Proceedings of the 5th International
IEEE EMBS Conference on Neural Engineering
Cancun, Mexico, April 27 - May 1, 2011

Experiments on Using Combined Short Window Bivariate Autoregression for EEG Classification

Tuan Hoang, Dat Tran, Phuoc Nguyen, Xu Huang and Dhamendra Sharma

Abstract— **In EEG-based classification problem, most of currently used features are univariate and extracted from single channels. However EEG signals recorded from multiple channels for a brain activity are correlated, features extracted from the EEG signals should reflect relationships among those channels. For this reason, we propose and apply a bivariate feature called Combined Short-Window BiVariate AutoRegressive model (CSWBVAR) for EEG classification problems. Given a pair of channels, we firstly divide each of them in to overlapping segments or short windows, and then estimate BVAR parameters for each pair of segments. CSWBVAR is formed by combining extracted BVAR parameters together with a pre-defined overlapping window parameter. We analyzed and compared CSWBVAR feature and univariate feature using the dataset III for motor imagery problem of BCI Competition II (2003). Preliminary results show that using CSWBVAR feature can improve classification accuracy up to 7% comparing with using univariate one with the same linear-kernel SVM classifier.**

I. Introduction

Brain-Computer Interface (BCI) is an emerging research field attracting a lot of research efforts from researchers around the world. Its aim is to build a new communication channel that allows a person to send commands to an electronic device using his/her brain activities [2]. A typical BCI system includes the following stages: data acquisition, data pre-processing, feature extraction, classification, device controller and feedback [8]. The performance of a BCI system depends on data pre-processing, feature extraction and classification methods used to build the classifier in that BCI system. This paper focuses on improving feature extraction to enhance the performance of EEG classification problems.

In EEG classification problems, most of features currently used are uni-variate ones which are extracted from single channels. However EEG signals recorded from multiple channels for a brain activity are correlated, features extracted from the EEG signals should reflect relationships among those channels. Some researchers used multivariate autoregressive model to formulate relationship among channels. Schlogl et. al. [14] used multivariate autoregressive (MVAR) parameters to analyze event-related EEG data. From MVAR parameters, they derived several measures including auto and cross spectra, phase relations, coherency, partial coherence, partial directed coherence, direct transfer function, and full frequency direct transfer function. Their work focused on investigating various aspects of multichannel spectral properties of EEG. Nolte et. al. [11] used imaginary part of coherency to find out brain interaction from EEG data when there is movement of subject. They concluded that it is possible to detect brain interaction during movement from EEG data. Anderson et. al. [1] extracted multivariate autoregressive model parameters and then fed them into a multilayer neural network for classification. They used a 256-point window to estimate MVAR parameters. They showed that multivariate feature was the best type of feature in their experiments. Recently, Brunner et. al. [4] conducted experiments to compare various types of autoregressive model features for BCI including univariate, multivariate, bilinear autoregressive (AR) parameters, and logarithmic band power (BP). They evaluated those features on the dataset IIa from BCI Competition 2008 and found that there was no significance between AR and BP features in improving BCI system accuracy. They also concluded that optimizing parameters individually for each subject yielded equally high results as using default parameters. One of the main problems of univariate and multivariate autoregressive models is how to identify a correct order. Akaike information criterion (AIC) [13] which represents a trade-off between estimated error and size of the model is usually used to estimate appropriate order. Beside of autoregressive models, tensor decompositions [12] have recently been used in feature extraction in high dimensional datasets. Also viewing EEG classification problem as matrix manipulation problem and admitting correlation among channels in EEG signal, Christoforos et. al. [5] proposed a method called Second Order Bilinear Discriminant Analysis. They showed that their method can outperform well known ones in single trial classification problem.

We proposed a feature extraction method that uses window as seen in the work of Anderson et. al. [1] and Brunner et. al. [4]. For the signal of a pair of channels, instead of using a single window, we use short moving windows having the same size and then combine them together with a pre-defined overlapping window parameter. We calculate bivariate autoregressive model parameters for the current window and then slide to the next window until end of the signal. All BVAR parameters are concatenated from the first window to the last one with a pre-defined moving window step or overlapping window to form a feature vector. Depending on the nature and well known biological knowledge of mental tasks, we can select several pairs of channels, and concatenate their corresponding feature vectors together to form an entire feature vector of a trial. It is

T. Hoang, D. Tran, P. Nguyen, X. Huang and D. Sharma are with Faculty of Information Sciences and Engineering, University of Canberra, ACT 2601, Australia Tuan.Hoang, Dat.Tran, Phuoc.Nguyen, Xu.Huang, and Dhamendra.Sharma@canberra.edu.au

978-1-4244-4141-9/11/$25.00 ©2011 IEEE

straightforward to see that the trial's feature vector is usually highly dimensional and may contain redundant irrelevant features. To eliminate those irrelevant features, we apply a feature selection algorithm before training a classifier. In this paper, we choose linear support vector machine (SVM) as our baseline classifier since SVM is one of state-of-the-art classifiers and linear-kernel SVM can efficiently deal with high dimensional feature vectors. The data set III for motor imagery problem in BCI Competition II [3] is used to evaluate our proposed method. Experimental results show that our proposed feature extraction method can improve classification accuracy up to 7% comparing with univariate one using the same linear-kernel SVM classifier.

II. Combined Short-Window Bivariate Autoregressive Feature (CSWBVAR)

A. Autoregressive model

Univariate autoregressive (UVAR) model is used to model single time-series with assumption that each value of the series can be estimated by taking a weighted sum of its previous values, plus white noise. Whereas, bivariate autoregressive (BVAR) model is used to model two time-series with assumption that each value of the two series can be estimated by taking a weighted sum of not only the previous values of the same series but also values of the other series.

Let $X(t) = [X_1(t), X_2(t), ..., X_n(t)]^T$ be n time-series in random process. In BCI system, n is number of channels used for collecting brain signals. Let d be the number of data samples of n channels. The pth-order MultiVariate AutoRegressive (MVAR) model is formulated as follows:

$$X(t) = \sum_{i=1}^{p} A_i X(t-i) + E(t) \quad (1)$$

In (1), A_i, $i = 1 \ldots p$ are $n \times n$ coefficient matrices and $E(t)$ is noise vector which is a zero mean uncorrelated with the covariance matrix Σ. We assume that $X(t)$ is a stationary process. If $n = 1$, we have univariate AR model, and $n = 2$, bivariate AR model.

To estimate A_i and Σ, we transfer (1) to Yule-Walker equations by multiplying (1) from the right by $X^T(t-i), i = 1 \ldots p$ and then taking expectation:

$$R(-k) + \sum_{i=1}^{p} A_i R(-k+i) = 0 \quad (2)$$

where $R(l)$ is the covariance matrix of lag l of $X(t)$.

To solve these equations with a specific order p, we use Levinson, Wiggins, Robinson (LWR) algorithm [7]. The correct order is identified by minimizing the Akaike Information Criterion[13] defined as in (3).

$$AIC(p) = 2log[det(\Sigma)] + 2\frac{n^2 p}{N_{total}} \quad (3)$$

where N_{total} is the total number of data points from all trials.

In this paper, after visually inspecting *AIC* curve, we chose the AR model with order 6 which results to a local minimum.

B. CSWBVAR Feature

Given a pair of channels and a time window, we calculate its BVAR coefficients. This sequence will form a feature vector of the window having size of $4 \times p$ coefficients. The feature vector of the trial is defined as the concatenation of these sequences together with a pre-defined moving window step. Let *w* be the window size and *s* be the moving window step, the feature vector size of one trial is equal to

$$4 \times \frac{d - w + 1}{s} \times p \quad (4)$$

Choosing correct window size and moving window step parameters is the most important task in our model. We can assume EEG signals are stationary by setting the window size to a value between 40ms and 100ms. With the sampling rate of 128Hz, each window has from 5 to 12 data points. They are too small to estimate AR model coefficients. Increasing window size requires us to use adaptive algorithm to estimate AR model parameters. Fortunately, LWR algorithm is an adaptive one. In our experimental design, we considered different window sizes including 12, 32, 64, and 128 data points and the overlapping part between two consecutive windows is 25%, 50% and 75% of the window size.

III. Experimental Results

The aim of our experiments is to compare classification results of BCI systems using univariate AR and bivariate AR coefficients as features and explain the results under multichannel viewpoint. We do not design experiments to compare with other classification methods as seen in the BCI competitions. Therefore the chosen data set which has appropriate number of channels was the well-known data set III provided by Department of Medical Informatics, Institute of Biomedical Engineering, Graz University of Technology for motor imagery classification problem in BCI Competition II [3]. In data collection stage, a female normal subject was asked to sit in a relaxing chair with armrests and tried to control a feedback bar by means of imagery left or right hand movements. The sequences of left or right orders are random. The experiment consisted of 7 runs with 40 trials in each run. There were 280 trials in total and each of them lasted 9 seconds of which the first 3 seconds are used for preparation. Collected data was equally divided into two sets for training and testing. The data was recorded in three EEG channels which were C3, Cz and C4, sampled at 128Hz, and filtered between 0.5 Hz and 30 Hz. Most of current algorithms only applied to the channels C3 and C4, and ignored the channel Cz. They argued that from brain theory, signals from channel Cz provide very little meaning to motor imagery problem. We truncated the first 3 seconds of each trial and used the rest for further processing. All trials are pre-processed by subtracting the ensemble mean of all trials.

For each trial we extracted both CSWUVAR and CSWBVAR parameters with different window sizes and moving window steps. We considered window sizes including 12, 32, 64, and 128 data points, corresponding to 100ms-, 250ms-, 500ms-, and 1s-segments. As with other previous work, we

TABLE I

AVERAGE ACCURACY IN PERCENTAGE (%) OVER ALL 3 MOVING WINDOW STEPS AND SEVEN COMBINATIONS OF PAIRS

Window Size	CSWUVAR	CSWBVAR
32	61.0	67.6
64	62.8	69.9
128	67.1	72.3

did not try experiments with segment's size greater than 1s due to keeping signal approximately stationary and being comfortable with nature of brain signal. The overlapping part between two consecutive windows was set to 25%, 50% (except window size 12), and 75% of the window size. From (4), we can see that extracted feature vectors have very large number of elements comparing with number of trials of the training set. This is the reason for choosing linear kernel SVM as the baseline classifier. All feature vector values were scaled to those in the range of $[-1, 1]$ and some irrelevant features were removed. We used 3-fold cross validation in the training phase.

Because the dataset has 3 channels ($C3, Cz$ and $C4$), we have 3 pairs of channels which are $p1(C3 - Cz)$, $p2(C3 - C4)$ and $p3(Cz - C4)$. We considered all possible combinations of these 3 pairs to form the following feature vectors from a trial: $(p1, p2, p3)$, $(p1, p2)$, $(p1, p3)$, $p1$, $(p2, p3)$, $p2$, and $p3$.

A. Performance comparison of CSWUVAR and CSWBVAR

To compare the performance of BCI system using CSWUVAR and CSWBVAR features, we calculated the average classification accuracy of all three moving window steps and all seven combinations of pairs. In total, we performed 21 experiments for each of the above-mentioned window size and each of the two features.

From Table 1 showing the average accuracy for all 21 possible combinations, it is easy to see that CSWBVAR feature outperforms CSWUVAR in all different moving window steps and combinations of pairs. Using CSWBVAR feature improves the accuracy from 5% to 7% comparing with using CSWAVAR. This accuracy improvement proves that extracting features from multi-channel model, especially bi-channel model in this paper, could gain more useful information from the EEG data than from single channels, leading to improving BCI system performance.

B. Affection of parameters to CSWBVAR feature

There are two important parameters that need to be set when extracting CSWBVAR which are window size and moving window step. Figure 1 shows experiments when moving window step is set to 50% of its corresponding window size. From the results, we see that the system usually achieves higher accuracy with window size of 128. The same conclusion is made for other moving window steps. It is proved by taking the average of accuracy of each window size as shown in column CSWBVAR of Table 1. We can say that window size significantly affects to the system

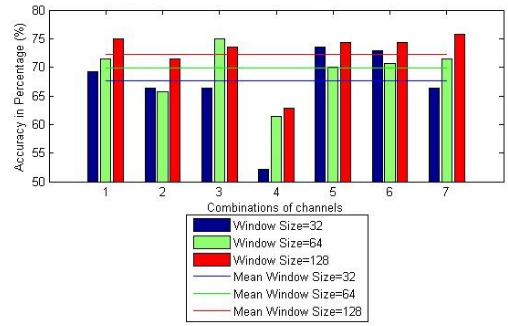

Fig. 1. Classification accuracy in percentage (%) of CSWBVAR feature. Blue bar: Window size = 32; Green bar: Window size = 64; Red bar: Window size = 128. Horizontal lines represent the average accuracies for all of these window sizes, respectively from bottom to top. Window step was set to 50% of the corresponding window size.

TABLE II

AFFECTION OF MOVING WINDOW STEPS ON THE SYSTEM PERFORMANCE. ACCURACIES ARE IN PERCENTAGE (%).

All	Overlap 75%	Overlap 50%	Overlap 25%
67.6	66.3	66.7	69.8
69.9	70.1	69.4	70.1
72.3	73.0	72.4	71.4

performance. However, moving window steps slightly affect to the system performance as seen in Table 2.

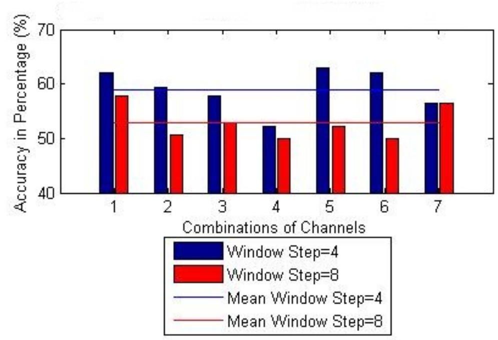

Fig. 2. System performance when window size is set to a small value (12 data points). Blue bar: moving window step = 4; Red bar: moving window step = 8. Horizontal lines represent average accuracy of moving window steps of 4 and 8, respectively from top to bottom. Accuracies are in percentage (%).

Figure 2 shows that with small window size (12 data points), AR parameters are poorly estimated degrading the system performance. Average accuracies of both experiments with window size of 12 are less than 60%. It suggests that using small window size is not a good choice for forming CSWBVAR features.

C. Combination of channels to form CSWBVAR feature

To analyze the mutual support of a channel to each other, we consider 3 experiments in which CSWBVAR features are

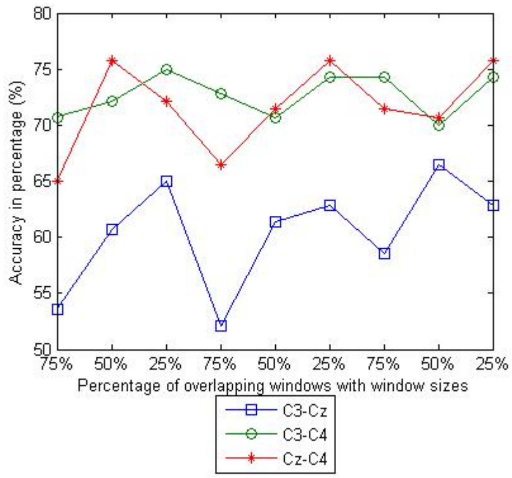

Fig. 3. Comparison of three pairs C3-Cz (blue line), C3-C4 (green line), and Cz-C4 (red line) in classification accuracy in percentage (%) of CSWBVAR feature.

extracted from single pair of channels only: C3-Cz, C3-C4, and Cz-C4. Figure 3 shows accuracies of these features with different moving window steps. We can see that the pair C3-Cz (blue line) always produces the worst performance, whereas the pair C3-C4 seems to produce the most stable and best performance. It reaffirms the judgment that C3 and C4 channels play important role in this motor imagery problem. The channel Cz, alone, although in theory plays less important role than C3 and C4 channels, but when combined with the C4 channel to form a pair from which CSWBVAR features are extracted, it contributes more significant role in improving the system performance.

IV. Conclusions and Future Work

We have presented a technique to form a feature which is called Combined Short Window BiVariate AutoRegressive (CSWBVAR) for BCI system. Experimental results on the data set III of BCI 2003 have shown that the proposed CSWBVAR could improve the classification accuracy up to 7% comparing with similar univariate version. It confirmed our hypothesis that using multi-channel model instead of single channel model, we can extract more useful information from EEG data that could improve BCI system performance. We also analyzed the affection of two parameters which are window size and moving window step on system accuracy. It showed that tuning window size of the model plays a more important role in improving or degrading system accuracy than tuning moving window step of the same one. We also found out good combinations of channels that can explain underlying mechanism of brain system.

One of drawbacks of our proposed CSWBVAR feature is that how we can select appropriate pairs of channels if there are many channels available in EEG data, leading to a very large number of pairs of channels. Our experimental results show that a poor channel in single channel processing can possibly turn to be a good channel in multi-channel processing. Finding a measure to estimate mutual support among channels is a very interesting and challenge problem. We will conduct more experiments on different well-known datasets with different suitable classifiers.

References

[1] C. W. Anderson, E. A. Stolz and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks", *IEEE Trans. Biomed. Eng.*, vol. 45, no.3, pp. 277-286, 1998.

[2] F. Babiloni, A. Cichocki, and S. Gao, "Brain-Computer Interfaces: Towards Practical Implementations and Potential Applications", *Special Issues in Computational Intelligence and Neuroscience*, vol. 2007, Article ID 62637, 2 pages, 2007.

[3] BCI Competition II, http://www.bbci.de/competition/ii/

[4] C. Brunner, M. Billinger, and C. Neuper, "A Comparison of Univariate, Multivariate, Bilinear Autoregressive, and Bandpower Features for Brain-Computer Interfaces", Fourth International BCI Meeting, Asilomar, CA, USA 2010.

[5] C. Christoforou, R. Haralick, P. Sajda, and L. C. Parra, "Second-Order Bilinear Discriminant Analysis, Journal of Machine Learning Research", vol. 11, pp. 665-685, 2010.

[6] A. Cichocki, Y. Washizawa, T. Rutkowski, H. Bakardjian, A. Phan, S. Choi, H. Lee, Q. Zhao, L. Zhang, and Y. Li, "Noninvasive BCIs: Multiway Signal-Processing Array Decompositions", *Computer*, vol. 41, no. 10, pp. 34-42, 2008.

[7] S. Haykin and S. Kesler, "Prediction error filtering and maximum entropy spectral estimation", *Non-Linear Methods of Spectral Analysis, in Topics in Applied Physics*, Springer-Verlag, ch. 2, pp.9-72, 1983.

[8] R. Krepki, G. Curio, B. Blankertz, and K.-R. Muller, "Berlin Brain-Computer Interface–The HCI communication channel for discovery", *International Journal of Human-Computer Studies*, vol. 65, issue 5, pp. 460-477, 2007.

[9] F. Lotte, "PhD thesis: Study of Electroencephalographic Signal Processing and Classification Techniques towards the use of Brain-Computer Interfaces in Virtual Reality Applications", 2008.

[10] M. M. Melody, "Real-World Applications for Brain-Computer Interface Technology", *IEEE Trans. Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 162-167, 2003.

[11] G. Nolte, U. Bai, L. Weathon, Z. Mari, S. Vorbach, and M. Hallet, "Identifying true brain interaction from EEG data using the imaginary part of coherency", *Neurophysiology*, vol. 115, pp. 2294-2307, 2004.

[12] A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets", *Nonlinear Theory and Its Applications*, IEICE (invited paper) October 2010 (in print).

[13] S. Sanei and J. A. Chambers, "EEG Signal Processing", 1st. Edition, ch. 2, pp. 42, John Wiley and Sons, 2007.

[14] A. Schlogl and G. Supp, "Analyzing event-related EEG data with multivariate autoregressive parameters", (Eds.) C. Neuper and W. Klimesch, *Progress in Brain Research: Event-related Dynamics of Brain Oscillations. Analysis of dynamics of brain oscillations: methodological advances. Progress in Brain Research*, vol. 159, pp. 135-147, 2006.