Optimizing the Channel Selection and Classification Accuracy in EEG-Based BCI.

Mahnaz Arvaneh et al. (Chai Quek*)

IEEE Transactions on Biomedical Engineering (2011)

Presenter : SeungChan Lee

GIST, Dept. of Information and Communication, INFONET Lab.

Gwangju Institute of Science and Technology

Background

- Channel selection problems in EEG-based BCI
  - A large number of EEG channels
    - It may include noisy and redundant signals. – degradation of performance
    - It needs a prolonged preparation time. – inconvenience in installation process
  - Selecting the least number of channels with required accuracy can balance both needs.
- Various channel selection methods
  - SVM based
    - Recursively eliminates the least-contributed channels based on classification accuracy.
  - Mutual information(MI) based
    - Rank the channels based on MI between channels and class labels
  - Common spatial filter(CSP) based
    - Directly select the channels according to their CSP coefficients
  - RCSP based
    - used sparse solutions of spatial filters

Background

- Research problems in EEG channel selection
  - How many channels are required for the best classification accuracy?
  - What is the minimum number of channels required to achieve the same accuracy as obtained by using all the channels?
- To address the research questions…
  - They proposed a sparse common spatial pattern(SCSP) algorithm.
  - The proposed algorithm minimizes the number of channels by sparsifying the common spatial filters within a constraint of classification accuracy.

CSP algorithm

- The CSP algorithm is effective in discriminating two classes of EEG data by maximizing the variance of one class while minimizing the variance of the other class.
- Summary of formula derivation
  - Let single trial EEG data $\mathbf{X} \in \mathbf{R}^{N \times S}$
    (N : the number of channels, S: the number of measurement samples)
  - The CSP algorithm projects **X** to spatially filtered **Z** as $\mathbf{Z} = \mathbf{W}\mathbf{X}$
    (the rows of $\mathbf{W}$: the spatial filters, the columns of $\mathbf{W}^{-1}$: CSP)
  - Normalized covariance matrix $\mathbf{C} = \dfrac{\mathbf{X}\mathbf{X}^T}{\mathrm{trace}(\mathbf{X}\mathbf{X}^T)}$
    trace(X) : sum of diagonal elements of **X**
  - $\mathbf{C}_C = \mathbf{C}_1 + \mathbf{C}_2 = \mathbf{F}_C \psi \mathbf{F}_C^T$
    $\mathbf{C}_1, \mathbf{C}_2$: Computed by averaging over multiple trials of EEG data
    $\mathbf{F}_C$: matrix of normalized eigenvectors
    ψ : diagonal matrix of eigenvalues
  - Whitening transformation matrix
  - Transformation of covariance matrices

CSP algorithm

- Summary of formula derivation
  - Whitening transformation matrix $\mathbf{P} = \sqrt{\psi^{-1}}\mathbf{F}_C^T$
  - Transformation of covariance matrices

$$\mathbf{C}_1' = \mathbf{P}\mathbf{C}_1\mathbf{P}^T, \quad \mathbf{C}_2' = \mathbf{P}\mathbf{C}_2\mathbf{P}^T$$

$$= \mathbf{U}\Lambda_1\mathbf{U}^T \qquad = \mathbf{U}\Lambda_2\mathbf{U}^T \qquad \Lambda_1 + \Lambda_2 = \mathbf{I}$$

$\mathbf{C}_1', \mathbf{C}_2'$: share common eigenvectors,

$\mathbf{U}$: eigenvectors matrix

Λ : diagonal eigenvalues matrix

  - Apply CSP projection matrix $\mathbf{W} = \mathbf{U}^T\mathbf{P}$

$$\mathbf{C}_1' = \mathbf{U}^T\mathbf{P}\mathbf{C}_1\mathbf{P}^T\mathbf{U} = \Lambda_1, \quad \mathbf{C}_2' = \mathbf{U}^T\mathbf{P}\mathbf{C}_2\mathbf{P}^T\mathbf{U} = \Lambda_2 \qquad \Lambda_1 + \Lambda_2 = \mathbf{I}$$

  - Because $\Lambda_1 + \Lambda_2 = \mathbf{I}$, the maximum variance of one class lead to the minimum variance of the another class.→ Optimal discrimination
  - Projection matrix **W** can be formulated as an optimization problem

$$\min_{\mathbf{w}_i}\left(\sum_{i=1}^{i=m}\mathbf{w}_i\mathbf{C}_2\mathbf{w}_i^T + \sum_{i=m+1}^{i=2m}\mathbf{w}_i\mathbf{C}_1\mathbf{w}_i^T\right)$$

Subject to : $\mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_i^T = 1, \; i = \{1, 2, \ldots, 2m\}$

$\mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_j^T = 1, \; i, \mathrm{j} = \{1, 2, \ldots, 2m\} \; i \neq j$

$\mathbf{C}_i$: covariance matrix of class i

$\mathbf{w}_i \in \mathbb{R}^{1\times N}, i = \{1, \ldots, 2m\}$ indicate first and last m rows of CSP projection matrix

SCSP algorithm

- Motivation
  - Sparsify the CSP spatial filters to emphasize on a limited number of channels with high variances between the classes
  - Discard the rest of the channels with low or irregular variances that may be due to noise or artifacts.
- Sparsity measurement
  - $l_1 / l_2 = \dfrac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$
  - The sparsest possible vector(only a single element is nonzero) has a sparseness of one.
  - Non-sparsity measurement : l_1 / l_2 norm increases when the sparsity decreases.
- Modification of CSP algorithm

SCSP algorithm

- Modification of CSP algorithm
  - Include regularization parameter in optimization problem

$$\min_{\mathbf{w}_i}(1-r)\left(\sum_{i=1}^{i=m}\mathbf{w}_i\mathbf{C}_2\mathbf{w}_i^T+\sum_{i=m+1}^{i=2m}\mathbf{w}_i\mathbf{C}_1\mathbf{w}_i^T\right)+r\sum_{i=1}^{i=2m}\frac{\|\mathbf{w}_i\|_1}{\|\mathbf{w}_i\|_2}$$

$$\text{Subject to : }\mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_i^T=1,\ i=\{1,2,\ldots,2m\}$$

$$\mathbf{w}_i(\mathbf{C}_1+\mathbf{C}_2)\mathbf{w}_j^T=1,\ i,\mathrm{j}=\{1,2,\ldots,2m\}\ i\neq j$$

  - Parameter $r(0\leq r\leq 1)$ controls the number of removed channels and classification accuracy.
  - Non-linear optimization problem → solved using sequential quadratic programming(SQP) and augmented Lagrangian methods

SCSP algorithm

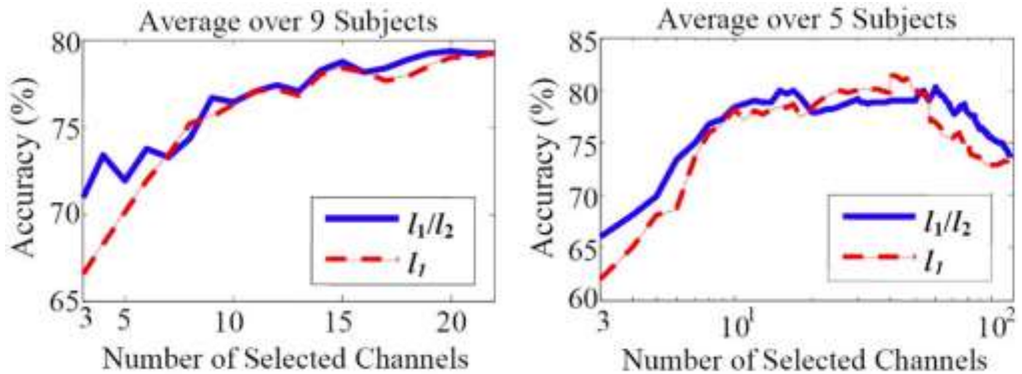
- Channel selection
  - From training set of two class motor imagery data, first two sparse spatial filters corresponding each class are obtained by solving the optimization problem.
  - Zero element channel → discard
    Non-zero element channel → select the channels
  - Importance order : apply ranking method(used maximum of the absolute values of the corresponding sparse spatial filter.

Datasets and processing

- Datasets
  - With a moderate number of channels (22 channels)
    - Dataset 2a from BCI competition 4
    - 9 subjects
    - Used only right and left hand motor imagery tasks
    - 72 trials training set + 72 trials testing set on each subjects
  - With a large number of channels (118 channels)
    - Dataset 4a from BCI competition 3
    - 5 subjects
    - Right hand and foot motor imagery tasks
    - 140 trials training set + 140 trials testing set on each subjects
- Data processing
  - Extract 0.5 ~ 2.5 seconds data samples after the visual cue
  - Apply 8 ~ 35Hz band-pass filter
  - (Training set) select optimal channels using first and last sparse spatial filter
  - (Test set) CSP retraining over selected channels and dataset spatially filtered using the first and last 3 spatial filters.
  - Variance of spatially filtered signal applied SVM classifier

Results and Discussion

- Performance comparison of l_1 and l_1/l_2 Regularization term
  - Varying r value(different number of channels)
  - l_1/l_2 norm based SCSP algorithm leads better classification accuracies when two different regularization based SCSP algorithm select same number of channels.



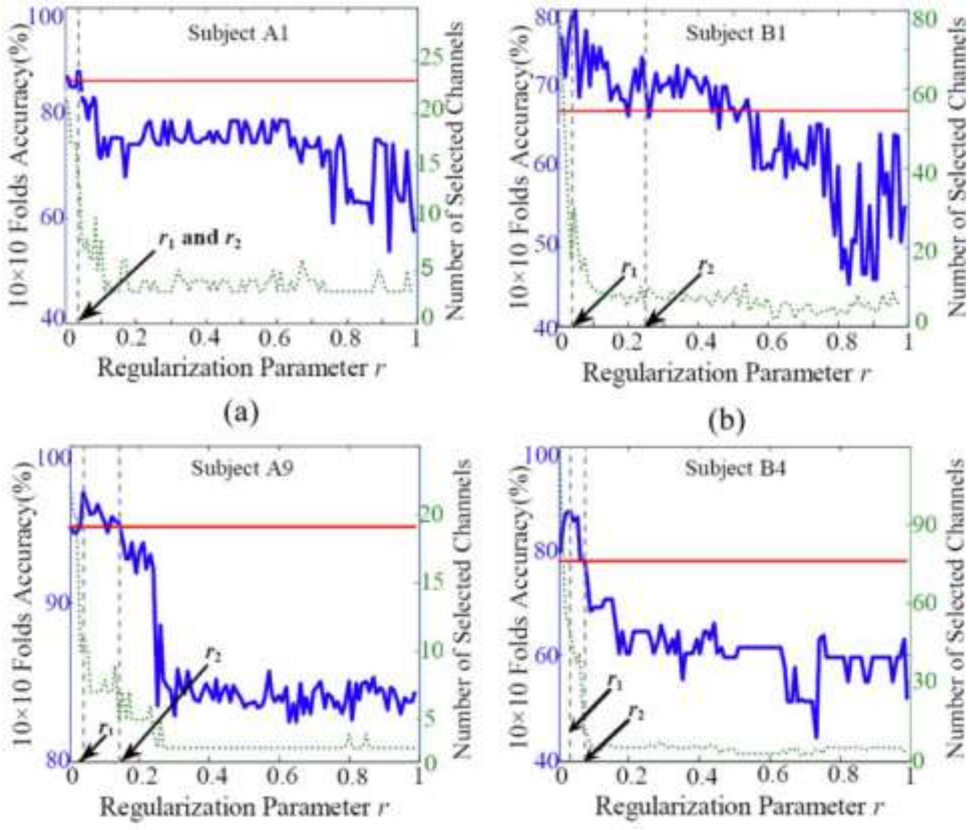
(a) (b)

Results and Discussion

- Channel selection with different criteria
  - Two channel selection criteria
    - First criterion : maximizes the accuracy by removing noisy and irrelevant channels.(SCSP1)
    - Second criterion : minimizes the number of selected channels while maintaining the classification accuracy.(SCSP2)
  - Procedure
    - r was chosen from 0.01 to 0.99.
    - For each r, a set of selected channels was determined.
    - Using 10x10 fold cross validation on training set, compute classification accuracy with each set of the selected channels.
    - Optimal r was selected based on the accuracy.

Results and Discussion

- Channel selection with different criteria



(a) (b) (c) (d)

– Summary

  - the use of small values of r improved the accuracy by removing some noisy and redundant EEG channels, while increased values of r reduced the number of channels but also decreased the classification accuracy.
  - further increase of the r value did not yield further reduction in the number of selected channels.

Results and Discussion

- Classification accuracy vs. number of selected channels.
- About bellow table (overall 22 channel subjects)
  - Decreasing the number of channels is very effective without accuracy degradation.(SCSP1: reduced 40% of the channels, SCSP2: reduced 61.2% of the channels)
  - the proposed SCSP algorithm using both criteria yielded significantly better classification accuracies (average 9.45% more) compared to the use of three typical channels.

Dataset IIa, BCI Competition IV						
			SCSP1		SCSP2	
Subject	All Ch Acc(%)	(C3,C4,Cz) Acc(%)	Acc (%)	♯ Selected Ch	Acc (%)	♯ Selected Ch
A1	90.97	75.69	91.66	13	91.66	13
A2	56.25	53.47	67.36	9	60.41	4
A3	96.52	93.05	97.91	14	97.14	12
A4	72.91	68.05	72.22	14	70.83	11
A5	63.88	53.47	65.27	11	63.19	9
A6	63.88	61.11	66.67	14	61.11	10
A7	79.86	57.63	84.72	19	78.47	15
A8	97.22	86.80	97.22	15	95.13	5
A9	91.66	88.88	91.66	10	93.75	5
Mean	79.23	70.90	81.63	13.22	79.07	8.55
Std	15.63	15.72	13.7	2.99	15.61	3.90
p-value	0.006	–	0.003	–	0.004	–

P-value denotes the paired T-test between results of (C3,C4,CZ) and other results.
(CH: Channels, ACC: Accuracy, ♯ : Number).

Results and Discussion

– Classification accuracy vs. number of selected channels.

– About bellow table (overall 118 channel subjects)

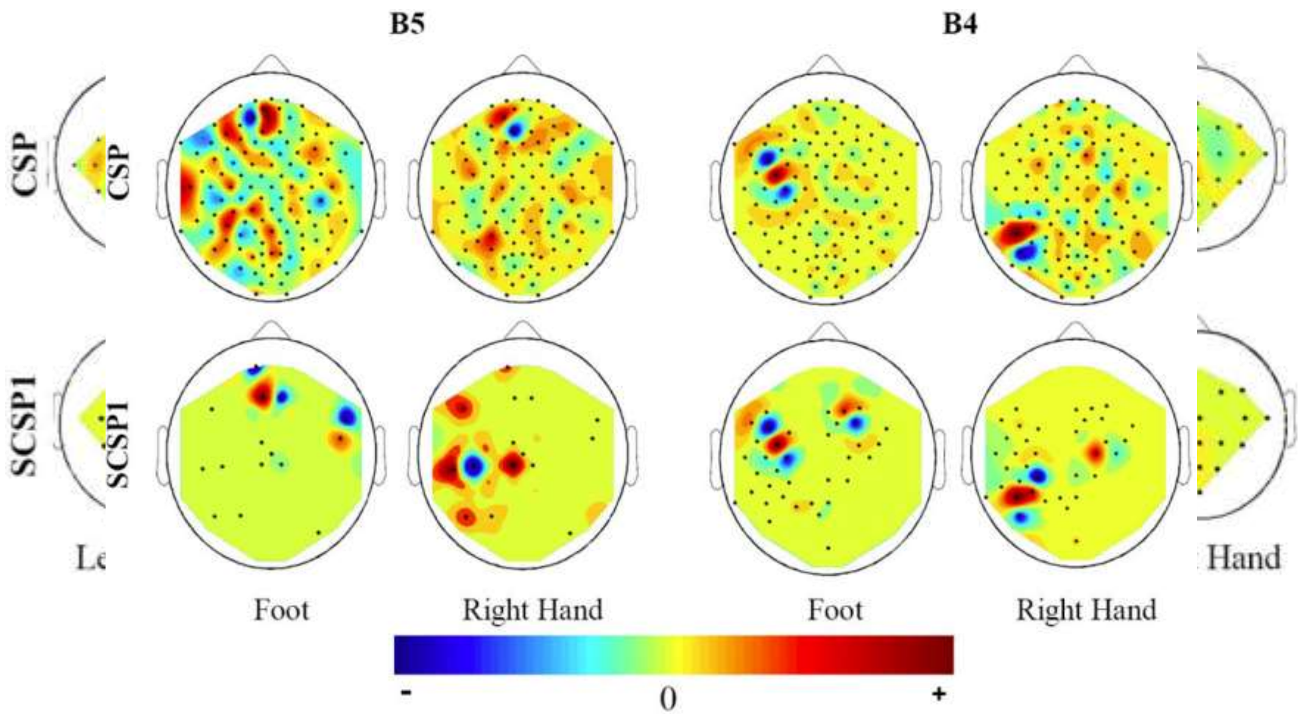
- Decreasing the number of channels is very effective without accuracy degradation.(SCSP1: reduced 81% of the channels, SCSP2: reduced 93% of the channels)
- The results also show an average improvement of 11.5% in the classification accuracy compared to the use of three typical channels.

Dataset IVa, BCI Competition III						
			SCSP1		SCSP2	
Subject	**All Ch Acc(%)**	**(C3,C4,Cz) Acc(%)**	**Acc (%)**	**♯ Selected Ch**	**Acc (%)**	**♯ Selected Ch**
B1	74.28	54.28	80.71	17	71.42	7
B2	94.28	80	97.14	12	95.71	10
B3	49.28	55	57.14	33	57.14	3
B4	77.14	70	85	36	77.85	10
B5	72.85	87.14	91.42	15	94.28	10
Mean	73.56	69.28	82.28	22.6	79.28	7.6
Std	16.06	14.69	15.38	11.05	16.19	3.08
p-value	0.535	–	0.043	–	0.023	–

P-value denotes the paired T-test between results of (C3,C4,CZ) and other results.
(CH: Channels, ACC: Accuracy, ♯ : Number).

Results and Discussion

- Spatial filter coefficient distribution
  - CSP filters have large weights in several unexpected locations.→ degradation of classification accuracies.
  - the SCSP filters have strong weights over the motor cortex areas and smooth weights over the other areas. → the proposed SCSP yielded filters that are neurophysiologically more relevant and interpretable.



Conclusion

- They investigated the reduction of channels whereby the classification accuracy is constrained to an acceptable range.
- Two criterions
  - Using the first criterion yielded the best classification.
  - Using the second criterion retained the least number of channels.
- The proposed SCSP algorithm yielded an average improvement of 10% in classification accuracy compared to the use of typical three channels
- A visualization of the obtained sparse spatial filters
  - The proposed algorithm improved the results by emphasizing on a limited number of channels with high variances between the classes.