Error Correction Output Coding Coupled with the CSP for Motor Imagery BCI Systems

Soroosh Shahtalebi[†] and Arash Mohammadi[†]

[†]Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC, Canada
Email: arash.mohammadi@concordia.ca

EUSIPCO 2017- 28 August-2 September, Kos island, Greece.

Outline

Introduction

Motivations and Contributions

Motivations

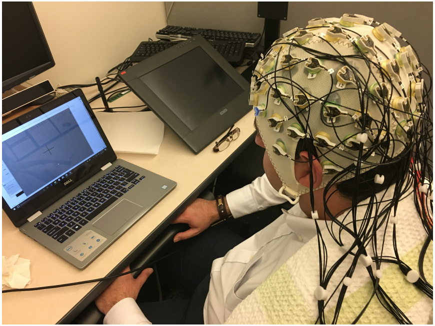
- In this work, we consider problem of classifying multiple Motor Imagery (MI) tasks by means of Error Correcting Output Coding (ECOC) coupled with Common Spatial Patterns (CSP) technique.
- The motivation of this paper is to apply the ECOC methodology, for the first time, to the multiple MI classification problem and evaluate its potential for classifying EEG signals.

Contributions

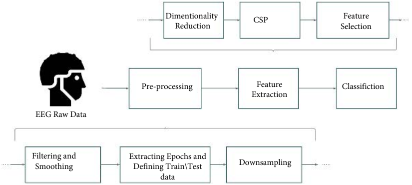
- We propose a novel architecture for analyzing MI tasks, referred to as the ECO-CSP.
- The ECO-CSP couples CSP with ECOC classifiers to extract discriminative features associated with the MI tasks.
- The ECO-CSP provides slight performance improvement in comparison to the previous approaches but with significantly lower computational complexity.

Electroencephalogram (EEG)

- EEG measures voltage fluctuations resulting from ionic current flows within the neurons of the brain.
- Human body is a volume conductor, any electrical activity in a certain spot influences all the neighboring recording channels resulting in a blurred and noisy image of brain activities.
- In this research, we analyze the EEG signals to classify Motor Imagery (MI) tasks.



Common spatial Patterns (CSP)



- During imagery movements, two related phenomena appear in electrical activity of the brain, i.e., Event Related Synchronization (ERS) and Event Related Desynchronization (ERD).
- These phenomena represent frequency specific changes of the ongoing EEG activity and consist either of decreases or of increases of power in given frequency bands.
- The CSP algorithm introduces spatial filters for multi-channel EEG recordings to better locate and extract ERD and ERS waveforms. Consequently, the CSP methodology enhances EEG channels containing higher weights for the ERD and the ERS.

Error Correcting Output Coding

- MI tasks are typically multiclass problems where conventional binary classifiers (SVM or LDA) are not directly applicable.
- Multiclass extensions of current binary classifiers e.g., Multiclass SVM) are developed. Or different techniques are devised to adapt current binary classifiers for multiclass usages (e.g., "one vs. all").
- One of the techniques which is intrinsically developed for multiclass classification, is Error Correcting Output coding (ECOC) which assigns predefined binary codewords to different classes and deploys a number of binary classifiers to predict the codeword of unseen data. The number of classifiers is equal to the length of the codeword
- ECOC is based on the idea that codeword of unseen trial must be one of predefined codewords, unless some error have affected the codeword and we have a noisy version of the codeword.
- The predefined codewords are designed in a way to handle the noise up to certain limits, so recovering the original codeword is plausible.
- A classifier predicts the value for each digit of the codeword, the classifiers are assumed to be independent.

Proposed Error Correcting Output Coding (ECOC) Coupled with Common Spatial Patterns (CSP)

Problem Formulation

- We consider that the EEG recordings are split into different epochs where each epoch is denoted by $\{\boldsymbol{X}_i\}_{i=1}^{N_{\text{Trial}}} \in \boldsymbol{R}^{N_{\text{ch}} \times N_{\text{t}}}$.
- Each trial has its own label which is denoted by $\{h_i\}_{i=1}^{N_{\text{Trial}}}$. Each label has its own corresponding bit-vector $\{\boldsymbol{b}_k\}_{k=1}^{N_{\text{c}}}$ with length of N_{Bits}.
- The ECOC forms a binary coding matrix of N_{Trial} bit-vectors of length N_{Bits}. This matrix is denoted by $\boldsymbol{\mathcal{C}}$.
- The coding matrix of $\boldsymbol{\mathcal{C}}$ determines that for classifying N_{c} classes, N_{Bits} number of classifiers ($\boldsymbol{\Lambda} = \{\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, \ldots, \boldsymbol{\gamma}^{(N_{\text{Bits}})}\}$) are constructed to decide on whether their corresponding bit is zero or one.
- To train the classifier $\boldsymbol{\gamma}^{(j)}$, two super-sets of $\boldsymbol{S}^{(j,0)}$ and $\boldsymbol{S}^{(j,1)}$ are formed, where $\boldsymbol{S}^{(j,1)}$ consists of all labels h_i for which $\mathcal{C}_{ij} = 1$, and $\boldsymbol{S}^{(j,0)}$ is the complement set.

Codeword Generation for ECOC

The following procedure is for the cases that $3 \leq N_\mathrm{c} \leq 7$ and codewords of length equal to $N_\mathrm{Bits} = 2^{N_\mathrm{c}-1} - 1$ are generated.

- All the entries in the first row are ones.
- The second row consists of $2^{N_\mathrm{c}-2}$ zeros followed by $2^{N_\mathrm{c}-2} - 1$ ones.
- The third row consists of $2^{N_\mathrm{c}-3}$ zeros, followed by $2^{N_\mathrm{c}-3}$ ones, followed by $2^{N_\mathrm{c}-3}$ zeros, followed by $2^{N_\mathrm{c}-3} - 1$ ones.
- The k-th row consists of alternating runs of $2^{N_\mathrm{c}-k}$ zeros and ones.

Extracted Codewords for 4 Classes of Data

Classes	Codeword							
Right Hand IM	1	1	1	1	1	1	1	
Left Hand IM	0	0	0	0	1	1	1	Class 2
Foot IM	0	0	1	1	0	0	1	
Tongue IM	0	1	0	1	0	1	0	
				Classifier 4				

Figure 2: The ECOC example with 7 bits for classifying 4 EEG classes.

Common Spatial Patterns Calculation

- The spatial covariance matrix for each trial is computed according to

$$\boldsymbol{C}_i = \frac{1}{N_{\mathrm{t}} - 1}\big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\big)\big(\boldsymbol{X}_i - \boldsymbol{\mu}_i\big)^T. \quad (1)$$

- According to super-sets of $\boldsymbol{S}^{(j,0)}$ and $\boldsymbol{S}^{(j,1)}$, the spatial covariance matrices are categorized in two sets.
- The average spatial covariance matrices for each super-set is computed and is denoted by $\bar{\boldsymbol{C}}^{(j,0)}$ and $\bar{\boldsymbol{C}}^{(j,1)}$, respectively.
- The composite spatial covariance matrix denoted by $\boldsymbol{C}^{(j,c)}$ is computed as follows

$$\boldsymbol{C}^{(j,c)} = \bar{\boldsymbol{C}}^{(j,0)} + \bar{\boldsymbol{C}}^{(j,1)}. \quad (2)$$

Common Spatial Patterns Calculation (continued)

- Eigenvalues and eigenvectors of composite covariance matrix $\boldsymbol{C}^{(j,c)}$ are calculated according to

$$\boldsymbol{C}^{(j,c)} = \boldsymbol{U}^{(j,c)}\boldsymbol{\lambda}^{(j,c)}[\boldsymbol{U}^{(j,c)}]^T. \tag{3}$$

- Then a whitening transform denoted by $\boldsymbol{P}_j$ is applied on $\boldsymbol{U}^{(j,c)}$ as

$$\boldsymbol{P}_j = \sqrt{\left[\boldsymbol{\lambda}^{(j,c)}\right]^{-1}}[\boldsymbol{U}^{(j,c)}]^T. \tag{4}$$

- From the properties of whitening transform we know that all the eigenvalues of $\boldsymbol{P}_j\boldsymbol{C}^{(j,c)}[\boldsymbol{P}_j]^T$ are equal to one. By decomposing $\boldsymbol{C}^{(j,c)}$ to its building parts, we have

$$\boldsymbol{S}^{(j,0)} = \boldsymbol{P}_j\bar{\boldsymbol{C}}^{(j,0)}[\boldsymbol{P}_j]^T \tag{5}$$

$$\text{and} \quad \boldsymbol{S}^{(j,1)} = \boldsymbol{P}_j\bar{\boldsymbol{C}}^{(j,1)}\boldsymbol{P}_j^T. \tag{6}$$

Common Spatial Patterns Calculation (Continued)

- $\boldsymbol{S}^{(j,0)}$ and $\boldsymbol{S}^{(j,1)}$ share common eigenvectors denoted by $\tilde{\boldsymbol{H}}_j$, i.e.,

$$\begin{aligned} \boldsymbol{S}^{(j,0)} &= \tilde{\boldsymbol{H}}_j \boldsymbol{\lambda}^{(j,0)} [\tilde{\boldsymbol{H}}_j]^T \quad (7) \\ \text{and} \quad \boldsymbol{S}^{(j,1)} &= \tilde{\boldsymbol{H}}_j \boldsymbol{\lambda}^{(j,1)} [\tilde{\boldsymbol{H}}_j]^T, \quad (8) \end{aligned}$$

- According to $\boldsymbol{C}^{(j,c)} = \bar{\boldsymbol{C}}^{(j,0)} + \bar{\boldsymbol{C}}^{(j,1)}$, we have $\boldsymbol{\lambda}^{(j,0)} + \boldsymbol{\lambda}^{(j,1)} = \boldsymbol{I}$, where $\boldsymbol{I}$ denotes an identity matrix of appropriate dimension.

Since the sum of two corresponding eigenvalues is always one, the eigenvector with largest eigenvalue corresponding to $\bar{\boldsymbol{S}}^{(j,0)}$ has the smallest eigenvalue for $\bar{\boldsymbol{S}}^{(j,1)}$ and vice versa. This property makes the eigenvectors $\tilde{\boldsymbol{H}}_j$ useful for classification of the two distributions. The projection of whitened EEG signals onto the first and last eigenvectors in $\tilde{\boldsymbol{H}}_j$ (i.e., the eigenvectors corresponding to the largest and smallest eigenvalues) will provide feature vectors that are optimal for discriminating two populations of EEG signals in the least square sense.

Feature Extraction\Selection

- The projection matrix corresponding to classifier $\boldsymbol{\gamma}_j$, for $(1 \leq j \leq N_{\text{Bits}})$, is $\boldsymbol{W}_j = \left[\boldsymbol{P}_j\right]^T \tilde{\boldsymbol{H}}_j$.
- The decomposition (mapping) of each trial $\boldsymbol{X}_i$, for $(1 \leq i \leq N_{\text{Trial}})$ is computed as $\boldsymbol{Z}_{i,j} = \left[\boldsymbol{W}_j\right]^T \boldsymbol{X}_i$.

Feature Selection

Since the eigenvalues in $\boldsymbol{\lambda}^{(j,c)}$ are sorted in descending order and the eigenvector with largest eigenvalue corresponding to $\bar{\boldsymbol{S}}^{(j,0)}$ has the smallest eigenvalue for $\bar{\boldsymbol{S}}^{(j,1)}$ and vice versa, the first and last m rows of $\boldsymbol{Z}_{i,j}$ could be used to extract the features. In other words, matrix $\boldsymbol{Z}^j_{i,p}$ is defined and is constructed from the first and last m rows of matrix $\boldsymbol{Z}_{i,j}$.

- The features for each trial are then extracted by

$$\boldsymbol{f}^j_{i,p} = \log\left(\frac{\operatorname{var}\left(\boldsymbol{Z}^j_{i,p}\right)}{\sum_{k=1}^{2m} \operatorname{var}\left(\boldsymbol{Z}^j_{k,p}\right)}\right), \tag{9}$$

which $\boldsymbol{f}^j_{i,p} \in \mathbb{R}^{2m}$

Simulation Results

Dataset

- Performance of the proposed ECO-CSP framework is evaluated based on BCI Competition IV-2a dataset.
- this dataset consists of four classes of motor imagery EEG measurements (Right hand IM, Left hand IM, Feet IM, and Tongue IM) obtained from nine subjects.
- Signals are recorded at sampling rate of 250Hz using 22 EEG channels and 3 monopolar electrooculogram (EOG) channels (with left mastoid serving as the reference). The original EEG signal recordings are already bandpass filtered (0.5-100Hz) and notch filtered.
- In total and for each subject, 288 trials for training and 288 trials for evaluation are available.
- kappa coefficient κ is used as a measure of error, i.e., $\kappa = \frac{\text{CCR}-P_{\text{rand}}}{1-P_{\text{rand}}}$, where CCR represents the correct classification rate, and the value of P_{rand} for this dataset is equal to 0.25.

Data Preparation

- Each epoch of EEG signal for each trial, is extracted and then bandpass filtered through two Chebychev type 2 filters of order 54 which extract the frequency contents of μ-band (8-13 Hz) and the β-band (13-30 Hz).
- Then a segment of 2 seconds starting from 0.5 second after presenting the cue to the subject is selected.
- The segments of 2 seconds are then imported to ECO-CSP algorithm to train $N_{\text{Bits}} = 7$ binary classifiers. In this experiment, commonly used SVM and LDA classifiers are trained.

Results of 10-fold Cross Validation for Training Stage

Subjects	Classifiers for each bit in class codewords													
	Classifier1		Classifier2		Classifier3		Classifier4		Classifier5		Classifier6		Classifier7	
	LDA	**SVM**	**LDA**	**SVM**	**LDA**	**SVM**	**LDA**	**SVM**	**LDA**	**SVM**	**LDA**	**SVM**	**LDA**	**SVM**
Subject 1	78.95	83.16	75.2	74.27	52.75	52.75	79.88	82.69	90.18	91.11	78.95	81.75	83.63	88.77
Subject 2	78.82	77.38	72.56	74.01	62.94	60.05	78.34	78.34	67.75	68.71	82.67	84.12	77.38	78.34
Subject 3	88.32	88.32	70.32	74.21	77.13	75.67	83.45	87.35	76.64	78.1	81.02	80.05	82.97	80.54
Subject 4	71.7	67.39	53	53.08	51.08	51.08	72.18	76.5	73.62	75.06	75.54	76.98	77.94	82.25
Subject 5	74.74	75.2	63.51	62.11	66.78	67.72	73.8	75.2	71.46	73.33	77.08	79.88	78.01	78.01
Subject 6	76014	79.42	67.72	66.78	68.19	66.78	78.48	80.82	65.38	65.38	80.35	80.82	76.61	77.54
Subject 7	85.45	85.94	69.94	68.48	74.79	74.79	78.18	83.52	97.09	98.55	87.39	86.42	84	84.97
Subject 8	83.27	83.78	82.26	82.76	83.27	81.75	89.86	92.4	80.23	77.19	73.17	75.16	87.83	88.34
Subject 9	88.85	89.33	75.76	77.21	68.48	68	80.12	80.12	72.85	73.82	75.27	81.09	77.21	78.18

Table 1: 10-fold Cross Validation results for training of 7 classifiers for 9 subjects of BCI Competition IV_{2a} dataset. The results are measured in Kappa Value

Results of Evaluation on Unseen Data

Subjects	**Different Classifiers**	
	LDA	**SVM**
Subject 1	48.61	48.14
Subject 2	20.83	20.37
Subject 3	50.46	49.07
Subject 4	34.25	31.49
Subject 5	18.05	16.2
Subject 6	8.33	12.5
Subject 7	57.87	63.88
Subject 8	44.44	38.42
Subject 9	37.05	37.96
Average	**35.54**	**35.34**

Table 2: Performance (in Kappa value (κ)) of 2 classifiers for 9 subjects of BCI Competition IV_{2a} dataset.

Conclusion

- This paper proposed a new architecture to couple a well regarded feature extraction technique for EEG signals with a successful multi-class classifying scheme.
- The BCI Competition IV-2a dataset is used to evaluate the performance of the proposed framework.
- The experiments indicate that the proposed ECO-CSP framework provides similar results when compared to other recently developed algorithms, but has extensively less computational complexity making it a practical alternative for real-time EEG motor imagery classification tasks.

Thank You

Please forward your questions to:

arash.mohammadi@concordia.ca