Evolutionary Fuzzy ARTMAP for autoregressive model order selection and classification of EEG signals

R. Palaniappan[+], P.Raveendran[+], Shogo Nishida[*] and Naoki Saiwaki[*]

[+]Faculty of Engineering, University Malaya, Kuala Lumpur, 50603 Malaysia

[*]Faculty of Engineering Science, Osaka University, Toyonaka, Osaka, 560 Japan

Abstract

A new technique of fusing genetic algorithms with Fuzzy ARTMAP is proposed. This method selects the appropriate autoregressive model order for EEG signals and consequently classifies these signals into their respective different mental tasks. The experimental results show that this method outperforms other statistical autoregressive model order selection methods like Akaike Information Criterion, Final Prediction Error and reflection coefficient.

1 Introduction

Linear autoregressive models (AR) have a broad spectrum of applications ranging from identification, prediction and control of dynamical systems but a problem with this method lies with the appropriate model order selection. As one can surmise, the AR model can be any order as desired. However, it should be as accurate as possible. From our intuition, we know that a model order, which is too small will not represent the properties of the signal, whereas a model order which is too high will also represent noise and inaccuracies and thus, will not be a reliable representation of the true signal. Therefore, methods that will determine the appropriate model order must be used.

In this paper, we propose a new method called Evolutionary Fuzzy ARTMAP (EFA) to select the appropriate AR model order for representing electroencephalogram (EEG) signals. The EFA network also classifies these signals for different mental tasks. The proposed EFA is actually a fusion of genetic algorithms (GA) and Fuzzy ARTMAP (FA) classifier where the optimum model order is selected by using GA, which uses FA classification as the fitness (objective) function. We show that this method outperforms other currently used methods, namely; Akaike Information Criterion (AIC), Final Prediction Error (FPE) and reflection coefficient (RC), which depend on regression analysis and nature of the modelling process.

2 Autoregressive systems and statistical model order selection methods

In this section, we'll discuss on AR modelling, AIC, FPE and RC, which are used to obtain the AR model order.

The EEG signals are represented by AR models of order p, which are given by

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n) \tag{1}$$

where p is the model order, $x(n)$ is the data of the signal at sampled point n, a_k are the AR coefficients and $e(n)$ represents the error term independent of past samples. To obtain a_k, we can solve the Yule-Walker equations directly using conventional linear equation solutions like Gaussian elimination but a shortcoming of this approach lies in its huge computational time. Thus, recursive algorithms have been developed which are based on the concept of estimating the parameters of a model of order p from the parameters of a model of order $p-1$. Some of these methods are Burg and Levinson-Durbin algorithm [12].

But before an AR model is proposed, we mentioned earlier that there is a prerequisite of having to know the order of the model. Methods pioneered by Akaike [1] are popular and two model order selection criteria developed by him i.e. AIC and FPE are based upon concepts in mathematical statistics.

Final Prediction Error (FPE) method gives the model order, which minimizes the function below

$$FPE = s_p^2 \frac{N+p+1}{N-p-1} \tag{2}$$

where p is the model order, N is the number of data points, s_p^2 is the total squared error divided by N and is given by

$$s_p^2 = \frac{1}{N-p} \sum_{p}^{N-1} e^2(n) \tag{3}$$

0-7803-6583-6/00/$10.00 © 2000 IEEE

The fractional portion of FPE increases with p and as such represent the inaccuracies in estimating the AR parameters.

The other criterion is called the Akaike's Information Criterion (AIC). Using this method, the order of the model is selected which minimizes the following function

$$AIC = N \ln s_p^2 + 2p \tag{4}$$

The term $2p$ represents the penalty for selecting higher orders.

Another concept used to select the model order is of the idea of partial correlation and this is sometimes known as refection coefficient. It is designated by $\pi_p = a_p$, the last coefficient of the AR model order p. It is a measure of the amount of correlation at lag p not accounted for by a $(p-1)$ order model. In this paper, we use a 95% confidence level and the order k is selected if $p \geq k$, where 95% of π_p values are less than $\pm 1.96 / \sqrt{N}$. After selecting the order of the model by the above criteria, we can proceed with the estimation of the AR coefficients. These coefficients are then used to obtain the power spectral density (PSD) function by using the equation

$$S(f) = \frac{s_p^2 T}{\left| \sum_{k=0}^{p} a_k e^{-i2\pi fkT} \right|^2} \tag{5}$$

where $S(f)$ represents the power spectral density function and T is the sampling period.

3 Fuzzy ARTMAP and Genetic Algorithms

This section introduces Fuzzy ARTMAP network and genetic algorithms. Fuzzy ARTMAP belongs to the ART family. There are several variations of these ART neural networks, namely ART1, ART2, ART3, Fuzzy ART, Fusion ART, Fuzzy ARTMAP and ART-EMAP. These systems were initially developed by Carpenter and Grossberg [3-5].

ARTMAP is a class of neural network that performs incremental supervised learning of recognition categories in response to input vectors presented in arbitrary order [5]. Earlier Adaptive Resonance Theory models like ART1 and ART2 consisted of unsupervised learning systems. In this paper, a more general ARTMAP system known as Fuzzy ARTMAP is used [4]. This system learns to classify inputs by using fuzzy set features i.e. the input features are from 0 to 1. This generalisation is accomplished by replacing the ART1 module of the binary ARTMAP system with Fuzzy ART module.

FA incorporates fuzzy set theory in its computation and as such it is able to learn stable responses to either analogue or binary valued input patterns. It consists of two Fuzzy ART modules (Fuzzy ART_a and Fuzzy ART_b) that create stable recognition categories in response to sequence of input patterns. During supervised learning, Fuzzy ART_a receives a stream of input features representing the pattern and Fuzzy Art_b receives a stream of output features representing the target class of the pattern. An Inter ART module links these two modules, which is actually an associative controller that creates a minimal linkage of recognition categories between the two Fuzzy ART modules to meet a certain accuracy criteria. This is accomplished by realising a Minimax Learning Rule that minimises predictive error and maximises predictive generalisation. It works by increasing the vigilance parameter ρ_a of Fuzzy ART_a by a minimal amount needed to correct a predictive error at Fuzzy ART_b. In this paper, we have used ρ_a of 0.9 for all the experiments.

Parameter ρ_a calibrates the minimum confidence that Fuzzy ART_a must have in a recognition category, or hypothesis that is activated by an input vector in order for Fuzzy ART_a to accept that category, rather than search for a better one through an automatically controlled process of hypothesis testing. Lower values of ρ_a enable larger categories to form and lead to a broader generalisation and higher code compression. A predictive failure at Fuzzy ART_b increases the minimal confidence ρ_a by the least amount needed to trigger hypothesis testing at Fuzzy ART_a using a mechanism called match tracking. Match tracking sacrifices the minimum amount of generalisation necessary to correct the predictive error. Match tracking leads to an increase in the confidence criterion just enough to trigger hypothesis testing which leads to a new selection of Fuzzy ART_a category. This new cluster is better able to predict the correct target class as compared to the cluster before match tracking.

Genetic Algorithms are a family of computational models inspired by evolution and are based on genetic processes of biological organisms. They are adaptive methods, which may be used to solve search and optimisation problems although the range of problems to which GA have been applied are quite broad. Over many generations, natural populations evolve according to the principles of natural selection and "survival of the fittest". By mimicking this process, GA are able to evolve solutions to real world problems, if they have been suitably encoded. For example, GA can be used in the online process control such as in a chemical plant or load balancing on a multi-processor computer system.

The basic principles of GA were first laid down by John Holland [8] and by students of Holland like DeJong [6]. GA work with a population of individuals each representing a possible solution to a given problem. Each individual is assigned a fitness score according to how good a solution is to the problem. For example, the fitness score might be the strength/weight ratio for a given bridge design. In nature, this is equivalent to assessing how effective an organism is at competing for resources. The highly fit individuals are given opportunities to reproduce, by cross breeding with other individuals in the population. This produces new individuals as offspring, which share some features taken from each parent. The least fit members of the population are less likely to get selected for reproduction and so die out.

A whole new population of possible solutions is thus produced by selecting the best individuals from the current generation and mating them to produce a new set of individuals. This new generation contains a higher proportion of the characteristics possessed by the good members of the previous generation. In this way, over many generations, good characteristics are spread throughout the population, being mixed and exchanged with other good characteristics as they go. By favouring the mating of more fit individuals, the most promising areas of the search space are explored. If the GA have been designed well, the population will converge to an optimal solution to the problem.

GA can deal with problems which are difficult for other techniques to solve due to their robustness but GA are not guaranteed to find the global optimum solution to a problem. However, they are generally good at finding solutions that are acceptably good. Therefore it is suitable for difficult areas where specialised techniques (which might be better and faster) do not exist. GA also fit well into improving existing systems.

During the reproductive phase of GA, individuals are selected from the population and recombined, producing offspring that will comprise the next generation. It starts with an initial population and selection is applied randomly from the initial population using a scheme that favours the more fit individuals (evaluated using a fitness function) to create the intermediate population. Good individuals will probably be selected several times in a generation while the poor ones may not be selected at all.

Despite the effectiveness of reproduction in increasing the percentage of superior representatives, the procedure is essentially sterile; it cannot create new and better strings. This function is left over to crossover and to a lesser but critical extent, to mutation. Crossover process is intended to simulate the exchange of genetic material that occurs during biological reproduction. Here, pairs in the breeding population are mated randomly with a crossover probability, ρ_c. In this paper, we have used a two-point crossover with ρ_c set at 0.3. In two-point crossover, chromosomes are regarded as loops formed by joining the ends together rather than linear strings as in single point crossover. To exchange a segment from one loop with that from another loop requires the selection of two cut points.

A surprisingly small role is played by mutation. Biologically, it randomly perturbs the population's characteristics, thereby preventing evolutionary dead ends. Most mutations are damaging rather than beneficial; mutation probability, ρ_m must be low to avoid the destruction of species. It works by randomly selecting a bit with a certain mutation probability in the string and reversing its value. The mutation probability is set at 0.03 in this paper.

4 Experimental Study

In this paper, we study the performance of five different types of order selection criteria for AR models to represent electroencephalogram (EEG) signals. The first three are AIC, FPE and RC. As a control, an experiment is also conducted with a fixed 6^{th} order model. Finally, we experiment with the proposed EFA model order selection system.

The subjects are seated in an Industrial Acoustics Company sound controlled booth with dim lighting and noiseless fans for ventilation. An Electro-Cap elastic electrode cap was used to record EEG signals from positions C3, C4, P3, P4, O1 and O2, defined by the 10-20 system of electrode placement [9]. The electrodes are connected through a bank of Grass 7P511 amplifiers and bandpass filtered from 0.1--100 Hz. The data was sampled at 250 Hz with a Lab Master 12 bit A/D converter mounted in an IBM-AT computer.

For this paper, the data from four subjects performing two different mental tasks are analysed. These tasks were chosen by Keirn and Aunon to invoke hemispheric brainwave asymmetry [11]. The tasks are:

- Math task, for which the subjects were given nontrivial multiplication problems, such as 72 times 38, and were asked to solve them without vocalizing or making any other physical movements;

- Geometric figure rotation, for which the subjects were asked to visualize a particular three-dimensional block figure being rotated about an axis.

Data was recorded for 10 seconds during each task and each task was repeated for two sessions. With a 250 Hz sampling rate, each 10 second trial produces 2,500 samples per channel. Overall, there are 16 different EEG files.

Each EEG signal is segmented with a half-second window, i.e. for a length of 125 points giving 20 patterns for each file with a total of 320 patterns. For all the experiments, 50% of available patterns are used for training while the rest 50% are for testing.

First, the different model order selection criteria are used to give the appropriate order of the model. Next, Burg algorithm is used (throughout the experiments) to derive the AR coefficients. This algorithm uses not only the forward prediction error but also the backward prediction error and minimises the sum of both of them. Therefore, in general it has proven to be a much more accurate method than other algorithms like Levinson-Durbin which uses only the forward prediction error. After this, we derive the PSD values in the range of δ, θ, α and β (i.e. 0-30 Hz) and using these spectral values, we train a FA network and classify the tasks into their respective categories. By following this procedure, we show the difference in the performance level of the different model order selection criteria for EEG signals.

Figure 1 shows the FA network structure as used in this paper. In the experimental study, we fixed the minimum order as 3 and the maximum order as 15. The lower limit is to avoid the PSD from being too flat when an extremely small model order is selected by any of the criteria. The higher limit will avoid the occurrence of peaks caused by spurious signals in the PSD function. The entire process is then repeated for the case with a fixed 6^{th} order AR model.

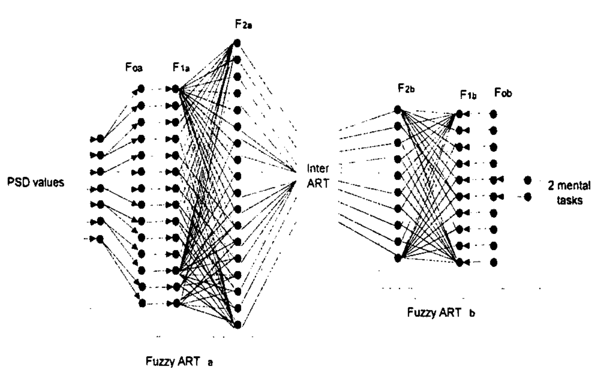

Figure 1: Fuzzy ARTMAP structure as used in the experimental study

Next, we obtain the appropriate model order given by the EFA network. Initially, GA generates randomly a total of 10 binary populations. Pairs of 4 binary numbers in each population are used to represent the model orders in each of the different patterns. Pairs of fours are chosen since they are sufficient to represent any number up to 15. Using these generated model orders, Burg AR coefficients are obtained; from which the spectral values are derived. The rest of the process of FA training and testing is similar to the earlier discussion. GA use the FA classification performance as a fitness function to reproduce, crossover and mutate the next generation of populations.

This entire cycle is iterated for 25 generations. Although 25 generations are quite low for general GA systems, we show that it is sufficient to improve the performance without putting too much load on the computational time. At the end of the 25^{th} generation, the optimum model orders for the different patterns are denoted by the population which gives the highest FA classification performance.

5 Results

Table 1 shows the classification performance of the EFA network and FA network with AIC, FPE and RC for 2 different mental tasks i.e. computing arithmetic and geometric figure rotation for the first session, second session and both sessions combined. FA classification results of another experiment with 6^{th} AR model are also shown as a comparison. The numbers of subjects are varied from 1 to 4. From these tabulated values, we can see that the EFA network gives a higher percentage of classification across all the subjects for all the experiments as compared to FA classification performance for AIC, FPE, RC and 6^{th} AR model. A percentage closer to 100 could have been obtained for all the EFA cases with more data for training the FA network and if the ocular artefacts like eye blinks are removed from the EEG data.

6 Conclusions

This paper proposes EFA, a successful combination of GA and FA. The proposed method is general and is applicable to any optimisation and/or classification problems. The results shows that the proposed EFA network can improve the EEG signal classification performance for different mental tasks as compared to statistical mathematical techniques like AIC, FPE and RC. This is since EFA selects the appropriate model order based on the improvement in classification and unlike the other statistical methods which selects the lowest appropriate model order based on the signal data only. The results also show that it is possible to recognise different mental tasks using EEG signals alone and this can be used, say as a mode of communication for paralysed patients [11].

Acknowledgement

We would like to thank Dr. Charles Anderson, Computer Science Dept., Colorado State University, USA for providing the EEG data.

Table 1: EFA and FA classification performance with AIC, FPE, RC and fixed order 6 for 1 to 4 subjects

Session 1				
	1 user	2 users	3 users	4 users
EFA	100	97.5	93.33	85
AIC	70	85	81.67	73.75
FPE	70	87.5	83.33	73.75
RC	80	87.5	85	78.75
6th order	70	85	88.33	78.75
Session 2				
	1 user	2 users	3 users	4 users
EFA	100	97.5	88.33	80
AIC	80	87.5	71.67	67.5
FPE	80	87.5	71.67	67.5
RC	85	82.5	81.67	65
6th order	70	92.5	76.67	70
Both session combined				
	1 user	2 users	3 users	4 users
EFA	87.5	80	74.17	71.25
AIC	75	66.25	66.67	56.88
FPE	75	65	66.67	60.63
RC	62.5	66.25	63.33	65.63
6th order	72.5	71.25	67.5	62.5

References

[1] H. Akaike, "Fitting Autoregressive Models for Prediction", *Annals of Institute Statistical Mathematics*, Vol.21, pp.243-247, 1969.

[2] J. Burg, "A new analysis technique for time series data", *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics*, Netherlands, August 1968.

[3] G. A. Carpenter and S. Grossberg, "A massively parallel architecture for a self-organizing neural pattern recognition machine," *Computer Vision, Graphics, and Image Processing*, Vol. 37, pp.54-115, 1987.

[4] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds and D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps", *IEEE Transactions on Neural Networks*, 3(5): 698-713, 1992.

[5] G.A.Carpenter, S. Grossberg, and J. H. Reynolds, "ARTMAP: Supervised real time learning and classification of nonstationary data by a self-organising neural network," *Neural Networks*, Vol. 4, 1991, pp. 565-588.

[6] K. DeJong, *The Analysis and Behaviour of a Class of Genetic Adaptive Systems*, Ph.D. thesis, University of Michigan, 1975.

[7] D. E. Goldberg, *Genetic Algorithm in Search, Optimization and Machine Learning*, Reading Mass., Addison–Wesley, 1989.

[8] J. H. Holland, *Adaptations in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, 1975.

[9] H. Jasper, "The ten twenty electrode system of the international federation", *Electroencephalographic and Clinical Neurophysiology,* 10: 371-375, 1958.

[10] G. Jenkins and D. Watts, *Spectral Analysis and Its Applications,* Holden-Day, San Francisco, 1968.

[11] Z. A. Keirn and J. I. Aunon, "A new mode of communication between man and his surroundings", *IEEE Transactions on Biomedical Engineering*, 37(12): 1209-1214, December 1990.

[12] R. Shiavi, *Introduction to Applied Statistical Signal Analysis,* Irwin and Aksen Associates, MA, 1991.