Classification of Stroke Patients' Motor Imagery EEG with Autoencoders in BCI-FES Rehabilitation Training System

Mushangshu Chen, Ye Liu, and Liqing Zhang*

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering, Department of Computer Science and Engineering, Shanghai Jiao Tong University, China
zhang-lq@cs.sjtu.edu.cn

Abstract. Motor imagery based Brain Computer Interface (BCI) system is a promising strategy for the rehabilitation of stroke patients. Common Spatial Pattern (CSP) is frequently used in feature extraction of motor imagery EEG signals and its performance depends heavily on the choice of frequency component. Moreover, EEG of stroke patients, which is full of noise, makes it hard for traditional CSP to extract discriminative patterns for classification. In order to deal with the subject-specific band selection, in this paper, we adopt denoising autoencoders and contractive autoencoders to extract and compose robust features from CSP features filtered in multiple frequency bands. We compare our method with traditional methods on data collected from two months clinical rehabilitation. The results not only demonstrate its superior recognition performance but also evidence the effectiveness of our BCI-FES rehabilitation training system.

1 Introduction

Brain Computer Interface (BCI), as an alternative communication channel between human brain and external devices, is a good way that combines Electroencephalography (EEG) signals with motor control [1]. Recently some studies have demonstrated that motor imagery based BCI is a very promising method in rehabilitation training of strokes [2]. One of the most effective algorithms for motor imagery based BCI is Common Spatial Pattern (CSP) [3]. The spatial filters generated by CSP reflect the specific activation of cortical areas. However, the performance of CSP heavily depends on the proper selection of frequency bands and channels [4][5].

It is generally considered that motor imagery of normal people attenuates EEG μ and β rhythm over sensorimotor cortices [6][7]. However, for special populations suffering from neurophysiological diseases (e.g., stroke), some studies recently found that the μ and β rhythm in motor imagery EEG of stroke patients have been modulated [8]. In our analysis, a similar regularity is observed

* Corresponding author.

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 202–209, 2014.
© Springer International Publishing Switzerland 2014

that the most informative frequency bands for classifying left and right motor imagery have deviated from normal α and β rhythms in different stages of the clinical experiments [9]. Furthermore, our analysis also illustrates that irregular discriminative patterns from impaired cortex is different from those of the normal subjects, and there frequently exists messy imagination contents in the motor imagery EEG of stroke patients [10]. In consequence, conventional methods suitable for normal subjects usually achieves a relatively low level classification accuracy [11].

To make full use of latent spectral information in EEG of stroke patients and extract robust features from noisy data, in this paper, we adopt denoising autoencoders [12] and contractive autoencoders [13] to obtain a encoding of the raw features of stroke patients' signals.

The rest part of this paper is organized as follows: a detailed introduction about our rehabilitation training system, experiment setup and data collection are given in Section 2. Section 3 describes the deep learning method for extracting EEG features. Section 4 demonstrates a comparative result when applying our method, frequency boosting and CSP-SVM on data collected from clinical experiments. Apart from the classification performance, we describe a phenomenon of contribution of band changing during rehabilitation, which may reveal some mechanisms of stroke patients' recovery in frequency aspect. Finally we give a brief conclusion in Section 5.

2 Experiment Paradigm

2.1 BCI-FES Rehabilitation Training System

Our BCI-FES rehabilitation system is consists of 5 modules: real-time data acquisition module, data storage and analysis module, visualization module, multimodal feedback module and human effect training module [14].

In general, the system aims at restoring motor functions of paralyzed limbs for post-stroke patients by active motor imagery directed by training tasks. EEG signal is collected by data acquisition module during subject's imagery and label of each segment of subject's motor imagery is recognized online after feature extraction and classification in data storage and analysis module. Multi-modal feedback module gives a corresponding feedback including visual, auditory and tactile response given the classification result and visualization module gives a real-time observation concurrently.

In order to improve training effect, we adopt a new rehabilitation training paradigm which attempts to reconstruct the motor sensory feedback loop [15]. During experiments the subject is required to reconfirm the label and can correct the label when necessary.

2.2 Experiment Setup and Data Collection

We conduct our clinical rehabilitation on seven participated subjects in hospital training with our BCI-FES rehabilitation system. Another three patients only receiving regular clinical treatments are considered as control group to assess the effectiveness of our system and rehabilitation paradigm.

In general, each subject is required to participate in 3 days' training per week. Each day's training consists of 8 sessions which contains 15 trials of motor imagery tasks. Each trial lasts for 4 seconds and is cut into 25 1s sliding windows with step length 0.125s for online classification. At the end of experiment cycle, a post-training section consists of 2 sessions will be conducted to evaluate rehabilitation efficacy. Raw EEG data is recorded by a 16-channel(FC3, FCZ, FC4, C1-C6, CZ, CP3, CPZ, CP4, P3, PZ and P4) g.USBamp amplifier under a sample rate of 256 Hz. After removing artifacts, we filter the EEG into α(8-12 Hz), β(12-30 Hz), γ(30-45 Hz) band for feature extraction.

3 Method

3.1 Common Spatial Pattern

The goal of CSP [3] is to design spatial filters that lead to optimal variances for the discrimination of two populations of EEG related to left and right motor imagery.

We denote raw EEG data as a $ch \times time$ matrix E, where ch is the number of channels and $time$ is the number of samples per channel. The filtered signal matrix S is $S = WE$ or $S(t) = We(t)$, where $W \in \mathbb{R}^{d \times ch}$ is spatial filter matrix.

First the sum spatial covariance can be eigen decomposed as $\Sigma_1 + \Sigma_2 = \frac{1}{n_L}\sum_{i=1}^{n_L} \frac{E_{L_i}E'_{L_i}}{trace(E_{L_i}E'_{L_i})} + \frac{1}{n_R}\sum_{i=1}^{n_R} \frac{E_{R_i}E'_{R_i}}{trace(E_{R_i}E'_{R_i})} = UDU^T$.

Then the whitened covariance by $P = \sqrt{D^{-1}}U^T$ can be decomposed as $\hat{\Sigma}_1 + \hat{\Sigma}_2 = P(\Sigma_1 + \Sigma_2)P^T = V(\Lambda_L + \Lambda_R)V^T$.

Finally, by selecting first and last m eigenvectors in V, the CSP filter is obtained as $W = P^T V \in \mathbb{R}^{2m \times ch}$. The filtered signal matrix is given by $s(t) = We(t) = (s_1(t) \dots s_d(t))^T, d = 2m$. And feature vector $x = (x_1, x_2, \dots, x_d)^T$ is calculated by $x_i = \log(\frac{var[s_i(t)]}{\sum_{j=1}^{d} var[s_j(t)]})$.

3.2 Autoencoder

A simplest autoencoder (AE) [16] is composed of two parts, an encoder and a decoder. The encoder is a function f that maps an input $x \in \mathbb{R}^d$ to a hidden representation $h \in \mathbb{R}^{d'}$ through a deterministic mapping $h = f_{\theta_1}(x) = s(Wx+b)$, parameterized by $\theta_1 = \{W, b\}$. The resulting latent representation h is mapped back to a reconstruction y, where $y = g_{\theta_2}(h) = s(W'h + b')$ with $\theta_2 = \{W', b'\}$.

Autoencoder training consists in optimizing parameters $\theta = \{W, b, W', b'\}$ to minimize the average reconstruction error on a training set D_n: $\mathcal{J}_{AE}(\theta) = \sum_{x \in D_n} L(x, g(f(x))) + \beta \sum_{j=1}^{d'} \mathrm{KL}(\rho || \hat{\rho}_j)$, where L is the reconstruction error, ρ is sparsity parameter, $\hat{\rho}_j$ is the average activation of hidden unit j and β controls the weight of the sparsity penalty term $\mathrm{KL}(\rho || \hat{\rho}_j)$. In case of s being the sigmoid, L is cross-entropy loss:$L(x, y) = -\sum_{i=1}^{d} x_i \log(y_i) + (1 - x_i)\log(1 - y_i)$.

Denoising Autoencoder(DAE). To enforce robustness to noisy inputs, denoising autoencoder [12] first corrupts input x, then train the autoencoder to reconstruct the clean version. The objective function is

$$\mathcal{J}_{DAE}(\theta) = \sum_{x \in D_n} \mathrm{E}_{\tilde{x} \sim q(\tilde{x}|x)}[L(x, g(f(\tilde{x})))] \tag{1}$$

where the expectation is over corrupted $\tilde{x}$ obtained from a corruption process $q(\tilde{x}|x)$.

Contractive Autoencoder(CAE). CAE [13] is obtained with the regularization term $||J_f(x)||_F^2 = \sum_{ij}(\frac{\partial f_j(x)}{\partial x_i})$, giving objective function

$$\mathcal{J}_{CAE}(\theta) = \sum_{x \in D_n} (L(x, g(f(x))) + \lambda ||J_f(x)||_F^2) \tag{2}$$

The basic autoencoders can be stacked into deep networks. Greedy layer-wise training is a good way to initialize a stacked autoencoder. The features from the stacked autoencoder can be used for classification by feeding the last layer's output to a softmax classifier.

In our analysis, EEG signals are filtered into α, β, γ band and form a $channel \times time \times window \times band$ format data. Then CSP is applied to each band's data $channel \times time \times window$ to extract frequency specific features. These features are normalized to $[0, 1]$ using $NormalizedFeature = \frac{Feature - \min(Feature)}{\max(Feature) - \min(Feature)}$ and then fed into a stacked autoencoder.

We adopt two kinds of autoencoder described above to pretrain each layer in turn with first week's data and finetune the whole model using the same data with labels. Then subsequent data are split into 2 parts: The first 7 sessions' data in each day is used to finetune the network in order to adapt the model to the pattern changing during rehabilitation and the last session's data is for testing. Figure 1 shows the structure of our model.

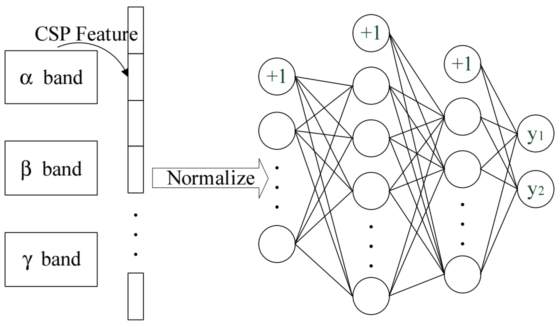

Fig. 1. The structure of our model. EEG data filtered by each band are transformed into a 4-dimensional feature by CSP, and fed into a neural network with $12 \times 25 \times 16 \times 2$ units after being normalized.

4 Result

Considering the training parameters of DAE for all the subjects, we set learning rate 0.5, scaling factor 0.99, momentum 0.5 and a binary masking noise with fraction 0.3 as the input. The sparsity constraint and penalty factor is subject specific.

In terms of CAE, we set λ 0.05, learning rate 0.5, scaling factor 0.99 and momentum 0.5 for all the subjects.

In order to evaluate the performance of our method, we also apply traditional CSP-SVM method and frequency boosting method on dataset of 4 patients. We calculate classification accuracies of stroke patients EEG in the chosen 6 weeks out of 2 months and finally the mean test session accuracy of each week is calculated. Table 1 shows test accuracies of 4 patients pretrained using DAE.

Table 1. Test accuracies of sliding windows using DAE. Note that the data collected in first week are used to pretrain each layer's autoencoder.

Subject	Age	2nd Week	3rd Week	4th Week	5th Week	6th Week
1	65	0.57	0.63	0.61	0.73	0.79
2	71	0.49	0.51	0.51	0.56	0.76
3	50	0.56	0.55	0.51	0.60	0.69
4	65	0.52	0.58	0.53	0.62	0.72

Compared with other methods, our method achieves a better accuracy (Fig. 2, Table 2). We consider the hidden layer's output as the hidden encoding of the original features. The hidden encoding of autoencoders provides a new representation of the original data in the subspace defined by the weights. Our experiment result shows that such representations are more robust to noisy data of stroke patients and give us much useful discriminative information for classification, which attributes to sparse and smooth representation.

Figure 2 also shows a rising tendency in terms of test accuracy. It's worth mentioning that the whole result also shows the feasibility and effectiveness of our BCI-FES rehabilitation training system.

Table 2. Best Test Session Accuracy. The best session for all subjects all appeals in the last week of our experiment.

Method	Subject 1	Subject 2	Subject 3	Subject 4
CSP-SVM	0.64	0.818	0.626	0.671
F-Boost	0.843	0.843	0.669	0.737
CSP-DAE	**0.941**	**0.913**	**0.779**	0.72
CSP-CAE	0.923	0.779	0.711	**0.749**

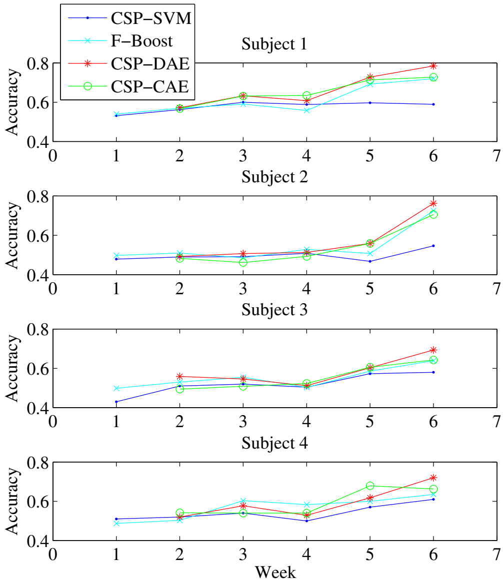

Fig. 2. Mean test accuracy of CSP-SVM, Frequency Boosting, CSP-DAE and CSP-CAE over time. Note that: (1) For autoencoders, data of first week is used for pretraining. (2) CSP-DAE achieves a higher accuracy in most cases. (3) A rising tendency can be observed over time.

To analyze the band contribution to classification, we calculate the relative importance of each feature using connection weights [17]. Figure 3 shows the gradual changes of importance during experiment of four subjects.

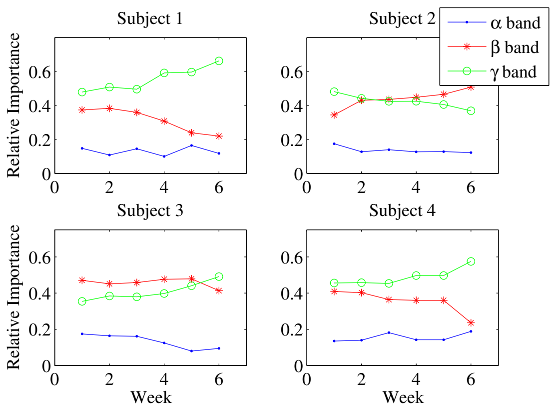

Fig. 3. The contribution changes of each band over rehabilitation process

The fact that the importance of gamma band increased for 3 patients while the contribution of beta band increased for subject 2 over time implies the gradual changes of motor imagery patterns of different stroke patients. Noting that oscillatory activity in gamma band is related to gestalt perception and cognitive functions and the oscillations in alpha and beta band are obvious indicators of movement, this phenomenon may reveal potential mechanisms about stroke recovery. We consider that the finetuning process of our last 5 weeks' training adapts the model to this migration so that the frequency modulations can be detected by CSP, thus improves classification accuracy comparing with traditional method.

Three patients receiving traditional clinical treatments get a lower clinical rehabilitation parameters in post assessment, which indicates that our system and active training paradigm accelerates the rehabilitation of impaired cortex.

5 Conclusion

In this paper, we propose a method which filters signal into multiple bands and use the autoencoder paradigm to train a network to classify two classes' motor imagery EEG of stroke patients. This method detects important structure in the raw common spatial patterns by using a local unsupervised criterion to pretrain each layer in the network and captures discriminative pattern changes during rehabilitation process by keeping finetuning the model. Compared with traditional CSP-SVM classifier, our method achieves a better result on both accuracy over time and optimal session accuracy. The analysis of band changing during rehabilitation provides a prior knowledge about motor imagery pattern of stroke patients. Furthermore, the comparison of experimental group with control group demonstrates the effectiveness of our multi-modal BCI-FES rehabilitation training system and active training paradigm.

Acknowledgement. The work was supported by the National Natural Science Foundation of China (Grant No. 91120305, 61272251).

References

1. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain–computer interfaces for communication and control. Clinical Neurophysiology 113(6), 767–791 (2002)
2. Pfurtscheller, G., Müller-Putz, G.R., Pfurtscheller, J., Rupp, R.: Eeg-based asynchronous bci controls functional electrical stimulation in a tetraplegic patient. EURASIP Journal on Applied Signal Processing 2005, 3152–3155 (2005)
3. Ramoser, H., Muller-Gerking, J., Pfurtscheller, G.: Optimal spatial filtering of single trial eeg during imagined hand movement. IEEE Transactions on Rehabilitation Engineering 8(4), 441–446 (2000)
4. Zhang, H., Zhang, L.: Spatial-spectral boosting analysis for stroke patients' motor imagery eeg in rehabilitation training. CoRR, Vol. abs/1310.6288 (2013)

5. Li, J., Zhang, L.: Regularized tensor discriminant analysis for single trial eeg classification in bci. Pattern Recognition Letters 31(7), 619–628 (2010)
6. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. Proceedings of the IEEE 89(7), 1123–1134 (2001)
7. Song, L., Gordon, E., Gysels, E.: Phase synchrony rate for the recognition of motor imagery in brain-computer interface. Advances in Neural Information Processing Systems 18, 1265 (2006)
8. Shahid, S., Sinha, R.K., Prasad, G.: Mu and beta rhythm modulations in motor imagery related post-stroke eeg: a study under bci framework for post-stroke rehabilitation. BMC Neuroscience 11(Suppl. 1), P127 (2010)
9. Liang, J., Zhang, H., Liu, Y., Wang, H., Li, J., Zhang, L.: A frequency boosting method for motor imagery EEG classification in BCI-FES rehabilitation training system. In: Guo, C., Hou, Z.-G., Zeng, Z. (eds.) ISNN 2013, Part II. LNCS, vol. 7952, pp. 284–291. Springer, Heidelberg (2013)
10. Wang, H., Liu, Y., Zhang, H., Li, J., Zhang, L.: Causal neurofeedback based BCI-FES rehabilitation for post-stroke patients. In: Lee, M., Hirose, A., Hou, Z.-G., Kil, R.M. (eds.) ICONIP 2013. LNCS, vol. 8226, pp. 419–426. Springer, Heidelberg (2013)
11. Liu, Y., Li, M., Zhang, H., Li, J., Jia, J., Wu, Y., Cao, J., Zhang, L.: Single-trial discrimination of eeg signals for stroke patients: A general multi-way analysis. In: EMBC 2013, pp. 2204–2207. IEEE (2013)
12. Vincent, P., et al.: Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. The Journal of Machine Learning Research 9999, 3371–3408 (2010)
13. Rifai, S., Vincent, P., Muller, X., Glorot, X., Bengio, Y.: Contractive auto-encoders: Explicit invariance during feature extraction. In: Proceedings of the 28th International Conference on Machine Learning (ICML 2011), pp. 833–840 (2011)
14. Liu, Y., Zhang, H., Wang, H., Li, J., Zhang, L.: Bci-fes rehabilitation training platform integrated with active training mechanism. In: IJCAI 2013 Workshop on Intelligence Science (2013)
15. Li, J., Zhang, L.: Active training paradigm for motor imagery bci. Experimental brain research 219(2), 245–254 (2012)
16. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.-A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine learning, pp. 1096–1103. ACM (2008)
17. Garson, G.D.: Interpreting neural-network connection weights. AI expert 6(4), 46–51 (1991)