Robust Sparse Representation based Classification Scheme for Non-stationary EEG Signal Classification

Presenter : PhD candidate Younghak Shin
Advisor : Professor Heung-No Lee

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

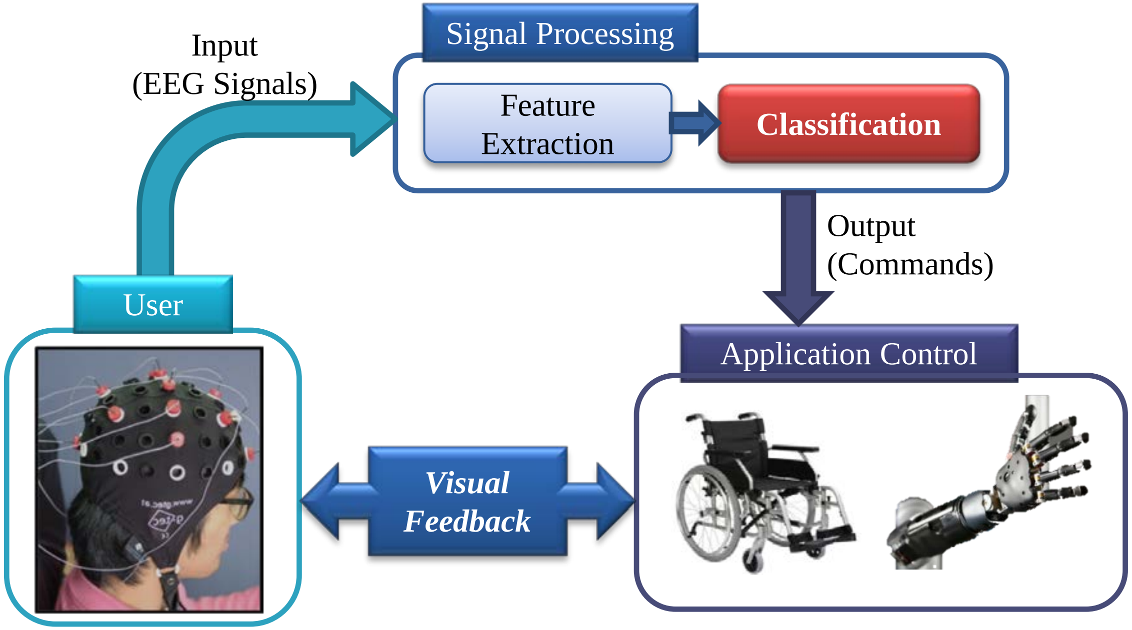
Gwangju Institute of Science and Technology

Outline of Presentation

- Introduction
- Sparse representation based classification for motor imagery BCI
  - Motivation
  - Methods
  - Results
  - Summary
- Evaluation of SRC for non-stationary EEG signals
  - Motivation
  - Methods
  - Results
  - Summary
- Simple adaptive SRC schemes
  - Motivation
  - Methods
  - Results
  - Summary
- Research Outcome

Brain Computer Interface

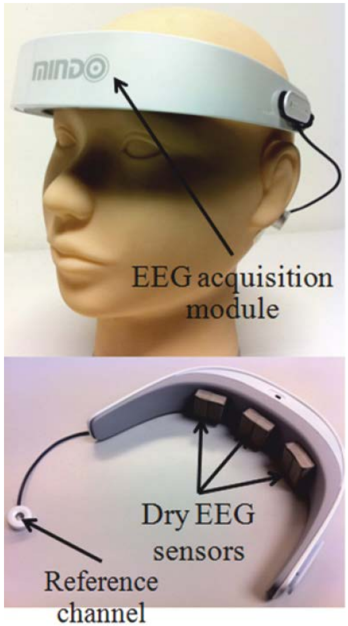
EEG based Brain-Computer Interface

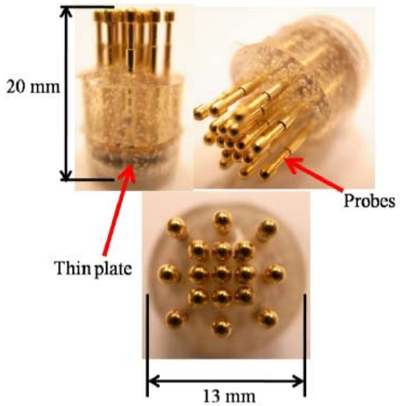


- EEG based BCI systems provide an alternative communication and control channel between human brain and external devices without any normal muscle movements.
- In the BCIs, signal processing is needed to transform the extracted feature of a user's intention into a computer command to control the external device.

Important issue in BCI

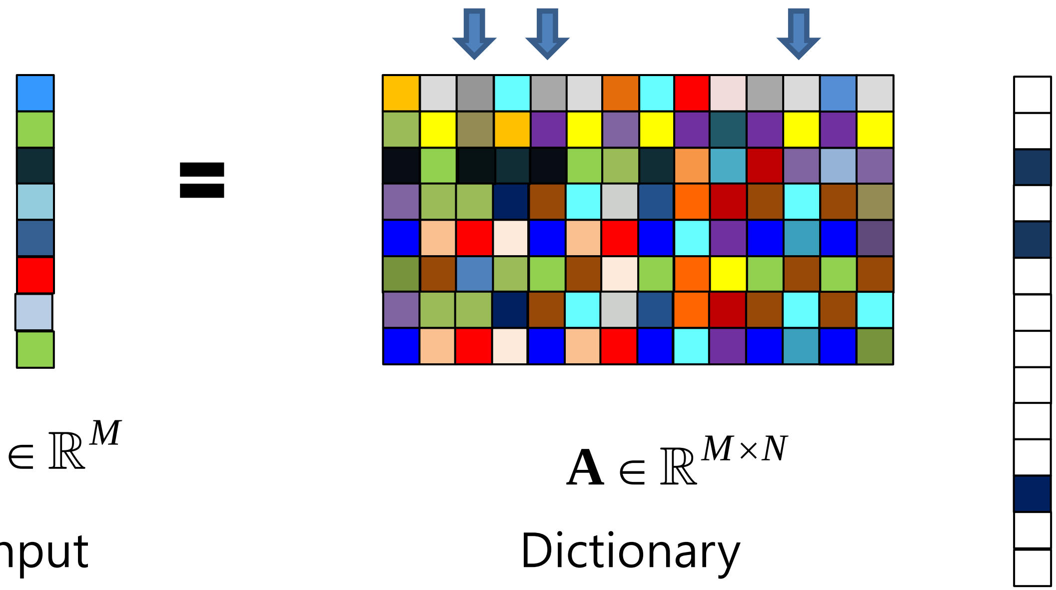
- Recently, much research effort focused on development of portable BCI systems for normal person by using headset shaped scalp electrodes.
- In addition, dry and active electrodes which do not need conductive gel for preparation of EEG recording are developed.
- However, for the BCI systems going beyond laboratory researches, the most important issue is stable classification performance.
- Therefore, powerful signal processing methods are needed.





Sparse Representation (SR)

- Recently, Sparse Representation has received a lot of attention in signal processing and machine learning field.
- The problem of SR is to find the most compact representation of a signal in terms of linear combination of atoms in an over-complete dictionary [Huang 2006].



$\mathbf{y} \in \mathbb{R}^{M}$

Input

$\mathbf{A} \in \mathbb{R}^{M \times N}$

Dictionary

$\mathbf{x} \in \mathbb{R}^{N \times 1}$

Sparse coefficient

$$\mathbf{y} = \mathbf{A}\,\mathbf{x}$$

Sparse Representation (SR)

- The problem of SR is to find the coefficient $\mathbf{x} \in \mathbb{R}^{N \times 1}$:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

where, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is known over-complete dictionary $(M \ll N)$

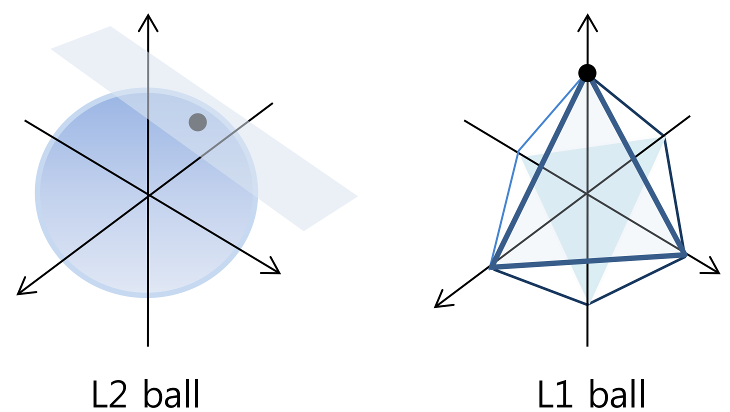
$\mathbf{y} \in \mathbb{R}^{M}$ is measured signal $\|\mathbf{x}\|_0$ denotes the L0 norm.

- Solving this under-determined problem is NP hard.
- The literature of compressive sensing (CS) reveals that if a solution is sparse enough, L1 norm minimization algorithm can solve this optimization problem effectively in polynomial time [Donoho2006].

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

L_p norm is defined by:

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)$$



Sparse representation based classification for motor imagery BCI

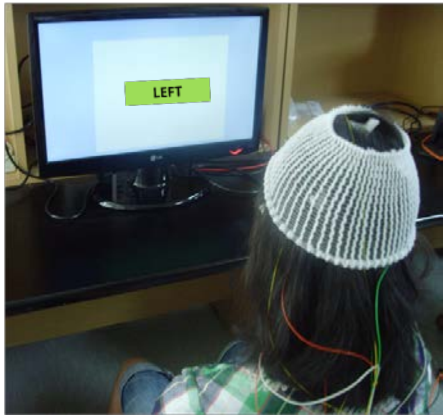
[Shin 2012, Journal of Neural Engineering]

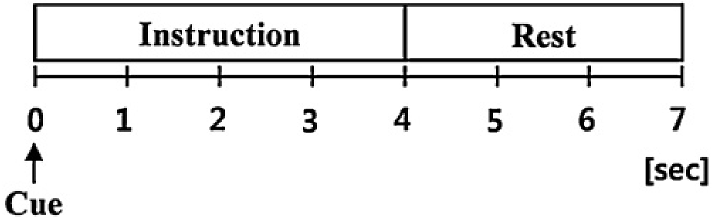
Motivation

- Sparse representation can be used for a number of applications including noise reduction, compression, and pattern recognition.
- Recently, classification based on SR has been studied in face recognition area and have shown robust classification performance [Wright 2009].
- In this study, we apply SR to the EEG signal classification.
- Using Mu and Beta rhythms as a feature of MI BCI, we aim to develop a new Sparse Representation based Classification (SRC) method.

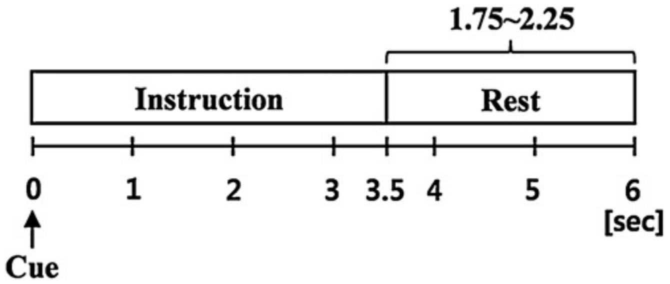
Data acquisition

- We use two different datasets
  - INFONET dataset
    - Five healthy subjects(average age = 22±6.85)
    - Right hand and left hand imaginations
    - 16 EEG channels
    - 80 trials per class

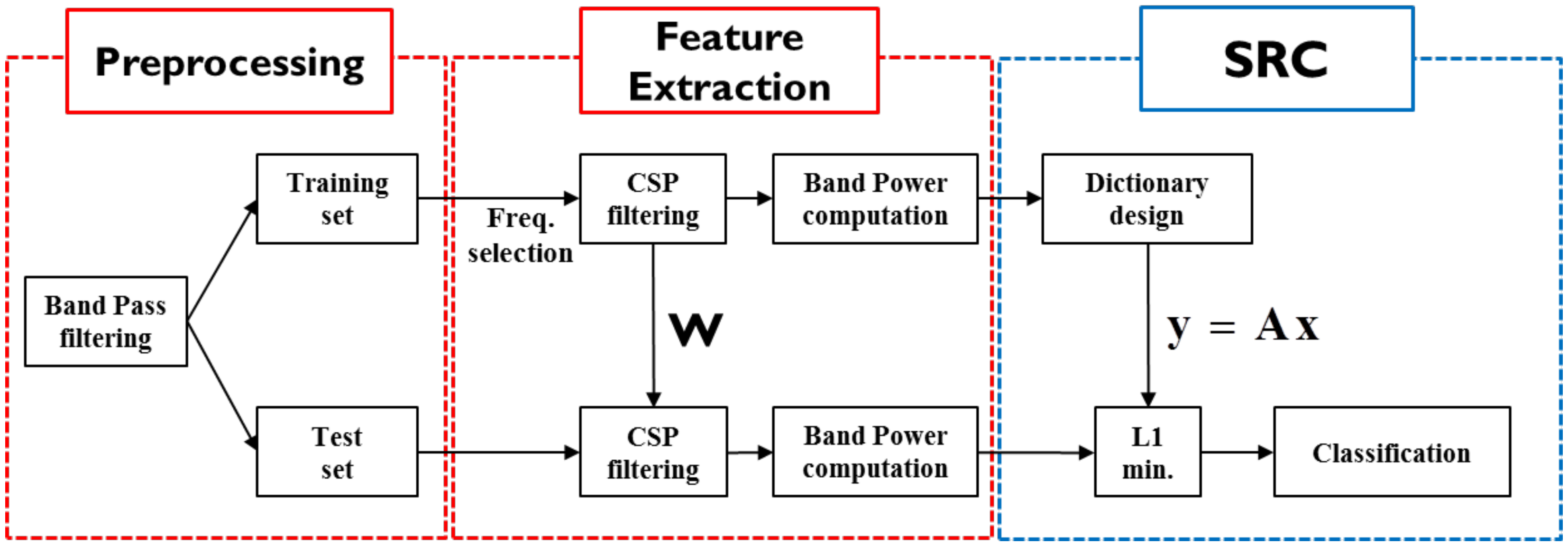




  - Berlin dataset
    - BCI competition dataset (Data set IVa)
    - Five healthy subjects
    - Right hand and right foot imaginations
    - 118 EEG channels
    - 140 trials per class



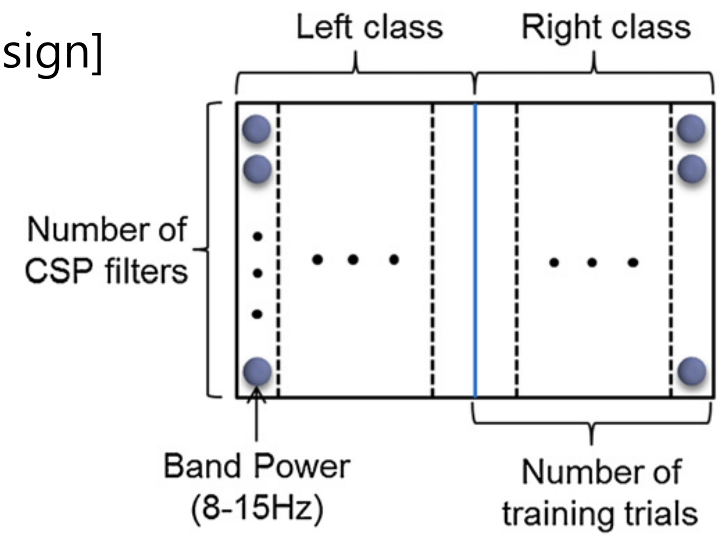
Proposed SRC scheme



- We use a band pass filtering as a preprocessing method.
- We designed dictionary **A** using CSP filtering.
- To use a mu rhythm as a BCI feature, we compute the power of mu band.
- To find sparse coefficient vector **x**, we use the L1 minimization tool for test signal **y**.

Incoherent Dictionary

[Dictionary design]



$$\mathbf{A} := [\mathbf{A}_L ; \mathbf{A}_R] \quad \mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, ..., \mathbf{a}_{i,N_t}]$$ where, i is class, N_t is total trials

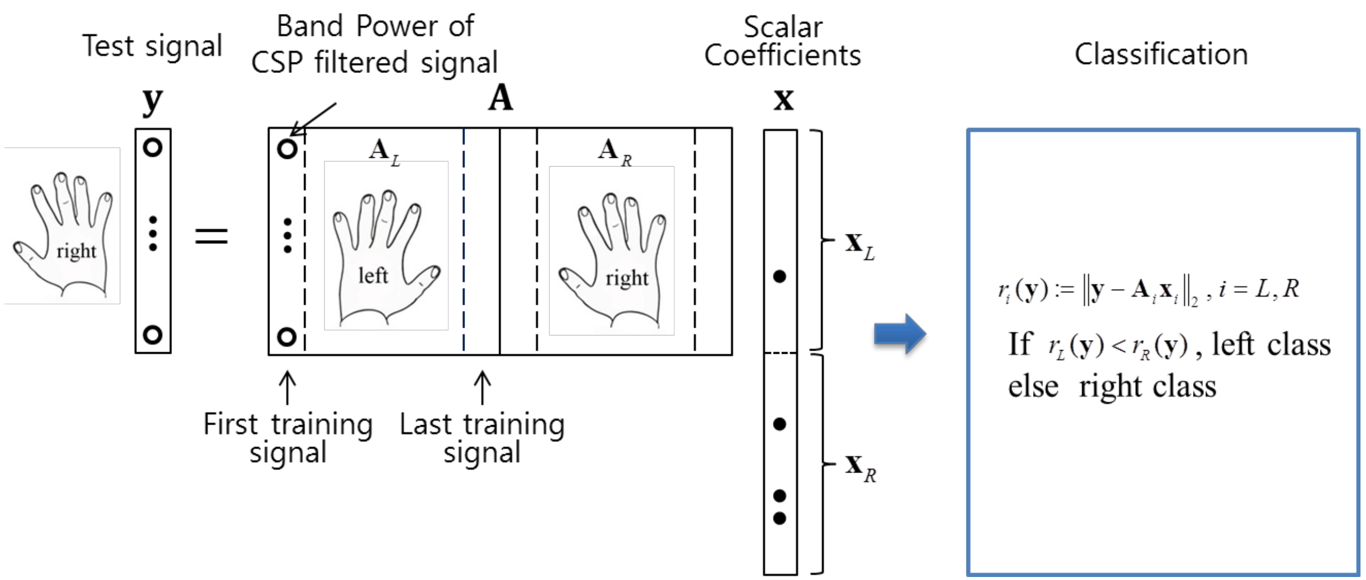
- We use the CSP filtering to design an *incoherent* dictionary.
- When a dictionary is incoherent, a test signal from one particular class can be predominantly represented by the columns of the same class [UP: Donoho 2001].
- Therefore, the incoherent dictionary promotes the sparse representation of the test signal under the L1 minimization.

Sparse Representation based Classification

- The SRC method can be categorized as sparsification step and identification step.
- Sparsification step is formulated as **y = Ax**.
- Where, **y** and **A** indicate a test feature vector and a collection of training feature vectors (**A** is dictionary), **x** is an unknown coefficient vector.
- In the sparsification step, **x** can be recovered by solving following optimization problem via L1 norm minimization tool:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}$$

- Using the recovered **x**, class identification is performed as follows: $\text{class}(\mathbf{y}) = \min_i r_i(\mathbf{y})$



Classification accuracy of INFONET dataset

Subject	SRC Accuracy [%]	LDA Accuracy [%]
A	95.63	93.13
B	63.75	61.87
C	68.14	67.50
D	80	76.25
E	71.25	68.12
Mean (SD)	75.75 (12.60)	73.37 (12.18)

- LOO cross validation is used to evaluate classification accuracy.
- We use 2 CSP filters out of 16.
- For all subjects, the accuracy of the proposed SRC is better than conventional LDA method.

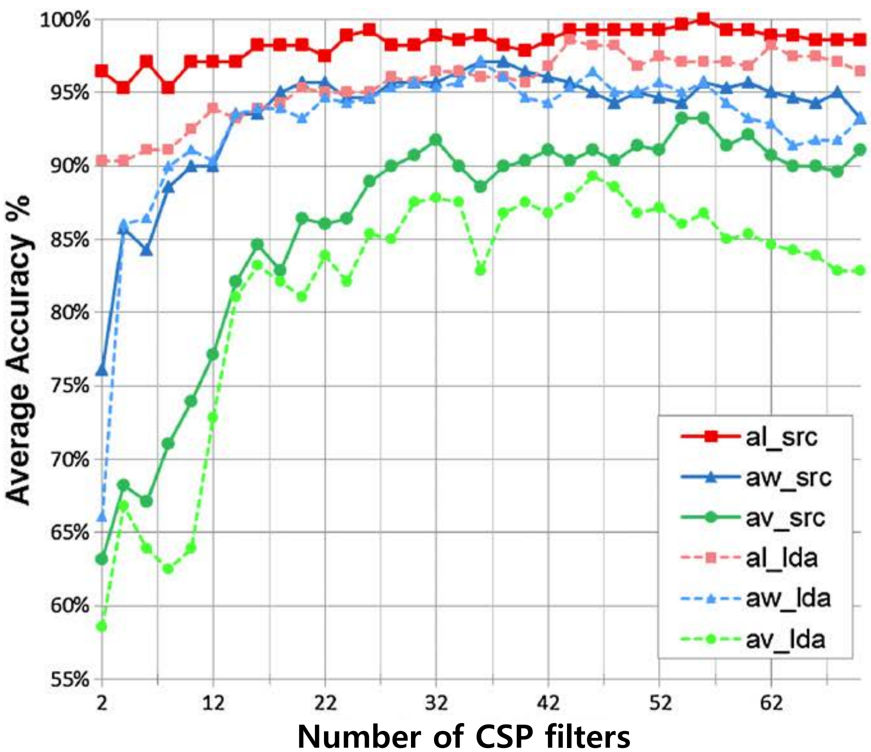
Classification accuracy of Berlin dataset

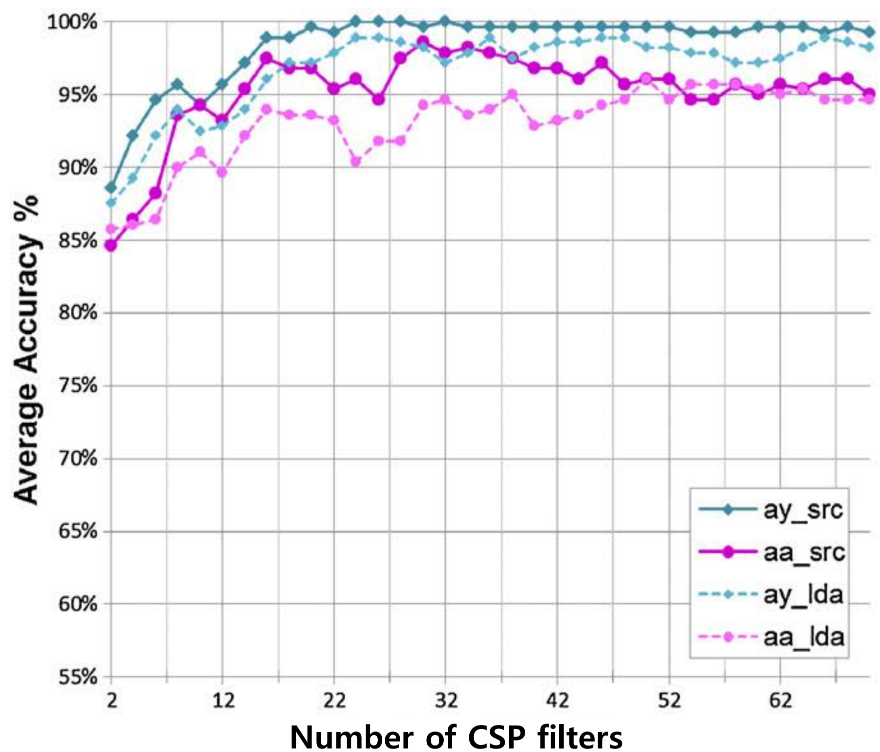
Subject	SRC Accuracy [%]	LDA Accuracy [%]
al	98.93	96.43
ay	100	97.14
aw	95.71	95.36
aa	97.86	94.64
av	91.79	87.86
Mean (SD)	96.85 (3.25)	94.29 (3.72)

- We use 32 CSP filters out of 118.
- For all subjects, the accuracy of the proposed SRC is better than conventional LDA method.

Classification results

- Berlin dataset
  - We examine classification accuracies of SRC and LDA as a function of the number of CSP filters (feature dimensions) for each subject.





Summary

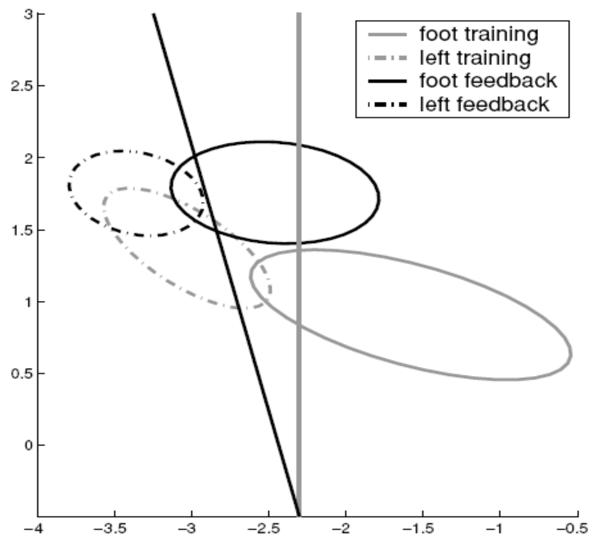
- We propose a sparse representation based classification (SRC) method for the motor imagery based BCI system.
- The SRC method needs a well-designed dictionary matrix made of a given set of training data.
- We use the CSP filtering to make the dictionary uncorrelated for two different classes.
- The SRC method is shown to provide better classification accuracy than the conventional LDA method.

Evaluation of SRC for non-stationary EEG signals

[Shin 2015, Biomedical Signal Processing and Control]

Motivation and purpose

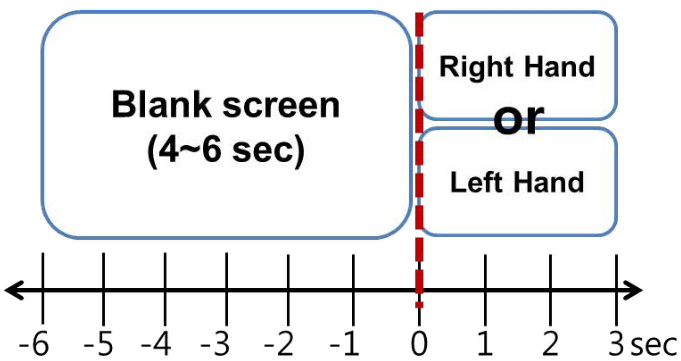
- Due to the non-stationarity of EEG, we can observe that test feature positions vary from original training feature positions in the feature space.
- This is one of the major obstacles in EEG signal classification.
- In this study, our aim is to evaluate the robustness of SRC for non-stationary EEG signal classification.
  - Evaluate the noise robustness of the SRC and SVM methods.
  - Examine working mechanism of SRC as the role of classifier compared with the conventional SVM



[Shenoy 2006]

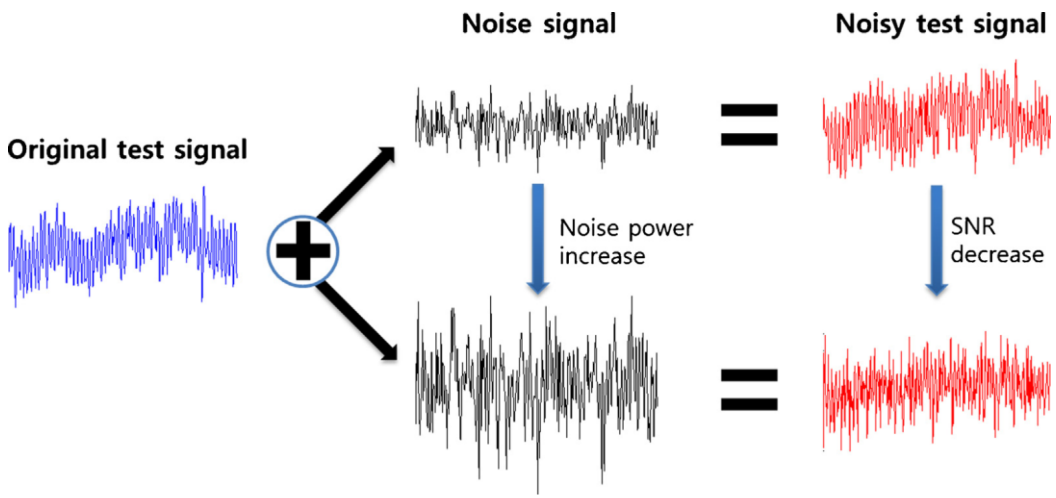
Data acquisition

- We use two-class MI dataset obtained from 20 subjects.
- Right hand and Left hand of motor imagery movements
- 64 EEG channels and 512 sampling rate
- 100 trials per class are collected
- We also record the resting state for each subject to estimate the background noise.
- For the resting state, subject just open their eyes.



Noise robustness analysis

- We generate the noisy test data by introducing two different noise sources into the original test data.
- Test data is contaminated by an additive random Gaussian noise and scalp recorded background noise.
- We generate five different noisy test data with various SNR levels. Thus, we control the noise power of each noise source in five levels.



[Noisy test signal generation using different power of noise signal]

Noise robustness analysis

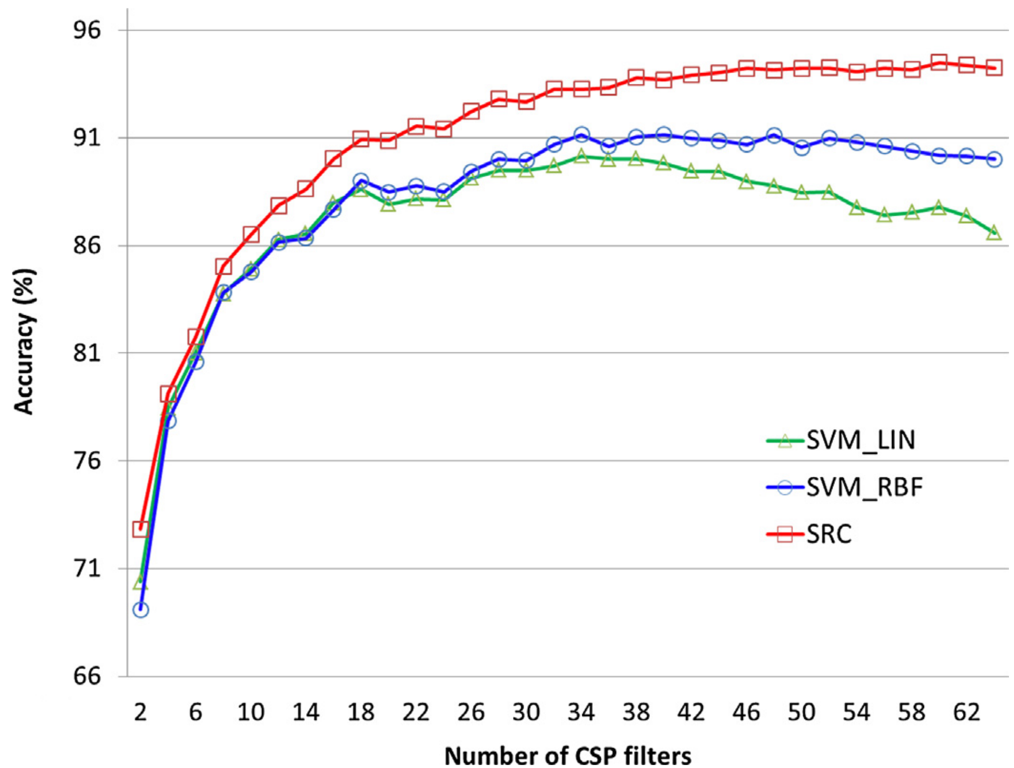
- Random Gaussian noise is artificially generated by m-dimensional Gaussian distribution, i.e., $N_m(\mu, \sigma^2)$ where μ and σ^2 are the mean and variance.
- For the Gaussian noise, we control the noise power by varying the standard deviation of Gaussian distribution.
- Subject-specific background noise is measured by the EEG recording of the resting state.
- In this recording, subject is instructed to just open their eyes without any task for one minute.
- For the background noise, we use a scale factor α to control the noise power as follows:

polluted test signal = test signal + α(resting noise)

- To evaluate and compare the classification accuracy of the SRC and SVM methods, we use the leave-one-out (LOO) cross-validation.

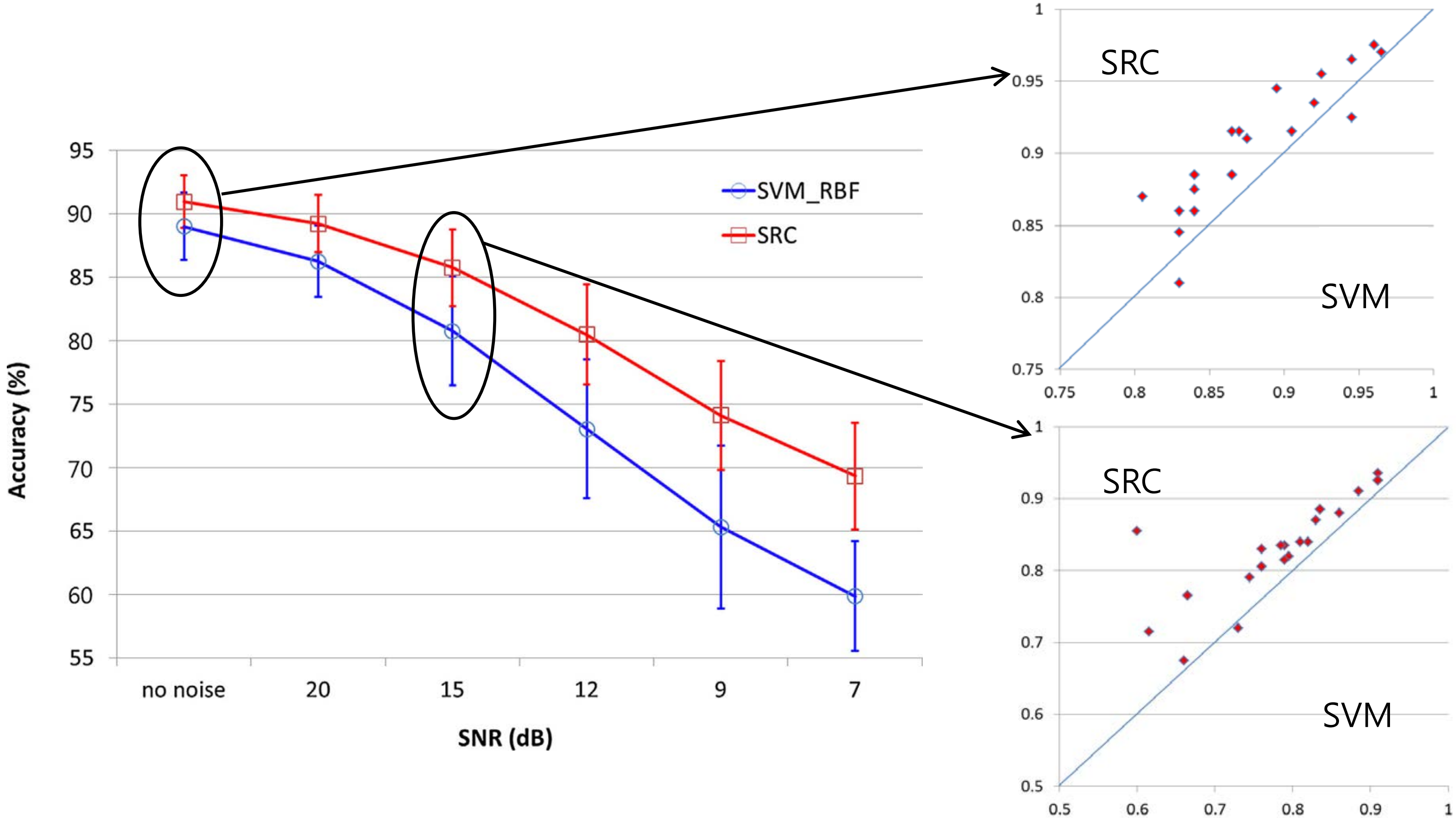
Comparison of classification results

- First, we compare classification accuracy of linear SVM, RBF kernel SVM, and SRC method using original (non-noisy) MI dataset.
- Classification accuracy as a function of the number of CSP filters.
- Regardless of feature dimension, SRC outperforms SVM.



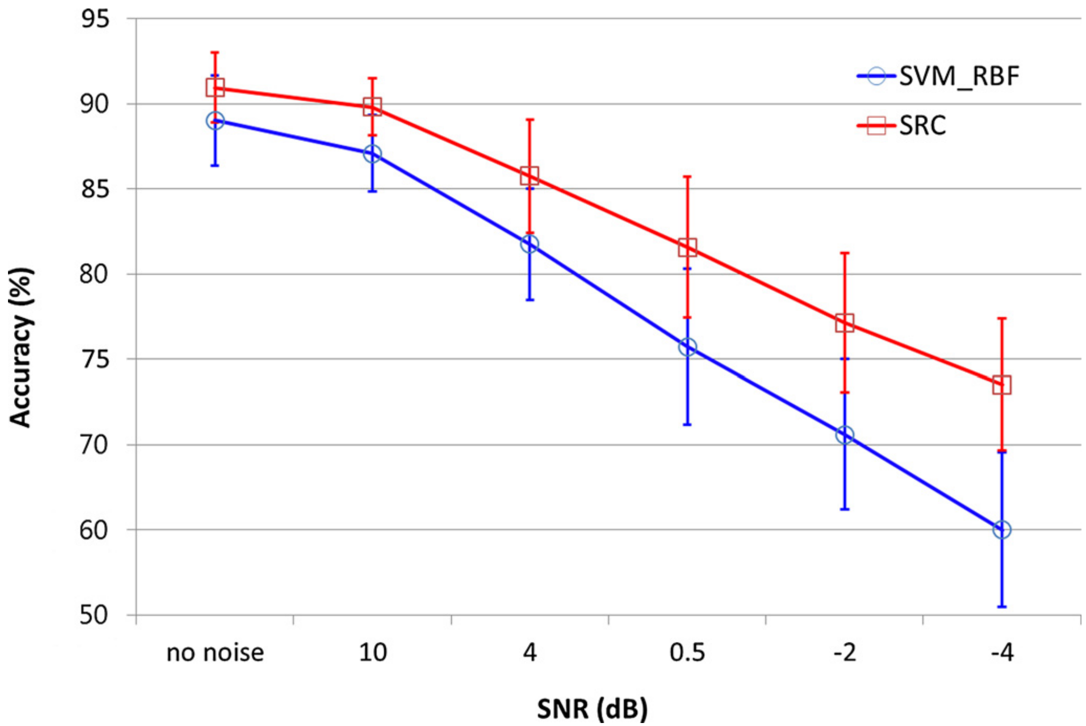
Noise robustness (Gaussian noise)

- We compare average classification accuracy of the SRC and RBF kernel based SVM for the noisy test data by the Gaussian noise.
- We found that the classification accuracy of SRC was higher than that of the RBF SVM for all SNR cases.



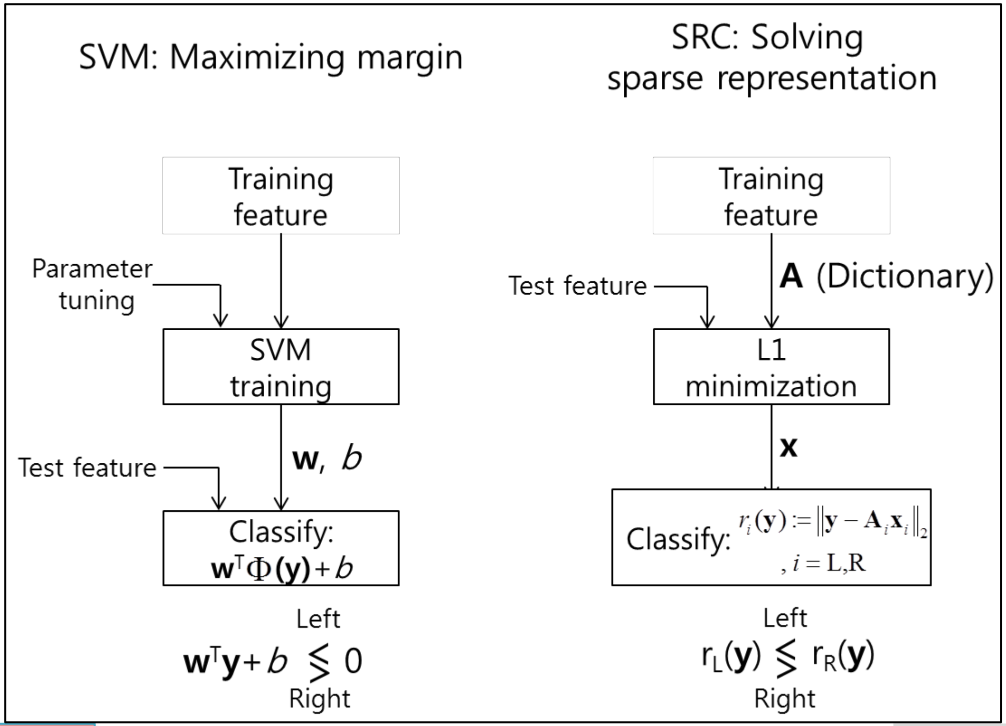
Noise robustness (Background noise)

- Average classification accuracy for noisy test data by the background noise is represented as a function of SNR.
- Classification accuracy of SRC was higher than that of the RBF SVM for all SNR cases.
- In addition, when the noise power increased, the accuracy difference between the SRC and SVM increased.



Comparison of classification mechanism

- In SVM (or LDA), a fixed decision boundary was obtained using all training signals. Then, for each test signal, the fixed decision rule was used for signal classification.
- In SRC, training(or parameter decision) of a classifier is not needed.
- Dictionary is simply formed by collecting the training features. Then, using all training features in the dictionary, sparsification step is performed for each test data adaptively.

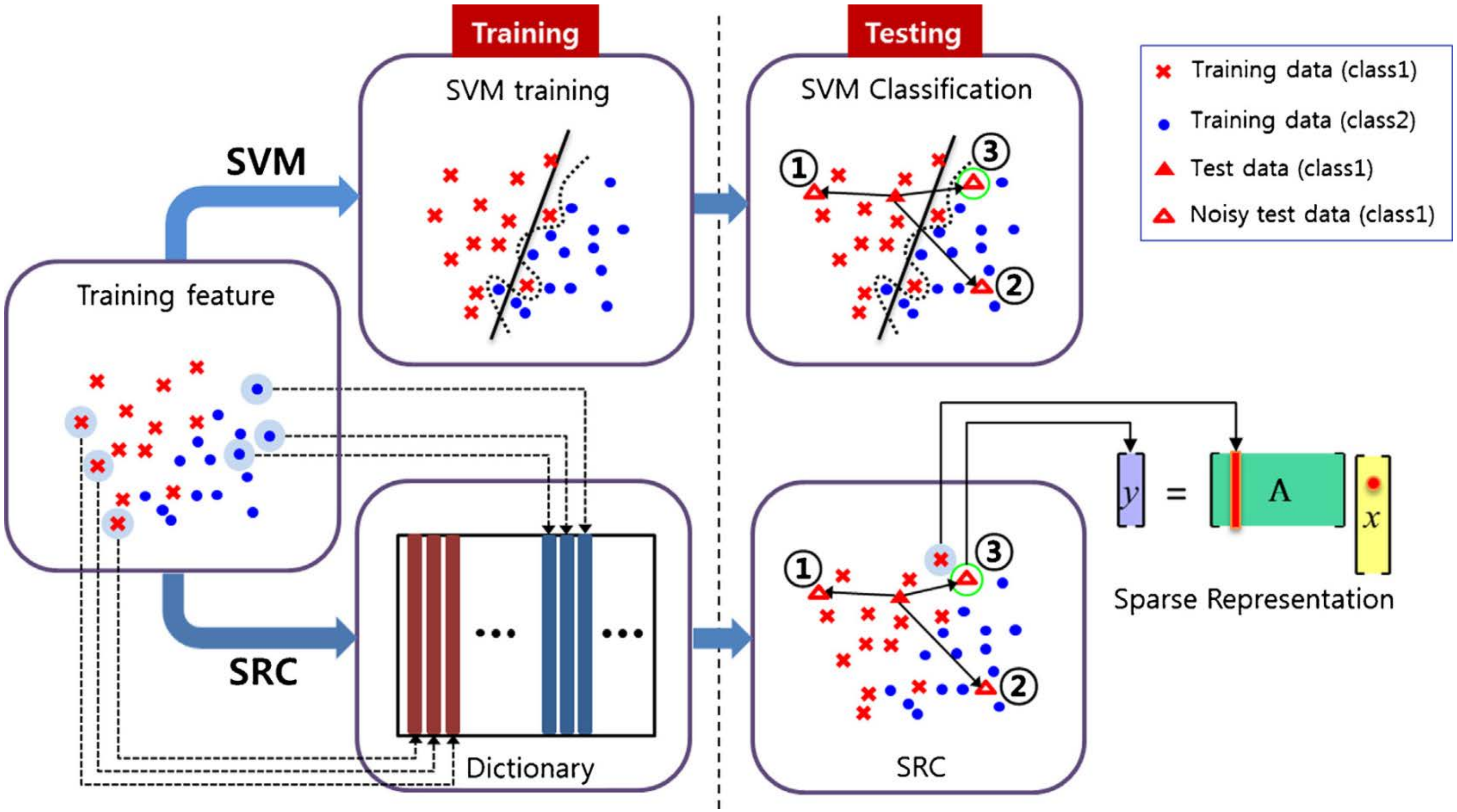


Toy example for polluted test data

① : both SVM and SRC correctly classified online test data

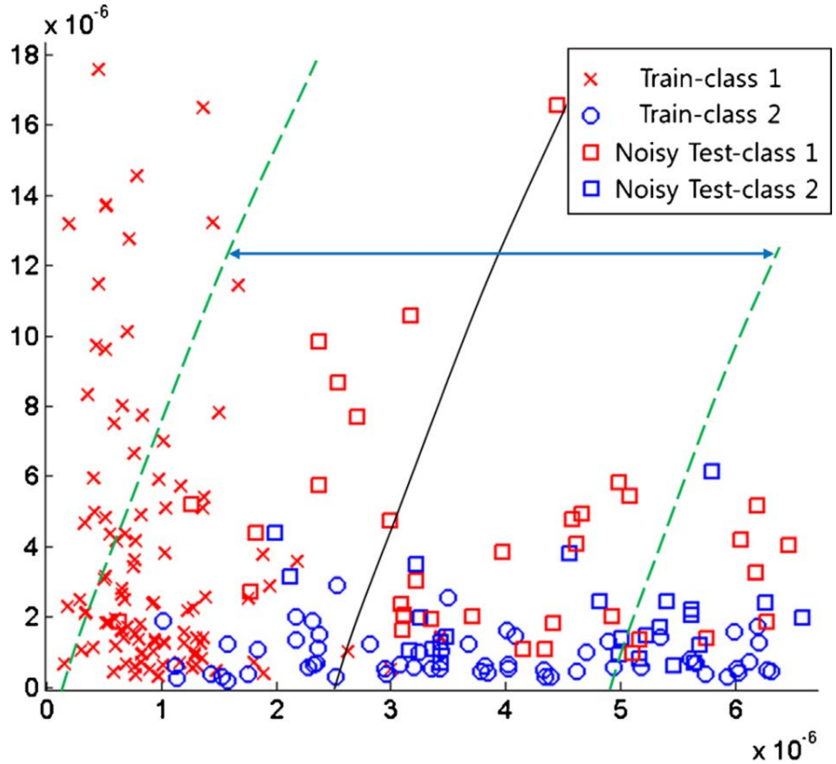
② : both SVM and SRC incorrectly classified online test data

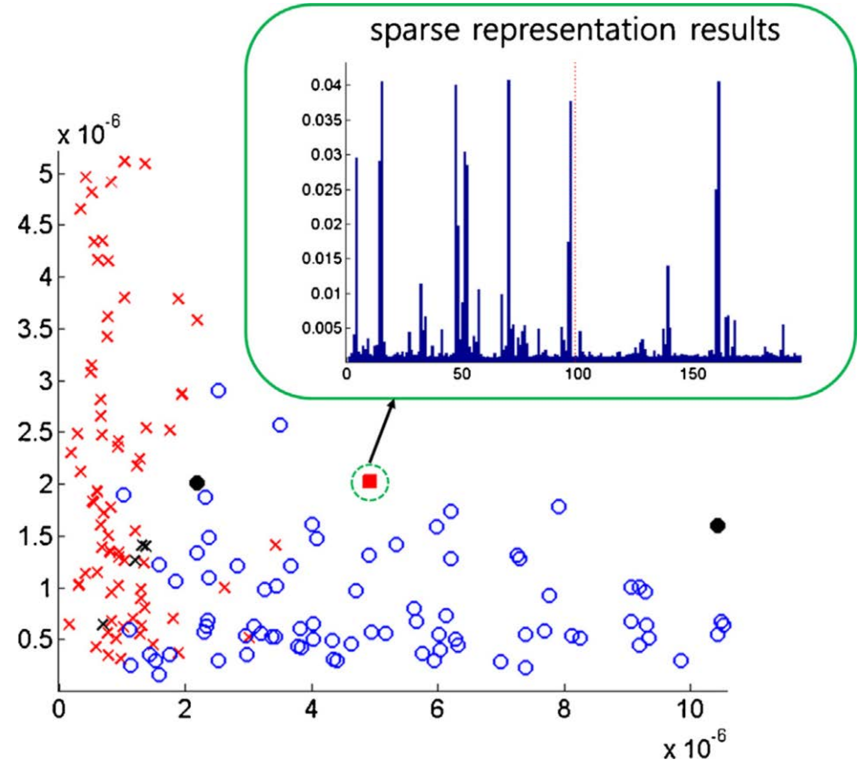
③ : based on the decision rule, the SVM resulted in wrong classification. On the other hand, SRC still had a chance for correct sparse representation with the same class training data



Example data analysis

- The region between the two green dotted lines is chosen as the near area of the decision boundary.
- When we considered case ③ examples, the RBF SVM had 18 miss-classification data. However, the SRC correctly classified 12 test data among 18 test data.
- Right figure shows one instance of the noisy test data that was not correctly classified by the SVM; however, was correctly classified by the SRC method.





Summary

- We evaluated and analyzed the noise robustness of the SRC method using Gaussian and background noise.
- We assessed the classification performance of the SRC when the noise power was varied.
- SRC showed superior noise robustness than the SVM for both Gaussian and background noise.
- The robust classification accuracy of the SRC was due to a different classification mechanism compared with the conventional decision rule based SVM.
- Thus, the SRC showed an inherent adaptive classification mechanism for each test trial via optimal sparse representation of the training trials.

Simple adaptive SRC schemes

[Shin 2015, Computers in Biology and Medicine]

Motivation and purpose

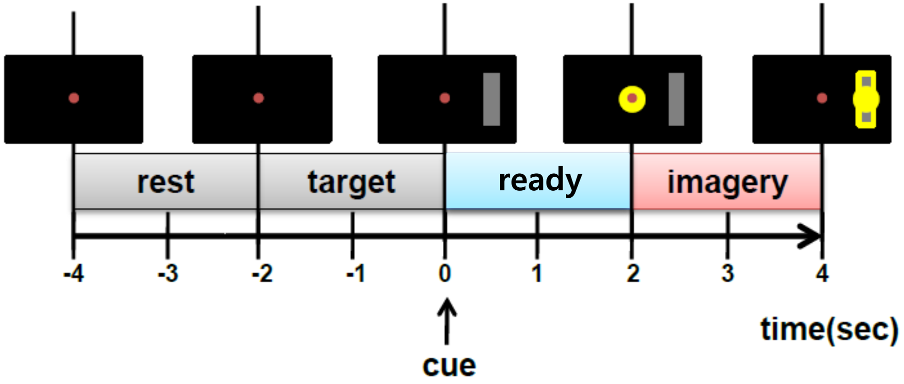
- To overcome the performance decrease caused by the non-stationarity of EEG signals, many adaptive signal classification methods are proposed.
- However, no research has been studied for adaptive SRC scheme for online BCI applications.
- In this study, we propose a simple dictionary update rule based adaptive SRC methods for real-time BCI systems.
- We suggest a dictionary update rule and an incoherence based dictionary modification (IDM) method.
- Using online motor imagery datasets, we evaluate classification performance of the proposed adaptive methods by comparing with the conventional methods.

Data acquisition

- We use online MI based BCI experimental dataset obtained from 10 subjects.
- Right hand(R), left hand(L) and foot(F) motor imagery signals are collected.
- Among R-F, L-F, L-R pairs, best pair is chosen for online classification.
- 64 EEG channels and 512 sampling rate were used.
- The same experimental paradigm was used for both calibration (training) and feedback (online) phases.
- In the one session experiment, total 60 training and 75 online testing trials per class were collected for each subject.

Experimental paradigm

- In each trial, the target bar was represented on 0sec at left, right or down side of monitor screen corresponding to the left, right or foot motor imagery.
- On 2sec after cue onset, subject was instructed to perform the MI task.
- In the training session, we just collected training trials for each MI signal.
- At that time, the classifier had not been designed. Therefore, the yellow ball (feedback) was set to move into the target direction automatically.
- However, in the online (feedback) session, the online feedback was provided in each trial.
- Thus, the yellow ball was controlled by the classified result which was analyzed from intention of each subject using the EEG data collected from 2 to 4sec.



Incoherence based dictionary modification method

- In IDM, we aim to eliminate some training trials that have a high average cross-coherence value with different class training trials.
- We expect to further increase the incoherence of the dictionary.
- Coherence value of the dictionary **A** can be simply estimated by each element of $\mathbf{G} = \mathbf{A}^{\mathrm{T}}\mathbf{A}$.
- We focus on the cross coherence part $\mathbf{G}_{CC}$ between the two classes.
- Using the $\mathbf{G}_{CC}$ we can easily check which trials of class 1 have large coherence values with trials from class 2 dictionary and vice versa.

1. Set *n* the number of elimination trials.
2. Compute the average value of each *column* of $\mathbf{G}_{CC}$.
3. Collect the indices of *column* numbers which have *n* highest average coherence values.
4. Eliminate *n* indices from original class-dictionary.
5. Repeat 2–4 steps for *row* of $\mathbf{G}_{CC}$.

Incoherence based dictionary modification method

- For example, if the number of training trials of each class-dictionary is five, then the dimension of **G** is 10 × 10 .
- We extract columns from 1-th to 5-th and rows from 6-th to 10-th of the **G** which are cross coherence part $\mathbf{G}_{CC}$.
- The values of last row and column represent the averaged value of five columns and rows respectively.
- The third row and column shows highest averaged value (n = 1).
- This means that 8-th row (8-th trial from class 2 dictionary) and third column (third trial from class 1 dictionary) shows high coherence value with many trials from the other class-dictionary.

$\mathbf{G}_{CC}$

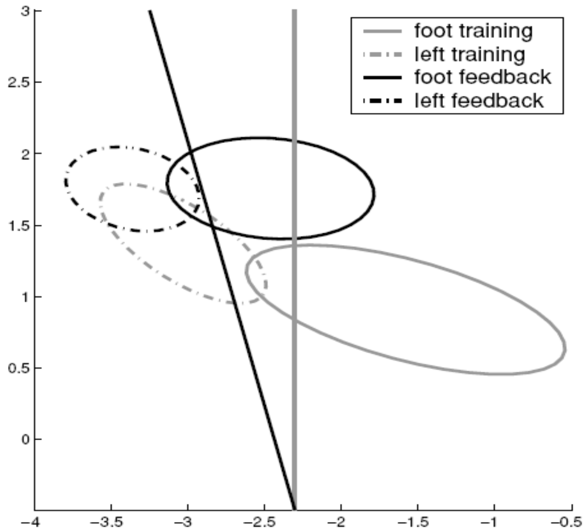
	1	2	3	4	5	Avg.
6	9	1	8	1	1	4
7	1	8	2	2	1	2.8
8	3	2	9	9	9	6.4
9	2	1	2	2	2	1.8
10	2	9	8	1	1	4.2
Avg.	3.4	4.2	5.8	3	2.8	

Eliminate 8th trial of $\mathbf{A}_2$

Eliminate 3rd trial of $\mathbf{A}_1$

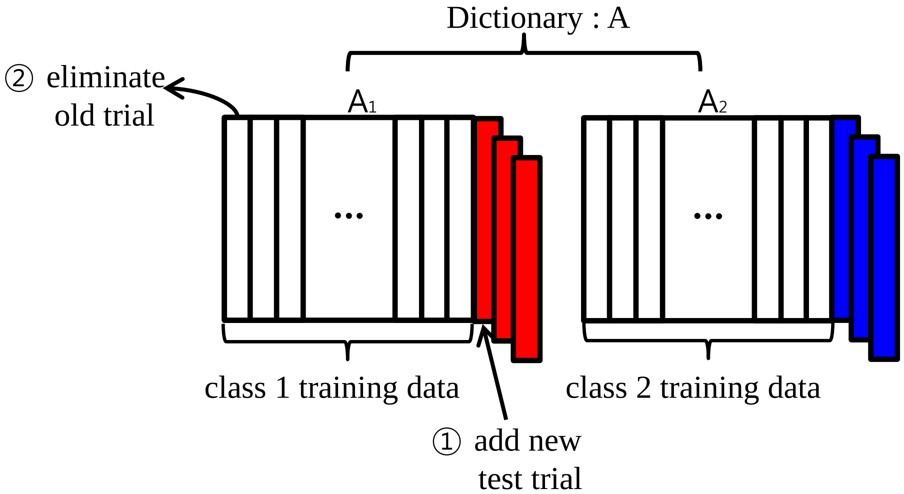
Dictionary update methods

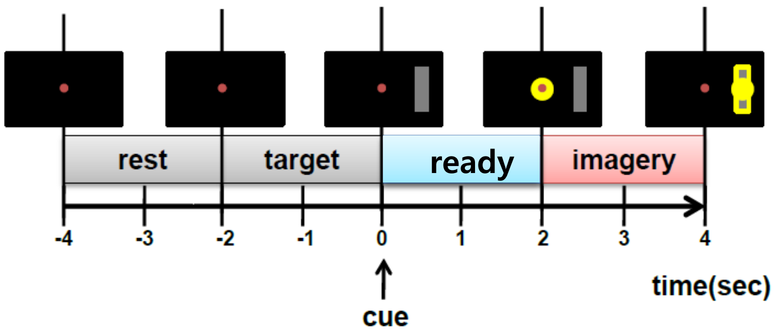
- The main concept of the conventional adaptive classification is re-adjustment(re-training) of the classifier for the new test data.
- In the SRC scheme, one important characteristic is that training of a classifier is not needed.
- Due to this unique classification mechanism, a simple intuitive method for adaptive SRC is dictionary update.
- In the online testing phase, a feature vector of a new test trial **y** can be easily updated as a new column of the dictionary.
- We consider four types of dictionary update rule:
  - Supervised accumulated update (SAU)
  - Supervised fixed update (SFU)
  - Unsupervised accumulated update (UAU)
  - Unsupervised fixed update (UFU)



Dictionary update rules

- In the supervised update rule, the target class label of test trials is used for updating the online test trials.
- In the unsupervised update rule, each test trial is updated into the corresponding class-dictionary based on the feedback result.
- In the accumulated update method, as shown in ① of left figure, all updated test trials are just stacked at the end (last column) of the class-dictionary.
- In the fixed update rule, the oldest training trial, i.e., the first training trial of the class-dictionary is eliminated as shown in ② when each new test trial is updated.





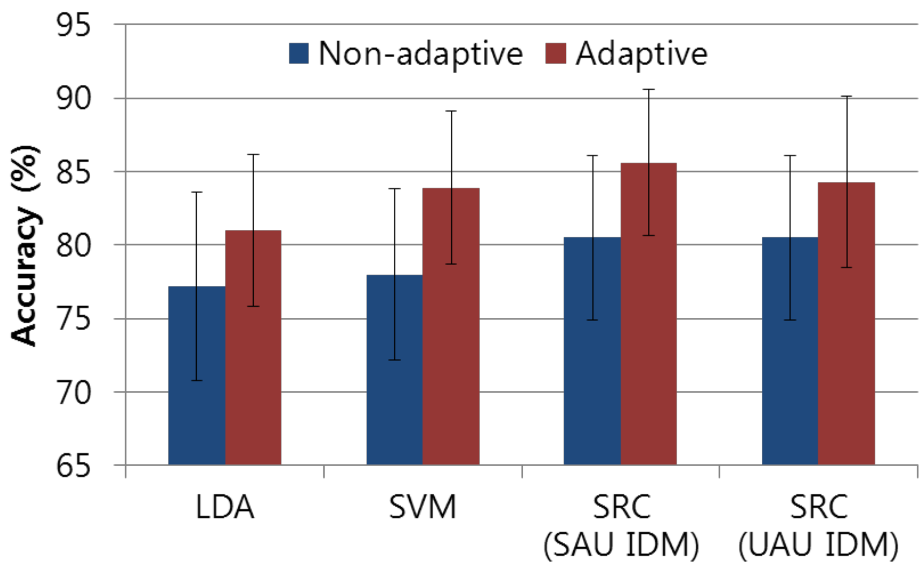
Comparison of classification results

- The proposed simple dictionary update methods with and without IDM show improved mean classification accuracy than the conventional SRC method.
- All methods with IDM show better mean accuracy than without IDM method.
- Supervised methods, i.e., SAU and SFU, show more improved results than the unsupervised methods, UAU and UFU. However, mean difference is very small.

Dataset	SRC		SRC_SAU		SRC_SFU		SRC_UAU		SRC_UFU	
	w/o IDM	w/ IDM	w/o IDM	w/ IDM	w/o IDM	w/ IDM	w/o IDM	w/ IDM	w/o IDM	w/ IDM
1	66	66.7	67.3	70.7	66.0	64.7	66.0	67.3	66.0	67.3
2	86	86.7	88.0	88.0	88.0	88.0	87.3	89.3	82.7	90.7
3	88.7	90.7	90.0	90.0	89.3	90.7	90.0	90.7	90.7	88.7
4	96.4	96.4	96.4	96.4	97.1	97.1	96.4	96.4	96.4	96.4
5	83.3	89.3	93.3	96.0	96.0	96.7	93.3	95.3	94.7	97.3
6	82.7	78.7	86.7	86.7	84.0	84.0	80.0	84.0	80.7	83.3
7	77.3	75.3	78.0	80.0	78.7	79.3	76.7	77.3	79.3	78.0
8	73.3	88.0	88.7	88.7	89.3	91.3	78.0	89.3	84.7	90.7
9	70.0	75.3	74.0	74.7	73.3	74.0	70.0	72.0	70.0	71.3
10	62.0	64.0	66.0	68.7	67.3	71.3	62.0	63.3	68.0	66.7
11	84.0	87.3	88.7	89.3	88.7	89.3	86.7	88.0	88.0	88.7
12	96.7	96.0	97.3	98.0	97.3	98.0	96.7	98.0	96.7	98.0
Mean	80.5	82.9	84.5	**85.6**	84.6	85.4	81.9	84.3	83.1	84.8

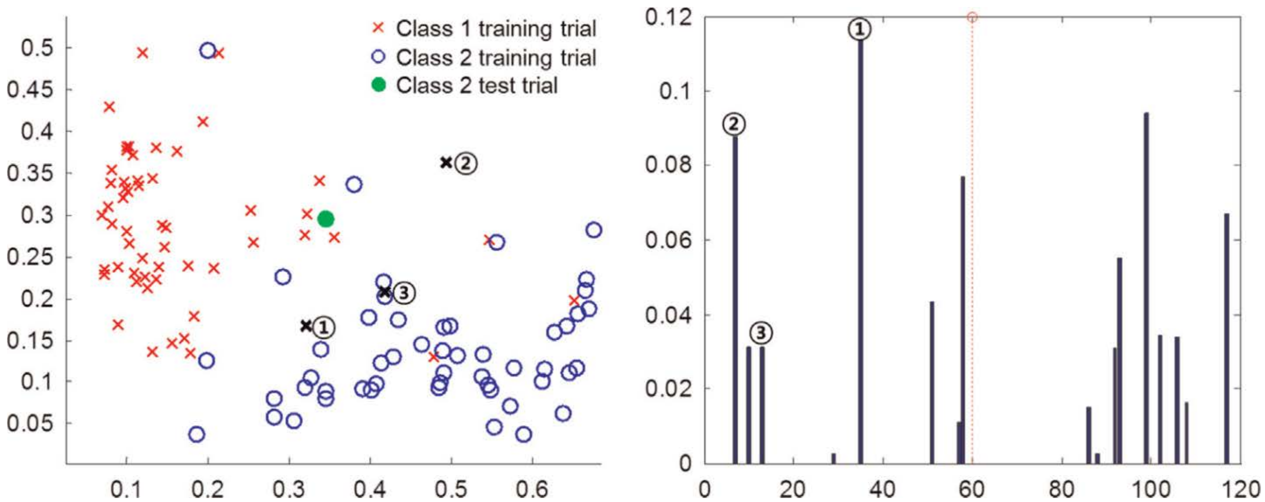
Comparison of classification results

- Proposed SAU and UAU with IDM show better mean classification accuracy than the other adaptive LDA and SVM methods.
- Even though the accuracy difference between the unsupervised SRC and adaptive SVM is not much, in the conventional adaptive methods, re-training (re-adjustment) of the hyper-plane is time consuming process.
- However, in the proposed methods, dictionary update for adaptation of each test sample is very simple process and re-training of classifier is not needed.

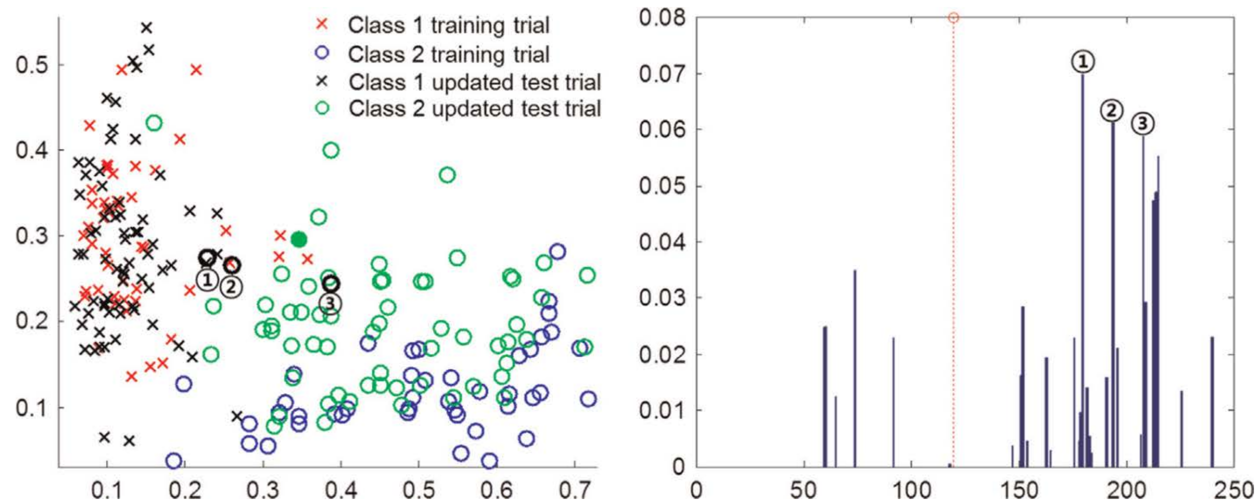


Example data analysis

- Classification results of conventional SRC for one test sample of dataset 5.



- Classification results of SRC_UAU IDM for the same test sample.



Summary

- We propose dictionary update methods with incoherence based dictionary modification (IDM) as adaptive SRC schemes.
- In the IDM, we try to create a maximally incoherent dictionary for SRC by using a simple incoherence measure of the training data.
- In the dictionary update methods, test data are easily updated and utilized for the classification of other new test data without requiring any additional computation.
- We find that proposed IDM based adaptive SRC schemes show improved classification results compared to the conventional SRC.
- Unsupervised adaptive SRC schemes which are more practically applicable show competitive classification accuracy than other adaptive LDA and SVM methods.

Research Outcome

Research outcomes		
International Journal paper	International conference paper	patent
3	6	1

- **Younghak Shin**, Seungchan Lee, Junho Lee and Heung-No Lee*, "Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems", *Journal of Neural Engineering*, no. 9, 056002, 2012**.**
- **Younghak Shin**, Seungchan Lee, Minkyu Ahn, Hohyun Cho Sung Chan Jun and Heung-No Lee*, "Noise Robustness Analysis of Sparse Representation based Classification Method for Non-stationary EEG Signal Classification*", Biomedical Signal Processing and Control* Vol.21, pp. 8-18, Aug. 2015.
- **Younghak Shin**, Seungchan Lee, Minkyu Ahn, Hohyun Cho, Sung Chan Jun and Heung-No Lee* "Simple Adaptive Sparse Representation based Classification Schemes for EEG based Brain-Computer Interface Applications" *Computers in Biology and Medicine* Vol.66, pp.29-38, Nov. 2015.
- Heung-No Lee, **Younghak Shin**, Seungchan Lee, "**뇌-컴퓨터 접속 장치, 그리고 그의 분류 방법**", patent number: 10-1380964, registration date: Mar. 27th, 2014.

Thank you for attention!

Reference

[Wolpaw 1991] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, “An EEG-based brain-computer interface for cursor control,” Electroencephalogr. Clin. Neurophysiol., vol. 78, pp. 252-259, 1991.

[Huang 2006] K. Huang and S. Aviyente, “Sparse Representation for Signal Classification,” Neural Information Processing Systems, 2006.

[Donoho 2006] D.Donoho, “Compressed sensing,” IEEE Trans. Information Theory, vol. 52, pp. 1289-1306, 2006

[Wright 2009] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, Yi Ma, “Robust Face Recognition via Sparse Representation” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210~227, February 2009

[Shenoy 2006] P.Shenoy, M.Krauledat, B.Blankertz, R.P.N.Rao, K.-R.Muller, “Towards adaptive classification for BCI”, J. Neural Eng. 3 (2006) R13.R23.

[Shin 2012] S. Younghak, L.Seungchan, L.Junho, L.Heung-No, “Sparse representation-based classification scheme for motor imagery-based brain–computer interface systems”, J. Neural Eng. 9 (2012) 056002.

[Blankertz 2008] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Müller. “Optimizing Spatial Filters for Robust EEG Single-Trial Analysis,” IEEE Signal Proc. Magazine, 25(1):41-56, January 2008.

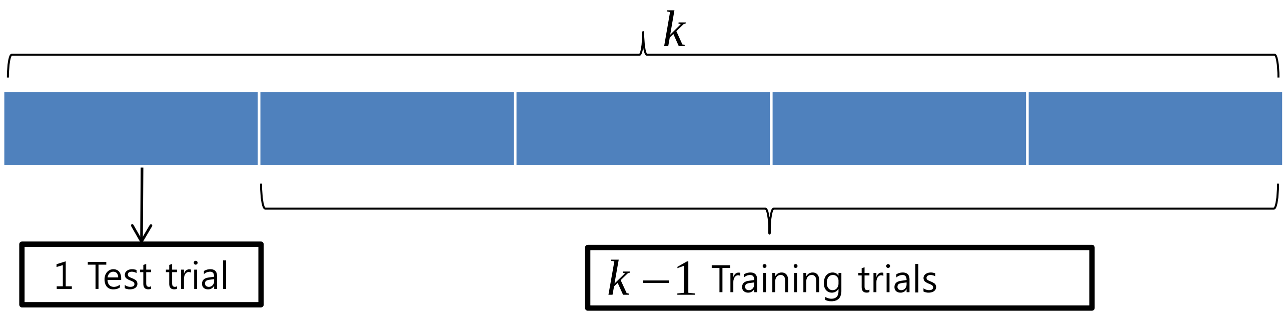
[Shin 2015-1] Younghak Shin, Seungchan Lee, Minkyu Ahn, Hohyun Cho Sung Chan Jun and Heung-No Lee*, “Noise Robustness Analysis of Sparse Representation based Classification Method for Non-stationary EEG Signal Classification”, Biomedical Signal Processing and Control Vol.21, pp. 8-18, Aug. 2015.

[Shin 2015-2] Younghak Shin, Seungchan Lee, Minkyu Ahn, Hohyun Cho, Sung Chan Jun and Heung-No Lee* “Simple Adaptive Sparse Representation based Classification Schemes for EEG based Brain-Computer Interface Applications” Computers in Biology and Medicine, Vol.66, pp.29-38, Nov. 2015.

[Donoho 2001] Donoho D L and Huo X Uncertainty principles and ideal atomic decomposition IEEE Trans. Inf. Theory, Vol.47, pp. 2845-2862, 2001.

Performance evaluation

- To evaluate the classification accuracy for each subject, we use the leave-one-out (LOO) cross-validation.
- LOO is useful for increasing the number of independent classification tests with a given limited data trials.
- Each time, one of the total trials is used as the test trial and the other trials are the training set.
- This method is repeated with all different combination of subsets.



- The classification accuracy is calculated as :

$$\text{Accuracy}(\%) = \frac{\text{correct test trials}}{\text{total test trials}} \times 100$$

CSP(Common Spatial Pattern)

Find vectors w satisfying the following optimization problems (Second order statistics)

$$\max_w \left(\frac{w_i^T C_R w_i}{w_i^T C_F w_i} \right), \ C_R = X_R X_R^T, C_F = X_F X_F^T$$

$\Rightarrow \min_w \left(-w_i^T C_R w_i \right)$ subject to $w_i^T C_F w_i = 1$ From Lagrangian method,

$$\Rightarrow L(w_i, \lambda_i) = -w_i^T C_R w_i + \lambda_i (w_i^T C_F w_i - 1)$$

$$\frac{d}{dw_i} L(w_i, \lambda_i) = -C_R w_i + \lambda_i C_F w_i = 0$$

$$\Rightarrow C_R w_i = \lambda_i C_F w_i$$

$$\therefore \max_w \left(\frac{w_i^T C_R w_i}{w_i^T C_F w_i} \right) = \max_w \left(\frac{w_i^T \lambda_i C_F w_i}{w_i^T C_F w_i} \right) = \max_w \lambda_i$$

$$C_R w_i = \lambda_i C_F w_i$$

$$\Rightarrow \left| C_R - \lambda_i C_F \right|_{\det} w_i = 0$$

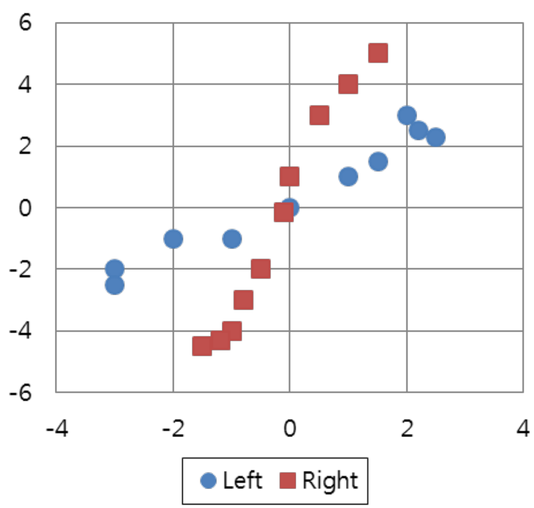
$$\left| C_R - \lambda_i C_F \right|_{\det} = 0$$

$$\left| X_R X_R^T - \lambda_i X_F X_F^T \right|_{\det} = 0 \quad \text{find max } \lambda_i$$

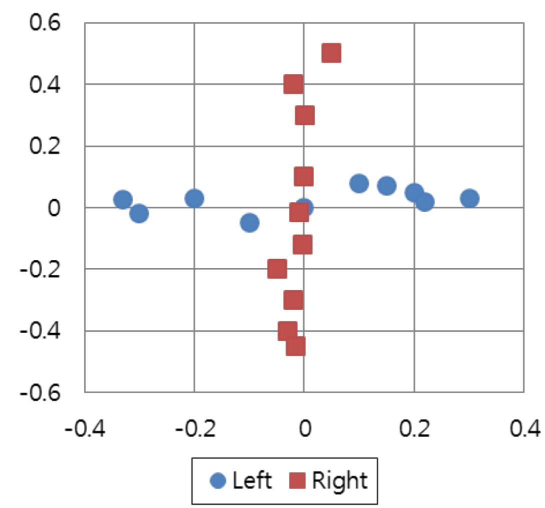
$\Rightarrow w$ is an eigen vector corresponding to the largest eigen value

CSP(Common Spatial Pattern) filtering

- CSP filtering is a powerful signal processing technique suitable for EEG-based BCIs [Blankertz 2008].
- CSP filters maximize the variance of the spatially filtered signal for one class while minimizing it for the other class.
- In our method, the CSP filtering was used to produce high *incoherence* between the two group of columns in the dictionary.
- Using the CSP filter, we form maximally uncorrelated feature vectors between the two classes.



[before CSP filtering]



[after CSP filtering]

LDA(Linear Discriminant Analysis)

Find the optimal direction **w** to project data upon so that between-class variance is maximized and within-class variance is minimized.

Let's define linear projection : $\mathbf{y} = \mathbf{w}^T\mathbf{x}$, where $\mathbf{w}$ is optimal projection vector, $\mathbf{x}$ is sample($\mathbf{x}_{R,}\mathbf{x}_F$)

Define sample average: $\mathbf{m}_R = \frac{1}{N_i}\sum_i \mathbf{x}_R,\ \mathbf{m}_F = \frac{1}{N_i}\sum_i \mathbf{x}_F$

And after projection average: $\bar{m}_R = \frac{1}{N_i}\sum_i \mathbf{y}_R = \frac{1}{N_i}\sum_i \mathbf{w}^T\mathbf{x}_R = \mathbf{w}^T\mathbf{m}_R$

$$\bar{m}_F = \frac{1}{N_i}\sum_i \mathbf{y}_F = \mathbf{w}^T\mathbf{m}_F$$

Then we can define Between class scatter: $\mathrm{B_S} = \left|\bar{m}_R - \bar{m}_F\right| = \left|\mathbf{w}^T\mathbf{m}_R - \mathbf{w}^T\mathbf{m}_F\right|$

Also, We can define Within class scatter: $\mathrm{W_S} = \bar{s}_R{}^2 + \bar{s}_F{}^2$

$$\bar{s}_R{}^2 = \sum_{y\in R}(y-\bar{m}_R)^2,\ \bar{s}_F{}^2 = \sum_{y\in F}(y-\bar{m}_F)^2$$

Then we can define objective function:

$$\max_{\mathbf{w}} J(\mathbf{w}) = \frac{\text{Between class scatter}}{\text{Within class scatter}} = \frac{\mathrm{B_S}^2}{\mathrm{W_S}} = \frac{\left|\bar{m}_R - \bar{m}_F\right|^2}{\bar{s}_R^2 + \bar{s}_F{}^2} = \frac{\mathbf{w}^T\mathbf{S_B}\mathbf{w}}{\mathbf{w}^T\mathbf{S_w}\mathbf{w}}$$

Linear discriminant analysis (LDA)

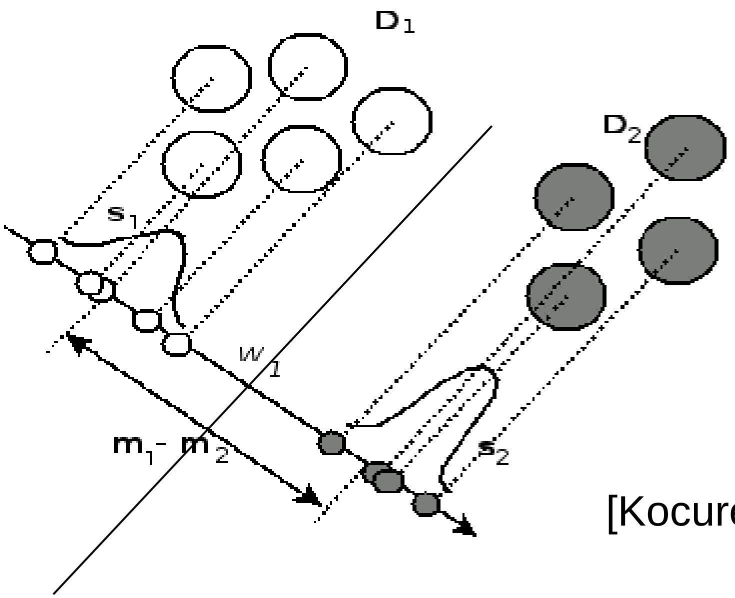
- The LDA (also known as Fisher's LDA) approach aims to find the optimal direction, $\mathbf{w_1}$, to maximize the Fisher ratio:

$$J(\mathbf{w}_1) = \frac{\mathbf{w}_1{}^T \mathbf{S_B} \mathbf{w}_1}{\mathbf{w}_1{}^T \mathbf{S_w} \mathbf{w}_1}$$

where, $\mathbf{S_B} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$

$$\mathbf{S_w} = \sum_i (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

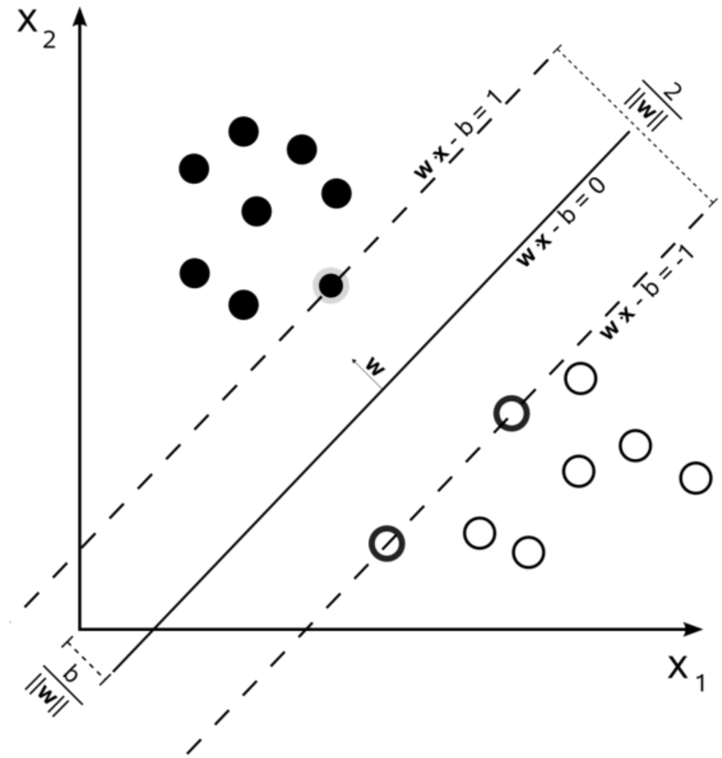
- The maximization of mean distances and minimization of class scatters.



[Kocurek 2007]

Support vector machine (SVM)

- The idea of SVM is proposed by Vapnik, aimed to find decision hyperplane with maximum margin, which is the distance between the hyperplane and the nearest training feature vectors (support vectors).



$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2,$$

$$\text{subject to} \quad y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1$$

$$n = 1, \ldots, N$$

- In the BCI field, SVM has shown robust classification performance in many studies.

Recovery algorithm

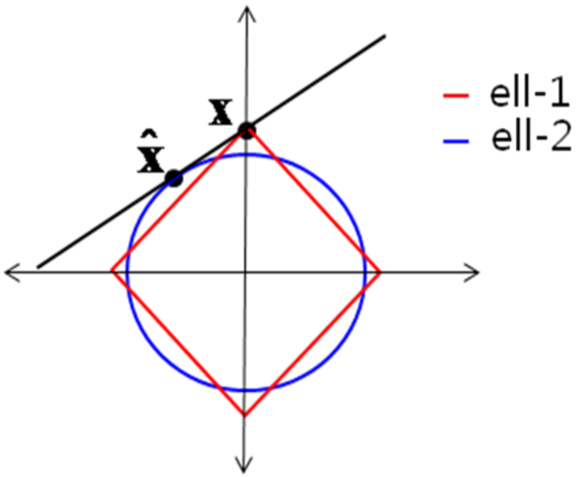
L_p norm is defined by: $\|x\|_P = \left(|x_1|^p + |x_2|^p + \cdots + |x_n|^p\right)$

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

- The L_0 norm is equivalent to the number of nonzero components in the vector **x**. This involves combinatorial search ; $\binom{N}{K}$

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

- The L_2 norm solution is $\bar{\mathbf{x}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}$
- This obtained by Least-square method



$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

- If solution is sparse enough, L_1 norm solution is equivalent to the L_0 norm solution.
- This problem can be solved by standard linear programming in polynomial time.