On the Use of Convolutional Neural Networks and Augmented CSP Features for Multi-class Motor Imagery of EEG Signals...

Conference Paper · August 2015

DOI: 10.1109/EMBC.2015.7318929

CITATIONS
10

READS
418

4 authors:

Huijuan Yang
Institute for Infocomm Research
46 PUBLICATIONS **493** CITATIONS
SEE PROFILE

Siavash Sakhavi
National University of Singapore
4 PUBLICATIONS **16** CITATIONS
SEE PROFILE

Kai Keng Ang
Institute for Infocomm Research
156 PUBLICATIONS **2,838** CITATIONS
SEE PROFILE

Cuntai Guan
Nanyang Technological University
304 PUBLICATIONS **5,740** CITATIONS
SEE PROFILE

On the Use of Convolutional Neural Networks and Augmented CSP Features for Multi-class Motor Imagery of EEG Signals Classification

Huijuan Yang, Siavash Sakhavi, Kai Keng Ang and Cuntai Guan

***Abstract*— Learning the deep structures and unknown correlations is important for the detection of motor imagery of EEG signals (MI-EEG). This study investigates the use of convolutional neural networks (CNNs) for the classification of multi-class MI-EEG signals. Augmented common spatial pattern (ACSP) features are generated based on pair-wise projection matrices, which covers various frequency ranges. We propose a frequency complementary feature map selection (FCMS) scheme by constraining the dependency among frequency bands. Experiments are conducted on BCI competition IV dataset IIa with 9 subjects. Averaged cross-validation accuracy of 68.45% and 69.27% is achieved for FCMS and all feature maps, respectively, which is significantly higher (4.53% and 5.34%) than random map selection and higher (1.44% and 2.26%) than filter-bank CSP (FBCSP). The results demonstrate that the CNNs are capable of learning discriminant, deep structure features for EEG classification without relying on the handcrafted features.**

***Index Terms*— multi-class motor imagery of EEG, deep learning, convolutional neural network, augmented CSP.**

I. INTRODUCTION

Learning layered, deep structures of high-dimensional data has generated great interest in the past decades. A deep neural network (DNN) with multiple learning layers is a powerful modeling tool in various applications [1], [2], [3], [4]. Recently, using DNNs to discover the unknown correlations in the EEG signals have been investigated [5], [6], [7]. These works include: stacked autoencoder-based emotion recognition using hierarchical power spectra density features [6]; deep belief net (DBN)-based classification of target and non-target images using uncorrelated features constructed by clustering, in an image rapid serial visual presentation task [5]; and unsupervised paradigm of DBNs-based classification and anomaly detection for EEG waveforms modeling [7]. Typically, the DBNs are generated by training each layer as a restricted boltzmann machine (RBM) in an unsupervised way (pre-training), these trained RBMs are then stacked together. The inferred states of hidden nodes from current layer are used as inputs to the next layer. To improve the generalization of the models, drop-out of certain amount of hidden nodes or sparse coding have been proposed [4]. Convolutional neural networks (CNNs) have found wide applications due to its regularization structure, good spatial locality and certain degree of translation invariance [1], [8]. The CNNs are widely used for two-dimensional data processing such as handwritten digit recognition [1], feature extraction based on stacked convolutional autoencoder [9].

The authors are with Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore 138632. Email: {hjyang, stusis, kkang, ctguan}@i2r.a-star.edu.sg

Common spatial pattern (CSP) is introduced for EEG analysis and classification, which intends to find the directions in the pattern space to optimally distinguish the two classes [10], [11], [12]. Let's denote Σ_1 and Σ_2 as the covariance matrices of the band-pass filtered EEG signal S_e (with the number of channels and trials being denoted as n_c and n_t, respectively) for the respective motor imagery action in a binary classification task, we have

$$W^T\Sigma_1 W = \varpi_1 \;\; and \;\; W^T\Sigma_2 W = \varpi_2 \tag{1}$$

where ϖ denote the diagonal elements of the matrix, W^T denote the transpose of the projection matrix W. Scaling W such that $\varpi_1+\varpi_2=I$, which can be achieved by solving the generalized eigenvalue problem by

$$\Sigma_1 w = \lambda\Sigma_2 w \tag{2}$$

The large λ_j corresponded to the spatial filter w_j that yields high variance in one motor imagery action and low variance in another action. The ith CSP features (F_c^i) are obtained by

$$F_c^i(w_{jj}, S_e) = log(\frac{var(Z_p^i)}{\sum_{i=1}^{2m}(var(Z_p^i))}) \tag{3}$$

where the first and last m (e.g., m=1) rows of the projected EEG signals (denoted as $\widetilde{Z}$) are employed. Here $\widetilde{Z}$ is obtained by projecting S_e to the characteristic patterns obtained from Eq. (2), which is given by $\widetilde{Z}=w_{jj}^T S_e^{n_c\times n_t}$. In this paper, we investigate the classification of multi-class motor imagery of EEG signals (MI-EEG) based on the CNNs. For this purpose, the augmented CSP features (ACSP) are employed. We hypothesize that the CNNs are capable of learning the deep structures from the augmented CSP features without employing the hand-crafted feature selection algorithms.

II. OUR PROPOSED METHOD

A schematic diagram illustrating our proposed method, namely, classification of multi-class motor imagery of EEG signal based on augmented CSP features and convolutional neural networks (CMI-ACSP-CNNs), is presented in Fig. 1. In what follows, the ACSP feature extraction and classification of MI-EEG based on CNNs will be elaborated in more details.

A. Extraction of Augmented CSP (ACSP) Features

The proposed augmented CSP (ACSP) features can be considered as extraction of the CSP features from a multi-level decomposition of the frequency ranges. The motivation of proposing the augmented CSP features is based on the observation that the fixed partition of the frequency in both

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

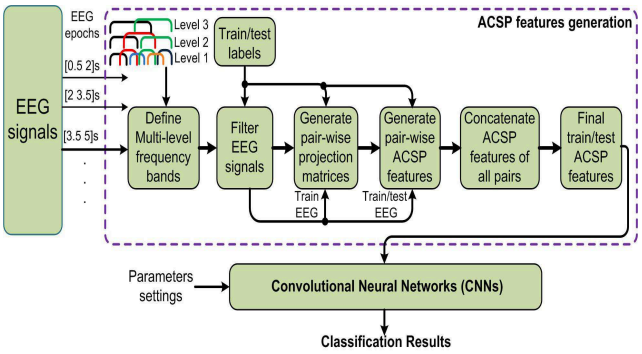

Fig. 1. Schematic illustration of the proposed *CMI-ACSP-CNNs* scheme.

the wide-band CSP [10] (e.g., computing CSP in a wide band of [4 30]Hz) and filter-bank CSP [13] (e.g, computing CSP in each small frequency band of width of 4 Hz, varying from 4Hz to 30Hz) may lose important frequency information. Hence, we propose to generate the ACSP features based on a varying partition of the frequency bands with different bandwidth to cover as many bands as possible. The following steps are carried out to generate the ACSP features.

1) Define the overall starting and ending frequency (f_s and f_e), width of each band (f_w), and width of each window shift (f_{ws}), with the constraint that different frequency bands with varying bandwidth should be included, e.g., f_s, f_e=[4 40]Hz, f_w=[4 7 7 13] and f_{ws}=[2 4 5 4] for a 4-level ACSP.

2) Filter the EEG signals for all the frequency bands with a band-pass filter such as butterworth filter to obtain the filtered EEG signal, denoted as S_e.

3) Generate the pair-wise projection matrix (denoted as $P_w(i,j)$) by solving Eq. (2) for a pair of classes of i and j ($i \neq j$, and i, j=1,2,...,n_s, n_s denotes the number of classes) based on S_e. Each projection matrix is generated to satisfy that the variance of the projected signal is maximized for the class under one condition and minimized for the class under another condition [10], [11], [12].

4) Generate the ACSP features based on the pair-wise projection matrices, which is detailed in Algorithm 1.

Data: $P_w(i,j), S_e(i,j), i,j = 1,2,...,n_s$
Result: F_t
F_t=Φ (empty set);
for i *in*$(1,2,...,n_s)$ **do**
  for j *in*$(1,2,...,n_s)$ **do**
    if $i \neq j$ **then**
      $F_c(i,j) = F_c(P_w(i,j), S_e(i,j))$;
      $F_t = [F_t; F_c(i,j)]$;
    end
  end
end
return F_t;

Algorithm 1: Generation of ACSP features

It is obvious that the features are of dimension of $n_f \times n_v$, where n_f and n_v denote the number of frequency bands and number of features obtained from n_m pair-wise projection matrices, respectively; n_v=$n_m \times n_{fs}$, and n_{fs} is the number of features for each frequency band.

III. Features Learning based on CNNs

A. Architecture of the CNNs

The backpropagation algorithm is generally employed to train the parameters of the CNNs. The hierarchical structure of the CNNs is the key to the successful learning and analysis. Typical architecture of CNNs are consisted of convolution (filtering) layers and subsampling layers which are alternated. The last several layers are usually consisted of fully-connected multi-layer networks. The features learned from the CNNs are concatenated and eventually fed to the network to obtain the classification results. The loss function employed is the mean square error (MSE) function. Let's name our proposed CNNs network for EEG signal classification as "CNNs-eegNet" with the architecture of the network being shown in Fig. 2. The CNNs-eegNet is

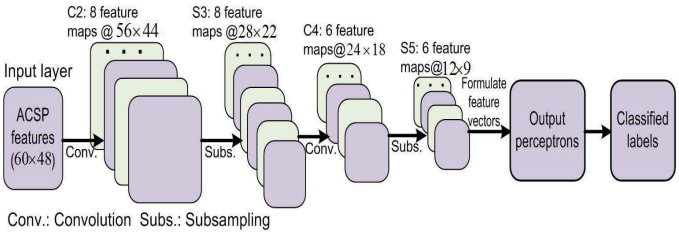

Fig. 2. Architecture of proposed CNNs-eegNet, each plane is a feature map, whose weights are to be identical.

consisted of 5 layers, where the first layer is the input layer of the ACSP features for the selected time segment. The input features are of dimension of $n_f \times n_v \times n_{tr}$, where n_f, n_v and n_{tr} denote the number of frequency bands, number of ACSP features and number of trials, respectively. The second and fourth layers are the convolution layers (denoted as "C2" and "C4") which are consisted of different numbers of output maps. Each feature map is connected to a $k \times k$ learnable kernel (e.g., 5×5). In a convolution layer, the feature maps from previous layer are convolved with the learnable kernels and fed to the activation function to generate the output feature map [8], which is given by

$$x_j^l = f\Big(\sum_{i \in F_p^l(j)} x_i^{l-1} * k_{ij}^l + b_j^l\Big) \tag{4}$$

where x_j^l and x_i^{l-1}, and $F_p^l(j)$ denote the output feature maps at levels l and l-1, and the set of selected feature maps for output map j at level l, respectively; k_{ij}^l and b_j^l are the kernel and bias at level l. The third and fifth layers are the subsampling layers (denoted as "S3" and "S5") by subsampling the feature maps at layers C2 and C4 by a factor of 2 in each dimension, i.e., each unit of a feature map at layers S3 and S5 is connected to a 2×2 neighborhood in the corresponding feature maps at layers C2 and C4 [8], which is given by

$$x_j^l = f(\beta_j^l \mathcal{D}_s(x_j^{l-1}) + b_j^l) \tag{5}$$

where $\mathcal{D}_s()$ is the subsampling function, which will sum over a 2×2 block to give one output; β_j^l and b_j^l are the multiplicative bias and additive bias, respectively.

B. Feature Map Selection

We observe that the classification performance is affected by how the feature map has been selected from the subsampling layer. Hence, we investigate the connection scheme between the convolutional and subsampling layers, with the aim to obtain a sparse and non-redundant representation of the features. A non-complete connection was suggested in [1], the objectives of which were to keep the number of connections within a reasonable bound, break the symmetries in the connection and finally generate the complementary features. In this paper, we investigate three ways to select the feature maps, namely, random map selection (RMS), our proposed frequency complementary map selection (FCMS), and the selection of all feature maps (SFM). Each output feature map is generated by convolving the kernels with a predefined number of feature maps which are randomly selected from the previous layer in the RMS scheme. It is noted that the randomness is not that good owing to the small numbers of maps. The FCMS scheme is illustrated in Fig. 2, the rationale of it is: the dependency and redundancy of the ACSP features among different frequency bands should be minimum in order to have good representation of the features. Let the ith feature map in the subsampling layer

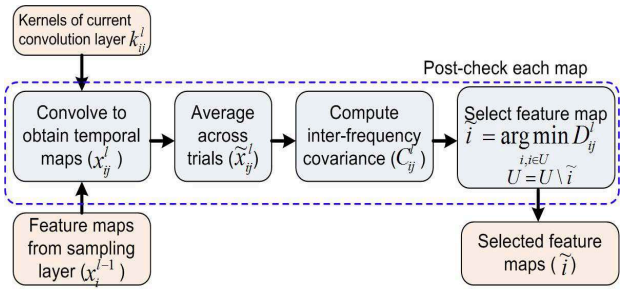

Fig. 3. Illustration of the frequency complementary map selection (FCMS) scheme.

(l-1) be denoted as x_i^{l-1}. The input feature map is firstly convolved with the kernel at the current convolution layer l for the output feature map j (denoted as k_{ij}^l) to obtain a temporal map x_{ij}^l. Averaging across trials is then carried out to obtain $\tilde{x}_{ij}^l$. Computing the inter-frequency covariance is given by

$$C_{ij}^l = \tilde{x}_{ij}^l * \tilde{x}_{ij}^{lT} \tag{6}$$

The inter-frequency dependency for the ith input feature map, jth output feature map was then measured by

$$\mathcal{D}_{ij}^l = \sum_{u=1}^{n_f} \sum_{v=1, v\neq u}^{n_f} C_{ij}^l(u,v) - diag(C_{ij}^l(u,v)) \tag{7}$$

where u (or: v) and n_f denote the u (or: v)th frequency band and the total number of frequency bands, respectively; $diag(X)$ gives the diagonals of the matrix X. Finally, the feature maps for each output feature map j are selected by

$$\tilde{i} = \arg\min_{i,i\in U} \mathcal{D}_{ij}^l \tag{8}$$

where U denote the indexes of feature maps in the subsampling layer. Updating U will be carried out once a feature map $\tilde{i}$ is selected, which is given by U=$U\backslash\tilde{i}$, where "\" denoted the "set difference". This process is iterated till the number of required feature maps has been selected. Note that the proposed scheme is an unsupervised method since no label information has been employed in the selection process. This is possible since the CSP features are consisted of features from both classes.

IV. Experimental Results

BCI IV competition IV, Data set IIa (available at: http://www.bbci.de/competition/iv/#dataset2a) is employed to evaluate the performance of proposed CNN-based classification scheme, i.e., CMI-ACSP-CNNs. The data set consisted of EEG data of 9 subjects who performed four motor imagery tasks of movement of left hand, right hand, both feet and tongue. Two sessions of data were collected with 288 trials for each session. The timing scheme consisted of a fixation of 2 s, cue time of 1.25 s, followed by motor imagery of 4 s. EEG signals were recorded with 22 Ag/AgCl electrodes and sampled at 250 Hz. To show the effectiveness of CNNs in learning the discriminant features, the classification accuracies of CNNs by employing different strategies for feature maps selection are compared with FBCSP algorithm, with the 5×5 cross-validation (CV) results shown in Table I. Noted that the time segment of [0.5 2.5] s is used. Eight levels of frequency bands are employed, where the starting and ending frequencies are chosen as f_s=4 Hz and f_e=40 Hz; the width of each band (f_w) is chosen as f_w=[3 4 7 8 11 12 13 15] Hz; and the width of each window shift (f_{ws}) are: f_{ws}=[2 2 4 5 6 6 5 5]. In this way, we have 60 frequency bands in total. Eight and six feature maps are selected for C2 and C4 layers, where each output feature map at layer C4 is computed based on 5 selected maps (e.g., for PCMS and RMS), or all the maps (for FMS) in layer S3, as shown in Fig. 3. The batch size and number of epochs are 38 and 30, and α=0.5 is used for updating the weights and bias in gradient-based learning. The MSE loss averaged across 5×5 folds is shown in Fig. 4.

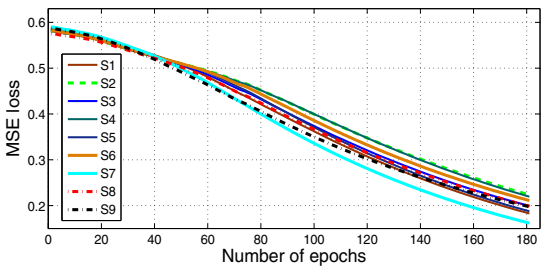

Fig. 4. Mean square error loss as a function of number of epochs for proposed PCMS scheme.

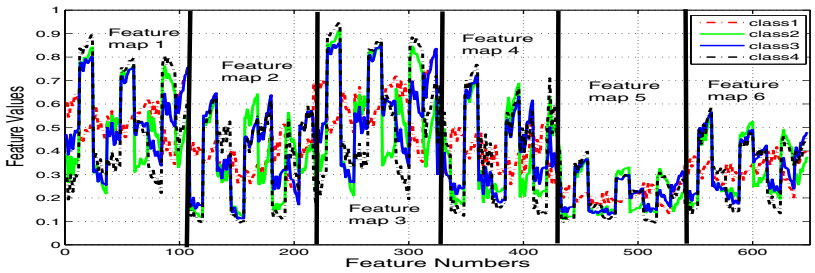

Fig. 5. Illustration of the 6 feature maps generated for subject "S9" during cross-validation.

TABLE I
THE 5×5 CV CLASSIFICATION ACCURACIES (%) FOR 4-CLASSES EEG

Sub.	Different strategies			Baseline algorithm
	$PCMS^a$	RMS^b	SFM^c	$FBCSP^d$
S1	75.86	72.69	77.14	79.16
S2	48.71	45.44	49.82	52.08
S2	79.01	74.11	80.41	83.33
S4	52.77	45.92	53.88	62.15
S5	64.00	57.65	65.47	54.51
S6	49.04	46.05	48.70	39.24
S7	80.88	79.08	81.37	83.33
S8	83.20	74.50	84.39	82.64
S9	82.57	79.97	82.29	66.67
A_{ac}	68.45	63.93	69.27	67.01
p-test	a vs. b	a vs. c	a vs. d	c vs. d
p-value	*(3.93e-4)	*(0.0077)	0.6190	0.4159

A_{ac}: average accuracy across subjects (%). *: significant at 5% significance level for paired t-test, i.e., p-test.

The results shown in Table I demonstrated that the CNNs are able to learn the discriminant features for multi-class EEG data classification. This is supported by the different feature maps generated during cross-validation, as shown in Fig. 5, where distinct features for the four classes were extracted for different feature maps. Hence, no feature selection scheme is needed after the CSP features are generated. It should also be noted that no significant improvements can be observed by our proposed PCMS scheme compared with full-map selection scheme and FBCSP algorithm due to the small number of feature maps. Nevertheless, the accuracy of our proposed scheme (i.e., selecting the feature maps by constraining the dependency among the frequency bands) is still significantly better than that obtained by randomly selecting the feature maps. Further, the accuracy of our proposed CNN-based classification method is still higher than that of FBCSP algorithm, where a feature selection scheme based on mutual information was implemented.

V. CONCLUSIONS

This paper investigates classification of multi-class motor imagery of EEG signals based on convolutional neural networks and augmented CSP features. Investigations on how to select the feature maps to form good representative features are made. Experiments are conducted on data set IIa of BCI competition IV. Averaged cross-validation accuracy of 68.45% and 69.27% is achieved for the proposed frequency complementary feature map selection scheme and using full maps, respectively. This is significantly higher (4.53% and 5.34%) than that obtained by randomly selecting the feature maps, and higher (1.44% and 2.26%) than the baseline FBCSP algorithm. These results further demonstrated the capability of CNNs in learning discriminant features for EEG classification.

REFERENCES

[1] Yann Lecun, Lon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, pp. 2278–2324, 1998.

[2] Honglak Lee, Chaitanya Ekanadham, and Andrew Y. Ng, "Sparse deep belief net model for visual area v2," in *Advances in Neural Information Processing Systems 20*. 2008, MIT Press.

[3] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[4] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *ArXiv e-prints*, vol. 1207.0580, 2012.

[5] S. Ahmed, L. M. Merino, Z. Mao, J. Meng, K. Robbins, and Y. Huang, "A deep learning method for classification of images rsvp events with eeg data," in *2013 IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*. IEEE, 2013, pp. 33–36.

[6] S. Jirayucharoensak, S. Pan-Ngum, and P. Israsena, "Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation," *The Scientific World Journal*, vol. 2014, pp. Article ID 627892, 10 pages, 2014.

[7] D. F. Wulsin, J. R. Gupta, R. Mani, J. A. Blanco, and B. Litt, "Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement," *J Neural Eng.*, vol. 8, no. 3, pp. 036015, June 2014.

[8] Jake Bouvrie, "Notes on convolutional neural networks," 2006.

[9] M. Jonathan, M. Ueli, C. Dan, and S. Jürgen, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *Proc. of the 21th Int. Conf. on Artificial Neural Networks-Volume Part I*, Berlin, Heidelberg, 2011, ICANN'11, pp. 52–59, Springer-Verlag.

[10] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movements," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 4, pp. 441–6, 2000.

[11] H. Yang, C. Guan, K. K. Ang, H. Zhang, and C. C. Wang, "Iterative clustering and support vectors-based high-confidence query selection for motor imagery eeg signals classification," in *21st Int. Conf. on Pattern Recognition (ICPR)*. IEEE, 2012, pp. 2169–72.

[12] H. Yang, C. Guan, K. K. Ang, Y. Pan, and H. Zhang, "Cluster impurity and forward-backward error maximization-based active learning for eeg signals classification," in *2012 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 569–72.

[13] K K Ang, Z Y Chin, C Wang, C Guan, and H Zhang, "Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b," *Front Neurosci.*, vol. 6, no. 39, pp. 1–9, 2012.