Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Creating a Nonparametric Brain-Computer Interface with Neural Time-Series Prediction Preprocessing

Damien Coyle, *Member, IEEE,* Thomas M. McGinnity, *Member, IEEE* and Girijesh Prasad, *MIEEE*

Abstract—**The issue of subject-specific parameter selection in an electroencephalogram (EEG)-based brain-computer interface (BCI) is tackled in this paper. Hjorth- and Barlow-based feature extraction procedures (FEPs) are investigated along with linear discriminant analysis (LDA) for classification. These are well-known nonparametric FEPs but their simplicity prevents them from matching the performance of more complex FEPs. Neural time-series prediction preprocessing (NTSPP) has been shown to enhance the separability of both time- and frequency-based features and is used in this work to improve the applicability of these FEPs. NTSPP uses a number of prediction modules (PMs) to perform *m*-step ahead prediction of EEG time-series recorded whilst subjects perform motor imagery-based mental tasks. Depending on the PMs, the NTSPP framework normally requires subject-specific parameters to be predefined. In this work each PM is a self-organizing fuzzy neural network (SOFNN). The SOFNN has a self-organizing structure and good nonlinear approximation capabilities however; a number of parameters must be defined prior to training. This is problematic therefore the practicality of a general set of parameters, previously selected via a sensitivity analysis (SA), is analyzed. The results indicate that a general set of NTSPP parameters may provide the best results and therefore a fully nonparametric BCI may be realizable.**

I. INTRODUCTION

PEOPLE suffering from severe neuromuscular disorders have, in many cases, severe difficulty in communicating. The electroencephalogram (EEG) provides an alternative communication pathway from the brain to a computer via a EEG-based brain–computer interface (BCI). A BCI requires advanced signal processing and pattern recognition techniques to extract reliable information from EEG. This information may be used for an alternative communication pathway which may be beneficial for disabled people [1][2][3]. The signal processing components are extremely important for the proficiency of a BCI. BCI research has progressed however; current state-of-the-art methods have a number of shortcomings such as low accuracy and speed, poor reliability as well as usability issues. Ideally, a desirable outcome for BCI research is a BCI which can self-organize and perform online autonomous adaptation to the perpetually varying, multifarious dynamics of each individual's EEG, so as to obtain, and subsequently maintain, fast and accurate communication [3][4][5].

This paper aims to build on previous work [6] to address the problem of self-organization and efficient algorithm adaptation for each subject. A framework for EEG preprocessing, referred to as neural time-series prediction preprocessing (NTSPP), has been recently proposed [3][4][5]. The NTSPP framework involves training prediction modules (PMs) to predict different types of EEG time-series data *m* steps ahead. Subsequently, features can be extracted from the predicted signals (Ys signals). Due to each PM's specialization on the type of data on which it was trained, feature separability and inter-session generalization is improved by NTSPP. In [5] the self-organizing fuzzy neural network (SOFNN) [6][7] was applied in the NTSPP framework and, although the results were similar to other PMs, it was very difficult to specify the SOFNN parameters so that the best network could be attained. If the predefined parameters are selected too stringently, too many neurons may be added thus over fitting may occur and training times are increased. This is problematic but the advantages of the SOFNN are very important. In real-world problems, even if assuming conditions are consistently maintained, signals recorded repetitively are often inconsistent, especially in the case of EEG therefore an algorithm which can autonomously adapt, such as the SOFNN, is desirable. However, to utilize the SOFNN more efficiently it was necessary to improve the transparency of its predefined parameters [5]. Computational enhancements and a sensitivity analysis (SA) are outlined in [6], where parameter choices that could enhance the SOFNNs applicability to the NTSPP framework were identified. The paper aims to demonstrate that the parameters chosen through the SA work well in a practical BCI based on NTSPP and simplistic, nonparametric FEPs and classification methods.

II. METHODOLOGY

A. Data acquisition

The data used in this work was recorded from 3 healthy subjects (aged 20-40) in a timed recording procedure (S1–S3) [2][8]. All signals were sampled at 125Hz and filtered between 0.5 and 30Hz. Two bipolar EEG channels were measured using two electrodes positioned at C3 and C4 (10/20 system). Data were recorded during virtual reality and gaming experiments, where the task was to control and perform decisions in various types of virtual environments or control a ball falling from the top to the bottom of the PC

Manuscript received April 24, 2006. The authors are with the Intelligent Systems Engineering Laboratory, School of Computing and Intelligent Systems, Faculty of Engineering, Magee Campus, University of Ulster, Northland Road, Derry, Northern Ireland, BT48 7JL, UK. (phone: +44 (0)28 7137 5170; fax: +44 (0)28 7137 5470; email: dh.coyle@ulster.ac.uk)

1-4244-0033-3/06/$20.00 ©2006 IEEE.

monitor using motor imagery [8]. The timing of the recording paradigm and feedback duration was similar for all subjects. The data sets for subjects S1-S3 consist of, respectively, 476, 1080, and 1080 trials. Each trial lasts ~10s, of which 4-5 seconds is event-related data.

B. Data Configuration

There are four different types of time-series data recorded during the previously described experiment i.e., left/right (C3/C4) – referred to as l3, l4, and r3, r4. For prediction the recorded EEG time-series data is structured so that the signal measurements from sample indices t to t-(L-1)τ are used to make a prediction of the signal at sample index $t+\tau+m$. The parameter L is the embedding dimension and

$$\hat{cx}_{t+\tau+m} = f(cx_t, ..., cx_{t-(L-1)\tau}) \tag{1}$$

where τ is the time delay, m is the prediction horizon (minus τ), f( ) is the SOFNN model, cx is the signal (i.e., l3, r3, l4, r4) and $\hat{cx}$ is the predicted signal. Even though a more extensive analysis was carried out in [6], results in this work involve τ ranging between 1 and 2 and m=0 or 49.

C. Neural Time Series Prediction Preprocessing

Four SOFNNs are used to perform prediction – two for left motor imagery data and two for right motor imagery data (L3, L4, R3, R4) (cf. Fig. 1). Details and analysis of the NTSPP framework using different types of PMs in conjunction with various types of feature extraction procedures (FEPs) are presented in [3][4][5]. The SOFNN can cope with EEG characteristics such as large dimensions, non-stationarity and noise contamination. The SOFNN is based on a hybrid recursive least squares estimator (RLSE), and an autonomous neuron adding and pruning structure, based on the optimal brain surgeon technique [6][7].

A problem with the SOFNN is that there are five parameters which must be chosen prior to training. These are the error tolerance (δ), the desired training RMSE (k_{rmse}), the distance threshold vector (k_d), the initial width vector (σ_0) and the error tolerance scalar (λ). These parameters govern the performance and complexity of the network [6]. The problem is exemplified if each SOFNN in the NTSPP framework is considered individually – resulting in 20 predefined parameters. The SA [6] was undertaken to determine if a general set of SOFNN parameters, which would work well for each signal across all subjects, could be defined. The analysis was limited to 3 subjects. It was observed that the average value of each parameter that achieved the lowest average prediction error is different for each signal type, indicating that each signal type has different dynamics. However, in the majority of cases, for all parameters the standard deviations overlap, indicating that the mean value of each parameter for all signal types may be a reasonable choice for each predefined parameter.

The analysis presented in this paper is to verify the practicality of the parameters selected in the SA [6]. To do

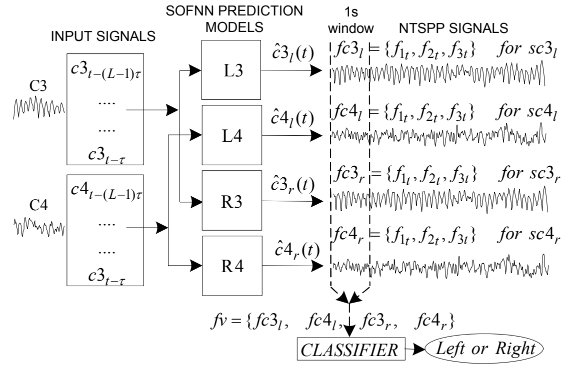

Fig. 1. Illustration of the NTSPP Framework and FEPs

TABLE 1: SELECTIONS FOR PREDEFINED SOFNN PARAMETERS

Para	δ	k_{rmse}	k_d	σ_0	λ
S	0.009	0.005	0.15	0.1	0.80
SA	0.100	0.070	0.57	0.45	0.75
NS	0.180	0.040	0.55	0.3	1.05

this, tests of a BCI involving NTSPP were carried out using SOFNNs trained with stringently selected parameters (S), non-stringently selected parameters (NS) and the parameters selected through the SA. (cf. Table 1). The amount of training data used to train the SOFNNs also has a significant effect on networks therefore, for each parameter combination, the SOFNNs were trained using a 4s segment of event-related data drawn from one randomly chosen trial for the first set of tests and from ten different trials for the second set of tests. When all networks are trained the features are extracted from both the original signals (Os) and the Ys signals (cf. Section II.D for FEPs). The main objective of NTSPP is to firstly improve feature separability and thus classification accuracy (CA) when compared to the features extracted from the Os signals. Secondly, when applying a multiple-step-ahead prediction the objective is not only to improve the CA but to reduce classification time (CT) by predicting the point of maximum separability prior to its occurrence and thus enhancing feature separability earlier than would be possible using the Os signals only. The final objective in this investigation was to determine if the parameters selected from the SA produced the best results.

D. Hjorth- and Barlow-based Feature Extraction

Hjorth introduced a method which involves using time domain information from the signal for EEG analysis [9][10]. The Hjorth method produces 3 features which are:

1. *Activity* which is the mean power in the time domain (i.e., variance) and is defined as

$$Activity = A^2 = \tfrac{1}{T}\int_{t-T}^{t} x(t)^2 dt. \tag{2}$$

2. *Mobility* which is obtained by firstly calculating the variance (standard deviation squared) of the slope

$$D^2 = \tfrac{1}{T}\int_{t-T}^{t} \left({dx(t)}/{dt} \right)^2 dt, \tag{3}$$

then

$$Mobility = \sqrt{D^2 \big/ Activity} = \sqrt{D^2 \big/ A^2}. \tag{4}$$

Mobility is thus influenced by the curve shape of the signal. It is the relative average slope and the dimension is cycles/s. If the unit is $\frac{1}{2\pi s}$ Mobility is the mean frequency.

3. *Complexity* is the Mobility of *dx(t)/dt* divided by the Mobility of *x(t)*.

$$Mobility\left(\frac{dx(t)}{dt}\right) = \sqrt{\frac{\frac{1}{T}\int_{t-T}^{t}\left(dx^2(t)\big/dt^2\right)^2 dt}{\frac{1}{T}\int_{t-T}^{t}\left(dx(t)\big/dt\right)^2 dt}}, \tag{5}$$

$$Complexity = \frac{Mobility\left(\frac{dx(t)}{dt}\right)}{Mobility\left(x(t)\right)}. \tag{6}$$

Complexity indicates the deviation of the slope and it can be seen as a measure of change in frequency of the input signal. If *x(t)* is a sine function the complexity is unity and as the signal deviates from a sinusoid the complexity increases.

Barlow-based feature extraction is somewhat similar to Hjorth-based feature extraction where again there are three features extracted from each signal [10][11]. These are:

1. *Mean Amplitude* (MA)

$$MA = \frac{1}{T}\int_{t-T}^{t}\left|x(t)\right|dt = E\left|x(t)\right| \tag{7}$$

2. *Mean Frequency* (MF) which is the mean frequency in the signal under analysis and is defined using

$$B = \frac{1}{T}\int_{t-T}^{t}\left|dx(t)\big/dt\right|dt = E\left|dx(t)\big/dt\right| \tag{8}$$

and

$$MF = \frac{B}{MA} = \frac{E\left|dx(t)\big/dt\right|}{E\left|x(t)\right|}. \tag{9}$$

This is the ratio of the running mean absolute slope to the running mean absolute amplitude.

3. *Spectral Purity Index* (SPI) is a measure of the irregularity in the signal where the squared running mean of the absolute slope is defined as

$$C = \left(E\left|dx(t)\big/dt\right|\right)^2 \tag{10}$$

and SPI is the ratio of C to the product of the absolute amplitude $E|x(t)|$ and the running mean absolute curvature

$$SPI = \frac{\left(E\left|dx(t)\big/dt\right|\right)^2}{E\left|d^2x(t)\big/dt^2\right|E\left|x(t)\right|}. \tag{11}$$

SPI will be less than 1 if a band of frequencies are present within the signal. If the signal is sinusoidal then SPI is equal to 1. The advantages of using a Barlow-based FEP are similar to those using the Hjorth i.e., frequency information can be obtained without iterating a complex function with multiple parameters. The feature extraction window width is the only parameter for these FEPs and was set to 1s. Features are extracted from each signal at the rate of the sampling interval. For Hjorth or Barlow the feature *fv* has dimension of 6 (2 signals, 3 features per signal) for Os signals, whereas the *fv* using Ys signals has a dimension of 12 (4 signals). The signal dynamics which are extracted by the FEPs are influenced by NTSPP thus it was anticipated that NTSPP would perform well with these FEPs.

The datasets for each subject was split into 3 sessions. A 5-fold cross-validation was carried out on session 1 for each subject, where the data was partitioned into a training set (80%) and a validation set (20%). Tests were performed five times using a different validation partition each time. The mean-CA (mCA) rates on the 5-folds of validation data and 95% confidence intervals (ci) were estimated. Subsequently, all session 1 data was utilized to train the system and the classifier was set up on the features which produced the highest mCA rate on session 1. The systems generalization abilities were then tested on a one-pass test on session 2. The best results were used to determine the best setup to use for a test on data from session 3.

III. Results & Discussion

The results are presented in Table 2 (details are presented in the caption). The light shaded markers (in the CA columns) indicate which signals, out of the Os or the Ys signals with either a short or long prediction horizon (m≥50) (see column 5), provided the best CA for each subject for each feature type. Considering one best CA result for all feature types and subjects (sessions 2 and 3) (i.e., the marked rows in the CA columns), 75% of the best results were obtained with Ys signals with m≥50, ~17% of the best results were obtained using Ys signals with m=1 or 2, and 8% of the best results were obtained with Os signals. Both Hjorth and Barlow features produced the best results on 50% of these cases. The CA is much lower than desired due to the simplicity of these FEPs (>90% CA is desirable for a BCI). However, it is clearly evident that the NTSPP framework improves feature separability and overall system performance which was the first objective of this analysis. More advanced FEPs utilized along with NTSPP have been shown to CA results between 90 and 100% but those

TABLE 2: RESULTS FROM OS AND YS SIGNALS USING HJORTH AND BARLOW FEPS. COLUMNS 6 AND 7 SPECIFY THE MCA ± 95%CI AND MEAN MUTUAL INFORMATION (MMI) [2] OBTAINED FOR SESSION 1. COLUMNS 8-11,(12-15), RESPECTIVELY, SPECIFY CA, CT, IT [12] AND MI FOR SESSIONS 2 AND 3.

					Session 1		Session 2				Session3			
Sub	**F**	**Sig**	**P**	***τ(m)***	**mCA%±ci**	**mMI**	**CA%**	**CT[s]**	**IT[bpm]**	**MI[bits]**	**CA%**	**CT[s]**	**IT[bpm]**	**MI[bits]**
S1	H	Os	*na*	*na*	79.96±6.46	0.23	80.15	3.99	4.22	0.13	80.26	3.34	5.08	0.2
		Ys	S_1	1(1)	80.21±5.0	0.22	85.29	3.66	6.52	0	77.63	3.2	4.37	0.13
			SA_{10}	1(50)	80.70±7.94	0.24	85.29	3.51	6.79	0.03	78.29	2.84	5.18	0.04
	B	Os	*na*	*na*	84.29±3.89	0.38	83.82	3.62	5.98	0.64	78.95	3.21	4.82	0.25
		Ys	SA_{10}	1(1)	82.42±7.41	0.41	86.03	3.75	6.66	0.23	80.26	2.88	5.9	0.2
			NS_{10}	1(1)	81.04±6.0	0.37	88.24	2.47	11.59	0.44	82.89	3.82	5.33	0.25
S2	H	Os	*na*	*na*	76.48±2.0	0.25	78.15	4.71	3.09	0.31	77.04	3.34	4	0.2
		Ys	SA_1	2(2)	77.96±4.3	0.29	81.85	3.32	5.72	0.35	78.7	2.87	5.28	0.25
			SA_{10}	1(50)	80.56±8.1	0.33	88.15	1.89	15.09	0.18	80.74	2.24	7.85	0.28
	B	Os	*na*	*na*	77.78±7.2	0.26	83.33	3.29	6.39	0.38	76.11	3.89	3.19	0.21
		Ys	SA_{10}	2(2)	80.19±3.4	0.36	83.70	3.03	7.1	0.38	80.19	2.21	7.66	0.32
			NS_{10}	2(51)	78.89±3.0	0.30	86.30	1.95	13.02	0.45	82.96	2.32	8.83	0.38
S3	H	Os	*na*	*na*	75.00±5.1	0.19	78.89	4.06	3.79	0.63	75.93	4.02	3.04	0.22
		Ys	SA_1	1(1)	78.33±4.4	0.21	82.96	3.64	5.63	0.24	81.3	4.02	4.54	0.25
			SA_1	1(50)	75.93±8.3	0.24	82.59	4.06	4.92	0.32	78.52	4.06	3.69	0.32
	B	Os	*na*	*na*	73.52±3.2	0.19	78.89	3.99	3.85	0.12	75.93	4.04	3.03	0.21
		Ys	SA_1	1(1)	76.11±6.5	0.21	78.15	3.54	4.11	0.18	75.74	4.02	2.99	0.28
			SA_1	1(50)	74.44±6.7	0.23	79.26	4.3	3.68	0.3	78.15	4.2	3.46	0.29

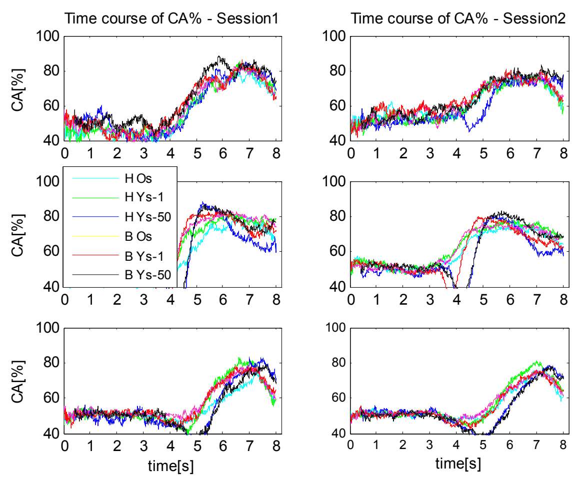

Fig. 2. Time courses of CA% for Os and Ys (sessions 1 & 2)

FEPs require subject-specific parameter selection and are thus not as apposite to an autonomous BCI. Improved CA is not the only advantage of the NTSPP framework. Using Barlow features and Ys signals with $m \geq 50$, on session 2 the CT is reduced resulting in a significantly higher information transfer rate [12]. Fig. 2 shows the time course of the CA (or separability) from the beginning to the end of the trial for each subject. A steeper rise in CA and a higher peak indicates that multiple step-ahead NTSPP can improve IT rate. From Table 2 the P column shows that 9 out of 12 of the best results were obtained using the SA parameters [6] which indicates that a general set may be chosen to obtain the best results (to verify statistical significance an analysis of more subjects would be required). If it was the case that a general set of parameters could be selected, then this BCI based on an SOFNN-based NTSPP framework which adapts its structure autonomously to suit each individual, used in conjunction with Hjorth- or Barlow-based feature extraction along with an LDA classifier, would indeed be a fully nonparametric BCI. The NTSPP framework is the only framework which enables features to be extracted from signals predicted multiple steps ahead. The approximation power of the SOFNN is advantageous in this respect. Further work will involve enhancing NTSPP by training the SOFNNs simultaneously with an objective function which may produce networks that influence, more so, the separable signal characteristics extracted by the FEPs.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, "Brain-computer interfaces for communication and control", *J. Clinical Neurophysiology*, vol. 113, pp. 767-791, 2002.

[2] A. Schlogl, C. Keinrath, R. Scherer, G. Pfurtscheller, "Estimating the mutual information of an EEG-based brain computer interface", *Biomed. Technik*, vol 47, pp. 03-08, 2002.

[3] D. Coyle, G. Prasad, and T.M. McGinnity. "A time-series prediction approach for feature extraction in a BCI", *IEEE Trans. on Neur. Syst. and Rehab. Eng*, vol. 13, no. 4, pp. 461-467, Dec. 2005.

[4] D. Coyle, G. Prasad, and T.M. McGinnity, "Improving signal separability and inter-session stability of a motor imagery-based BCI by neural-time-series-prediction-preprocessing". *Proc. of the 27th Int. IEEE Eng. in Med. and Bio. Conf,* pp. 2294-2299, Sept. 2005.

[5] D. Coyle, G. Prasad, and T.M. McGinnity, "Improving information transfer rates of a BCI by self-organising fuzzy neural network-based multi-step-ahead time-series prediction", *Proc. of the 3rd IEEE Syst, Man and Cybern, (UK&RI Chapter) conf.*, pp. 230-235, Sept. 2004.

[6] D. Coyle, G. Prasad and T.M. McGinnity, "Enhancing autonomy and computational efficiency of the SOFNN for a BCI", *IEEE World Congress on Computational Intelligence,* 2006. (to be published)

[7] G. Leng, G. Prasad, T.M. McGinnity, "An on-line algorithm for creating self-organising fuzzy neural networks", *Neural Networks*, vol. 17, issue 10, pp 1477-1493, 2004.

[8] Graz dataset IIIb for the BCI 2005 data classification competition. http://ida.first.fhg.de/projects/bci/competition_iii/desc_IIIb.pdf

[9] B. Hjorth, "EEG analysis based on time-domain properties," *Electroencephal. and C. Neurophysiol.*, vol. 29, pp. 306-310, 1970.

[10] C. Guger, *Real-time Data Processing Under Windows for an EEG-based Brain-Computer Interface*, PhD Thesis, Technical University Graz, Austria, Sept. 1999.

[11] I. I. Goncharova, and J. S. Barlow, "Changes in EEG mean frequency and spectral purity during spontaneous alpha blocking," *Electroencephal. and C. Neurophysiol*, vol. 76, pp. 197-204, 1990.

[12] J.R. Wolpaw, H. Ramouser, D.J. McFarland, and G. Pfurtscheller, "EEG-based communication: Improved accuracy by response verification", *IEEE Trans. Rehab. Eng.* vol. 6. pp. 326-333, 1998.