AUTOREGRESSIVE ORDER SELECTION CRITERIA: A CASE STUDY FOR EEG SIGNALS

R. Palaniappan and P.Raveendran

Dept. of Electrical and Telecommunication
Faculty of Engineering
University of Malaya
Kuala Lumpur 50603 Malaysia

Abstract: Autoregressive (AR) models have a broad spectrum of applications ranging from identification; prediction and control of dynamical systems and digital spectral analysis using these models have proven to be superior to classical Fourier transform techniques. However, there is a parameter that must be selected to utilize an AR technique properly i.e. the model order. Therefore, methods that will determine the appropriate model order must be used. In this paper, we perform a case study of the currently available statistical methods like Akaike Information Criterion, Final Prediction Error, Residual Variance, Minimum Description Length, Criterion Autoregressive Transfer and Hannan-Quinn. These methods depend on the statistical properties of the data, which selects the lowest order that is optimal to represent the signal. We use a Fuzzy ARTMAP (FA) neural network to study the performance of the different statistical criteria to select the appropriate AR model order for EEG signal representations. We perform this by training the FA to classify different mental tasks using the spectrum obtained from AR analysis with the selected model order. In addition to the other statistical methods, we have also studied the performance of a fixed 6th order AR model.

Keywords: Autoregressive, Model order, Fuzzy ARTMAP, Spectral analysis, EEG, Mental tasks.

I. INTRODUCTION

Autoregressive (AR) models have proven to be superior to Fourier methods due to the ability of AR models to handle short segments of data while giving better frequency resolution and smoother power. In addition, AR methods need only one or more cycles of sinusoidal-type activity to be present in the segment to produce good spectral peaks and they also provide the ability to observe small shifts in peak frequencies, which are not easily observed with Fourier derived spectra. The AR model coefficients can be easily estimated by solving a set of linear equations using the Yule-Walker method or solving recursively for higher orders using Levinson-Durbin or Burg method.

Successful applications of AR models are abundant in literature. It has been used in radar applications, geophysical applications, medical signal processing like Electroencephalogram (EEG) and Electrocardiogram (ECG) and speech processing [2,6,13].

However, there is a problem or rather a parameter that must be selected in order to utilize an AR technique properly i.e. the model order. As one can surmise, the AR model can be any order as desired. However, it should be as accurate as possible in terms of signal representation. 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 perform a case study of the currently available statistical methods like Akaike Information Criterion, Final Prediction Error, Residual Variance, Minimum Description Length, Criterion Autoregressive Transfer and Hannan-Quinn. These methods depend on the statistical properties of the data, which selects the lowest order that is optimal to represent the signal. We use a Fuzzy ARTMAP neural network to study the performance of the different statistical criteria to select the appropriate AR model order for EEG signals. We perform this by training the FA to classify different mental tasks using the spectrum obtained from AR analysis with the selected model order. In addition to the other statistical methods, we have also studied the performance of a fixed 6th order AR model.

II. AUTOREGRESSIVE SYSTEMS

A real valued, zero mean, stationary, non-deterministic, autoregressive process of order p is given by

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

where p is the model order, $x(n)$ is the data of the signal at sampled point n, a_k are the real valued AR coefficients and *e(n)* represents the error term independent of past samples.

The most common method of estimating a_k is to use the autocorrelation technique of solving the Yule-Walker equations [3,10]. 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 like Burg's algorithm [4] and Levinson–Durbin algorithm [3].

Burg's method is more accurate than Levinson-Durbin since it uses the data points directly unlike the latter method, which relies on the estimation of the autocorrelation function, which is generally erroneous for small data segments. The earlier method also uses more data points simultaneously by minimizing not only a forward error (as in the Levinson-Durbin case) but also a backward error. Details of this algorithm can be found in [4].

III. STATISTICAL MODEL ORDER SELECTION METHODS

We mentioned earlier that before an AR process could be used, there is a prerequisite of having to know the order of the model. Most order selection criteria are transformations of the mean squared error[1], σ_e^2 which is computed as a function of the order in model estimation. These techniques employ a multiplication of this error and a cost function, which increases monotonically with order *p*. Methods pioneered by Akaike [1] are popular and two model order selection criteria developed by him i.e. Akaike Information Criterion (AIC) and Final Prediction Error (FPE) are based upon concepts in mathematical statistics. FPE method gives the model order, which minimizes the function below

$$FPE(p) = \sigma_e^2(p)\frac{N+p+1}{N-p-1}, \tag{2}$$

where *p* is the model order, *N* is the number of data points, $\sigma_e^2(p)$ is the estimated error variance for the model.

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

Akaike then extended this model selection criterion to any maximum likelihood situation. Using AIC, the order of the model is selected which minimizes the following function

$$AIC(p) = N\ln\sigma_e^2(p) + 2p. \tag{3}$$

The term *2p* represents the penalty for selecting higher orders. The two criteria are asymptotically equivalent and in the limit of large N, FPE and AIC will predict the same optimal order [12].

Residual Variance (RV) criterion is based on the fact that if an insufficient number of terms have been fitted in the AR model given by Eq. (1), the estimate of the error variance $\sigma_e^2(p)$ will be inflated by those terms not yet included. Only when the correct number of terms has been included will a valid estimate of $\sigma_e^2(p)$ be obtained. Jenkin and Watts [10] suggest that if the Residual Variance (RV) estimate

$$RV(p) = \frac{N-p}{N-2p-1}\sigma_e^2(p), \tag{4}$$

is plotted versus *p*, the curve will flatten out or show a minimum at the point corresponding to the correct order of the AR process.

Since AIC has a tendency to overestimate the optimal order, Rissanen [15] has suggested MDL, which is given by

$$MDL(p) = \ln(\sigma_e^2(p)) + \frac{\ln(N)}{N}p. \tag{5}$$

MDL increases the penalty factor for higher orders as compared to AIC and as such favors the selection of lower orders.

Another criterion which counteracts the overfitting tendency of AIC is given by Hannan and Quinn [7]. HQ criterion chooses the minimum of

$$HQ(p) = \ln(\sigma_e^2(p)) + \frac{2\ln(\ln N)}{N}p. \tag{6}$$

Criterion Autoregressive Transfer (CAT) by Parzen [14] selects the order which minimize the criterion

$$CAT(p) = \frac{1}{N}\left(\sum_{k=1}^{p}\frac{1}{\sigma_e^2(k)}\right) - \frac{1}{\sigma_e^2(p)}. \tag{7}$$

[1] In this paper, mean squared error is used interchangeably with error variance. This is since the error is assumed to be white noise where the mean value is zero.

A point must be noted here. There are generally two different methods by which these criteria have been applied. The common method is to select the order, which gives the global minimum by using the specific criterion. The other method is to select the order, which gives the first minimum of the criterion. In this paper, we use the earlier method since it is more common and actually follows more closely to the original work behind these criteria.

After selecting the order of the model by the any one of the discussed criterion, we can proceed with the estimation of the AR coefficients using Burg's algorithm. For a detailed algorithm on Burg's algorithm, refer to [4]. These coefficients are then used to obtain the power spectral density (PSD) values by using the equation

$$S(f) = \frac{\hat{\sigma}_p^2 T}{\left| \sum_{k=0}^{p} a_k e^{-i2\pi fkT} \right|^2}, \tag{8}$$

where $S(f)$ represents the power spectral density function, T is the sampling period, $\hat{\sigma}_e^2(p)$ is the unbiased estimated variance of the residuals. The term T is included so that the true power of the corresponding analog signal will be represented digitally.

IV. FUZZY ARTMAP

This section introduces Fuzzy ARTMAP (FA) network. ARTMAP is a class of neural network that performs incremental supervised learning of recognition categories in response to input vectors presented in arbitrary order. 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. This system learns to classify inputs by using fuzzy set features i.e. the input features are from 0 to 1. This generalization 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 analog 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 realizing a learning rule that minimizes predictive error and maximizes predictive generalization. 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.

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 generalization 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 generalization 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. Further details of this method can be found in [5].

V. EXPERIMENTAL STUDY

In this paper, we study the performance of six different types of order selection criteria for AR models to represent EEG signals. These criteria are AIC, FPE, RV, HQ, CAT and MDL. In addition, an experiment is also conducted with a fixed 6^{th} order model since many authors like Anderson et. al. [2] and Keirn and Aunon [13] have used it in their experiments.

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 amplifiers and bandpass filtered from 0.1--100 Hz. The data was sampled at 250 Hz with a 12-bit A/D converter mounted on a computer.

For this paper, the data from three subjects performing two different mental tasks are analyzed. These tasks were chosen by Keirn and Aunon to invoke hemispheric brainwave asymmetry [13]. The tasks are:

- Math task, for which the subjects were given nontrivial multiplication problems, such as 22 times 49, 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. The patterns for each data set are chosen randomly at the beginning and are fixed for all the experiments. We fixed the maximum order as 15 for all the experiments to avoid the occurrence of peaks caused by spurious signals in the PSD function when a very high model order is used.

First, the different model order selection criteria like AIC, FPE, RV, HQ, CAT and MDL are used to give the appropriate order of the model. Next, Burg algorithm is used (throughout the experiments) to derive the AR coefficients. After this, we derive the PSD values in the range of 0-50 Hz per channel and using these spectral values (the PSD values for all the 6 channels are concanated into one vector), we train a FA network and classify the tasks into their respective categories. The entire process is then repeated for the case with a fixed 6th order AR model. By following this procedure, we are able to obtain the difference in the performance level of the different model order selection criteria for EEG signals. Figure 1 shows this procedure.

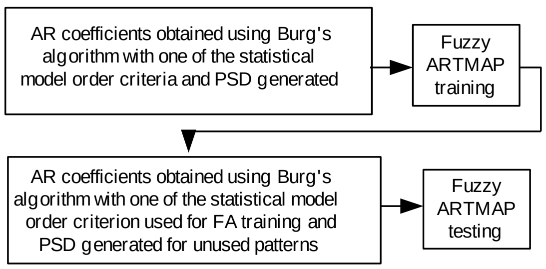

Figure 1: FA training and testing for evaluating the performance of statistical model order criteria

VI RESULTS

Table 1 shows the classification performance for subjects 1, 2 and 3. It also shows the average performance for all subjects. These results are for classification of two different mental tasks i.e. computing arithmetic and geometric figure rotation for combined two sessions. Experiments with Fuzzy ART_a vigilance parameter, ρ_a values of 0.0 are conducted. This is to maximize generalization ability and network compression. As far as the statistical model order criteria are concerned, it is difficult to conclude which criterion is best since the performance varies for different subjects. However, from the average results for all the subjects, we can see that MDL gives the best performance with a 88.89% classification. We are also able to conclude that subject 3 performs better than the other two subjects in most of the cases.

Table 1: Results of experimental study

Criteria	Subject 1	Subject 2	Subject 3	Average
AIC	79.17	70.83	95.83	81.94
FPE	79.17	75.00	91.67	81.95
RV	79.17	75.00	95.83	83.33
MDL	87.50	79.17	100.00	88.89
HQ	87.50	70.83	100.00	86.11
CAT	75.00	75.00	91.67	80.56
6th order	79.17	83.33	91.67	84.72

Finally, before we end the discussion on the results, we would like to point out the fact that a higher percentage can be obtained with more data for training and if the ocular artefacts like eye blinks are removed from the EEG data.

VII CONCLUSION

In this paper, we have studied EEG signal classification performance using AR models with fixed 6th order model and statistical techniques like AIC, FPE, RV, MDL, HQ and CAT. It is difficult to conclude which criterion is best since the performance varies for different subjects. However, the results also show that it is possible to recognize different mental tasks using EEG signals alone and this can be used, say as a mode of communication for paralyzed patients as suggested in [13] .

Acknowledgement

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

References

1. Akaike, H., “A New Look at the Statistical Model Identification,” *IEEE Transactions on Automatic Control*, vol. 19, pp. 716-723, December 1974.

2. Anderson, C.W., Stolz, E.A., and Shamsunder, S., “Multivariate Autoregressive Models for

Classification of Spontaneous Electroencephalogram During Mental Tasks," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 3, pp. 277-286, 1998.

3. Box, G.E.P., and Jenkins, G.M., *Time Series Analysis: Forecasting and Control,* Holden Day, San Francisco, 1976.

4. Burg, J.P, "A new analysis technique for time series data," *NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics,* Netherlands, August 1968. Also in Childers, D.G. (Ed)., *Modern Spectrum Analysis,* New York, IEEE Press, 1978.

5. Carpenter, G.A., Grossberg, S., and Reynolds, J.H. A Fuzzy ARTMAP Nonparametric Probability Estimator for Nonstationary Pattern Recognition Problems. *IEEE Transactions on Neural Networks*, vol. 6, no. 6, pp. 330-1336, 1995.

6. Childers, D.G. (Ed)., *Modern Spectrum Analysis,* New York, IEEE Press, 1978.

7. Hannan, E.J., and Quinn, B.G., "The determination of the order of an autoregression," *Journal of the Royal Statistical Society,* vol. B41, no.2, pp. 190-195, 1979.

8. Jansen, B.H., Bourne, J.R., and Ward, J.W., "Autoregressive Estimation of Short Segment Spectra for Computerized EEG Analysis," *IEEE Transactions on Biomedical Engineering*, vol. 28, no. 9, pp. 630-638, September 1981.

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

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

11. Jones, R.H., "Autoregression Order Selection," *Geophysics*, vol.41, pp.771-773, Aug. 1976. Also in Childers, D.G. (Ed), *Modern Spectrum Analysis*, New York, IEEE Press, 1978.

12. Jones, R.H., "Identification and Autoregressive Spectrum Estimation," *IEEE Transactions on Automatic Control*, vol. 19, pp. 894-897, December 1974. Also in Childers, D.G. (Ed), *Modern Spectrum Analysis*, New York, IEEE Press, 1978.

13. Keirn, Z.A., and Aunon, J.I., "A new mode of communication between man and his surroundings," *IEEE Transactions on Biomedical Engineering*, vol. 37, no.12, pp. 1209-1214, December 1990.

14. Parzen, E., "Some Recent Advances in Time Series Modelling," *IEEE Transactions on Automatic Control*, vol. 19, pp. 723-730, December 1974. Also in Childers, D.G. (Ed). *Modern Spectrum Analysis*, New York, IEEE Press, 1978.

15. Rissanen, J., "Modelling by shortest data description," *Automatica*, vol.14, pp.465-471, 1978.