A BCI System Based on Orthogonalized EEG Data and Multiple Multilayer Neural Networks in Parallel Form

Kenji Nakayama, Hiroki Horita, and Akihiro Hirano

Graduate School of Natural Science and Technology, Kanazawa University
Kanazawa, 920-1192, Japan
nakayama@t.kanazawa-u.ac.jp
http://leo.ec.t.kanazawa-u.ac.jp/~nakayama

Abstract. A BCI system, using orthogonalized EEG data sets and multiple multilayer neural networks (MLNNs) in a parallel form, is proposed. In order to emphasize feature of multi-channel EEG data, Gram-Schmidt orthogonalization has been applied. Since there are many channel orders to be orthogonalized, many kinds of orthogonalized data sets can be generated for the same EEG data set by changing the channel order. These data sets have different features. In the proposed method, different channel orders are assigned to the multiple MLNNs in a training phase and in a classification process. A good solution can be searched for by changing the channel orders within a small number of trials. By using EEG data for five mental tasks, a correct classification rate is increased from 88% to 92%, and an error classification rate is decreased from 4% to 0%.

Keywords: BCI, EEG, Brain waves, Neural network, Mental task, Orthogonal components, Gram-Schmidt.

1 Introduction

Approaches to BCI systems include nonlinear classification by using spectrum power, adaptive auto-regressive model and linear classification, space patterns and linear classification, hidden Markov models, and so on [1]. Furthermore, neural networks have been also applied [2]. In our previous works, FFT of EEG data and a multilayer neural network (MLNN) have been applied to the BCI. Efficient pre-processing techniques to extract features have been also employed [5]. Furthermore, the generalization learning methods have been applied [4],[6]. Effects of sensor locations has been analyzed for BCI using MEG data [7].

Methods to extract essential features of the multi-channel EEG data have been proposed. In our previous work, Gram-Schmidt orthogonalization has been applied to generate the orthogonal components [8]. The orthogonalized data sets have different features for the different channel order to be orthogonalized, resulting in different classification performances. For this reason, the optimum channel order should be searched for [8].

K. Diamantaras, W. Duch, L.S. Iliadis (Eds.): ICANN 2010, Part I, LNCS 6352, pp. 205–210, 2010.

2 BCI System Based on Multiple MLNNs

2.1 Gram-Schmidt Orthogonalization

The vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M\}$, which express the brain waves at M-channels, are usually linearly independent. This set can be transferred into the orthogonal vector set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_M\}$ by Gram-Schmidt orthogonalization [9]. $\{\boldsymbol{v}_i\}$ are Fourier transformed and their amplitude are pre-processed [5], and are used as the MLNN input data [8].

2.2 Proposed BCI System Using Orthogonalized EEG and Multiple MLNNs in Parallel Form

In order to overcome the above channel order problem, a BCI system using multiple MLNNs in a parallel form, shown in Fig.1, is proposed in this paper.

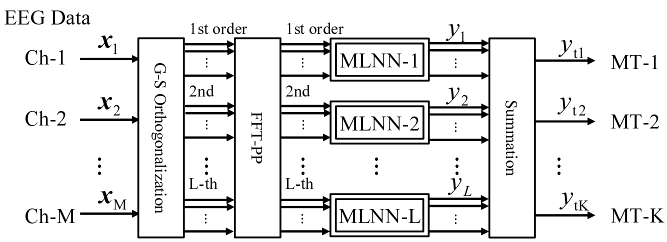

Fig. 1. A BCI system using orthogonalized EEG data and multiple MLNNs in parallel form. L kinds of channel orders are used for MLNN-1∼MLNN-L.

Let $\boldsymbol{v}_{ij}$ be the i-th orthogonalized input data set generated by using the i-th channel order. L input data sets, $\boldsymbol{v}_{ij}, i = 1 \sim L, j = 1 \sim M$, are generated from the same EEG data set, and are applied to MLNN-i, $i = 1 \sim L$ individually. They are trained independently so as to output the desired response. In the classification process, letting the output of MLNN-i be $\boldsymbol{y}_i = [y_{i1}, y_{i2}, \cdots, y_{iK}]^T, i = 1 \sim L$, the total output y_{tk} is given by Eq.(1). The mental task is classified based on the maximum element in $\boldsymbol{y}_t = [y_{t1}, y_{t2}, \cdots, y_{tK}]^T$ [5].

$$y_{tk} = \frac{1}{L} \sum_{i=1}^{L} y_{ik}, \qquad k = 1 \sim K \tag{1}$$

2.3 Conventional Multiple MLNNs in Parallel Form

A similar structure has been proposed as shown in Fig.2 [10], denoted 'Method-I' in this paper. MLNN-i, $i = 1 \sim L$ receive the same input data, that is $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_M$, and provide the outputs, $\boldsymbol{y}_i = [y_{i1}, y_{i2}, \cdots, y_{iK}]^T, i = 1 \sim L$, which represent the corresponding mental task. In order to realize high generalization performances, different initial connection weights are assigned to MLNN-1 ∼MLNN-L. Each MLNN is trained so as to output the desired targets. The final outputs are also given by Eq.(1)

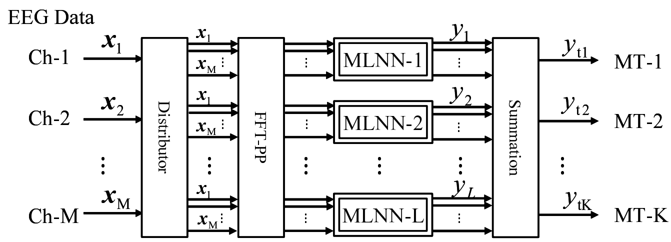

Fig. 2. Conventional Method-I: EEG data are not orthogonalized. Initial connection weights of MLNN-1 ∼ MLNN-L are different.

Another conventional approach using multiple MLNNs is 'Bagging Method' [11], called 'Method-II' in this paper. MLNN-1 ∼MLNN-L are trained by using different EEG data for the same mental task. Each MLNN is trained to output the desired targets, and the final outputs are given by Eq.(1). In the classification process, a single EEG data set is applied to all MLNNs.

In the proposed method (Fig.1), the EEG data are orthogonalized. On the other hand, the conventional methods, that is Method-I and Method-II, use the original EEG data. As described in Sec.2.1, the Gram-Schmidt orthogonalization process can generate different orthogonalized data sets by changing the channel order. This means that feature of each channel can be emphasized, at the same time, different kinds of feature sets can be generated, which can be effectively applied to MLNNs. These two points of the proposed method can improve the mental task classification performance.

3 Simulations and Discussions

In this paper, EEG data, available from the web site of Colorado State University [3], are used. The following five kinds of mental tasks are employed. (1)Baseline-relaxed situation, (2)Multiplication, (3)Letter composing, (4)Rotation of a 3-D object, (5)Counting numbers.

3.1 Simulation Setup

The EEG data with a 10 sec length for five mental tasks were measured 10 times. Therefore, 10 data sets are available. Among them, 8 data sets are used for training and the remaining 2 data sets are used for testing. Five different combinations of 2 data sets are used for the testing. Thus, five independent trials are carried out. Classification accuracy is evaluated based on the average over five trials [1].

Mental task classification is evaluated based on a correct classification rate (P_c), an error classification rate (P_e) and a rate of correct and error classification (R_c) as follows:

$$P_c = \frac{N_c}{N_t} \times 100\%, \quad P_e = \frac{N_e}{N_t} \times 100\% \tag{2}$$

$$R_c = \frac{N_c}{N_c + N_e}, \quad N_t = N_c + N_e + N_r \tag{3}$$

N_c, N_e and N_r are the numbers of correct and error classifications and rejections, respectively. When the MLNN outputs are smaller than the threshold, no estimation is provided, that is 'Rejected'. N_t is the total number of the testing data. R_c is used to evaluate a correct classification rate except for 'Rejection'.

The number of hidden units is 20. The threshold for rejection is set to be 0.7. The MLNNs are trained by the error back propagation learning algorithm.

3.2 Classification Performances

Table 1 shows classification performances of the proposed BCI system shown in Fig,1. In the learning process, small random numbers, uniformly distributed in $[-0.05, 0.05]$, are added to the MLNN input in 'Generalization' and not added in 'No Generalization'. The channel orders, with which good classification performances are obtained in the BCI system using a single MLNN shown in Fig.1, are used. 'L' is the number of the MLNNs in a parallel form. For example, in the case of '$L = 5$', five kinds of the above channel orders are used for MLNN-1∼MLNN-5. 'Average' means average values of P_c, P_e and R_c of the BCI system for all channel orders. In this table, P_e is always zero. P_c can be improved by the generalization method.

Table 1. Classification performance of proposed method

	No Generalization			Generalization		
L	P_c[%]	P_e[%]	R_c	P_c[%]	P_e[%]	R_c
Average	71.5	12.2	0.855	82.3	8.1	0.911
5	76	0	0.947	80	0	0.957
10	80	0	0.929	78	0	0.978
20	78	0	1.0	84	0	1.0
30	76	0	1.0	84	0	1.0

Table 2 shows classification performance of Method-I. Even though P_c can be improved from the proposed method, P_e stil remains $4 \sim 6\%$. Table 3 shows classification performances of Method-II. Compared to the previous two methods, the classification accuracy is not so good.

Table 2. Classification performance of Conventional Method-I

	No Generalization			Generalization		
L	P_c[%]	P_e[%]	R_c	P_c[%]	P_e[%]	R_c
Average	76.1	7.3	0.912	86.8	5.4	0.942
5	76	6	0.927	88	4	0.957
10	76	6	0.927	88	4	0.957
20	76	6	0.927	88	4	0.957
30	76	6	0.927	88	4	0.957

Table 3. Classification performance of Conventional Method-II (Bagging Method)

	No Generalization			Genaralization		
L	P_c[%]	P_e[%]	R_c	P_c[%]	P_e[%]	R_c
Average	65.5	13.9	0.825	75.9	11.5	0.867
5	60	8	0.882	76	0	1.0
10	66	6	0.917	76	0	1.0
20	66	6	0.917	76	2	0.974
30	68	6	0.919	76	2	0.974

3.3 Searching for Good Solution in Proposed BCI System

The following searching method is proposed in this paper. In one trial, the combination of the channel orders, which are randomly determined and are used in the BCI system shown in Fig.1, is changed 10 times. 10 kinds of solutions for the BCI system can be obtained. Among them, the best solution is selected. This kind of the trial is repeated 5 times (Trial: 1st, 2nd, 3rd, 4th, 5th), in order to confirm general efficiency. The simulation results are shown in Table 4. Good solutions with the highest R_c are selected in each trial, and are listed in this table. As shown in this table, very high P_c, that is more than 90%, and very low P_e, that is zero, can be obtained for $L = 10$. These results are very superior to the results of the conventional methods.

This result means (1) the channel orders assigned to multiple MLNNs can be determined randomly, and (2) the good solution can be searched for by changing the combination of the channel orders only 10 times. The same results were obtained for another subjects.

Table 4. Classification performance of proposed method. Best solution is searched for.

	$L = 5$			$L = 10$		
Trial	P_c[%]	P_e[%]	R_c	P_c[%]	P_e[%]	R_c
1st	88	0	1.0	92	0	1.0
2nd	88	0	1.0	92	0	1.0
3rd	88	2	0.978	92	0	1.0
4th	92	0	1.0	92	0	1.0
5th	88	0	1.0	92	0	1.0

4 Conclusion

A BCI system, which uses multiple MLNNs in a parallel form with the orthogonalized EEG data sets, is proposed. Different channel orders are assigned to the multiple MLNNs. By searching for good solutions for different combinations of the channel orders, the correct classification of $P_c = 92\%$ and the error classification of $P_e = 0\%$ can be obtained. These results are very superior to those of the conventional methods.

References

1. Anderson, C., Sijercic, Z.: Classification of EEG signals from four subjects during five mental tasks. In: Bulsari, A.B., Kallio, S., Tsaptsinos, D. (eds.) EANN 1996, Systems Engineering Association, PL34, FIN-20111, Finland, pp. 407–414 (1996)
2. Muller, K.R., Anderson, C.W., Birch, G.E.: Linear and non-linear methods for brain-computer interfaces. IEEE Trans. Neural Sys. Rehab. Eng. 11(2), 165–169 (2003)
3. Colorado State University, http://www.cs.colostate.edu/eeg/
4. Robert, J., Burton, M., Mpitsos, G.J.: Event-dependent control of noise enhances learning in neurla networks. Neural Networks 5(4), 627–637 (1992)
5. Nakayama, K., Inagaki, K.: A brain computer interface based on neural network with efficient pre-processing. In: Proc. IEEE, ISPACS 2006, Yonago, Japan, pp. 673–676 (December 2006)
6. Nakayama, K., Kaneda, Y., Hirano, A.: A brain computer interface based on FFT and multilayer neural network-Feature extraction and generalization. In: Proc. IEEE, ISPACS 2007, Xiamen, China, pp. 101–104 (December 2007)
7. Nakayama, K., Kaneda, Y., Hirano, A., Haruta, Y.: A BCI using MEG vision and multilayer neural network - Channel optimization and main lobe contribution analysis. In: Proc., IEEE ISPACS 2008, Bangkok, Thailand, pp. 316–319 (February 2009)
8. Nakayama, K., Horita, H., Hirano, A.: Neural network based BCI by using orthogonal components of multi-channel brain waves and generalization. In: Kůrková, V., Neruda, R., Koutník, J. (eds.) ICANN 2008, Part II. LNCS, vol. 5164, pp. 879–888. Springer, Heidelberg (2008)
9. Strang, G.: Linear Algebra and its Applications. Academic Press, Inc., N.Y. (1976)
10. Caruana, R.: Multitask learning. Machine Learning 28(1), 41–75 (1997)
11. Bauer, E., Kohavi, R.: An empirical comparison of voting classification algorithms: Bagging, Boosting, and Variants. Machine Learning 36, 105–142 (1999)