Volume 4 No. 9, September 2014 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Fractal Dimension and Linear Discriminant Analysis for faster and more Accurate Brain Computer Interface

Abbas. Najafizadeh, Abdolreza Asadi Ghanbari
Department of Computer engineering, Shahr-e-Qods Branch,Islamic Azad University,Tehran,Iran

ABSTRACT

Feature extraction is the process of accurately simplifying the representation of data by reducing its dimensionality while extracting its relevant characteristics for the desired task. It has a substantial effect on the classification accuracy and speed since classification carried out without a successful feature extraction process on a high dimensional and redundant data would be computationally complex and would overfit the training data. Fractal dimension is a statistical measure indicating the complexity of an object or a quantity that is self-similar over some region of space or time interval. It has been successfully used in various domains to characterize such objects and quantities but its usage in BCI has been more recent. There are several fractal dimension estimation methods, some of which are not applicable to all types of data exhibiting fractal properties. In order to achieve a higher classification accuracy and speed, the fractal dimension estimation method that is most suitable to the data at hand should be chosen. In this study, after preprocess the EEG data by the coherence average, principal component analysis (PCA), and independent component analysis (ICA) commonly used fractal dimension estimation methods Katz's method, Higuchi's method, the rescaled range (R/S) method, were evaluated for feature extraction in EEG based BCI by conducting offline analyses of a two class EEG dataset. Support vector machine (SVM) and linear discriminant analysis (LDA) were tested in combination with these methods to determine the methodology with the best performance and result compare with wavelet feature extraction method.

Keywords: *Feature extraction; Fractal dimension; Principal Component Analysis (PCA); Independent Component Analysis (ICA).*

1. INTRODUCTION

The cutting-edge research fields between bioinformatics and computer science have been developing dramatically recently. Brain computer interface (BCI) is now becoming one of hot research topics due to the following three reasons. First, it provides a new approach to understand neurophysiologic mechanism of how brain is executing specific task [1]. Second, BCI is of practical significance in real applications. The technique can be used for helping people with severe motor disabilities [2, 3] and applied to clinical rehabilitation of motor functions [4,-6]. Third, it also provides a possibility to combine brain intelligence and machine (e.g., computer) intelligence, and people could enhance their ability of manipulating objects in the world by combining themselves' intelligence and machine's intelligence.

Generally speaking, there are a number of ways of measuring brain activity, such as electrical signal and blood oxygen level dependent (BOLD) signal. Near-Infrared Reflectance Spectroscopy (NIRS) [7] and Functional Magnetic Resonance Imaging (fMRI) [8] are based on the measure of BOLD signal. Electrical signal is acquired from electrodes mounted on the surface of scalp or implanted into the tissue of brain, namely, noninvasive or invasive manners. For the invasive manner, electrodes should be implanted into the brain or laid on the cortex surface of the brain. For example, [9] recorded movement-related cortical potentials through an invasive system. Although such an invasive system has the higher spatial resolution, it is easy to cause the damage of users' brain tissue. In this paper, we rather concentrate on noninvasive BCI based on Electroencephalogram (EEG) measurements [10]. EEG signal has an advantage in temporal resolution compared with BOLD signal. But it also suffers a number of drawbacks for BCI such as temporal variation of EEG, redundancies and artifact [11].

Three major components of the BCI system noise reduction (preprocessing) feature extraction and feature translation or classification [12]. Functional model of a BCI system for EEG signal classification is depicted in Fig.1.

International Journal of Information and Communication Technology Research

http://www.esjournals.org

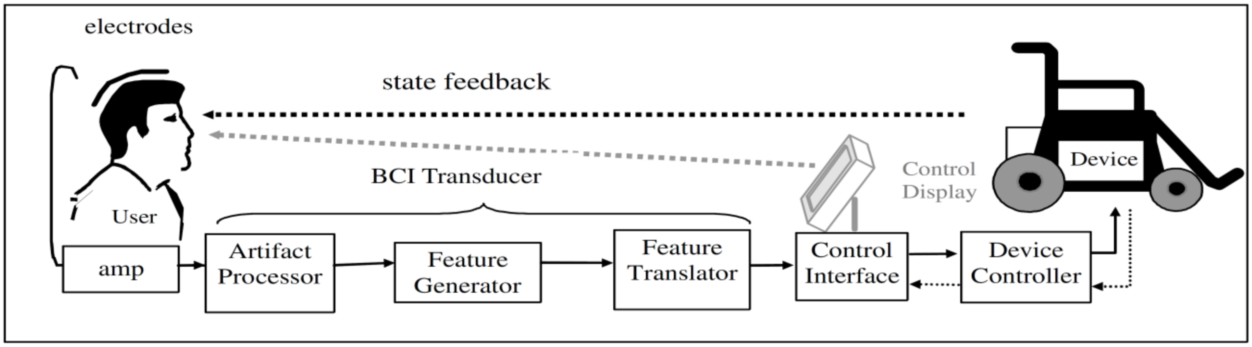

Fig. 1. Functional model of a BCI system depicting its principle functional components [12].

Feature extraction is the process of accurately simplifying the representation of data by reducing its dimensionality while extracting its relevant characteristics for the desired task. It has a substantial effect on the classification accuracy and speed since classification carried out without a successful feature extraction process on a high dimensional and redundant data would be computationally complex and would overfit the training data. Fractal dimension is a statistical measure indicating the complexity of an object or a quantity that is self-similar over some region of space or time interval. In this study, commonly used fractal dimension estimation methods to characterize time series (Katz's method, Higuchi's method and the rescaled range method) were evaluated for feature extraction in EEG based BCI and result compare with wavelet feature extraction method. For preprocessing, the coherence average, principal component analysis (PCA), and independent component analysis (ICA) were used to reduce dimensions and improve signal to noise ratio (SNR). After constructing the feature vectors, the test samples were classified using SVM and LDA classifiers.

2. MATERIALS AND METHODS

In this research, EEG signal used as the basic data for classification. The EEG data is from an open EEG database of University of Tuebingen. Two types of the EEG database are employed as [13].

a. Dataset I

The datasets were taken from a healthy subject. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 6s. During every trial, the task was visually presented by a highlighted goal at either the top or bottom of the screen to indicate negativity or positivity from second 0.5 until the end of the trial. The visual feedback was presented from second 2 to second 5.5. Only this 3.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 3.5s results in 896 samples per channel for every trial. This dataset contain 266 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

b. Dataset II

The datasets were taken from an artificially respirated ALS patient. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received auditory and visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 8s. During every trial, the task was visually and auditorily presented by a highlighted goal at the top or bottom of the screen from second 0.5 until second 7.5 of every trial. In addition, the task ("up" or "down") was vocalised at second 0.5. The visual feedback was presented from second 2 to second 6.5. Only this 4.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 4.5s results in 1152 samples per channel for every trial. This dataset contain 200 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

3. PRE PROCESSING

It would be difficult to identify the effective brain signals in a single trial without pre-processing. In this study, a Butterworth filter was used as the low-pass filter with a cut-off frequency of 30 Hz. The signals were then processed using coherence average, PCA and ICA by the analysis tool EEGLAB 5.02 (http://sccn.ucsd.edu/eeglab/).

Volume 4 No. 9, September 2014 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

a. Coherence Average

The coherence average is commonly used to process weak signals, such as EEG, with a strong noise and to improve SNR of signals. SNR was defined as follows:

$$SNR = \frac{P}{\delta^2} \tag{1}$$

Where P is the power of ideal EEG signal and δ^2 is the power of the noise. If the noise is assumed as a stationary random signal with a mean value of 0, then the variance of the noise is δ^2. After a coherence average of N samples with the same Stimulus-Code, the variance of the noise will be reduced to δ^2/N, so the new EEG SNR will becomes N times larger.

$$SNR = \frac{P}{\delta^2 / N} = \frac{NP}{\delta^2} \tag{2}$$

b. Principle Component Analysis

Principle component analysis (PCA) involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables termed principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA is theoretically an optimal linear scheme (in terms of least mean square error) for compressing a set of high dimensional vectors into a set of lower dimensional vectors and then reconstructing the original set. PCA is a non-parametric analysis and the answer is unique and independent of any hypothesis regarding data probability distribution. Importantly, PCA presents a method of compressing the high resolution data into a format for ICA to extract the required information–increasing computational efficiency.

The steps to process the EEG data by PCA are as follows:

Step 1: estimate the sample covariance matrix of the high-dimensional EEG signal X(t) after processed by coherence average.

$$\hat{\sum} = X(t)X^T(t) \tag{3}$$

Where $X(t) = [x_1(t), x_2(t), ..., x_p(t)]$ and $x_i(t)$ is a normalized time series from ith sampling channel of EEG with zero mean value and p is the total number of sampling channels of EEG.

Step 2: calculate the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and eigenvectors $e_1, e_2, \ldots, e_p$ of the covariance matrix $\sum$. Here $\lambda_1, \lambda_2, \ldots, \lambda_p \geq 0$. We can select p characteristic signals $y_1, y_2, \ldots, y_p$. Here,

$$Y = [y_1, y_2, ..., y_p]^T = E^T X \tag{4}$$

$$E = [e_1, e_2, ..., e_p] \tag{5}$$

$$E^T \sum E = \Lambda \tag{6}$$

$$\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_p\} \tag{7}$$

Step 3: choose principle component. $\lambda_i \Big/ \sum_{i=1}^{p} \lambda_i$ is the weight of y_i in Y. Find a suitable n, where $\sum_{i=1}^{n} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i \geq 80\%$.

Step 4: form the new low-dimensional signal Y*. Y* = $[y_1, y_2, \ldots, y_n]^T$.

c. Independent Component Analysis

ICA is a statistical and computational technique for revealing an observed multidimensional random vector into components that are statistically as independent from each other as possible. In practical situations, we cannot generally find a representation where the components are really independent, although we can at least find components that are as independent as possible. This leads us to the following definition of ICA. Given a set of observations of variables $(y_1(t), y_2(t). . . y_n(t))$, such as above-mentioned low-dimensional signal Y*, where t is the time or sample index, assume that the observations are generated as a linear mixture of independent components:

$$\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = A \begin{pmatrix} S_1(t) \\ S_2(t) \\ \vdots \\ S_n(t) \end{pmatrix} \tag{8}$$

Where A is a matrix determined by the Infomax ICA based on stochastic gradient learning rules. Infomax explicitly tries to maximize the joint entropy of a nonlinear function of the separated outputs; however, it implicitly minimizes the mutual information between the separated outputs so as to make them mutually independent [14]. Independent component analysis now consists of estimating both the matrix A and the $s_i(t)$, when we only observe the $y_i(t)$. Note that we assumed here that the number of independent components was equal to the number of observed variables.

Alternatively, we could define ICA as follows: find a linear transformation given by a matrix W, so that the random variables $y_i, i = 1, 2, ..., n$ are as independent as possible.

International Journal of Information and Communication Technology Research

http://www.esjournals.org

$$\begin{pmatrix} S_1(t) \\ S_2(t) \\ \vdots \\ S_n(t) \end{pmatrix} = W \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} \qquad (9)$$

This formulation is not that different from that described above, since after estimating A, its inverse A^{-1} gives W. It can be shown that the problem is well defined, that is, the model can be estimated if all the components $s_i(t)$ are non-Gaussian or only one component is Gaussian. This is a fundamental requirement that also explains the main difference between ICA and factor analysis, in which the non-Gaussian nature of the data is not taken into account. In fact, ICA could be considered as non-Gaussian factor analysis, since in factor analysis we are also modeling the data as linear mixtures of some underlying factors [14].

4. FEATURE EXTRACTION

Feature extraction is the process of accurately simplifying the representation of data by reducing its dimensionality while extracting its relevant characteristics for the desired task. It has a substantial effect on the classification accuracy and speed since classification carried out without a successful feature extraction process on a high dimensional and redundant data would be computationally complex and would overfit the training data. Fractal dimension using Katz's method [15], Higuchi's method [16], the rescaled range (R/S) method [17] extracted and compare results with the time–frequency features using wavelet transform were then.

a. Fractal Dimension

In Katz’s method, Higuchi’s method and the R/S method, the fractal dimension of the samples from selected electrodes were concatenated into feature vectors. In the TDFD [18], DFD and DS methods [19], the fractal dimensions were estimated using the fractal dimension estimation method of the methodology with the best performance.

4.1.1. Katz’s Method

Katz’s method calculates the fractal dimension of a sample as follows: The sum and average of the Euclidean distances between the successive points of the sample (*L* and *a*, respectively) are calculated as well as the maximum distance between the first point and any other point of the sample (*d*). The fractal dimension of the sample (*D*) then becomes:

$$D = \frac{\log(L/a)}{\log(d/a)} = \frac{\log(n)}{\log(n) + \log(d/L)} \qquad (10)$$

Where n is L divided by a.

4.1.2. Higuchi’s Method

Higuchi’s method calculates the fractal dimension of a sample as follows: First, subsample sets (X_k) are constructed from the sample (X) as:

$$X_k^m = \{X(m+ik)\}_{i=0}^{[(N-m)/k]} \qquad (11)$$

Where $k \in [1, k_{\max}], m \in [1, k]$ and N is the sample size.

Then, the length of each $X_k(L_m)$ is calculated as:

$$L_m(k) = \frac{\left(\sum_{i=1}^{[(N-m/k]} |X(m+ik) - X(m+(i-1)k)| (N-1) / ([(N-m)/k]k) \right)}{k}$$

(12)

Finally, the fractal dimension of the sample (D) is solved from:

$$\langle L(k) \rangle \propto k^{-D} \qquad (13)$$

Where <L> is the average of L_m. Three k_{max} values from the range of 8 to 18 [20] (8, 13 and 18) were tested.

4.1.3. R/S Method

The R/S method calculates the fractal dimension of a sample by iteratively dividing it into non-overlapping subsamples with decreasing subsample size and performing the following operations at each iteration: For each subsample, a new subsample (X) is constructed from its zero mean (ξ) such that the *n*th point of X is the cumulative sum of the first n points of ξ. Then, the difference between the maximum and the minimum values, and the standard deviation of X (R and S, respectively) are calculated in order to obtain their ratio (R/S). Finally, R/S of each X is averaged ($(R/S)_{avg}$). After obtaining $(R/S)_{avg}$ at each iteration, the Hurst exponent (H) becomes the slope of the log-log plot of $(R/S)_{avg}$ versus subsample size. The fractal dimension then becomes 2 – H.

4.1.4.. TDFD Method

In TDFD method, a window (with size s) is slid over a sample by a time step and the fractal dimension of the part of the sample inside the window is estimated. The fractal dimensions were concatenated into feature vectors. Different window sizes were tested using a time step of one second.

4.1.5. DFD and DS Methods

The DFD method is a variation of the DS method. In the DFD method, first, the fractal dimensions of the samples from selected electrodes are estimated and then, the pairwise differences of the fractal dimensions are calculated. However, in the DS method [19], first, the pairwise differences of the samples from selected

International Journal of Information and Communication Technology Research

http://www.esjournals.org

electrodes are calculated and then, the fractal dimensions of the pairwise differences are estimated. In both methods, the resultant values were concatenated into feature vectors.

5. CLASSIFICATION APPROACHES

a. Support Vector Machine

The SVM is a relatively new classification technique developed by Vapnik [21] which has shown to perform strongly in a number of real-world problems, including BCI.

The invention of SVM was driven by underlying statistical learning theory, i.e., following the principle of structural risk minimization that is rooted in VC dimension theory, which makes its derivation even more profound. The SVMs have been a topic of extensive research with wide applications in machine learning and engineering.

b. Linear Discriminant Analysis (LDA)

In this paper, we used linear transformation since linear discriminant analysis frequently achieves good performances in the tasks of face and object recognition, even though the assumptions of common covariance matrix among groups and normality are often violated [22] In addition, kernel tricks can be used with linear discriminant analysis for non-linear transformation [23]. The basic idea of LDA is to find a linear transformation that best discriminate among classes and the classification is then performed in the transformed space based on some metric such as Euclidean distance. Mathematically a typical LDA implementation is carried out via scatter matrix analysis [24].

6. RESULTS

The classification accuracies (Table 1 and Table 2) were evaluated for each fractal dimension calculation method and classifier combination. Katz's method was the fastest method and combining it with KNN, the highest classification accuracy of 89.87% as well as the second highest classification accuracy of 88.79% were achieved using SVM and wavelet transform. On the other hand, R/S method with any classifier performed the worst with the classification accuracies. The performances of the rest of the combinations were similar (Table 1 and 2).

7. CONCLUSION

Since all fractal dimension estimation methods are not applicable to all types of data exhibiting fractal properties, commonly used fractal dimension estimation methods to characterize time series with different classifiers were evaluated to find the most suitable method for EEG data. Katz's method with KNN was determined to be the best methodology and the results. The results warrant further research to use this methodology in online analysis of EEG data and analysis of other signals.
The joint analysis of fractal dimension estimation methods demonstrated the validity of our proposed method for feature extraction integrating coherent average, principal component analysis (PCA), independent component analysis (ICA). It would provide useful information to develop new BCI system.

Table 1. results of the dataset type I

FEATURES	FRACTAL DIMENSION			WAVELET TRANSFORM
CLASSIFIER	*Katz's Method*	*Higuchi's Method*	*R/S Method*	*Daubechies wavelet*
LDA	89.73%	84.19%	81.45%	85.63%
SVM	87.43%	86.01%	82.15%	87.25%

Table 2. results of the Dataset type II

FEATURES	FRACTAL DIMENSION			WAVELET TRANSFORM
CLASSIFIER	*Katz's Method*	*Higuchi's Method*	*R/S Method*	*Daubechies wavelet*
LDA	89.87%	85.61%	80.05%	85.75%
SVM	87.93%	88.25%	82.92%	88.79%

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Acknowledgment

This study was supported by Islamic Azad University, Shahr-e-Qods Branch, Iran. The authors would like to acknowledge staff of university.

REFERENCES

[1] Pfurtscheller G, Neuper C. Motor imagery activates primary sensorimotor area in man. Neurosci Lett 1997;239:65–8.

[2] Müller-Putz GR, Scherer R, Pfurtscheller G, Rupp R. EEG-based neuroprosthesis control: a step towards clinical practice. Neurosci Lett 2005;382:169–74.

[3] Kübler A, Kotchoubey B, Hinterberger T, Ghanayim N, Perelmouter J, Schauer M, et al. The thought translation device: a neurophysiological approach to communication in total motor paralysis. Exp Brain Res 1999;124:223–332.

[4] Duffau H. Brain plasticity: from pathophysiological mechanisms to therapeutic applications. J Clin Neurosci 2006;13(9):885–97.

[5] Daly JJ, Wolpaw JR. Brain–computer interfaces in neurological rehabilitation. Lancet Neurol 2008;7(11):1032–43.

[6] Cooper RA, Dicianno BE, Brewer B, LoPresti E, Ding D, Simpson R, et al. A perspective on intelligent devices and environments in medical rehabilitation.MedEng Phys 2008;30(10):1387–98.

[7] Xu M, Takata H, Ge S, Hayami T, Yamasaki T, Tobimatsu S, et al. NIRS measurement of hemodynamic evoked responses in the primary somatosensory cortex by finger stimulation. In: IEEE/ICME international conference on complex medical engineering; 2007. p. 1425–9.

[8] Ciuciu P, Abry P, Rabrait C,WendtH. Log wavelet leaders cumulant based multifractal analysis of EVI fMRI time series: evidence of scaling in ongoing and evoked brain activity. IEEE J Sel Top Signal Process 2008;2(6):929–43.

[9] Ikeda A, Shibbasaki H. Invasive recording of movement-related cortical potentials in humans. J Clin Neurophysiol 1992;9(4):409–520.

[10] Keirn ZA, Aunon JI. A new mode of communication between man and his surroundings. IEEE Trans Biomed Eng 1990;37(12):1209–14.

[11] Abdolreza Asadi Ghanbari, Karim Adinehvand, Mousa Mohammad Nia," Overhead Reduction in EEG signals using Particle Swarm Optimization and Independent Component Analysis", *International Journal of Information and Communication Technology Research Volume 4 No. 5, May 2014*

[12] Mason SG, Birch GE. A general framework for brain–computer interface design. IEEE Trans Neural Syst Rehabil Eng 2003;11:70–85.

[13] BCI Competition 2003. http://ida.first.fraunhofer.de/projects/bci/competition.

[14] H. Aapo, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (4–5) (2000) 411–430.

[15] M.J. Katz, Fractals and the analysis of waveforms, Computers in Biology and Medicine, vol. 18, 1988, pp. 145- 156.

[16] T. Higuchi, Approach to an irregular time series on the basis of the fractal theory, Physica D Nonlinear Phenomena, vol. 31, 1988, pp. 277-283.

[17] Hurst, H.E, Long-term storage capacity of reservoirs, American Society of Civil Engineers, vol. 116, 1951, 770-799.

[18] S. Sabanal and M. Nakagawa, A study of time-dependent fractal dimensions of vocal sounds, Journal of the Physical Society of Japan, vol. 64, 1995, pp. 3226-3238.

[19] I. Takuya and N. Masahiro, An application of EEG analyses based on fractal theory to emotion information orocessing, IEIC Technical Report (Institute of Electronics, Information and Communication Engineers), vol. 104, 2005, pp. 53-58.

[20] S. Spasic, A. Kalauzi, M. Culic, G.Grbic, L. Martac, Estimation of parameter kmax in fractal analysis of rat brain activity, Annals of the New York Academy of Sciences, vol. 1048, 2005, pp. 427-429.

[21] S. Avidan, "Support Vector Tracking," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp.1064-1072, Aug. 2004.

[22] R. O. Duda, et al. (2001). Pattern Classification. John Wiley & Sons, Inc.

[23] S. Mika, et al. (1999). 'Fisher Discriminant Analysis with Kernels'. In Y.-H. Hu, J. Larsen, E. Wilson, & S. Douglas (eds.), Neural Networks for Signal Processing IX, pp. 41–48. IEEE.

[24] K. Fukunaga (1990). Introduction to statistical pattern recognition. Academic Press.