KALMAN FILTER PARAMETERS AS A NEW EEG FEATURE VECTOR FOR BCI APPLICATIONS

Amir H. Omidvarnia, Farid Atry, S. Kamaledin Setarehdan, and Babak N. Arabi

Control and Intelligent Processing Centre of Excellence, ECE Department, Faculty of Engineering, University of Tehran
Tehran, Iran
email: ksetareh@ut.ac.ir

ABSTRACT

With recent advances in signal processing and biomedical instrumentation, EEG[1] signals can be used as a new communication channel between human and computers. Implementation of this channel is possible by recording and analyzing brain waves. Such a system translates human thoughts for a computer thus it is called a "Brain Computer Interface" or BCI

In this paper, a new feature vector for each EEG channel is introduced using the Kalman filter. This feature vector has equal or in some cases, better performance than the other commonly used features. Different classifiers were used to classify EEG signals using the new features and the results are compared.

1. INTRODUCTION

For several years, people were seeking for a non muscular channel between the brain and the out world so that they can control peripherals by thinking. With production of advanced bioinstruments for recording and amplifying EEG signals as well as cheap and powerful personal computers, this dream was realized and Brain-Computer Interface was developed. Main applications of this system are related to severe neural disordered patients. These disorders include ALS[2], brain stem stroke, cerebral palsy, muscular dystrophies, brain or spinal cord injury and … BCI system enables these patients to control other devices such as computers without any physical activity. Current BCI systems use different kinds of EEG signals such as SCP (Slow Cortical Potentials), P300 component and Mu or Beta rhythms from sensorimotor cortex which are recorded by using non-invasive methods from scalp. Also they use cortical neural activities which are recorded by using implanted electrodes in invasive manners [1]. There are usually two methods for implementation of a BCI system: Pattern recognition approach and Biofeedback approach [2,3].

In the first method, subject thinks to some particular mental tasks such as Visual Counting, Mental Multiplication, … and during this period, his or her EEG signal is recorded. Then, features of the EEG signals are extracted and classified by means of a classifier. The result of the classification is used for controlling such devices as a mouse and/or a keyboard. The commonly used EEG features for this purpose include AR[3] coefficients, power of EEG signal in different frequency bands and wavelet coefficients. Defining and extracting a new and effective feature is an open problem in pattern recognition based BCI approaches.

In the second method, subject learns to control some characteristics of his or her EEG signal during a training period. These characteristics are then used as the features to be extracted and classified for control purposes.

A pattern recognition based BCI system includes different blocks, which is shown in figure 1. Biofeedback can also be used in the first method for improving the classification results [3].

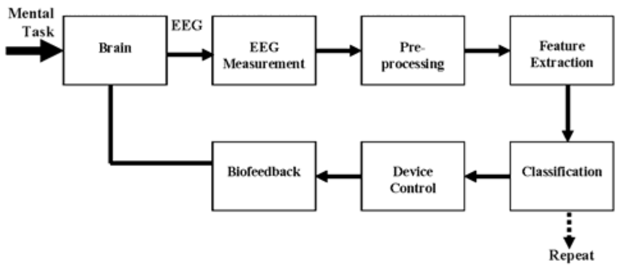


Figure 1 – General construction of a typical pattern recognition based BCI system

The characteristics of the data and the pre-processing methods used in this study are described in section 2. In section 3, classification algorithms are explained. Section 4 discusses the feature extraction scheme using the Kalman Filter. Results obtained in this study are displayed in section 5. Finally section 6 concludes the paper.

2. METHODOLOGY

2.1 The data set

The data set used in this study was obtained from the Colorado University [4]. The data have been recorded according to the 10-20 standard from the C3, C4, P3, P4, O1, O2 electrodes. Each recorded signal has a length of 10 second with a sampling rate of 250 Hz. Seven subjects participated in the recordings In this study however only the signals of subjects 1,3 and 6 were used because only these three subjects had participated in 2 recording sessions. In each session, 5 records and in each record 5 different mental tasks

---

[1] Electroencephalogram
[2] Amyotrophic Lateral Sclerosis
[3] Auto Regressive

have been used. The mental tasks considered in the data collection includes: Baseline (B), Visual Counting (C), Mental Letter Writing (L), Mental Multiplication (M) and Mental 3-D Geometric Rotation (R). Our goal is to discriminate all possible combinations of the pairs of mental tasks from each other using the corresponding EEG signals [5].
The EOG[4] signal was also recorded simultaneous to the EEG signal using the standard positions used for the EEG electrodes which is called 10-20 standard. This is for removing the artefacts caused by the eye movements from the recorded EEG signals in the data processing stage.

2.2 Data Processing

Initially, using the EOG signals and an appropriate empirical threshold value, sections of the EEG signals coincident with the eye movements are identified and removed from the data set. The pre-processed signals of length 10 seconds were then divided into windows of 1 second length each with an overlap of 0.8 seconds between the neighbouring windows. For feature extraction, a multi input-single output Kalman filter was used. The extracted features were then fed into the MLP[5] neural network and Bayesian classifier with Gaussian, k-NN[6] and Parzen kernels at separate studies. In the MLP neural network, Steepest Decent optimization algorithm was used. For test, the holdout method was used where 70% of the data is used for training and 30% for test.

3. CLASSIFICATION APPROACHES

The basic principles of Bayesian, k-NN and Parzen classifiers and MLP neural network will briefly be discussed throughout this section, whereas a detailed description can be found in [6]. These classifiers were used for discriminating EEG features in different classes.

3.1 Bayesian classifier

Suppose ω_1 and ω_2 are the 2 classes that the data belongs to them. First, we consider two known probability functions $P(\omega_1)$ and $P(\omega_2)$ for these classes. If N is the number of all training data and N1 and N2 are the number of data in the two classes ω1 and ω2 respectively then:

$$P(\omega_2)\approx\frac{N_2}{N} \quad and \quad P(\omega_1)\approx\frac{N_1}{N} \qquad (1)$$

Also $p(x|\omega_i)$, i =1,2 are considered to be known values that determine distribution of the data in each class. If these functions are unknown, we can estimate them with different methods such as ML[7] and … Based on the Bayes rule, we have:

$$P(\omega_i|x)=\frac{p(x|\omega_i)P(\omega_i)}{p(x)} \qquad (2)$$

where p(x) is probability density function :

$$p(x)=\Sigma_{i=1}^{2} p(x|\omega_i)P(\omega_i) \qquad (3)$$

Now, Bayesian classification rule can be shown as below [6]:

$$P(\omega_1|x)>P(\omega_2|x)\Rightarrow x\in\omega_1 \qquad (4)$$

$$P(\omega_1|x)<P(\omega_2|x)\Rightarrow x\in\omega_2 \qquad (5)$$

3.2 Parzen Estimation

In a multidimensional classification problem, ℓ dimensional space is divided into hyper cubes with length h and volume h^l, instead of columns with width h.
Suppose xi , i =1,2,..,N are feature vectors. We define Φ(x) as below:

$$\Phi(x_i) = 1 \quad \text{if } |x_{ij}|\le 1/2$$
$$\Phi(x_i) = 0 \quad \text{Otherwise} \qquad (6)$$

where x_{ij}'s (j=1, 2,…, l) are x components. Value of function for all points inside unit hyper cube with the centre on origin is equal to 1 and for outside points is equal to zero. The relation in (6) can be rewritten as:

$$\hat{p}(x) = \frac{1}{h^l}\left(\frac{1}{N}\sum_{i=1}^{N}\phi\left(\frac{x_i - x}{h}\right)\right) \qquad (7)$$

Value of sigma is equal to k_N or number of points that fall into the hyper cube. Then pdf estimation is resulted from division of k_N on N and hl.
In a decision making problem with M classes $\omega_1,\omega_2,\dots,\omega_M$ for estimation of conditional pdf, related to i'th class, at first consider a constant value (or a constant h), then count the number of points belong to ω_i in this value. If this number is equal to k_i, estimated pdf is:

$$\hat{p}(\omega_i|x) = \frac{k_i}{N_i V} \qquad (8)$$

where V is the volume of the hyper cube and is a constant, and N_i is the number of all data in the i'th class. $p(\omega_i \mid x)$ can be used as the kernel of Bayesian classifier [6].

3.3 K-Nearest Neighbour pdf Estimation

In Parzen estimation the value around x point was fixed (h^l) and the number of points inside the volume (k_i) was changed from one point to another. But in the k-NN estimation method k_N is equal to a constant k and the size of the volume around x changes as it contains k points. Also we

[4] Electro-Occulogram
[5] Multi Layer Perceptron
[6] k-Nearest Neighborhood
[7] Maximum Likelihood

can use other types of volumes such as hyper sphere. Then , the Pdf is estimated as follows [6]:

$$\hat{p}(\omega_i|x) = \frac{k}{N\ V(x)} \qquad (9)$$

3.4 Multi Layer Perceptron

MLP is a general function approximator which can be used as a simple and powerful classifier. For training the network, Back Propagation approach with Steepest Decent optimization algorithm has been used. Suppose Q is the number of all data and E_q is error of feeding x^q to network. N is the dimension of the input data, M and J are the number of neurons of the hidden layer and the output layer respectively. Now we determine and update coefficients u_{mj} and w_{mn} that minimize error function, which is defined as below [6]:

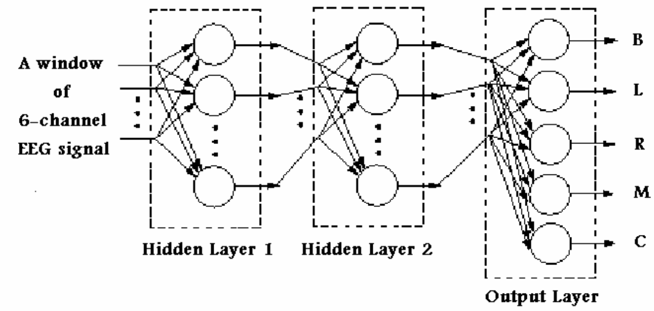


Figure 2 – A typical neural network for classification of 5 mental tasks .B, C, L, M and R stand for Baseline, Counting, Letter writing, Multiplication and Rotation, consequently.

$$E^q = \sum_{j=1}^{J}(z_j^q - t_j^q)\hat{}2 \qquad (10)$$

$$\overset{+}{w}_{mn} = \overset{-}{w}_{mn} - \rho \frac{\partial E^q}{\partial w_{mn}} \qquad (11)$$

$$\overset{+}{u}_{mj} = \overset{-}{u}_{mj} - \rho \frac{\partial E^q}{\partial u_{mj}} \qquad (12)$$

In this study we used a MLP with two hidden layers of 10 and 3 neurons each. Its Output layer had 5 neurons equal to the number of classes or the mental tasks.

4. KALMAN FILTER PARAMETERS AS A NEW EEG FEATURE VECTOR

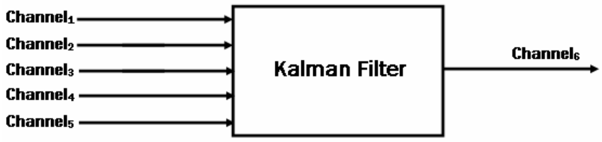
Consider a dynamic system y(t) = X(t)θ(t), where X(t) is the vector of the input values at time t , y(t) as the output value of the system at time t, and θ(t) is the state of the dynamic system at time t (figure 3). The Kalman filter is a tool for estimating the state of a dynamic system by means of a recursive approach and using the input and output values of the system. The main problem in using the Kalman filter is to find the optimum estimator $\hat{\theta}(t)$ for $\theta(t)$ by using the observations x(1),x(2),…,x(t). It has been shown in [7] that this quantity is equivalent to the conditional expectation $\hat{\theta}(t)$ = E{θ(t)|x(1),x(2),…,x(t)}. Thus the system output at time t can be written as:

$$y(t) = x^T(t)\ \hat{\theta}(t) + n(t) \qquad (13)$$

where n(t) is the additive noise at the output of the system. For a system with the constant parameters one can write ($\hat{\theta}(t+1) = \hat{\theta}(t)$). Therefore $\hat{\theta}(t)$ can be estimated using the recursive algorithm of the Kalman Filter .

In the problem of the mental tasks classification, the EEG signal was divided into many short segments in the preprocessing stage. In general, EEG is a non stationary signal and its statistical parameters are not constant [8,9]. However for very short segments of the signal it can be considered stationary, with constant statistical parameters.

In the new feature extraction algorithm, we aim to identify a Kalman Filter, which can approximate the sixth EEG channel as its output using the other 5 channels as its inputs with the least possible error. In other words, it is equivalent to feed one channel as the output and the other 5 channels as the inputs into the Kalman filter and to optimise the filter coefficients to achieve the minimum estimation error. In this study, all possible combinations of 5 input signals and one output signal from the six recording electrodes (C3, C4, P3, P4, O1, O2) were tested (6 combinations in total).



$$y_{[N\times1]} = x_{[N\times5]} \cdot \theta_{[5\times1]}$$

Figure 3 – Using Kalman filter for feature extraction in the Time domain

For each segment, a θ vector is obtained and used as feature vector in classification problem.

5. RESULTS

The results of this study are summarized in Table 1. As Table 1 shows, the results of new extracted features in more cases are better than the other best known and commonly used features such as AR coefficients, power of signal in different EEG bands $(\delta, \theta, \alpha, \beta, \gamma)$ and wavelet coefficients. Most of classifiers have better results on the new feature rather than the other features except Bayesian classifier with a Gaussian kernel. Thus it seems that the probability density function (pdf) of this feature is too complicated to be estimated by fitting Gaussian curves. In this case, k-NN and Parzen estimators are more effective than Bayesian classifier with a Gaussian kernel for estimating the pdf of feature statistical distribution. The k-NN classifier has the best performance between the all other classifiers. MLP neural network as a classifier has almost the same results on different features. There are completely nonlinear boundaries among classes and MLP classifier has the ability of estimating these boundaries, thus these results show that complexity order of boundaries among different classes in different features are approximately equal.

New feature which has been extracted using the Kalman filter includes the linear parameters among different channels. As these parameters had good results in classification problem, a linear relation between the different EEG channels can be concluded. Best results for this linear relation are obtained in the 4-15 Hz frequency band. The use of the frequencies out of this band decreases the correct classification rate. Thus α and lower β bands are the best EEG bands for extraction of this new feature.

Feature	Classifier	Bayes (Gaussian)	Parzen	k-NN	MLP
AR Coefficients		90.44	87.69	93.16	83.88
Wavelet Coefficients		88.58	75.33	81.33	85.19
Power of signal		83.00	77.27	81.72	82.27
Mixed features		91.02	83.13	86.63	87.22
New feature		78.29	92.23	96.13	93.86

Table 1 – Average of discrimination percent of classifiers for different features in 3 subjects

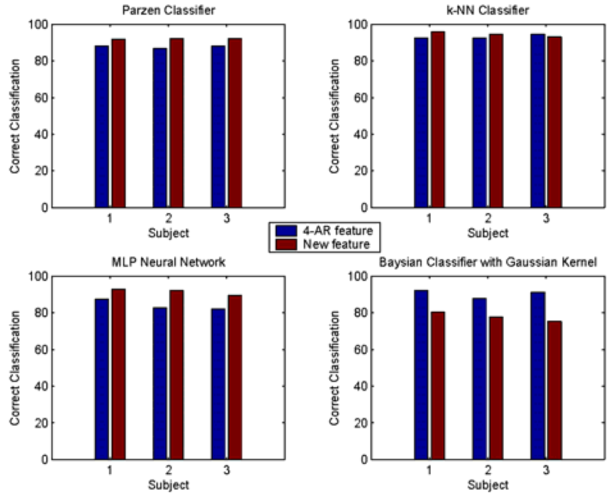


Figure 4- Classification results of 4-AR coefficients in contrast with new Kalman feature for 3 subjects

6. CONCLUSION

A new feature vector is developed in this study using a multi input-single output Kalman filter. It was shown that the classification results of this feature vector is comparable or better than the other commonly used features such as the AR coefficients, power of the signal at different frequency bands and the wavelet coefficients. Since the new features are the linear parameters of the Kalman filter, a linear relationship between the different EEG channels is expected. The method of selection of optimum combination for feature extraction with the Kalman filter can be used as a procedure for reduction of the EEG channels and source localization of mental tasks too.

7. REFERENCES

[1] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, Theresa M. Vaughan*," Brain–computer interfaces for communication and control ",* Invited review, Clinical Neurophysiology 113 (2002) 767–791, 2 March 2002

[2] C. Neuper, A. Schlögl, G. Pfurtscheller (1999) Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery. *J Clin Neurophysiol*. 16(4):373-82.

[3] Pfurtscheller G, Neuper C, Schlögl A, Lugger K. (1998) Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Trans Rehabil Eng*. 6(3):316-25.

[4] Dr. Charles Anderson’s website: http://www.cs.colostate.edu/~anderson/res/eeg/#Contents

[5] Charles W. Anderson, Saikumar V. Devulapalli, Erik A. Stolz: EEG as a means of communication: preliminary experiments in EEG analysis using neural networks. ASSETS 1994: 141-147

[6] S.Theodoridis, K.Koutroumbas,"Pattern Recognition", Academic Press, 1999

[7] J. Melsa and D. Cohn, “Decision and Estimation Theory”, McGraw-Hill, 1978.

[8] P. Karjalainen, \Estimation theoretical background of root tracking algorithms with application to EEG." Phil. Lic. thesis, University of Kuopio, Department of Applied Physics, 1996.
Available:
http://venda.uku._/research/biosignal/publications/.

[9] M.P. Tarvainen, J.K. Hiltunen, P.O. Ranta-aho, and P.A. Karjalainen, "Estimation of Nonstationary EEG with Kalman Smoother Approach: an Application to Event-Related Synchronization (ERS)", submitted to IEEE Trans Biomed Eng, June 19, 2003