Multi-class Motor Imagery Classification by Singular Value Decomposition and Deep Boltzmann Machine

Zhongliang Yu, Jinchun Song
School of Mechanical Engineering and Automation
Northeastern University
Shenyang, China
zhon.liangyu@gmail.com

***Abstract*—Motor function rehabilitation is very urgent for patients. Motor imagery is an efficient way for rehabilitation. To achieve the supervision of multiple rehabilitation targets simultaneously, the promotion of multi-class motor imagery classification accuracy is critical. In this paper, a multi-class classification method is proposed by utilizing singular value decomposition and deep boltzmann machine. Singular value decomposition is applied to suppress the artifacts and acquire the channel-individual characteristics. The deep boltzmann machine is employed to extract and model the characteristics and achieve the motor imagery classification. Results demonstrate that the proposed method has achieved a 14.2% higher classification accuracy than the common spatial pattern on average. This results are further validated by the statistical methods, which present a significant difference ($p < 0.05$). The proposed method is favorable for promoting the multi-class motor imagery classification efficiency.**

Keywords—Brain computer interface; Motor imagery; Deep boltzmann machine; Multi-class; Classification

I. INTRODUCTION

Brain computer interface (BCI) is a kind of communication between human brain and outside interface, independent of the common path of nerves and muscles, aiming to augment, assist and repair the human cognitive or sensory-motor functions [1]. It plays an important role in function and neuro reestablishment [2] and rehabilitation systems [3]. Recently, many BCI signal modulation methods have been studied, such as electroencephalogram (EEG) [4], electrocorticogram (ECOG) [5], positron emission tomography (PET) [6] and magnetoencephalogram (MEG) [7]. Due to its non-invasion and low cost, EEG becomes popular for BCI studies. Moreover the excellent time resolution of EEG conforms to the real time requirements of rehabilitation.

Motor imagery (MI) is a widely used BCI signal modulation, implemented by brain activity without any physical movement. MI and motor execution share common cognitive process and motor-related nerves [8], which is beneficial for motor function and neuro rehabilitations based on BCI. EEG-based MI-BCI has been extensively studied and applied in motor function and neuro rehabilitations, such as in Parkinson's disease rehabilitation [9] and in stroke induced gait rehabilitation [10]. Due to the problems of long time training, individual difference and low signal to noise ratio induced by the nonlinear and weak EEG signals, the improvement of single-trial MI classification accuracy of multi-class is urgent for real-time feedback in rehabilitation.

For decades, many researches have been devoted to improving the single-trial MI classification accuracy. The widely applied common spatial patterns (CSP) is one of the most efficient methods to diagonalize the MI covariance matrices from two classes simultaneously [11]. However, the characteristics of CSP depend on the frequency bands that selected artificially by experience. A lot of studies have investigated to improve the way of characteristic selection, such as common sparse spectral spatial pattern (CSSSP) [12] and regularized CSP [13]. Among these algorithms, filter bank common spatial patterns (FBCSP) is one of the most excellent methods. The optimal frequency band and time range of FBCSP are selected via mutual information [14]. Nevertheless, FBCSP requires to manually select the size of subbands to achieve good performance without any elasticity [15]. The methods discussed above attempt to improve the classification accuracy of two classes. When referring to multi-class classification, multi binary classification is always needed, and the binary-class method may decrease the classification capacity.

In this paper, to achieve better multi-class classification performances, a data-driven and self-adapting method (SVD^2BM) without any assumptions is proposed by utilizing the singular value decomposition (SVD) and the deep boltzmann machine (DBM). SVD is used to suppress the artifacts and acquire the channel-individual characteristics. DBM is applied to extract and model the potential characteristics and achieve the multi-class classification of MI. The proposed SVD^2BM presents better performances than conventional ones. It is promising for practical MI applications.

II. MATERIALS AND METHODS

A. Experiments

The publically available dataset of BCI competition IV, dataset IIa, is used to evaluate the MI classification performances of the SVD^2BM. The data from dataset IIa are recorded by twenty-two electrodes at a sampling of 250 Hz. The subjects are instructed to imagine left hand, right hand, both feet and tongue, by the random cue. The data collected are online bandpass filtered and notch filtered to remove artifacts. The dataset includes nine subjects. The imagination lasts from t = 3 s to t = 6 s. Before this period, the concentration lasts from

978-1-5090-5363-6/17/$31.00 ©2017 IEEE

t = 0 s to t = 2 s, and the cue lasts from t = 2 s to t = 3.25 s. The break time before next trial lasts from t = 6 s to t = 7.5 s.

B. Methods

1) Common spatial patterns

Initially, CSP was proposed to solve the projection problem of two class classification with different spatial distribution. To maximize the differences from two classes, the variance of one class is maximized by CSP, while the variance of the other one is minimized. $\boldsymbol{X}_k = [\boldsymbol{x}_1(t), \boldsymbol{x}_2(t),..., x_n(t)]^T$, $t = t_0, ..., T$, is the kth EEG segment, n the number of electrodes. The normalization average covariance matrices from the two classes can be derived as:

$$\boldsymbol{C}_{1/2} = \frac{1}{N}\sum \frac{(\boldsymbol{X}_{1/2} \times \boldsymbol{X}_{1/2}^{T})}{\text{trace}(\boldsymbol{X}_{1/2} \times \boldsymbol{X}_{1/2}^{T})} \tag{1}$$

Whiten the matrix $\boldsymbol{C}$ which is the sum of average covariance matrices of the two classes, then a matrix $\boldsymbol{P}$ can be obtained below:

$$\boldsymbol{P}\boldsymbol{C}\boldsymbol{P}^{T} = \mathrm{I} \tag{2}$$

Let $\boldsymbol{S}_1 = \boldsymbol{P}\boldsymbol{C}_1\boldsymbol{P}^T$ and $\boldsymbol{S}_2 = \boldsymbol{P}\boldsymbol{C}_2\boldsymbol{P}^T$, then the orthogonal matrix $\boldsymbol{R}$ and the diagonal matrix $\boldsymbol{D}$ can be derived as:

$$\boldsymbol{S}_i = \boldsymbol{R}\boldsymbol{D}_i\boldsymbol{R}^T \tag{3}$$

where, the class number index i = 1, 2. Simultaneously, the average covariance matrices is diagonalized. Namely, the separability of the two classes is maximized. The spatial filter $\boldsymbol{W}$ can be derived as:

$$\boldsymbol{W} = \boldsymbol{R}\boldsymbol{T}\boldsymbol{P} \tag{4}$$

Thus, the $\boldsymbol{X}_k$ can be projected as:

$$\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{X}_k \tag{5}$$

2) Singular value decomposition

SVD is a kind of dimensionality reduction methods, and has been widely used for EEG analysis [16], especially for high dimension characteristic extractions. Let the matrix $\boldsymbol{X}$ with the size of $m \times n$ is the EEG segment, then the matrix $\boldsymbol{X}$ can be decomposed as:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V} \tag{6}$$

where $\boldsymbol{U}$, $\boldsymbol{V}$ are the unitary matrices with the size of $m \times m$ and $n \times n$, respectively. The positive semidefinite matrix $\boldsymbol{\Sigma}$ is the singular matrix with the size of $m \times n$, and elements of diagonal on $\boldsymbol{\Sigma}$ is the singular value of matrix $\boldsymbol{X}$. The matrix $\boldsymbol{\Sigma}$ that may not be a non-singular matrix is sorted by the singular value as:

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_i \cdots \geq \lambda_m \tag{7}$$

where λ_i indicates the singular value.

3) Restricted boltzmann machine

Restricted boltzmann machine (RBM) is a kind of random neuron networks, proposed by Hinton and Sejnowski in 1986. It has been successfully used in modeling the distribution of binary valued data [17]. RBM includes a hidden binary valued layer and a visible binary valued layer. There is no existing connection between the units on the same layer. The hidden and visible layers are fully connected. The sketch of the RBM is shown in Fig. 1. RBM obeys energy-based model. The joint configuration model of visible and hidden layers is defined by the following energy function E:

$$E(\boldsymbol{v},\boldsymbol{h};\theta) = -\sum_{ij} W_{ij} v_i h_j - \sum_i b_i v_i - \sum_j a_j h_j \tag{8}$$

where b_i, a_j are the bias of the visible and the hidden layers. θ the parameter of the model; W_{ij} the weight. Then, the joint probability density of $\boldsymbol{v}$ and $\boldsymbol{h}$ can be described as:

$$p(\boldsymbol{v},\boldsymbol{h};\theta) = \frac{1}{Z(\theta)} exp(-E(\boldsymbol{v},\boldsymbol{h};\theta)) \tag{9}$$

where $Z(\theta)$ is the partition function. To simultaneously equate Eq. (8) and Eq. (9), the joint probability density can be derived as:

$$p(\boldsymbol{v},\boldsymbol{h};\theta) = \frac{1}{Z(\theta)} exp(\sum_{ij} W_{ij} v_i h_j + \sum_i b_i v_i + \sum_j a_j h_j) \tag{10}$$

To achieve the maximization of the observation, the logarithm of the marginal probability density function $\log(p(v))$ should be maximized:

$$log(p(v)) = \frac{1}{N}\sum_n log\ p(v^n) \tag{11}$$

In the past, due to the restriction of computer science and hardware equipment, the units and layers number of the neuron network was limited. This leaded to the impossibility

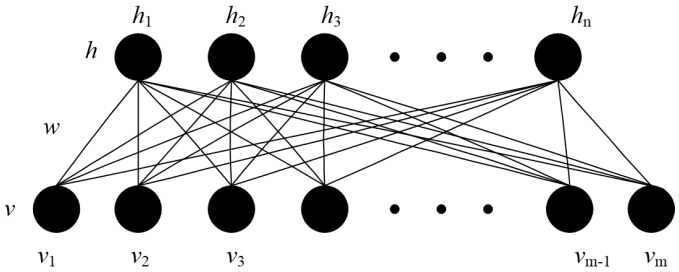

Fig. 1. The sketch of the RBM model.

of the extraction of deep characteristics, especially in non-linear signal analysis, such as EEG. With the rapid development, it is possible to add the number of layers and units to extract the deep characteristics now. In this paper, the multi-class MI is to be analyzed and classified with multi-layer RBM model.

4) Data processing

Before analysis, each trial of the dataset is stripped from the time sequence. All the trials are bandpass filtered from 5 Hz to 30 Hz. Subsequently, the MI process on each channel is segmented and reconstructed. The window length of 0.5 s is chosen without overlap. SVD is implemented to obtain singular value on every channel. The last non-zero singular value is set to zero to remove artifacts, then the artifacts removed matrix is reconstructed to acquire the channel-individual characteristics. DBM is applied to extract and model the potential characteristics and achieve the multi-class classification of MI. The DBM applied includes three hidden layers, and the units of them are 500, 500 and 2000 respectively. The number of units chosen is half of the one in [18], for the reduction of the computing time. The epochs of backpropagation are set as 100, and the deep learning toolbox applied is from [18]. To evaluate the performances of the SVD^{2}BM, the conventional CSP and FBCSP are involved. The computing and statistical analysis are executed by MATLAB.

III. Results

The multi-class classification accuracy of the SVD^{2}BM, the CSP and the FBCSP after the K-fold cross validation (K = 10) is illustrated in Table I successively. The training and testing datasets, which are half of the trials, are selected in random without overlap. The results demonstrate that the SVD^{2}BM presents a higher classification accuracy when compared with each of the CSP subjects and six of the FBCSP subjects. The classification accuracy of the SVD^{2}BM is respectively 14.2% and 13.4% higher than the CSP and the FBCSP on average. The accuracy of the three algorithms is further analyzed with the one-way analysis of variance (ANOVA). The statistic results indicate that there is a significant difference (F (1, 27) = 3.49, p < 0.05) among these methods. The multi pairwise comparison results performed by the least-significant difference (LSD) show a significant difference (p < 0.05) between the SVD^{2}BM and the CSP. Besides, the SVD^{2}BM exhibits the lowest standard deviation among the three methods.

TABLE I. 4 Classes Classification Accuracy (%) of SVD^{2}BM, CSP and FBCSP Based on Dataset IIa

Dataset	**Subjects**	**SVD^{2}BM**	**CSP** [19]	**FBCSP** [20]
IIa	S1	90.8	78.3	79.2
	S2	79.5	44.7	52.1
	S3	81.4	82.2	83.3
	S4	80.5	59.1	62.2
	S5	84.6	39.7	54.5
	S6	72.2	50.1	39.2
	S7	81.5	81	83.3
	S8	69.1	65.5	82.6
	S9	83.8	77.4	66.7
MEAN		80.4	66.2	67.0
SD		6.5	16.5	16.2

IV. Discussion

In this study, the SVD is applied for the suppression of the artifacts and the acquisition of the channel-individual characteristics. The DBM is used to extract and model the characteristics and classify the multi-class MI. The results reveal that the SVD^{2}BM presents a better performance on multi-class MI classification and a lower standard deviation that is favorable for decreasing the individual difference of the subjects. Namely, the training datasets of the subjects without any training can be substituted by the ones of the well trained subjects, to some extent, for assistance. This is beneficial for the actual application of MI.

Here, we mainly focus on the influences of the number of backpropagation epochs and units on the comprehensive classification performances. Firstly, the units of the hidden layers are fixed as 500, 500 and 2000, while the number of backpropagation epochs varies from 1 to 150. The variation of the classification accuracy with the number of epochs is shown in Fig. 2. In the beginning, with the increasing of the backpropagation epochs, the classification accuracy is synchronously enhanced for all subjects. When the backpropagation epochs reach upwards of 80, most of the subjects have reached the stable classification accuracy. The classification accuracy does not increase with the enhancement of backpropagation epochs further. That is, appropriate backpropagation epochs are critical for high classification accuracy with optimal time complexity. Secondly, the backpropagation epochs are fixed as 100, while the units of the hidden layers vary. For the convenience of discussion, four groups of the units of the three hidden layers are distributed as: (1) 1000, 1000, 4000; (2) 500, 500, 2000; (3) 250, 250, 1000;

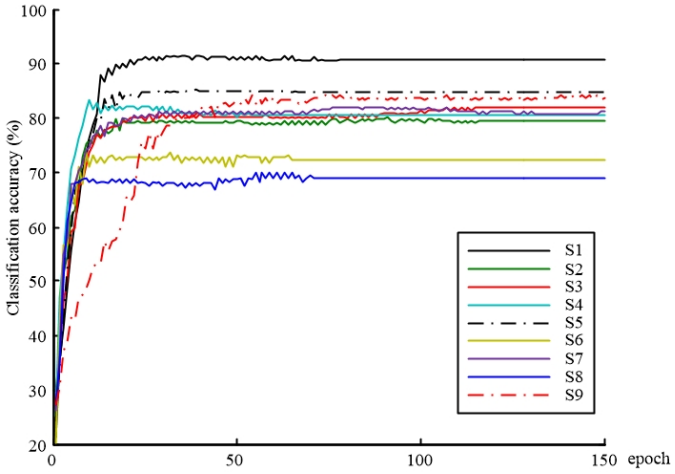

Fig. 2. The variation of the classification accuracy with the backpropagation epochs.

TABLE II. CLASSIFICATION ACCURACY (%) OF THE FOUR GROUPS OF UNITS OF THREE HIDDEN LAYERS

Dataset	Subjects	Group 1	Group 2	Group 3	Group 4
IIa	S1	91.9	90.8	90.1	90.8
	S2	81.7	79.5	79.9	79.9
	S3	79.9	81.4	80.7	78.4
	S4	84.4	80.5	80.2	80.9
	S5	87.3	84.6	88.1	83.1
	S6	71.2	72.2	70.8	69.8
	S7	80.4	81.5	80.4	80.8
	S8	68.0	69.1	66.0	66.8
	S9	80.3	83.8	84.2	83.8
MEAN		80.6	80.4	80.0	79.4
SD		7.4	6.5	7.6	7.2

(4) 100, 100, 400. The classification accuracy of the four groups is shown in Table II. The differences of the four groups on the average classification accuracy are little. When it comes to the standard deviation, group 2 exhibits the best performance that is favorable for the reduction of the subject-individual difference. Besides, the computing time of group 1 is nearly twice of group 2. Thus, among the four groups, group 2 is more suitable for the multi-class MI classification due to the comprehensive performances on classification accuracy, standard deviation and time complexity. That is, the improvement of the overall classification performances does not always depend on the enhancement of units.

V. CONCLUSION

This work attempts to explore the methods for improving the multi-class MI classification performances. The SVD^{2}BM method has been proposed by utilizing SVD and DBM. SVD has been used to suppress the artifacts and acquire the channel-individual characteristics. DBM has been applied to extract and model the characteristics and classify the multi-class MI. The quantitative analysis results have demonstrated that the proposed SVD^{2}BM presents a better average classification accuracy with better stability when comparing with CSP and FBCSP. Due to the employment of DBM, the time complexity of SVD^{2}BM is higher than CSP. Besides, the classification performances of SVD^{2}BM are affected by the selection of the parameters of DBM model. To solve the problem, the automatic parameter optimization algorithm concerning DBM model will be studied in our further work.

ACKNOWLEDGMENTS

This work is supported by the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 51605085) and the Postdoctoral Science Foundation of China (Grant No. 2016M590229).

REFERENCES

[1] M. O. Krucoff, S. Rahimpour, M. W. Slutzky, V. R. Edgerton, and D. A. Turner, "Enhancing Nervous System Recovery through Neurobiologics, Neural Interface Training, and Neurorehabilitation," Frontiers in Neuroscience, vol. 10, 2016.

[2] K. K. Ang, C. Guan, K. S. Phua, C. Wang, L. Zhou, K. Y. Tang, *et al.*, "Brain-computer interface-based robotic end effector system for wrist and hand rehabilitation: results of a three-armed randomized controlled trial for chronic stroke," Front. Neuroeng, vol. 7, 2014.

[3] K. K. Ang, K. S. G. Chua, K. S. Phua, C. Wang, Z. Y. Chin, C. W. K. Kuah, *et al.*, "A randomized controlled trial of EEG-based motor imagery brain-computer interface robotic rehabilitation for stroke," Clinical EEG and neuroscience, vol. 46, pp. 310-320, 2015.

[4] R. Xu, N. Jiang, C. Lin, N. Mrachacz-Kersting, K. Dremstrup, and D. Farina, "Enhanced low-latency detection of motor intention from EEG for closed-loop brain-computer interface applications," IEEE Transactions on Biomedical Engineering, vol. 61, pp. 288-296, 2014.

[5] A. Rouse, J. Williams, J. Wheeler, and D. Moran, "Spatial co-adaptation of cortical control columns in a micro-ECoG brain–computer interface," Journal of Neural Engineering, vol. 13, p. 056018, 2016.

[6] Y. Zhu, K. Xu, C. Xu, J. Zhang, J. Ji, X. Zheng, *et al.*, "PET Mapping for Brain-Computer-Interface-Based Stimulation in a Rat Model with Intracranial Electrode Implantation in the Ventro-posterior Medial Thalamus," Journal of Nuclear Medicine, p. jnumed. 115.171868, 2016.

[7] N. Heinze, T. Pfeiffer, A. Schoenfeld, and G. Rose, "Towards an estimation of ECoG decoding results based on fully non-invasive MEG acquisition," Clinical Neurophysiology, vol. 126, pp. e156-e157, 2015.

[8] N. Sharma and J.-C. Baron, "Does motor imagery share neural networks with executed movement: a multivariate fMRI analysis," Frontiers in human neuroscience, vol. 7, p. 564, 2013.

[9] D. Caligiore, M. Mustile, G. Spalletta, and G. Baldassarre, "Action observation and motor imagery for rehabilitation in Parkinson's disease: A systematic review and an integrative hypothesis," Neuroscience & Biobehavioral Reviews, 2016.

[10] R. Dickstein, S. Levy, S. Shefi, S. Holtzman, S. Peleg, and J.-J. Vatine, "Motor imagery group practice for gait rehabilitation in individuals with post-stroke hemiparesis: A pilot study," NeuroRehabilitation, vol. 34, pp. 267-276, 2014.

[11] C. Park, C. C. Took, and D. P. Mandic, "Augmented complex common spatial patterns for classification of noncircular EEG from motor imagery tasks," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 22, pp. 1-10, 2014.

[12] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K.-R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," IEEE transactions on biomedical engineering, vol. 53, pp. 2274-2281, 2006.

[13] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," IEEE Transactions on biomedical Engineering, vol. 58, pp. 355-362, 2011.

[14] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs," Pattern Recognition, vol. 45, pp. 2137-2144, 2012.

[15] J. S. Kirar and R. Agrawal, "Optimal Spatio-spectral Variable Size Subbands Filter for Motor Imagery Brain Computer Interface," Procedia Computer Science, vol. 84, pp. 14-21, 2016.

[16] X. Zhang, W. Diao, and Z. Cheng, "Wavelet transform and singular value decomposition of eeg signal for pattern recognition of complicated hand activities," Digital Human Modeling, pp. 294-303, 2007.

[17] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," Advances in neural information processing systems, Barcelona, Spain, 2012, pp. 2222-2230.

[18] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.

[19] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Multiclass brain–computer interface classification by Riemannian geometry," IEEE Transactions on Biomedical Engineering, vol. 59, pp. 920-928, 2012.

[20] H. Yang, S. Sakhavi, K. K. Ang, and C. Guan, "On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification," Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, 2015, pp. 2620-2623.