer application or a neuroprosthesis.

Brain-Computer Interfacing

- Brain-Computer Interface systems aim to translate the intent of a subject measured from brain signals e.g. EEG into control commands for a computer application or a neuroprosthesis.

- Very popular paradigm is motor imagery.

Brain-Computer Interfacing

- Brain-Computer Interface systems aim to translate the intent of a subject measured from brain signals e.g. EEG into control commands for a computer application or a neuroprosthesis.

- Very popular paradigm is motor imagery.
- Challenges: High-dimensional EEG signal (e.g. 128 electrodes) has low spatial resolution, is noisy and intrinsically non-stationary.

Brain-Computer Interfacing

- Brain-Computer Interface systems aim to translate the intent of a subject measured from brain signals e.g. EEG into control commands for a computer application or a neuroprosthesis.

- Very popular paradigm is motor imagery.
- Challenges: High-dimensional EEG signal (e.g. 128 electrodes) has low spatial resolution, is noisy and intrinsically non-stationary.
  $\Rightarrow$ Use spatial filtering method e.g. Common Spatial Pattern (CSP) to reduce dimensionality and increase signal-to-noise ratio.

Brain-Computer Interfacing

- Brain-Computer Interface systems aim to translate the intent of a subject measured from brain signals e.g. EEG into control commands for a computer application or a neuroprosthesis.

- Very popular paradigm is motor imagery.
- Challenges: High-dimensional EEG signal (e.g. 128 electrodes) has low spatial resolution, is noisy and intrinsically non-stationary. $\Rightarrow$ Use spatial filtering method e.g. Common Spatial Pattern (CSP) to reduce dimensionality and increase signal-to-noise ratio.
- Problem: CSP does not tackle the non-stationarity issue.

Brain-Computer Interfacing

- Brain-Computer Interface systems aim to translate the intent of a subject measured from brain signals e.g. EEG into control commands for a computer application or a neuroprosthesis.

- Very popular paradigm is motor imagery.
- Challenges: High-dimensional EEG signal (e.g. 128 electrodes) has low spatial resolution, is noisy and intrinsically non-stationary.
  $\Rightarrow$ Use spatial filtering method e.g. Common Spatial Pattern (CSP) to reduce dimensionality and increase signal-to-noise ratio.
- Problem: CSP does not tackle the non-stationarity issue.
  $\Rightarrow$ We propose an extension called stationary CSP (sCSP) which extracts stationary features.

Sources of Non-stationarities

- Non-stationarities are changes of the signal properties over time.

Sources of Non-stationarities

- Non-stationarities are changes of the signal properties over time.
- They may have very different sources and time scales.

Sources of Non-stationarities

- Non-stationarities are changes of the signal properties over time.
- They may have very different sources and time scales.
    - Changes in electrode impedance when an electrode gets loose or the gel dries out.

Sources of Non-stationarities

- Non-stationarities are changes of the signal properties over time.
- They may have very different sources and time scales.
  - Changes in electrode impedance when an electrode gets loose or the gel dries out.
  - Muscular activity or eye movements lead to artefacts.

Sources of Non-stationarities

- Non-stationarities are changes of the signal properties over time.
- They may have very different sources and time scales.
    - Changes in electrode impedance when an electrode gets loose or the gel dries out.
    - Muscular activity or eye movements lead to artefacts.
    - Changes of task involvement as subjects get tired.

Sources of Non-stationarities

- Non-stationarities are changes of the signal properties over time.
- They may have very different sources and time scales.
  - Changes in electrode impedance when an electrode gets loose or the gel dries out.
  - Muscular activity or eye movements lead to artefacts.
  - Changes of task involvement as subjects get tired.
  - Differences between sessions e.g. no feedback in the calibration session vs. feedback in later sessions or small differences in electrode positions between sessions.

Sources of Non-stationarities

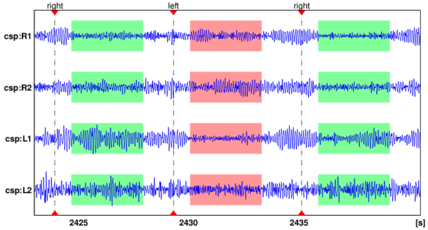
- Non-stationarities are changes of the signal properties over time.
- They may have very different sources and time scales.
  - Changes in electrode impedance when an electrode gets loose or the gel dries out.
  - Muscular activity or eye movements lead to artefacts.
  - Changes of task involvement as subjects get tired.
  - Differences between sessions e.g. no feedback in the calibration session vs. feedback in later sessions or small differences in electrode positions between sessions.

$\Rightarrow$ Extraction of invariant features and adaptation to changes are key challenges in BCI research.

Methods

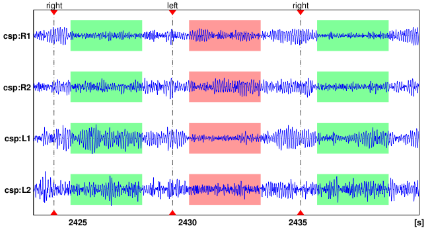
Common Spatial Patterns

- CSP maximizes the variance of bandpass filtered EEG signals from one class while minimizing the variance from the other class.



Common Spatial Patterns

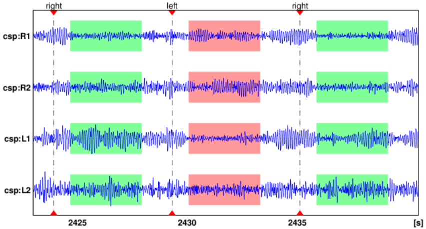
- CSP maximizes the variance of bandpass filtered EEG signals from one class while minimizing the variance from the other class.



- Thus it is suited to discriminate between mental states in BCI.

Common Spatial Patterns

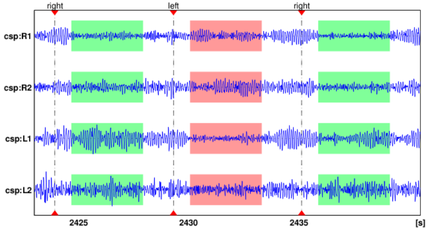
- CSP maximizes the variance of bandpass filtered EEG signals from one class while minimizing the variance from the other class.



- Thus it is suited to discriminate between mental states in BCI.
- CSP can be solved as generalized eigenvalue problem.

Common Spatial Patterns

- CSP maximizes the variance of bandpass filtered EEG signals from one class while minimizing the variance from the other class.



- Thus it is suited to discriminate between mental states in BCI.
- CSP can be solved as generalized eigenvalue problem.

$$\max_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{\Sigma}_+ \mathbf{w}}{\mathbf{w}^\top \{\mathbf{\Sigma}_+ + \mathbf{\Sigma}_-\} \mathbf{w}},$$

where $\mathbf{w}$ is the spatial filter and $\mathbf{\Sigma}_+$ and $\mathbf{\Sigma}_-$ are the average covariance matrices from class 1 and 2, respectively.

Stationary Common Spatial Patterns

Idea: We want to make features stationary i.e. we want to penalize non-stationary projection directions. This can be done by adding a penalty term $P(\mathbf{w})$ to the denominator of the Rayleigh quotient.

Stationary Common Spatial Patterns

Idea: We want to make features stationary i.e. we want to penalize non-stationary projection directions. This can be done by adding a penalty term P($\mathbf{w}$) to the denominator of the Rayleigh quotient.
Ideally P($\mathbf{w}$) should be the following quantity

$$P(\mathbf{w}) = \sum_k \left| \mathbf{w}^\top \boldsymbol{\Sigma}_c^{(k)} \mathbf{w} - \mathbf{w}^\top \boldsymbol{\Sigma}_c \mathbf{w} \right|,$$

where $\boldsymbol{\Sigma}_c^{(k)}$ is the covariance matrix of the k-th chunk of class c and $\boldsymbol{\Sigma}_c$ is the average covariance matrix of class c. A chunk consists of ≥ 1 trials.

Stationary Common Spatial Patterns

Idea: We want to make features stationary i.e. we want to penalize non-stationary projection directions. This can be done by adding a penalty term P($\mathbf{w}$) to the denominator of the Rayleigh quotient.
Ideally P($\mathbf{w}$) should be the following quantity

$$P(\mathbf{w}) = \sum_k \left| \mathbf{w}^\top \boldsymbol{\Sigma}_c^{(k)} \mathbf{w} - \mathbf{w}^\top \boldsymbol{\Sigma}_c \mathbf{w} \right|,$$

where $\boldsymbol{\Sigma}_c^{(k)}$ is the covariance matrix of the k-th chunk of class c and $\boldsymbol{\Sigma}_c$ is the average covariance matrix of class c. A chunk consists of ≥ 1 trials.

Problem: If we add this quantity to the denominator, it won't be a generalized eigenvalue problem of the form $\frac{\mathbf{w}^T A \mathbf{w}}{\mathbf{w}^T B \mathbf{w}}$ anymore.

Stationary Common Spatial Patterns

Thus we add a different, but related quantity as penalty term

Stationary Common Spatial Patterns

Thus we add a different, but related quantity as penalty term

$$P(\mathbf{w}) = \mathbf{w}^\top \sum_k \mathcal{F}\left(\mathbf{\Sigma}_c^{(k)} - \mathbf{\Sigma}_c\right)\mathbf{w},$$

where $\mathcal{F}$ is an operator to make symmetric matrices be positive definite, i.e. to flip the sign of all negative eigenvalues.

Stationary Common Spatial Patterns

Thus we add a different, but related quantity as penalty term

$$P(\mathbf{w}) = \mathbf{w}^\top \sum_k \mathcal{F}\left(\boldsymbol{\Sigma}_c^{(k)} - \boldsymbol{\Sigma}_c\right)\mathbf{w},$$

where $\mathcal{F}$ is an operator to make symmetric matrices be positive definite, i.e. to flip the sign of all negative eigenvalues.

Intuition: Treat variations to both sides in a similar way before projecting. This gives an upper bound to the quantity which we wanted to minimize.

Experimental Results

Dataset and Setup

- We tested sCSP and CSP on a data set of 80 subjects performing motor imagery with the left and right hand or with the feet.

Dataset and Setup

- We tested sCSP and CSP on a data set of 80 subjects performing motor imagery with the left and right hand or with the feet.
- For each user we select the best binary task-combination and estimate parameters like frequency band or time interval of interest in a calibration session (150 trials) without providing feedback.

Dataset and Setup

- We tested sCSP and CSP on a data set of 80 subjects performing motor imagery with the left and right hand or with the feet.
- For each user we select the best binary task-combination and estimate parameters like frequency band or time interval of interest in a calibration session (150 trials) without providing feedback.
- After that we perform a test session consisting of 300 trials and provide 1D visual feedback i.e. a moving arrow on a screen.

Dataset and Setup

- We tested sCSP and CSP on a data set of 80 subjects performing motor imagery with the left and right hand or with the feet.
- For each user we select the best binary task-combination and estimate parameters like frequency band or time interval of interest in a calibration session (150 trials) without providing feedback.
- After that we perform a test session consisting of 300 trials and provide 1D visual feedback i.e. a moving arrow on a screen.
- We used recordings of 68 preselected electrodes, log-variance features, a LDA classifier and error rate to measure performance.

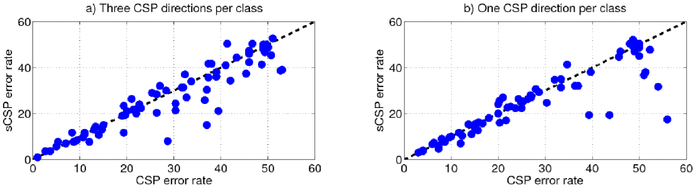
Dataset and Setup

- We tested sCSP and CSP on a data set of 80 subjects performing motor imagery with the left and right hand or with the feet.
- For each user we select the best binary task-combination and estimate parameters like frequency band or time interval of interest in a calibration session (150 trials) without providing feedback.
- After that we perform a test session consisting of 300 trials and provide 1D visual feedback i.e. a moving arrow on a screen.
- We used recordings of 68 preselected electrodes, log-variance features, a LDA classifier and error rate to measure performance.
- We selected a fixed number of filters per class.

Dataset and Setup

- We tested sCSP and CSP on a data set of 80 subjects performing motor imagery with the left and right hand or with the feet.
- For each user we select the best binary task-combination and estimate parameters like frequency band or time interval of interest in a calibration session (150 trials) without providing feedback.
- After that we perform a test session consisting of 300 trials and provide 1D visual feedback i.e. a moving arrow on a screen.
- We used recordings of 68 preselected electrodes, log-variance features, a LDA classifier and error rate to measure performance.
- We selected a fixed number of filters per class.
- The regularization parameter and the chunk size (1, 3 and 5) of sCSP were selected by cross-validation.

Classification Performance



Chunk size	Error rate			
	0 – 15	15 – 30	> 30	all
CV selected	0.0599	0.6642	**0.0069**	**0.0068**
1	0.0567	0.5194	**0.0085**	**0.0066**
3	0.1749	0.3547	**0.0121**	**0.0087**
5	0.1568	0.6690	**0.0076**	**0.0091**

Table: Overview of p-values for different chunk sizes and different error regions when using one-sided t-test with the hypothesis that sCSP performs better than CSP. Bold values are significant when $\alpha = 0.05$.

Interpretation

Why does sCSP perform better ?

Interpretation

Why does sCSP perform better ?

- CSP fails to extract the correct pattern when affected by artefacts.
  $\Rightarrow$ sCSP is more robust as it treats artefacts as non-stationarity.

Interpretation

Why does sCSP perform better ?

- CSP fails to extract the correct pattern when affected by artefacts.
  $\Rightarrow$ sCSP is more robust as it treats artefacts as non-stationarity.
- CSP may overfit.
  $\Rightarrow$ sCSP penalizes filters with large norms.

Interpretation

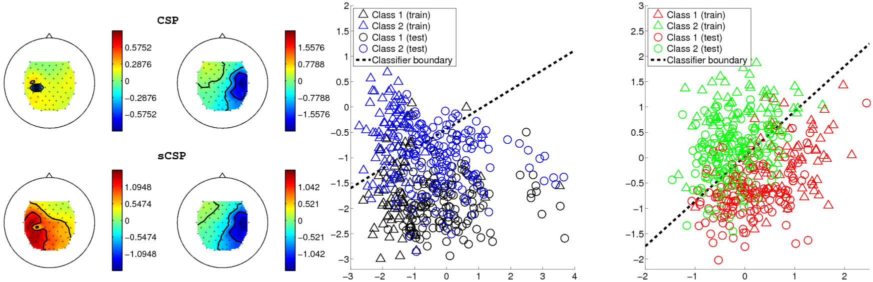
Why does sCSP perform better ?

- CSP fails to extract the correct pattern when affected by artefacts.
  $\Rightarrow$ sCSP is more robust as it treats artefacts as non-stationarity.
- CSP may overfit.
  $\Rightarrow$ sCSP penalizes filters with large norms.
- CSP produces non-stationary features.
  $\Rightarrow$ sCSP reduces changes (also between calibration and feedback).

Interpretation

Why does sCSP perform better ?

- CSP fails to extract the correct pattern when affected by artefacts.
  $\Rightarrow$ sCSP is more robust as it treats artefacts as non-stationarity.
- CSP may overfit.
  $\Rightarrow$ sCSP penalizes filters with large norms.
- CSP produces non-stationary features.
  $\Rightarrow$ sCSP reduces changes (also between calibration and feedback).



Summary

Summary

- We proposed an extension of the popular Common Spatial Patterns method which extracts stationary features.

Summary

- We proposed an extension of the popular Common Spatial Patterns method which extracts stationary features.
- sCSP significantly increases classification accuracy, especially for subjects who perform badly.

Summary

- We proposed an extension of the popular Common Spatial Patterns method which extracts stationary features.
- sCSP significantly increases classification accuracy, especially for subjects who perform badly.
- Unlike other methods our approach is completely data-driven and does not require additional recordings or models of the expected change that occurs in the EEG.

Summary

- We proposed an extension of the popular Common Spatial Patterns method which extracts stationary features.
- sCSP significantly increases classification accuracy, especially for subjects who perform badly.
- Unlike other methods our approach is completely data-driven and does not require additional recordings or models of the expected change that occurs in the EEG.
- We also showed (not presented here) that a combination of stationary features and unsupervised adaptation further improves classification performance.

Questions ?