Proceedings of the 4th International
IEEE EMBS Conference on Neural Engineering
Antalya, Turkey, April 29 - May 2, 2009

FrD3.6

Minimising Prediction Error For Optimal Nonlinear Modelling of EEG Signals Using Genetic Algorithm

Tugce Balli and Ramaswamy Palaniappan

School of Computer Science and Electronic Engineering
University of Essex
Colchester, United Kingdom
<tballi, rpalan>@essex.ac.uk

Joydeep Bhattacharya[1,2,3]
[1]Department of Psychology, University of London
London, United Kingdom
[2]Goldsmiths College, Vienna, Austria
[3]Comission for Scientific Visualization,
Austrian Academy of Sciences,
Tech Gate, Vienna, Austria
j.bhattacharya@gold.ac.uk

Abstract— **Genetic algorithm (GA) is used for jointly estimating the embedding dimension and time lag parameters in order to achieve an optimal reconstruction of time series in state space. The conventional methods (false nearest neighbours and first minimum of the mutual information for estimating the embedding dimension and time lag, respectively) are also included for comparison purposes. The performance of GA and conventional parameters are tested by a one step ahead prediction modelling and estimation of dynamic invariants (i.e. approximate entropy). The results of this study indicated that the parameters selected by GA provide a better reconstruction (i.e. lower root mean square error) of EEG signals used for a Brain-Computer Interface (BCI) application. Additionally, GA based parameters are found to be computationally less intensive since both parameters are jointly optimised. In order to further illustrate the superiority of the embedding parameters estimated by GA, approximate entropy (ApEn) features using embedding parameters estimated by GA and conventional methods were computed. Next these ApEn features were used to classify the EEG signals into two classes (movement and non-movement) for BCI application. These results show that the embedding parameters estimated by GA are more appropriate than those estimated by the conventional methods for nonlinear modelling of EEG signals in state space.**

Keywords-component; EEG; Embedding dimension; Genetic Algorithm; Nonlinear Prediction Error; State space reconstruction; Time lag

I. INTRODUCTION

Traditionally, linear methods are used for the analysis of electroencephalogram (EEG) signals. Although linear analysis is simple to implement and easier to interpret, it can only approximate the underlying nonlinear properties of EEG signals. Hence, nonlinear analysis methods would be a logical step in obtaining an improved characterisation of EEG signals.

The constructions of state space trajectory is a crucial step in the nonlinear analysis of EEG signals, where the univariate potential data (i.e. a single channel EEG) are transformed to its trajectory in a multidimensional state space by Taken's method of time delayed embedding for state space reconstruction [1].

However, the performance of the reconstruction procedure crucially depends on two embedding parameters: time lag (τ) and embedding dimension (m).

Many suggestions [2] are available for estimating these embedding parameters separately, such as false nearest neighbours for m and local minimum of correlation or of mutual information for τ. These estimations procedure are often independent of each other, yet little is known how to jointly estimate these two parameters. Considering an important role of embedding window (m-1)τ in time delay embedding [3], a joint estimation of m and τ seems more appropriate than two separate independent estimators.

In this study, we have used genetic algorithm (GA) with nonlinear prediction algorithm [4, 5] for jointly estimating the two embedding parameters. Here, the GA works by minimising the first order nonlinear prediction error for embedding dimension and time lag pairs; thus providing an optimal reconstruction that represents the dynamic evolution of time series in state space. The performances of embedding parameters were tested for both modelling and estimation of dynamic invariants by employing EEG signals used for a Brain-Computer Interface (BCI) application. The embedding parameters estimated by conventional methods were also included for comparison purposes.

II. DATA SET

The signals used in this study were acquired using a Guger Technologies g.BSamp EEG recording device with a sampling rate of 256Hz. The EEG signals were recorded over the motor cortex from five bipolar channels located at C3, C1, Cz, C2 and C4 as shown in Figure 1. The EMG signals (for labelling of movement and non-movement related EEG) were recorded from the flexors of the left arm and right mastoid was used as the reference.
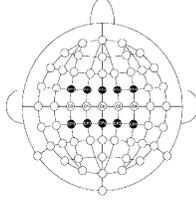
Figure 1. Electrode Layuot

Data were recorded from nine right handed subjects where two of the subjects were female and the ages of subjects ranged from 23 to 46.

In each run, the subjects were asked to perform self paced flexion or extension of the left index finger. They were instructed to perform each movement for 5 - 10 seconds and to rest for a minimum of 10 seconds between each trial. Each subject performed three runs in one session. Each run lasted for 610 seconds where the subjects had 5 seconds of pre-waiting and post-waiting periods before and after the fixation cross appeared on the screen for 600 seconds.

III.   METHODOLOGY

A.   Reconstructed State Space Analysis

Nonlinear time series analysis is essentially based on dynamical systems theory, according to which any dynamical system is represented by its state vectors and these states change in time**.** The evolution of states is defined by specific rules, often considered deterministic, in which the future states are defined as a function of previous states of the system. The transition of these states is defined by Equations (1) − (3), where $s(t)$ stands for the state of the system at time t, M is the representation of k dimensional state space and Φ is the evolution operator:

$$s(t+1) = \Phi(s(t)) \tag{1}$$

$$\Phi : M \to M \tag{2}$$

$$s(t) \in M \subseteq R^k \tag{3}$$

Suppose that a single scalar measure $\{x(t)\}$ can be measured at a time from the system using an observation function g such that; $x(t) = g(s(t))$, $g : M \to R$ and $t = 1$, $2,\ldots, N$. The observation function cannot provide a complete representation of the underlying properties of the dynamical system. According to Takens theorem [1] this can be achieved by representing time series as time lagged versions of itself such that;

$$f : M \to R^d \tag{4}$$

$$s(t) \to y(t) = f(s(t)) = [x(t), x(t-\tau), x(t-2\tau),\ldots \\ , x(t-(m-1)\tau)] \tag{5}$$

$$S = \begin{bmatrix} x(1) & x(1+\tau) & \cdots & x(1+(m-1)\tau) \\ x(2) & x(2+\tau) & \cdots & x(2+(m-1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-(m-1)\tau) & x(N-(m-1)\tau+1) & \cdots & x(N) \end{bmatrix}$$

where τ is the time lag, m is the embedding dimension and S is the complete representation of the single scalar time series in state space.

The selection of embedding dimension, m, and time lag, τ, parameters is important to achieve a good reconstruction of time series in state space.   In the following sections, the conventional methods of embedding parameter selections and the proposed method of selection will be described.

B.   Conventional Methods for Parameter Selection

The most commonly used methods for estimating the time lag are first zero crossing of the autocorrelation function and first minimum of the mutual information [2].

The autocorrelation function only detects the linear dependencies of time series data. However the mutual information detects both linear and nonlinear dependencies in the time series, for this reason we have employed the mutual information to estimate the time lag in this study.

The mutual information is a measure that quantifies the amount of information possessed about $x(t+\tau)$ based on $x(t)$. For estimation of time lag τ, mutual information is calculated as a function of τ and the τ value leading to the first local minimum of mutual information is chosen as the optimum value. If τ is too small each data point is too close together and the attractor tends to stretch out along the diagonal, on the other hand if τ is too large this leads to excessive folding of the attractor. The first local minimum mutual information (MMI) represents the time lag where $x(t)$ and $x(t+\tau)$ are maximally decorrelated.

The most common approach to the selection of optimum embedding dimension is false nearest neighbors (FNN) method. The idea behind FNN is that if the embedding dimension is not large enough the points in the embedding space become close due to projection (false neighbors) rather than system dynamics. The FNN method identifies the false neighbors by calculating the distance of nearest neighbors in dimension m and the distance of these neighbors in dimension $m+1$, if the ratio of the distances is larger than a threshold the neighboring states are identified as false neighbors. The embedding dimension is estimated by calculating the fraction of false neighbors for increasing embedding dimensions and the dimension where no false neighbors left is chosen as the optimum value.

C.   Parameter Selection Using Genetic Algorithm (GA)

In this study we used binary GA [6] for jointly estimating the embedding dimension m, and time lag parameters τ, for state space reconstruction. The candidate embedding

parameters were estimated by the GAs and the quality of reconstruction was assessed using nonlinear prediction error (NLPE). As the utilised data set was too long to process all at once, the EEG time series from each run was segmented into corresponding movement and non-movement classes (each with length of one-eight of a second) and one-third of the trials from each class, which were randomly selected, were used for training the GA. The embedding parameters m and τ were estimated using the algorithm presented below. The GA parameters used are as listed in Table 1. Our method was consisted of the following steps:

1. Generate the initial population of chromosome pairs (m and τ) with random binary values.

2. Calculate the fitness function f_i for each chromosome pair by minimising the mean nonlinear prediction error from all the segments.

3. Generate the new generation of chromosomes by selecting the chromosomes based on their fitness function – half of the new population was selected by tournament selection and the other half using roulette wheel selection methods.

4. Perform two point crossover between randomly selected chromosomes (with a crossover probability).

5. Perform mutation operation for randomly selected bit (with a mutation probability).

6. Repeat steps (3)-(6) until the maximum number of generations reached or the root mean square error falls below 0.001.

TABLE I. GENETIC ALGORITHM PARAMETERS

Length of chromosomes	4
Minimum embedding dimension value	1
Maximum embedding dimension value	15
Minimum time lag value	1
Maximum time lag value	15
Population size	40
Crossover probability	0.5
Mutation probability	0.01
Maximum generations	50

D. Performance Assesment of GA Based Embedding

First, we used the nonlinear prediction error as a fitness function which is a locally linear forecasting method that exploits the deterministic structure in the signal. This method works by deriving the local neighborhood relations from the data and using these relations to predict the future values [4]. By using this method we tried to obtain an embedding that spreads the data in the state space based on the deterministic dynamic evolution of the system [5]. The nonlinear prediction algorithm is as follows [2, 4]:

1. Divide the time series into training, $x_1,...,x_{tr}$, and testing sets, $x_{tr+1},...,x_N$

2. Reconstruct the time series with embedding dimension, m, and time lag of τ

3. Choose a delay vector x_i from the testing set for T step ahead prediction

4. Calculate the distance, d_{ij}, between the delay vector x_i and x_j, where $j=1,...,tr$

5. Order the distances d_{ij} and find k nearest neighbours $\mathbf{x}_{j(1)}$, $\mathbf{x}_{j(2)},...,\mathbf{x}_{j(k)}$ of $\mathbf{x}_i$ and fit a linear model to the state vectors of the form,

$$x_{j(l)} + T = a_o + \sum_{n=1}^{m} a_n * x_{j(l)-(n-1)*\tau}$$

where T is the number of steps ahead for prediction, and $l=1,...,k$. Note that the coefficients $a_0,...,a_m$ can be calculated using an ordinary recursive least squares algorithm. We chose one step ahead prediction, $T=1$.

6. Use fitted model to estimate x_{i+T} for the test vector x_i

7. Repeat steps (3)-(7) for all vectors in the test set and compute the final prediction error $E = \sqrt{\dfrac{1}{N} \sum_{i=tr}^{N} e_i^2}$

Next, we assessed our proposed method in terms of approximate entropy (ApEn). Since the primary aim of time delay embedding is to estimate dynamic invariants such as Approximate Entropy, Kolmogorov-Smirnov (KS) entropy, correlation dimension and largest Lyapunov exponents, it makes more sense to show direct comparison in any of these invariants. Accordingly ApEn is estimated using the state space vectors reconstructed with corresponding GA and standard parameters. For further demonstration of the performance comparison, the seperability of estimated ApEn feature for movement and non-movement EEG segments are investigated.

IV. RESULTS AND DISCUSSION

The mean and the standard deviation of root mean square error for one step ahead prediction using GA based parameters and standard methods based parameters is shown in Figure 2. The graphs clearly illustrate that the GA parameters improved the one step ahead prediction, demonstrating that the dynamic reconstruction is improved using the GA parameters compared to standard parameters.

Next we calculated ApEn using a window of one-eighth second with an overlap of one-sixteenth second from each run. In addition to GA and FNN & MMI parameters, the ApEn was also calculated for embedding dimension ranging from 1 to 10 with a fixed time lag of 1 (based from the study in [8]). The separability of features (for the movement and non-movement classes) were tested using a ten fold linear discriminant analysis (LDA) classifier. The maximum LDA classification result along with associated electrode site and the mean of best classification from all channels are presented in Table 2. The separability of the features estimated with three groups of parameters were comparable demonstrating that the GA parameters were as good as conventional methods for estimation of dynamic invariants.
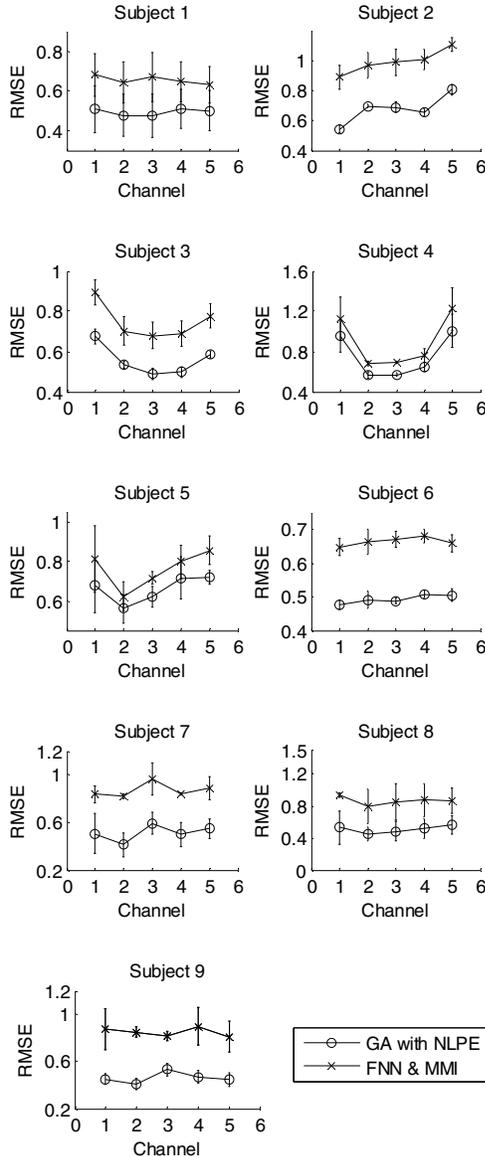
Figure 2. The mean and standard deviation of the root mean square error (RMSE) for one step ahead prediction with parameters estimated using genetic algorithm with nonlinear prediction error (GA with NLPE) and false nearest neighbors along with minimum mutual information (FNN & MMI) methods.

TABLE II.   LDA CLASSIFICATION WITH APPROXIMATE ENTROPY FEATURES

	FNN & MMI			Exhaustive Search			GA with NLPE		
Subject	Best Acc	Site	Mean Acc	Best Acc	Site	Mean Acc	Best Acc	Site	Mean Acc
1	81.63	Cz	76.67	74.60	C3	72.32	86.03	C3	81.25
2	78.35	C4	70.78	82.79	C4	74.14	77.99	C4	71.09
3	73.35	C1	72.55	73.13	C4	70.75	75.01	C4	75.01
4	72.37	Cz	70.02	73.74	C4	72.57	72.76	C4	64.77
5	64.58	Cz	63.32	62.38	C1	62.28	67.03	C3	62.75
6	67.24	Cz	64.61	63.68	Cz	62.71	67.37	C2	63.18
7	60.93	C4	59.93	62.20	C4	60.04	62.07	C4	60.50
8	60.99	C1	56.72	60.57	C1	59.05	61.28	C4	60.37
9	64.73	C4	62.11	66.43	C4	63.20	63.53	C4	60.73
Average	69.35		66.30	68.83		66.34	70.34		66.62

V.   CONCLUSION

In this study, the problem of optimal embedding parameter selection was approached using GA with NLPE. The proposed method's performance compared to standard methods like false nearest neighbours, for selection of embedding dimension, and first minimum of mutual information, for selection of time lag, was tested from both modelling and estimation of dynamic invariants perspective. However our primary aim was selection of parameters that will provide optimal reconstruction of the underlying dynamics of the time series (i.e. modelling).

The results of one step ahead prediction showed that the GA parameters were able to achieve a better prediction compared to standard methods. This demonstrates that these parameters are able to provide a better model of the dynamics compared to standard methods that aim to achieve a topological unfolding in the state space.

The performances of the methods were further illustrated by estimating the ApEn feature, where the features were used for measuring the class separability for movement and non-movement EEG segments. The results of classification showed that the performance of features estimated using GA parameters were comparable to standard parameters.

In the literature, the standard methods have been successfully applied to the synthetic time series such as Lorenz and Henon attractors. However the estimation of embedding parameters for noisy signals such as EEG and ECG has been a problem. In this study we have shown that the embedding parameters estimated by GA are more appropriate than those estimated by standard methods for nonlinear modelling of EEG signals in state space.

REFERENCES

[1] F. Takens. Detecting strange attractors in turbulence. Dynamical Systems and Turbulence, Lecture Notes in Mathematics, 898:366-381, 1981.

[2] H. Kantz and T. Schreiber. Nonlinear Time Series Analysis. Cambridge University Press, 2003.

[3] D. Kugiumtzis. State space reconstruction parameters in the analysis of chaotic time series – the role of the time window length. Physica D, 95: 13-28, 1996.

[4] M. Casdagli. Chaos and Deterministic versus Stochastic Non-linear Modelling. J.R. Statist. Soc., 54(2):303-328, 1991.

[5] M. Small. Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance. Nonlinear Science Series A, vol. 52, World Scientific, 2005.

[6] D. E. Goldberg. Genetic Algorithms in Search, Optimisation, and Machine Learning. Addison-Wesley, 1989.

[7] S.M. Pincus. Approximate entropy as a measure of system complexity. Prc. Natl. Acad. Sci., 88:2297-2307, 1991.

[8] M. Dyson, T. Balli, J.Q. Gan, F. Sepulveda and R. Palaniappan. Approximate Entropy for EEG Based Movement Detection. Proceedings of the 4rd International Brain-Computer Interface Workshop and Training Course 2008. Verlag der Technischen Universität Graz, 2008. To appear.