INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF NEURAL ENGINEERING

J. Neural Eng. **1** (2004) 212–217 PII: S1741-2560(04)77616-8

A new approach in the BCI research based on fractal dimension as feature and Adaboost as classifier

Reza Boostani and Mohammad Hassan Moradi

Amir Kabir University of Technology, Faculty of Biomedical Engineering, Tehran, Iran

Received 13 March 2004
Accepted for publication 6 October 2004
Published 17 November 2004
Online at stacks.iop.org/JNE/1/212
doi:10.1088/1741-2560/1/4/004

Abstract

High rate classification of imagery tasks is still one of the hot topics among the brain computer interface (BCI) groups. In order to improve this rate, a new approach based on fractal dimension as feature and Adaboost as classifier is presented for five subjects in this paper. To have a comparison, features such as band power, Hjorth parameters along with LDA classifier have been taken into account. Fractal dimension as a feature with Adaboost and LDA can be considered as alternative combinations for BCI applications.

1. Introduction

In order to provide alternative communication channels for patients with locked-in-syndrome and for patients with severe spinal cord injury, different brain computer interface (BCI) systems have been designed [1, 2]. A vision for the future is that a BCI might someday help a tetraplegic patient to move his limbs by functional electrical stimulation controlled with thoughts. In the meantime we have to use non-invasive methods for studying the feasibility of the EEG as input signals for a BCI and to optimize pre-processing, feature extraction and feature classification methods.

In this way, the Graz-BCI team have employed different features such as band power [3], adaptive autoregressive coefficients [4] and some classifiers including LDA, FIRMLP [5], LVQ [6] and HMM [7] to improve the classification rate between different motor imagery tasks. They have also used DSLVQ and G.A. [8] for feature selection. Deriche and Al-Ani [9] selected the best feature combination among the variance, AR coefficients, wavelet coefficients, and fractal dimension by a modified mutual information method and classified them by a MLP classifier. Cincotti *et al* [10] compared the performance of three classifiers (ANN, Mahalonobis distance and HMM) to classify the band power feature for six subjects. The aim of this paper is to improve the classification rate of a cue-based BCI by a new approach based on fractal dimension and Adaboost classifier. As a comparison, band power and Hjorth parameters along with LDA are assessed on the same data set (five subjects).

2. Subjects and data acquisition

Five healthy subjects (L1, o3, k3, f8 and o8), all familiar with the Graz-BCI, participated in this study. Each subject sat in a relaxing chair with armrests about 1.5 m in front of the computer screen. Three bipolar EEG-channels were recorded from 6 Ag/AgCl electrodes placed 2.5 cm anterior and 2.5 cm posterior to the standardized positions C3, Cz and C4 (international 10–20 system).

The EEG was filtered between 0.5 and 70 Hz and recorded with a sample frequency of 128 Hz. The training consisted of a repetitive process of triggered movement imagery trials. Each trial lasted 8 s and started with the presentation of a blank screen. A short acoustical warning tone was presented at second 2 and a fixation cross appeared in the middle of the screen. From second 3 to second 7 an arrow (cue), representing the mental task to perform, was prompted. An arrow pointing either to the left or to the right indicated the imagination of a left hand or right hand movement. The order of appearance of the arrows was randomized and at second 7 the screen content was erased. The trial finished with the presentation of a randomly selected inter-trial period (up to 2 s) beginning at second 8. Figure 1 shows the timing scheme. Three sessions were recorded for each subject on three different days. Each session consisted of three runs with 40 trials each.

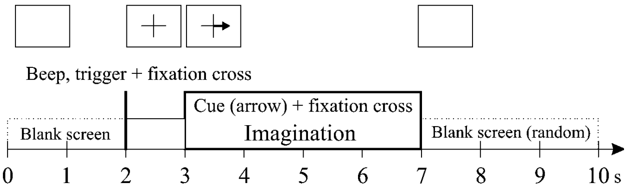

Figure 1. Training paradigm.

3. Feature extraction and classification

The goal of feature extraction is to find a suitable representative (signal features) of the data that simplify the subsequent classification or detection of brain patterns. Band power, Hjorth parameters and fractal dimension were employed as discriminative features which describe the signal in terms of specific frequency bands, morphological characteristics and entropy, respectively.

3.1. Band power (BP)

The EEG contains different specific frequency components, for example, alpha and beta bands which are particularly important to classify the different brain states, especially for discriminating motor imagery tasks. The BP was calculated in two standard frequency bands (10–12 Hz and 16–24 Hz) for 1 s time window and 250 ms overlap [1].

3.2. Hjorth parameters

The Hjorth [12] parameters describe the signal characteristics in terms of activity (variance (VAR) of signal), mobility (a measure of mean frequency) and complexity (a measure of the deviation from sine shape) which are briefly described as follows:

$$\text{Activity}(y) = \frac{\sum_{i=1}^{N} (y(i) - \mu)^2}{N}$$

$$\text{Mobility} = \sqrt{\frac{\text{VAR}(y')}{\text{VAR}(y)}}$$

$$\text{Complexity} = \frac{\text{Mobility}(y')}{\text{Mobility}(y)}$$

where y is the signal, y' is the derivative of the signal, N is the number of samples in the window and μ is the mean of the signal in the window. The calculation window for these parameters is 500 ms without overlapping. Finally, an exponential window has been applied to the features for smoothing.

3.3. Fractal dimension (FD)

Fractal dimension has a relation with the entropy and entropy has a direct relation with the amount of information inside a signal. Fractal dimension can be interpreted simply as the degree of meandering (or roughness or irregularity) of a signal. Katz, Higuchi and Peterson are well-known methods for calculating fractal dimension but the Katz method is more robust [11] than others, and therefore is implemented in this research. Katz fractal dimension is derived directly from the waveform. The FD of a curve can be defined as

$$FD = \frac{\text{Log}_{10} L}{\text{Log}_{10} d}$$

where L is the total length of the curve and d is the diameter estimated as the distance between the first point of the sequence and the point of the sequence that provides the farthest distance. FD is calculated every second with 250 ms overlap.

3.4. Linear discriminant analysis (LDA)

LDA classifier is still one of the most powerful methods and is very robust. Fisher's linear discriminant [13] is maximizing the between-group variance to the within-group variance ratio, which in this case is measured by the ratio of the determinants of the preceding two matrices.

3.5. Adaboost

The principle of the Adaboost [14] is that a committee machine can adaptively adjust to the errors of its components, the so-called weak learners. The classification rate of each weak learner should exceed 50%. Here, a neural network with one hidden layer is selected as the weak learner [15]. First, the first neural network trains with the equal error weight for all the samples:

$$D_1(i) = 1/N, \qquad i = 1, \ldots, N$$

where $D_1(i)$ is the error weight for the samples in the first iteration and N is the number of input samples. For the next iteration, the error weight of the samples is changed regarding their error in the previous iteration by the following relation:

$$D_n(i) = \frac{D_{n-1}(i)}{Z_{n-1}} * \begin{cases} \beta_{n-1} & \text{if} \quad F_{n-1}(x_i) = d_i \\ 1 & \text{otherwise} \end{cases}$$

$$\varepsilon_{n-1} = \sum_{F_{n-1}(x_i) \neq y_{n-1}} D_{n-1}(i) \qquad 0 < \varepsilon_{n-1} < 0.5$$

$$\beta_{n-1} = \frac{\varepsilon_{n-1}}{1 - \varepsilon_{n-1}} \qquad 0 < \beta_{n-1} < 1$$

where $F_{n-1}(x)$ is the output of the $(n-1)$th weak learner and ε_{n-1} is the error of the weak learner in the $(n-1)$th step, and d_i is the label of the ith input sample. β_{n-1} measures the importance of the hypothesis of $F_{n-1}(x)$ and it decreases with error. It is obvious that the misclassified samples in each stage are given a high value of $D_n(i)$ (error weight) for the next stage. This iterative procedure repeats till the time that n reaches T (the maximum considered value for the number of weak learners). After the training phase, total output of the Adaboost is calculated as follows:

$$\phi(x) = \arg\max \sum_{n=1\ldots T, F_n(x)=y} \log \frac{1}{\beta_n}.$$

For this classifier, all the feature values must be normalized in the interval $[-1, 1]$. A schematic diagram of the mentioned Adaboost is shown in figure 2.

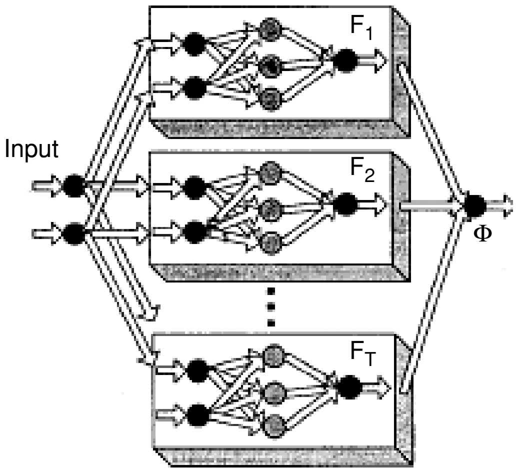

Figure 2. A committee of neural networks generated using Adaboost. F_1, F_2 and F_T are the output of weak learners and the output ϕ is the total output of the Adaboost [14].

Table 1. Error rate and the latency of its minimum of cross validation data for subject L1.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
20%	35%	20%	25%	30%	22%
4.75	5.25	4	4	4.25	4.75

4. Evaluation of classification performance

Features have been extracted from five subjects and applied to the LDA and Adaboost classifiers. The training set is evaluated ten times, by tenfold cross validation. The best classifiers from the evaluation phase have been selected and applied to the test data. In total, 360 trials have been recorded for each subject; then artefact trials were removed and for each subject 240 trials were selected for cross validation and the rest for the testing phase.

5. Results

After the extraction of features (BP, FD and Hjorth parameters) from the signals, they have been applied to the LDA and Adaboost classifiers. The first two features have been extracted from the signals every 1 s (window length) with 250 ms overlap but Hjorth parameters have been extracted every 500 ms with no overlap. No combination of the features has been considered in this paper. The cross validation and test results for five subjects are shown in tables 1–5 and 6–10, respectively. In each table, the minimum classification errors for all the features and classifiers and their latency of minimum error rate (in seconds) are shown. From the results, it is obvious that BP with LDA yielded the best performance for four subjects (L1, o3, o8, f8), the exception being case k3 for which Hjorth parameters along with both classifiers showed a significant result. In the test phase, FD and Adaboost showed the best result (16.5% error) for subject L1; also fairly

Table 2. Error rate and the latency of its minimum of cross validation data for subject o3.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
15.4%	22%	25%	30%	18%	20%
4.75	4.75	4	4	4.5	4.25

Table 3. Error rate and the latency of its minimum of cross validation data for subject o8.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
13.6%	16%	20%	26%	16.5%	18%
4.75	4	4	4.5	4.75	4.5

Table 4. Error rate and the latency of its minimum of cross validation data for subject k3.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
15%	16%	9.6%	10%	19.7%	21%
5.25	5	4.5	4	3.5	5.5

Table 5. Error rate and the latency of its minimum of cross validation data for subject f8.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
16.5%	16%	21.5%	20%	23%	26.5%
5.5	4	4.5	3.5	4.5	5

Table 6. Error rate and the latency of its minimum of test data for subject L1.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
28%	32%	26%	35%	25%	16.5%
4.75	4.25	4	4	4.75	4.5

Table 7. Error rate and the latency of its minimum of test data for subject o3.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
9.5%	25%	23%	28%	19.1%	22%
5.75	4.5	5	4	5	6.5

similar results were obtained in the cases o8 and k3 with LDA and BP. FD along with LDA has shown good results in cases k3 and L1. But for the three other subjects (o3, o8 and f8) the test results confirm the cross validation phase. To have significant results, the F_test and the T_test were performed on the test results. Test and training data were randomly chosen from the pure data for 20 times. All results were

Table 8. Error rate and the latency of its minimum of test data for subject o8.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
18%	20%	24%	30%	24%	20%
6.25	4	4	4.5	4.75	4.5

Table 9. Error rate and the latency of its minimum of test data for subject k3.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
10.2%	20.4%	23.5%	17%	10.2%	14.3%
4.5	5.5	5.5	5.5	5	5

Table 10. Error rate and the latency of its minimum of test data for subject f8.

Band power		Hjorth parameters		Fractal dimension	
LDA	Adaboost	LDA	Adaboost	LDA	Adaboost
16.4%	20.71%	23.5%	22.86%	18.5%	25%
6.5	4.5	4.5	4	4.25	4.25

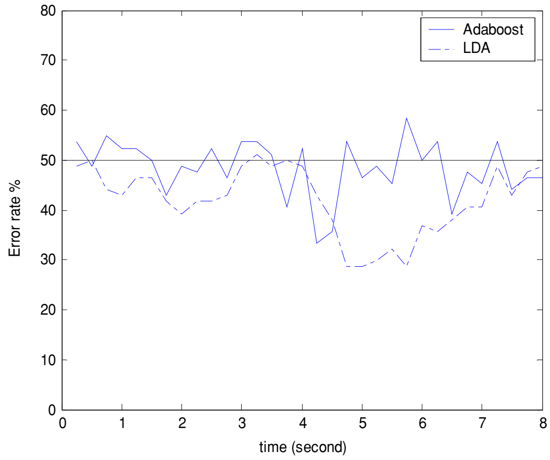

Figure 3. Error rate of BP with Adaboost and LDA for subject L1.

significant, which means that the *P* value was lower than 0.05 for all evaluated results. To have a better representation of results, the test curves for FD and BP along with two classifiers (LDA and Adaboost) for the whole paradigm are depicted in figures 3–12 for our subjects. Hjorth results are eliminated from the graphs, because these results are not the best in any case, but for clarity for comparing BP and FD with the mentioned classifiers just two curves have been shown in all the graphs.

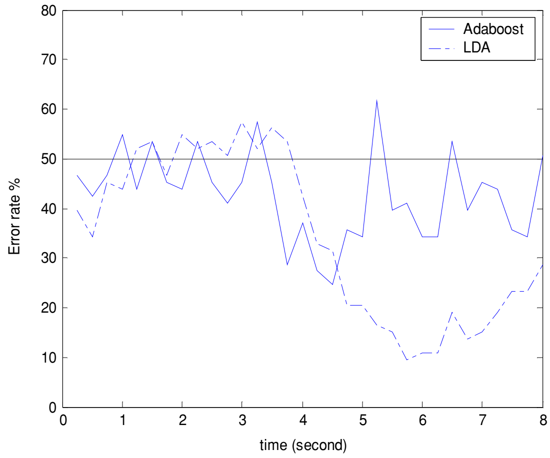

Figure 4. Error rate of BP with Adaboost and LDA for subject o3.

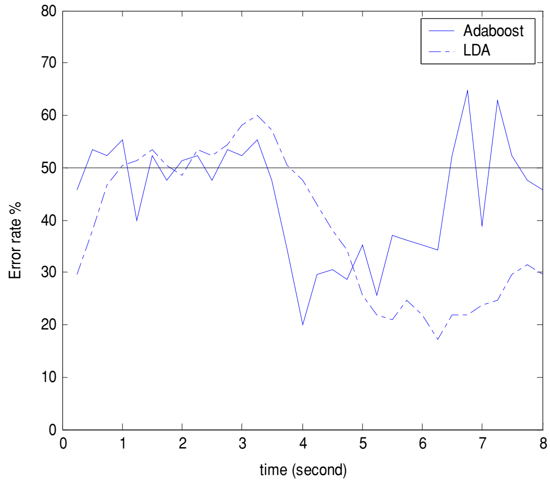

Figure 5. Error rate of BP with Adaboost and LDA for subject o8.

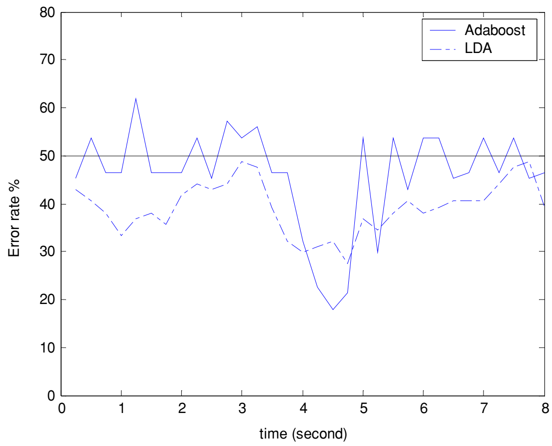

Figure 6. Error rate of FD with Adaboost and LDA for subject L1.

6. Discussion

In order to increase the performance of a cue-based BCI system, FD, BP and Hjorth parameters along with Adaboost and LDA are assessed. Cross validation results in four (out of five) cases indicate that the best combination for classifying the imagery tasks is BP with LDA. But in the test phase, results did not completely confirm the cross validation results.

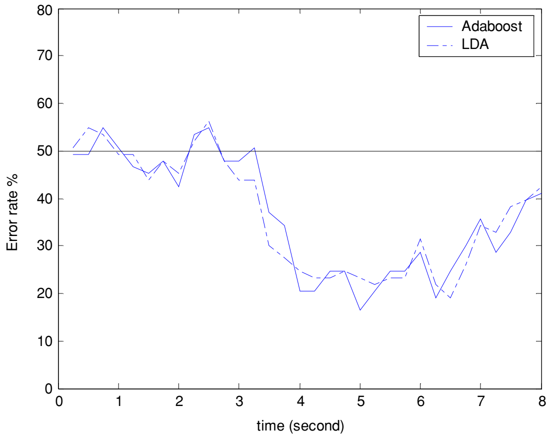

Figure 7. Error rate of FD with Adaboost and LDA for subject o3.

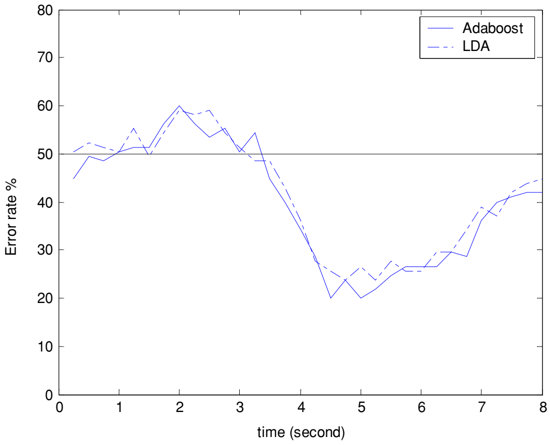

Figure 8. Error rate of FD with Adaboost and LDA for subject o8.

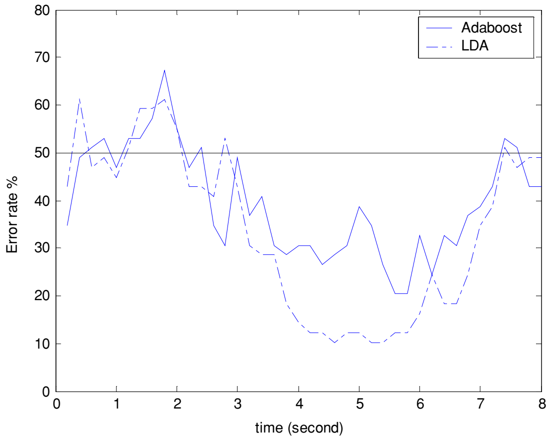

Figure 9. Error rate of BP with Adaboost and LDA for subject k3.

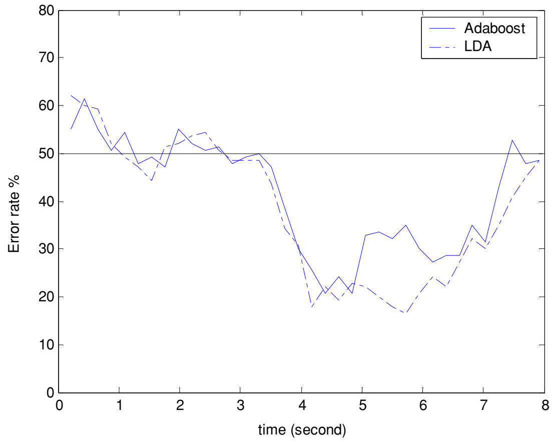

Figure 10. Error rate of BP with Adaboost and LDA for subject f8.

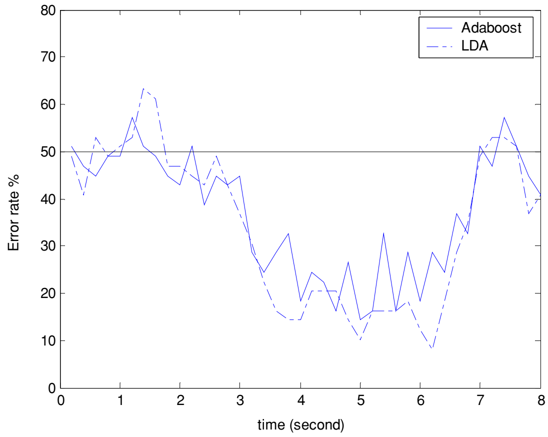

Figure 11. Error rate of FD with Adaboost and LDA for subject k3.

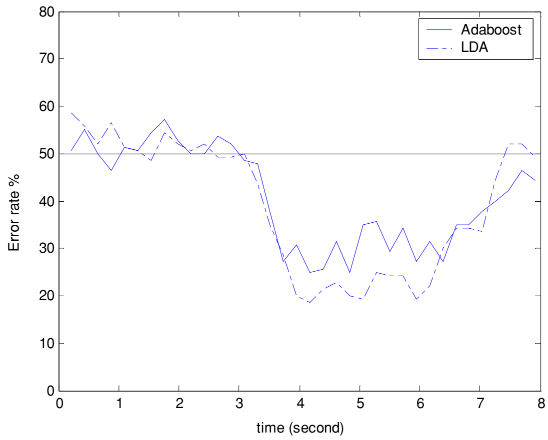

Figure 12. Error rate of FD with Adaboost and LDA for subject f8.

FD and Adaboost showed a significant result for subject L1 in comparison with other combinations. This successful combination yielded good results for the cases k3 and o8. FD has also shown very good results with LDA in the cases k3 and L1. It can be claimed that among our cases FD with both classifiers can be an acceptable alternative with the combination of BP and LDA. In many articles [1, 4, 7] BP and LDA are presented as a gold-standard technique for

BCI applications. This paper showed that the selection of a feature and a classifier is extremely dependent on the case. In the above-mentioned articles results were shown for a maximum of two or three cases and no appropriate comparison was made. FD is also applied in the detection of epileptic seizure onset [16]. Calculation of FD is very fast and can be employed for on-line BCI applications. The significant property of FD is that FD and LDA, in three cases, give a faster feedback of imagery action than BP–LDA as is shown in tables 6–10. Hjorth parameters also demonstrate discrimination between EEG patterns with a relative short latency (second 4 and 4.5) and best classification rate for subject k3 in the cross validation phase but these significant results have not been repeated in the test results. There is a trade-off between the minimal error rate and its latency. Regarding our paradigm, latency of minimal error rate after 2.5 s of the cue stimulus is acceptable. Therefore, 9.5% error for case o3 (happening in 6.25 s) cannot be acceptable, because, in the real application, a subject might lose concentration if he does not see the feedback on the screen. The evaluation was also performed for different window lengths; in the 250 ms interval, FD shows supremacy to BP and Hjorth parameters. As the time window for calculating the features increases, the classification rate improves but it causes a delay in reporting the changes in the EEG. Of the two classifiers, LDA shows a very reliable and robust behaviour. In contrast, Adaboost is more time consuming for the training (because the small neural networks must be trained), especially for the cross validation phase and because of its non-linearity property, it is not as robust as LDA. One of the reasons why the FD along with Adaboost showed a better result just in one case is that the extracted EEG features are really scattered in the feature space and have a chaotic behaviour in the case L1, therefore, FD can present it much better than two other features, and the non-linear classifier (Adaboost) could find a more flexible margin through the classes. From another angle, BP–LDA showed the best classification rate in four cases in the test phase, but in case k3, FD–LDA results were exactly the same. Nevertheless, a combination of BP–LDA is still one of the best approaches in the BCI applications. The point is that the biological properties of every human are unique, therefore, gold-standard combination does not make sense for all the cases. Thus, for each individual, we have to find the best combination of feature and classifier or on some occasions, a combination of the features by evolutionary algorithms or a tree combination of classifiers which can lead to the best result.

Acknowledgments

The authors wish to warmly express their appreciation to Professor Pfurtscheller, Dr Graimann and Dr Schloegl for their assistance and cooperation in our paper. Our data were recorded in the Graz BCI laboratory.

References

[1] Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain–computer communication *Proc. IEEE* **89** 1123–34
[2] Wolpaw J R, McFarland D J and Vaughan T M 2000 Brain–computer interface research at the Wadsworth center *IEEE Trans. Rehab. Eng.* **8** 222–6
[3] Kalcher J, Flotzinger D and Pfurtscheller G A 1992 New approach to a brain–computer-interface (BCI) based on learning vector quantization (LVQ3) *Proc. Annual Int. Conf. IEEE* vol 4 pp 1658–9
[4] Pfurtscheller G, Neuper C, Schlogl A and Lugger K 1998 Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters *IEEE Trans. Rehab. Eng.* **6** 316–28
[5] Haselsteiner E and Pfurtscheller G 2000 Using time-dependent neural networks for EEG classification *IEEE Trans. Rehab. Eng.* **8** 457–63
[6] Bozorgzadeh Z, Birch G E and Mason S G 2000 The LF-ASD brain computer interface: on-line identification of imagined finger flexions in the spontaneous EEG of able-bodied subjects *IEEE Int. Conf. Acoustic and Speech Proc.* vol 6 pp 2385–8
[7] Obermaier B, Neuper C, Guger C and Pfurtscheller G 2001 Information transfer rate in a five-classes brain–computer interface *IEEE Rehab. Eng.* **9** 283–8
[8] Flotzinger D, Pregenzer M and Pfurtscheller G 1994 Feature selection with distinction sensitive learning vector quantisation and genetic algorithms *IEEE Int. Conf. on Computational Intelligence* vol 6 pp 3448–51
[9] Deriche M and Al-Ani A 2001 A new algorithm for EEG feature selection using mutual information *Proc. ICASSP: IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* vol 2 pp 1057–60
[10] Cincotti F, Scipione A, Tiniperi A, Mattia D, Marciani M G, Millan J, Salinari S, Bianchi L and Babiloni F 2003 Comparison of different feature classifiers for brain computer interfaces *IEEE Conf. EMBS on Neural Engineering* pp 645–7
[11] Esteller R, Vachtsevanos G, Echauz J and Lilt B 1999 A comparison of fractal dimension algorithms using synthetic and experimental data ISCAS *Proc. IEEE Int. Symp. on Circuits and Systems* vol 3 pp 199–202
[12] Hjorth B 1970 EEG analysis based on time domain properties *Electroencephalogr. Clin. Neurophysiol.* **29** 306–10
[13] Webb A 1999 *Statistical Pattern Recognition* (New York: Oxford University Press)
[14] www.boosting.org
[15] Murphey Y L, Chen Z and Guo H 2001 Neural learning using AdaBoost *Proc. IJCNN: Int. Joint Conf. on Neural Networks* vol 2 pp 1037–42
[16] Esteller R, Vachtsevanos G, Echauz J, Henry T, Pennell P, Epstein C, Bakay R, Bowen C and Litt B 1999 Fractal dimension characterizes seizure onset in epileptic patients *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing* vol 4 pp 2343–6