Canonical Correlation Analysis Neural Network for Steady-State Visual Evoked Potentials Based Brain-Computer Interfaces

Ka Fai Lao, Chi Man Wong, Feng Wan, Pui In Mak, Peng Un Mak, and Mang I Vai

Department of Electrical and Computer Engineering
Faculty of Science and Technology
University of Macau, Av. Padre Tomás Pereira Taipa, Macau, China
{da92050,fwan}@umac.mo

Abstract. Canonical correlation analysis (CCA) is a promising feature extraction technique of steady state visual evoked potential (SSVEP)-based brain computer interface (BCI). Many researches have showed that CCA performs significantly better than the traditional methods. In this paper, the neural network implementation of CCA is used for the frequency detection and classification in SSVEP-based BCI. Results showed that the neural network implementation of CCA can achieve higher classification accuracy than the method of power spectral density analysis (PSDA), minimum energy combination (MEC) and similar performance to the standard CCA method.

Keywords: Canonical Correlation Analysis (CCA), Neural Network, Brain Computer Interface (BCI), Steady-State Visual Evoked Potential (SSVEP).

1 Introduction

Non-invasive BCIs have been widely investigated in recent decades. It provides an alternative communication and control channel between human and environment through electroencephalographic (EEG) produced by brain activities [1]. This communication channel brings great benefit to patients with severe motor disabilities, which enable them to express their wishes to caregivers or operate an intelligent wheelchair by a BCI without any brain's normal pathway of peripheral nerves and muscles. A general BCI includes three working processes: signal acquisition, feature extraction and feature classification. The brain activity then is translated to a device command or message. Different types of EEG signals such as P300 evoked potential, sensorimotor mu/beta rhythms, slow cortical potential (SCP), movement-related cortical potential (MRCP) and visual evoked potential (VEP), have been paid much attention in the research of BCI [2].VEP is an evoked potential over occipital area elicited by an external visual stimulus. It can be classified into transient VEP (TVEP) and steady-state VEP (SSVEP). SSVEP is elicited by a consecutive, stable and periodic stimulus with repetition rate higher than about 4 Hz. The main characteristics of SSVEP are frequency-locked and phase-locked. SSVEP is highly spoken of among

C. Guo, Z.-G. Hou, and Z. Zeng (Eds.): ISNN 2013, Part II, LNCS 7952, pp. 276–283, 2013.

all the BCI applications because of its high signal to noise ratio (SNR), relative immunity to artifacts, high information transfer rate (ITR) and low training requirements. Basically, most BCIs based on SSVEPs utilize the frequency information of SSVEPs for identification. It means that the computer can detect which target the subject desires to select by checking the frequency information of SSVEPs [3]. Up to now, the common feature extraction methods include power spectral density analysis (PSDA), minimum energy combination (MEC) and CCA. For PSDA method, PSD is estimated by fast Fourier Transform (FFT) and its peak is detected to recognize the target stimulus. MEC method combines multiple electrode signals to cancel the noise as much as possible, and then the target stimuli is recognized according to the maximum signal power. Researches have demonstrated that CCA method has lower deviation, higher detection accuracy and higher insensitive to SNR than the tradition method [3] [4].

Although the standard CCA method has provided great performance on SSVEP-based BCI, there are still many possible improvements. Neural network (NN) implementation for BCI is one of the promising research directions. The important characteristics of NN include its self-adaptive structure, universal function approximation and the expansibility of network. In other words, NN implementation for BCI provides a more flexible and extendable environment for researchers. Moreover, some researchers have applied the neural network based approaches to achieve excellent performance on the classification accuracy [6] – [8]. Inspired by their work, we aim to apply the NN based CCA (NNCCA) models proposed by *Hsieh et al.* [9] [10] in SSVEP-based BCI to recognize SSVEPs. The performance of NNCCA is verified with the real EEG data from five healthy subjects and compared with the standard CCA, PSDA and MEC method. The preliminary offline experiment results show that NNCCA can provide the similar classification accuracy to standard CCA and the better accuracy than MEC and PSDA.

2 Methodology

2.1 CCA

CCA is a statistical technique that applied to two datasets which we believe there are some potential relationships between them. Their relationship is determined by correlation coefficient. Consider two multidimensional datasets $X \in \mathbb{R}^{\mathrm{p}\times\mathrm{n}}$ and $Y \in \mathbb{R}^{\mathrm{q}\times\mathrm{n}}$ as well as their correlation is wanted to be found, in other points of view, two inputs are wanted to be reproduced each other through linear combination of their variables:

$$U = W_1^{\mathrm{T}} X \tag{1}$$

$$V = W_2^{T} Y \tag{2}$$

$W_1 \in \mathbb{R}^{p\times 1}$ and $W_2 \in \mathbb{R}^{q\times 1}$ denote the correlation coefficient that can maximize the correlation between the canonical variates *U* and *V*. Furthermore, the canonical correlation coefficient of the variates is defined by Eq. (3):

$$\eta = \frac{cov(U,V)}{\sqrt{cov(U,U)cov(V,V)}} = \frac{W_1^T S_{XY} W_2}{\sqrt{W_1^T S_{XX} W_1 W_2^T S_{YY} W_2}} \tag{3}$$

$$F = W_1^T S_{XY} W_2 - \frac{\lambda_1}{2}(W_1^T S_{XX} W_1 - 1) - \frac{\lambda_2}{2}(W_2^T \textstyle\sum_{YY} W_2 - 1) \tag{4}$$

S denotes the sample covariance matrix. The maximum canonical correlation corresponds to the maximum canonical correlation coefficient. To solve the maximum canonical coefficient, Lagrange Multiplier method is applied to Eq. (3) and the cost function in Eq. (4) is obtained. The maximum correlation coefficient then is solved from the cost function. According to the method presented by *Lin et al.* [3], EEG signals from multiple channels are used to calculate the canonical correlation coefficients with all stimulus frequencies in the system. In other words, the EEG signal and stimulus signal are considered as *X* and *Y*. Assume there are *I* stimulus frequencies and all of them are square-wave periodic, the *i* th stimuli signal can be decomposed into Fourier series:

$$Y_i = \begin{pmatrix} sin\,(2\pi f_i/f_s) & \cdots & sin\,(2\pi f_i n/f_s) \\ cos(\,2\pi f_i/f_s) & \cdots & cos(\,2\pi f_i n/f_s) \\ \vdots & \ddots & \vdots \\ sin\,(2\pi f_i/f_s) & \cdots & sin\,(6\pi f_i n/f_s) \\ cos(\,2\pi f_i/f_s) & \cdots & cos(\,6\pi f_i n/f_s) \end{pmatrix} \tag{5}$$

where *n* is the number of sample and f_s is sampling rate. Since the analysis is based on temporal and spatial information, phase different between inputs is also the influence factor of correlation. Linear combination of sine and cosine signal in Eq. (5) can match the phase between *X* and *Y* and the phase information of *X* is reflected on W2. Finally the frequency with the largest coefficient is identified as the stimulus frequency. CCA provides several superiorities on feature extraction of SSVEP-based BCI. First, signal preprocessing and feature extraction can be simultaneously done by CCA. Second, it provides another technique for the feature extraction. Especially most of the analysis methods are performed in frequency domain. Those superiorities provide a strong interest on this research field.

2.2 NNCCA

The core of CCA is a maximization problem. The cost function as similar to Eq. (4) thus can be used to develop the learning rule of NN and construct the NN based CCA. The NNCCA model presented by *Hsieh* [9] [10] is showed in Fig. 1. Each functional network is a three layer feed forward network which contains two hidden layers and one output layer. From the figure it can be observed that the network is combined by three parts. The double-barreled network on the left maps the input {*x*, *y*} to {*u*, *v*}.

The neuron function of its first layer can be either performed by hyperbolic tangent function, as shown in Eq. (6) and Eq. (7), or the identity function. Second layer is usually the identity function. According to the choice of neuron function (in first layer), this network can perform either the linear (equivalent to standard CCA) or nonlinear CCA. In this study, we focus on the implementation of linear NNCCA.

$$h_k^{(x)} = \tanh\left[\left(W^{(x)}x + b^{(x)}\right)_k\right] \tag{6}$$

$$h_l^{(y)} = \tanh\left[\left(W^{(y)}x + b^{(y)}\right)_k\right] \tag{7}$$

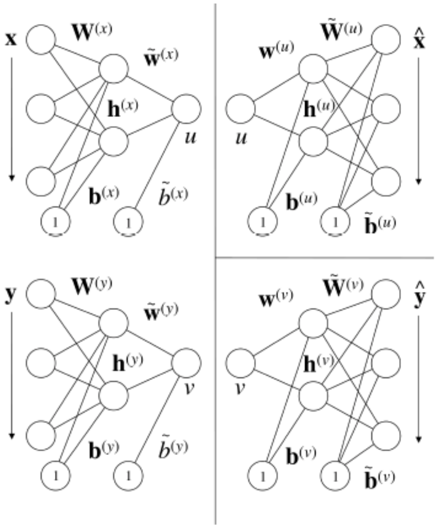

Fig. 1. NNCCA model proposed by the research of *Hsieh* [9] [10]

The networks show on the right of figure try to inversely map $\{u, v\}$ to the corresponded input $\{x', y'\}$. However, the inverse mapping is unnecessary in the implementation of SSVEP based-BCI therefore it will be ignored. The cost function of the NNCCA model is given by Eq. (8). The second, third, fourth and fifth terms of Eq. (8) are normalization constraints that force u and v to have zero mean and unit variance; the sixth term is a weight penalty whose relative magnitude is controlled by the parameter P_1. Larger values of P_1 lead to smaller weights (i.e. fewer effective model parameters), which results in a more linear model. Compare to the cost function of standard CCA in Eq. (4), Eq. (8) provides more constraints to ensure the convergence of neural network model. Due to the effect of weight penalty and the convergence precision, linear NNCCA gives slightly different canonical correlation coefficient when compared with standard CCA. The difference of them can be found in the experiment, which is discussed in the later sections.

$$J = -\eta(U,V) + \langle U\rangle^2 + \langle V\rangle^2 + \left(\langle U^2\rangle^{\frac{1}{2}} - 1\right)^2 + \left(\langle V^2\rangle^{\frac{1}{2}} - 1\right)^2 + P_1\left[\sum_{ki}\left(W_{ki}^{(x)}\right)^2 + \sum_{lj}\left(W_{lj}^{(y)}\right)^2\right] \quad (8)$$

3 Offline Experiment

There were 5 healthy subjects ranging from the ages of 22 – 27 participated in this experiment. They were asked to continuously focus on the target stimulus (showed on a LCD monitor with 60 Hz refresh rate) for every 10 seconds, with 4 seconds of rest between each trial. There were 5 stimuli (flashing white squares) as shown on the LCD screen, with flashing frequency of 7.5 Hz, 8 Hz, 10Hz, 15Hz and 20Hz. Each subject carried out 9 trials for each stimulus. That is, each trial contains a 6 seconds useful EEG data and totally 45 trials were completed for each subject. All EEG signals were recorded by a g.USBamp biosignal amplifier at 600 Hz sampling rate ($f_s = 600$ Hz) from six channels PO3, PO4, POz, Oz, O1 and O2 placed on the standard position of the 10-20 international system. In order to verify whether NNCCA can perform the same functions as standard CCA, NNCCA and standard CCA are compared on the above datasets. Moreover, PSDA and MEC are also compared here.

4 Offline Experimental Result and Discussion

There are 45 trials of EEG data for each subject. Standard CCA method and linear NNCCA method are first applied to those trials. After this, 225 pairs of canonical correlation coefficients are obtained (one EEG signal is calculated with 5 stimulus signal). To eliminate the effect of individual differences, those coefficients are grouped by subject. 100 samples are randomly drawn from each group and analysis of variance (ANOVA) is used to check the differences between the results of both CCA methods. Results of ANOVA from each group of samples are 0.8824, 0.8285, 0.8961, 0.909 and 0.8709. Hence, there are no significant differences between the two methods ($p > 0.05$). On the other hands, their consistency can be directly verified through the observation of data. Fig. 2 shows the average correlation coefficient for each group. It can be found that the average canonical correlations obtained by NNCCA have a slightly larger value than standard CCA. However, the difference between them is very small (no larger than 0.005). As we have introduced in the previous part, the difference is caused by the effect of weight penalty and the convergence precision (we will just call it 'effect' in later discussion). Due to the 'effect', the canonical correlation calculated by NNCCA will not be same as the one calculated by standard CCA, but will be a similar value (can be larger or small) as the standard one. After all the average error percentage between the correlations calculated by the two methods is estimated, which is about 2 to 3%. This result indicates that linear NNCCA is a good estimator of standard CCA.

After considering the consistence of linear NNCCA and standard CCA, we then focus on the classification accuracy of linear NNCCA. Table 1 presents the classification accuracy (which successfully recognizes the SSVEP to its corresponded stimulus frequency, according to maximum canonical correlation coefficient or corresponded standard of the method) that obtains from the method of standard CCA, linear NNCCA, MEC and PSDA. The best average accuracy is given by both CCA method, where linear NNCCA method obtains 83.12% accuracy and standard CCA obtain 82.66% accuracy. The different of accuracy is induced by the 'effect'. The interesting thing is the 'effect' might coincidentally generate correct (or wrong) classification but it rarely happens. Moreover, the average computation time of linear NNCCA method for single run is around 2 seconds on an Intel Core 2 Duo E8500 CPU. The number of run depends on the complexity of the network. For instance, if the hidden layer of the network contains 5 neurons, it needs more run to obtain a convergence and true result than the 4 neurons case. In general, linear NNCCA with single neuron in each functional network (or in other words, for two data set input) can give a convergence result in about 5 to 8 runs, which means at least 10 seconds are needed to recognize SSVEP. This computation time is still too long for the online application. The computation rate is probably enhanced by the hardware that supports parallel computation. Recent work has used General-Purpose Graphic Processing Units (GPGPU) for the computation of BCI system [5], it has demonstrated that such hardware can efficiently increase the computation rate compared to the CPU system. With GPGPU receiving more and more attention, it will be the promising direction to increase the rate and enable the online application of NN implementation BCI.

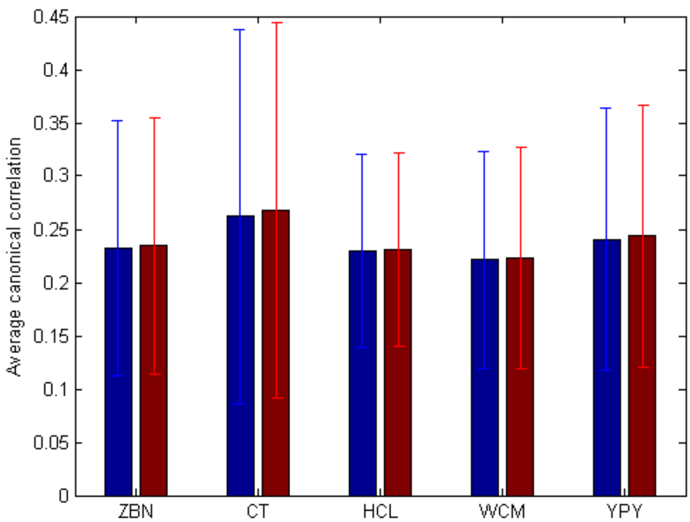

Fig. 2. Comparison of the average correlation coefficient for each group. The left bar is the average result of standard CCA and the right bar is the average result of linear NNCCA.

Table 1. The classification accuracy of feature extraction methods

Subject	CCA	NNCCA	MEC	PSDA
WCM	**91.1%**	**91.1%**	78.1%	77.8%
YPY	**93.3%**	**93.3%**	76.2%	55.8%
HCL	73.3%	**75.6%**	**97.4%**	73.3%
CT	86.7%	**88.9%**	68.3%	**88.9%**
ZBN	**68.9%**	66.7%	64.1%	55.6%
Average	82.66%	**83.12%**	76.81%	70.27%

5 Conclusion

This study attempts to implement CCA algorithm with NN for the frequency detection in SSVEP-based BCI. Results have proposed that NNCCA method can be applied to SSVEP based-BCIs as it leads to higher classification accuracy than those with the methods of PSDA, MEC, and similar performance to the standard CCA method. For future research it would be interesting to make good use of the parallel computing characteristic of NN approach. By using the parallel computing-supported hardware, such as GPGPU, the computation speed can be enhanced and the online application of NN implementation BCI can hopefully be achieved.

Acknowledgement. This work is supported in part by the Macau Science and Technology Development Fund under grant FDCT 036/2009/A and the University of Macau Research Committee under grants RG059/08-09S/FW/FST, RG080/09-10S/WF/FST and MYRG139(Y3-L2)-FST11-WF.

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, 767–791 (2002)
2. Wolpaw, J.R., Boulay, C.B.: Brain-Computer Interfaces, The Frontiers Collection. Springer, Heidelberg (2010) ISBN 978-3-642-02090-2
3. Lin, Z., Zhang, C., Wu, W., Gao, X.: Frequency recognition based on canonical correlation analysis for SSVEP-Based BCIs. IEEE Transactions on Biomedical Engineering 54(6), 1172–1176 (2007)
4. Nan, W., Wong, C.M., Wang, B., Wan, F., Mak, P.U., Mak, P.I., Vai, M.I.: A comparison of minimum energy combination and canonical correlation analysis for SSVEP detection. In: 2011 5th International IEEE/EMBS Conference on Neural Engineering, NER 2011, pp. 469–472 (2011)
5. Wilson, J.A.: Using general-purpose graphic processing units for BCI systems. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2011, pp. 4625–4628 (2011)

6. Cecotti, H.: A time-frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses. Pattern Recognition Letters 32(8), 1145–1153 (2011)
7. Coyle, D.: Neural network based auto association and time-series prediction for biosignal processing in brain-computer interfaces. IEEE Computational Intelligence Magazine 4(4), 47–59 (2009)
8. Hsu, W.-Y.: Fuzzy Hopfield neural network clustering for single-trial motor imagery EEG classification. Expert Systems with Applications 39(1), 1055–1061 (2012)
9. Hsieh, W.W.: Nonlinear canonical correlation analysis by neural networks. Neural Networks 13(10), 1095–1105 (2000)
10. Cannon, A.J., Hsieh, W.W.: Robust nonlinear canonical correlation analysis: application to seasonal climate forecasting. Nonlinear Processes in Geophysics 12, 221–232 (2008)