Inter subject inconsistency measures of EEG data on the basis of correlation dimension

Md. Zakir Hossain
Dept. of Electrical and Electronic Engineering
Khulna University of Engineering and Technology
Khulna, Bangladesh
Zakireee62@yahoo.com

Md. Asadur Rahman
Dept. of Electrical and Electronic Engineering
Khulna University of Engineering and Technology
Khulna, Bangladesh
eeeasadur@yahoo.com

***Abstract*—Careful attentions must be required for analyzing high dimensional Electroencephalographic (EEG) signals. There have many discrepancy or linearity when signals are recorded at different stimulations for a subject. When these signals are collected for different trails of a subject, there have similarity also. In this paper, we want to measure this correspondence on the basis of correlation aspect. For this reason, we incorporate neural network with canonical correlation analysis (CCA) as a linear and nonlinear function. The effectiveness of the network capabilities is tested with a sine-cosine reference signals.**

***Keywords*- Electroencephalographic (EEG); canonical correlation analysis (CCA); neutral network; inconsistency**

I. INTRODUCTION

The brain is the centre of the nervous system. It generates one kind of electrical signals which is related to the body functions called electroencephalogram (EEG). These signals are recorded from the scalp placing electrodes on the right position. When these signals are collected with a sufficient time for different trials on the repetitive flicker of visual stimulation [1], it forms high dimensional set. All of these data are not equally important; also there have much correlated sets with them. One can easily use one of the correlated set for reducing computational cost. In these sense we want to search correlations between subjects.

In this paper, we execute neural network (NN) with canonical correlation analysis (CCA) to calculate correlation, since NN is well known for their powerful capacity. It also gives the best linear compression of data set in the sense of least mean square error [2]. The Neural CCA method is advantageous because of i) low capacity machine can be used, ii) better correlation can be obtained than statistical methods [3], iii) data is entered sequentially in the network for reducing computational cost.

There are different types of potential of brain signal; Steady State Visual Evoked Potential (SSVEP) is one of them. When subjects are focuses on the repetitive flicker of a visual stimulus [4], then these types of potential is recorded using brain-computer interface (BCI). Due to noise contamination, frequency components detection of EEG signal with a shorter time window (TW) is a challenging task [5]. There is also inconsistency among the subjects and trials. In this aspect, we want to measure correlation between different subjects to assure the reliable analysis of EEG data. For that reason, we include neural network with non linear and linear CCA for doing this task easily with a low computational task.

An established and broadly used method for EEG signal identification is the Power spectral density (PSD) analysis, which describes the frequency distribution with power. To evaluate PSD from the EEG signals within a TW Fast Fourier Transform (FFT) can be used [6]. To recognize SSVEP of the EEG signals based on canonical correlation analysis (CCA) is proposed by Lin et al. [7], where underlying frequency components are recognized through maximizing the correlations between the EEG signals of multiple channels and sine-cosine reference signals at different stimulus frequencies. Our concentration here is to scrutinize inter subjects and trial to trial inconsistency of EEG signals by incorporating neural network with CCA.

The comparative study of CCA and PSD analysis is explored in [8]. Where CCA maximizes correlation between two sets of variables, but statistical CCA takes longer time to search this correlation. So we incorporate NN with CCA for reducing computational cost as well it provides more flexibility in the optimization. In this sense we want to realize inconsistency between different subjects as well as between different trails of a subject.

For the linearity test of EEG data, in this paper we add linear function and nonlinear function with neural CCA. Due to sequential presentation of input data it requires very less memory, whereas statistical CCA requires huge amount of memory because data is fed at once.

The paper is structured as follows. In section II, methodology of the experiment is described into four subsections as Characteristics of collected EEG signals, neural CCA network, sine-cosine signal generation and discrepancy recognition accordingly. Experimental results are discussed in section III. Finally we conclude the work in section IV.

II. METHODOLOGY

There are various types of potential that's generated from human brain. SSVEP is one of the brain signals related to visual which identify the brain condition at various modes such as reading, watching TV etc. There have consistency when these signals are collected for various trials at different stimulus. In this sense we search correlation of EEG signals. We also want to calculate the possible correlation between the EEG signals and a set of reference signals to test the effectiveness of our network. Usually the reference signals are consisting of sine-cosine signals. The overview of the entire methodology is explored in Fig. 1.

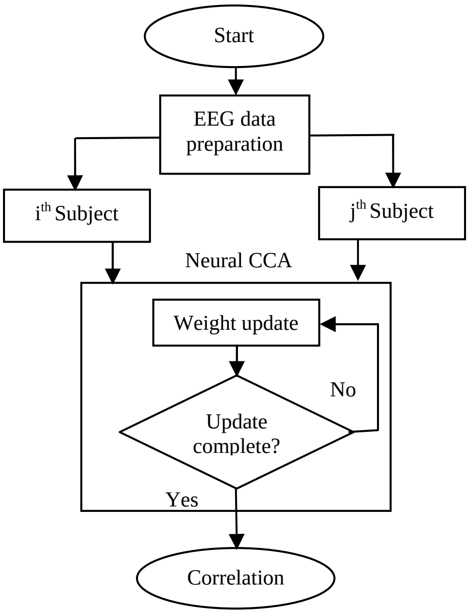


Fig.1: Overview of the methodology for inconsistency measures between subjects.

A. Characteristics of collected EEG data

The high dimensional EEG dataset are collected from SSVEP database (EEG) [9]. For the clarity, descriptions of dataset are presented here as shown in the site. SSVEP acquisition was performed with 128 active electrodes (patterns) at a sampling rate of 2048Hz. Four healthy subjects with normal or corrected-to-normal vision were participated in this study and there have no any neurological disorders. There were used small reversing black and white checkerboards with dimensions of $1.8^o \times 1.8^o$ arc and 6×6 checks for SSVEP stimulation. Photo sensitive epilepsy also tested for the subjects before each experiment.

To cover each of the three SSVEP response regions a single small checkerboard stimulus was displayed for three frequencies sequentially (8, 14 and 28Hz) [10]. At the time of data collection, subjects were seated 0.9m from a 21inch CRT computer display operated at a high vertical refresh rate. At every stimulus frequency there were 5 trials for each subject. Each subject had total 15 trials at 3 different frequencies. Therefore, for four subjects a total of 60 trials were found at 8, 14 and 28Hz frequencies in the database. Where one trial had 128 rows (channels) with more than 6000 columns (attributes) and almost above 30,000 attributes for a subject. When these high dimensional data are analyzed with statistical CCA it needs high computational cost. So, we want to avoid this problem using neural CCA where data are entered sequentially rather than at once.

B. Neural CCA network

CCA can be explored to expose the underlying correlation between two sets of data which is a multivariable statistical method [11]. We incorporate linear and nonlinear into neural CCA network accordingly to realize intra and inter subject inconsistency. Consider $\mathbf{x_1}$ and $\mathbf{x_2}$ as two sets of EEG signals. Now we try to find the linear combination of the signals that gives us maximum correlation between the combinations as described in Fig. 2. Let

$$\mathbf{y_1} = \mathbf{w_1 x_1} = \textstyle\sum_j w_{1j} x_{1j} \qquad (1)$$

$$\mathbf{y_2} = \mathbf{w_2 x_2} = \textstyle\sum_j w_{2j} x_{2j} \qquad (2)$$

Where j is the number of attribute in every sample, there were total 128 samples for 128 electrodes. Now we search those values of $\mathbf{w_1}$ and $\mathbf{w_2}$ that maximize as a correlation profile between y_1 and y_2.

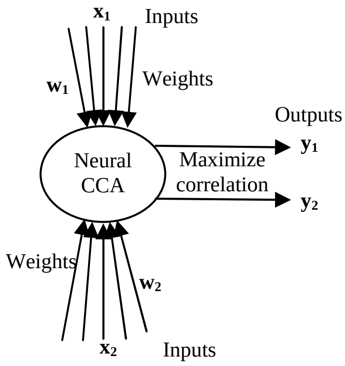


Fig. 2: The CCA network for correlation determination between two sets of EEG data.

Fed forward activation is used for each input to the corresponding output through the respective weights, $\mathbf{w_1}$and $\mathbf{w_2}$. The following joint learning rules [3] are utilized in here.

$$\Delta w_{1j} = \eta x_{1j}(y_2 - \lambda_1 y_1) \qquad (3)$$

$$\Delta \lambda_1 = \eta_0(1 - y_1^2) \qquad (4)$$

$$\Delta w_{2j} = \eta x_{2j}(y_1 - \lambda_2 y_2) \qquad (5)$$

$$\Delta \lambda_2 = \eta_0(1 - y_2^2) \qquad (6)$$

Where w_{1j} is the jth element of weight vector, $\mathbf{w_1}$ etc, λ_1 and λ_2 are Lagrange multipliers. To find representative result $\eta_0 = 0.5$ and $\eta = .001$ are chosen.

To search the nonlinearity between different subjects, various trials and with sine-cosine reference signals nonlinearity is incorporated with linear combinations. When $\mathbf{x_1}$ and $\mathbf{x_2}$ are two sets of signals then maximum nonlinear correlation is found as described in Fig. 2, let

$$\mathbf{y_1} = \mathbf{w_1 f_1} = \textstyle\sum_j w_{1j} \tanh(v_{1j} x_{1j}) \qquad (7)$$

$$\mathbf{y_2} = \mathbf{w_2 f_2} = \textstyle\sum_j w_{2j} \tanh(v_{2j} x_{2j}) \qquad (8)$$

The joint learning rules for nonlinear correlation are

$$\Delta w_{1j} = \eta f_1 (y_2 \quad \lambda_1 y_1) \qquad (9)$$

$$\Delta v_{1j} - \eta x_{1j} w_{1j}(y_2 - \lambda_1 y_1)(1 - f_1^2) \qquad (10)$$

$$\Delta \lambda_1 = \eta_0(1 - y_1^2) \qquad (11)$$

$$\Delta w_{2j} = \eta f_2(y_1 - \lambda_2 y_2) \qquad (12)$$

$$\Delta v_{2j} = \eta x_{2j} w_{2j}(y_2 - \lambda_1 y_1)(1 - f_2^2) \qquad (13)$$

$$\Delta \lambda_2 = \eta_0(1 - y_2^2) \qquad (14)$$

To evaluate the relationship between two time series, correlation computation plays important role. Experiments were done using time series of EEG signals of different subjects and sine-cosine reference signals with different harmonics. When correlation is high then they are greatly matched signals. As linear correlation is greater than nonlinear correlation then they lies in the linear regions otherwise they lies in the nonlinear regions. We will test this region on the basis of correlation measures.

C. Sine-cosine reference signals

To determine the inconsistency of EEG signals we use a sine-cosine signal set as an input with other EEG signals. These measures how linearly or nonlinearly match an EEG signal with known reference signal. The reference signals are constructed with fundamentals as well as harmonics of sine-cosine waves at the k^{th} stimulus frequency f_k (k = 1, 2… K) as follows [12].

$$Y_k = \begin{pmatrix} \sin(2\pi f_k 1/f_s) \dots\dots\dots \sin(2\pi f_k J/f_s) \\ \cos(2\pi f_k 1/f_s) \dots\dots\dots \cos(2\pi f_k J/f_s) \\ \vdots \\ \sin(2\pi H f_k 1/f_s) \dots\dots\dots \sin(2\pi H f_k J/f_s) \\ \cos(2\pi H f_k 1/f_s) \dots\dots\dots \cos(2\pi H f_k J/f_s) \end{pmatrix} \quad (15)$$

Where, H, J and f_s denote the number of used harmonics, sampling points and sampling rate respectively.

D. Inconsistency measure

The inconsistencies of EEG signals are recognized on the basis of linear and nonlinear function of CCA. The correlation is very important in order to assess the relationship between two time series. The experiment was completed using time series of two different subjects, different trials, trails with reference signal and also between reference signals and subject.

A1: CCA algorithm for inter subject inconsistency measure

Input: *EEG signals $x_1, x_2, x_3 \dots x_k \in R^{IxJ}$ and sine-cosine signals y_k (k=1, 2… K)$\in R^{2HxJ}$ corresponding to K stimulus frequencies, respectively.*
Output: *Correlation C_k*
for *m=1 to K do*
    Random initialization for $\mathbf{w_1}$ and $\mathbf{w_2}$
    repeat
        Find $\mathbf{w_1}$ and $\mathbf{w_2}$ which maximize correlation between $\mathbf{y_1}$ and $\mathbf{y_2}$ by the CCA
    until *the maximum number of iteration is reached*
    Compute the optimized signals y_1 and y_2
end

Compute correlation C_k

The relationship between y_1 and y_2 is calculated as correlation coefficient Z_k which is given bellow:

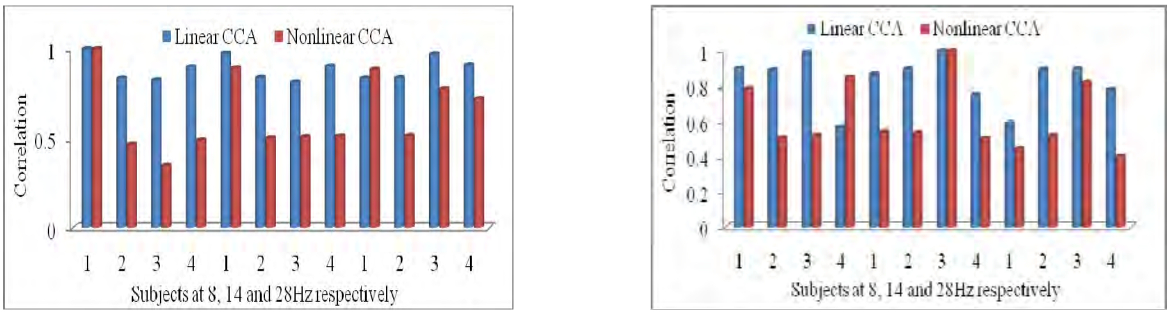
$$Z_m = \sqrt{1 - \frac{||y_1 - y_2||^2}{||y_1 - E[y_1]||^2}} \quad (16)$$

Where, ||.|| denotes norm. Larger C_k implies more significant relationship between $y_1$and y_2.

III. Results and Discussions

Inter and intra subject inconsistencies are analyzed using the neural CCA with MATLAB 7.5. Also discrepancy with reference sine-cosine signal of different subjects is studied. For that reason linearity and nonlinearity are included accordingly. Random weight is generated initially and used to update for final output. Before every update weight is normalized and finally correlation coefficient is found as output. The computation is performed by an Intel(R) Core(TM) i5-2450M CPU with 2.5GHz, 4.00 GB of RAM, Operating system 64-bit computer.

At first S1 (first subject) of 8Hz stimulus frequency is selected as input x_1 and rests of subject are entered in the network as input x_2 accordingly. It is seen that correlation is 100 percent, when same subject at same stimulation is correlated. But it shows different correlation for others. Here it is seen from Fig. 3(a) that linearity is higher than nonlinearity for almost every case.

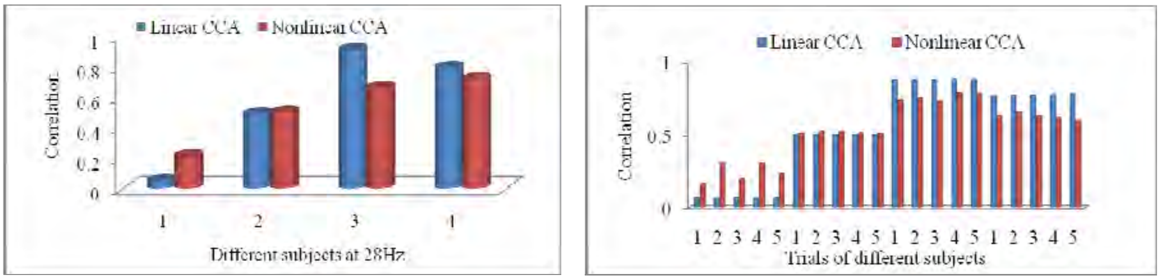


(a) (b)

Fig. 3: Inter subject inconsistency (a) with S1 at 8Hz and (b) with S3 at 14Hz

We also test the inconsistency for every other subject. Fig. 3(b) show the discrepancy of different EEG subjects with S3 (third subject) at 14Hz stimulation. It is seen that 100 percent correlation for same subject at same stimulus frequency. But it is also analyzed that nonlinearity is higher only between S3 at 14Hz and S4 at 8Hz. For other cases linear function shows higher correlation than nonlinear function.

We also measure different subject inconsistency with reference signal. We reserve the same frequency as EEG at stimulus frequency. For total four subjects of 28Hz stimulation, we take the stimulus frequency of reference signal is 28Hz. Fig. 4(a). Show the various subject discrepancies with reference signal as a function of correlation dimensions. It is seen that nonlinearity is higher for S1, but lower for S3 and S4. So we may include that S3 and S4 are the linear function of 28Hz reference signal.



(a) (b)

Fig. 4: Inconsistency with (a) different subject at 28Hz reference and (b) among trials at14Hz reference

Trial to trial inconsistency also calculated for different trials of a subject. We analyze the dependency of different trials with reference signal as function of linear and nonlinear correlations. For 14Hz stimulus frequency it is shown in Fig. 4(b). From this figure it is also exposed that where S1 is the nonlinear function of reference signal but S3 and S4 is the linear function.

Finally we concatenated all subject for a specified stimulation. Then it is tested with reference signal using linear and nonlinear CCA which is shown in Fig. 5. It is seen that all stimulation show higher linearity than nonlinearity for same stimulation which is expected.

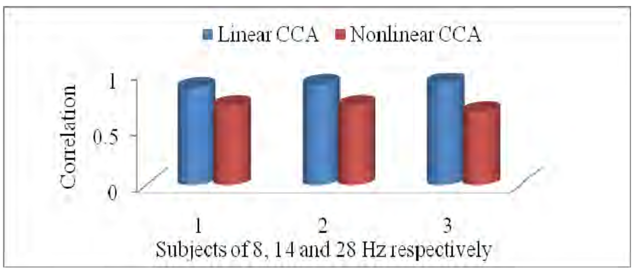


Fig. 5: Linearity test for different stimulus frequency

Though the different subjects of EEG show linear function in between them, they may linear or nonlinear function for reference sine-cosine signal.

IV. CONCLUSION

In this paper, we test the linearity of EEG signals with other signals as a function of correlation dimensions. There were four subjects and five trials for each subject. The analyzing of such high dimensional data takes enough time to execute. When it is seen that these signals are linear or nonlinear and show high correlation, then analyzing of one signal can reflect the whole result. It is seen from this experiment that trial to trial discrepancies are almost negligible. The first subject shows higher nonlinearity, but third and fourth subjects show higher linearity. These tests are done using neural CCA, because statistical standard CCA needs high computational cost than neural CCA.

REFERENCES

[1] D. H. Zhu, J. Bieger, G. G. Molina, R. M. Aarts, "A survey of stimulation methods used in SSVEP-based BCIs," Comput. Intell. And Neurosci,vol. 2010.
[2] E. Oja, "A simplified neuron model as a principle component analyzer". Journal of Mathematical Biology, vol. 16, pp. 267-273, 1982.
[3] P. L. Lai and C. Fyfe, "A neural implementation of canonical correlation analysis," Elsevier on Neural Networks, vol. 12, pp. 1391-1397, 1999.
[4] G. R. MÄuller-Putz, R. Scherer, C. Brauneis, and G. Pfurtscheller, "Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components". J. Neural. Eng. vol. 2, pp. 123-130, 2005.
[5] Y. Zhang, J. Jin, X. Qing and B. Wang, X. Wang, "LASSO based stimulus frequency recognition model for SSVEP BCIs". Biomed. Signal Process. Control, 2011.
[6] M. Cheng, X. Gao, S. Gao and D. Xu, "Design and implementation of a brain-computer interdace with high transfer rates". IEEE Trans. Biomed. Eng, vol. 49, pp. 1181-1186, 2002.
[7] Z. Lin, C. Zhang, W. Wu and X. Gao. "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs". IEEE Trans. Biomed. Eng, vol. 53, pp. 2610-2614, 2006.
[8] H. G. Hakvoort, B. Reuderink and M. Obbink, "Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system". Technical Report TR-CTIT-11-03, EEMCS, pp. 1381-3625, 2011
[9] H. Bakardjiana, T. Tanakaa, A. Cichocki, "Optimization of SSVEP brain responses with application to eight-command Brain–Computer Interface," Neuroscience Letters, vol. 469, pp. 34–38, 2010. Available: http://www.bakardjian.com/work/ssvep_data_Bakardjian.html
[10] D. Regan, "Steady-state evoked potentials", J. Opt. Soc. Am, vol. 67, pp. 1475–1489, 1977.
[11] T. R. Knapp, "Canonical correlation analysis: A general parametric significance-testing system". Psychological Bulletin vol. 85, pp. 410–416, 1978.
[12] Y. Zhang, G. Zhou, Q. Zhao, A. Onishi, J. Jin, X. Wang, and A. Cichocki, "Multiway Canonical Correlation Analysis for Frequency Components Recognition in SSVEP-based BCIs", Lecture Notes in Computer Science, vol. 7062, pp 287-295, 2011.