Automatic Estimation of the Optimal AR Order for Epilepsy Analysis Using EEG Signals

Evangelia Pippa, Iosif Mporas, Vasileios Megalooikonomou

***Abstract*— In this paper, we propose a computationally efficient method to estimate the optimal order of the autoregressive (AR) modeling of electroencephalographic (EEG) signals in order to use the AR coefficients as features for the analysis of EEG signals and the automatic detection of epileptic seizures. The estimation of the optimal AR-order is made using regression analysis of statistical features extracted from the samples of the EEG signals. The proposed method was evaluated in both background and ictal EEG segments using recordings from 10 epileptic patients. The experimental evaluation showed that the mean absolute error of the estimated optimal AR order is approximately 4 units.**

I. Introduction

ELECTROENCEPHALOGRAPHY (EEG) is a technology developed to investigate the brain by measuring the electric fields generated by neurons in the cortex [1]. Electrodes, which act as the sensors to detect the electrical activity, are attached to the surface of the cerebral cortex and are connected by wires to a computer to capture the brain activities. EEG is a non-invasive procedure and provides information which is easily recorded and is obtainable using inexpensive technology compared to other brain activity diagnosis methods [2]. EEG has two main advantages, the direct measure of neuronal activity and the best temporal resolution by milliseconds [3]. In addition to the profound impact on our understanding of the brain [4],[5] EEG signals are essential in studying brain disorders. EEG signals are useful in analyzing the cognitive activity of the brain and diagnosing types of seizure and potential mental and neurological health problems, such as epilepsy [2].

Epilepsy is a chronic disorder of central nervous system that predisposes individuals to experiencing recurrent seizures [6]. Approximately 1% of the world population suffers from seizures, while more than 20% of the epileptic patients suffer from seizures that are refractory to medication [7],[8]. Epileptic seizures are brief episodes of abnormal excessive or synchronous neuronal activity in the brain of patients suffering from epilepsy [9].

Due to the difficulty of manual investigation of multi-parametric recordings and in combination with the progress of signal processing and pattern recognition technology, approaches for automatic detection of seizures have been proposed in the literature [6],[8],[10],[11]-18]. In many of the approaches found in the literature the autoregressive filter coefficients are used as features for the analysis [10],[12],[13],[16],[17]. This is owed to the fact that AR model parameters are invariant to scaling changes in the data that can arise from inter-subject variations, such as scalp and skull thickness [19] as well as AR reduce the spectral loss problems and gives better frequency resolution [20]. AR-based approaches require the estimation of the model order which should be such that the model estimated spectrum fits the signal spectrum [21]. Although in other signal processing areas there are practical rules for selecting the AR-order (e.g. in speech processing [22] AR-order is practically selected to be equal to the sampling frequency in kHz plus 4), in EEG signals no such practical knowledge has been reported.

In order to define the optimum model order the Akaike Information criterion (AIC) is usually used [12],[16]. In particular, the AIC criterion is computed for each candidate AR order and the order that minimizes the AIC criterion is chosen as optimum. However, such computations are time consuming and computationally inefficient. Furthermore, given the fact that the computation of AIC criterion requires the estimation of the error of the model and consequently the computation of the AR coefficients for each candidate AR order, such a practice is prohibitive in practical applications, especially in online applications, where a fast setup is needed [23].

In this paper, we present an efficient way to estimate the optimum order of the autoregressive model that avoids iterative and time consuming computations of the AIC criterion value. The proposed methodology estimates the optimum order using regression analysis on statistical features extracted from the EEG signal samples and can be used as a preprocessing step in applications where EEG signals are parameterized using autoregressive model coefficients. The proposed methodology is especially useful in online systems for automatic detection of epileptic events such as [24] where the computation time is crucial.

The rest of this paper is organized as follows. Section II defines the autoregressive model and presents the existing methods for obtaining the optimum AR order. In Section III

Manuscript received June 15, 2015. This research was partially supported by the EC-funded ARMOR Project (FP7-ICT-2011-5.1 - 287720) “Ad-vanced multi-paRametric Monitoring and analysis for diagnosis and Optimal management of epilepsy and Related brain disorders”, and Greek national funds through the Operational Program "Education and Lifelong Learn-ing" of the NSRF - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

E. Pippa is with the Dept. of Computer Engineering and Informatics, University of Patras, 26500 Rion-Patras, Greece (corresponding author to provide e-mail: pippa@ceid.upatras.gr).

I. Mporas, is with the Dept. of Computer Engineering and Informatics, University of Patras, 26500 Rion-Patras, Greece (e-mail: imporas@upatras.gr).

V. Megalooikonomou is with the Dept. of Computer Engineering and Informatics, University of Patras, 26500 Rion-Patras, Greece (e-mail: vasilis@ceid.upatras.gr).

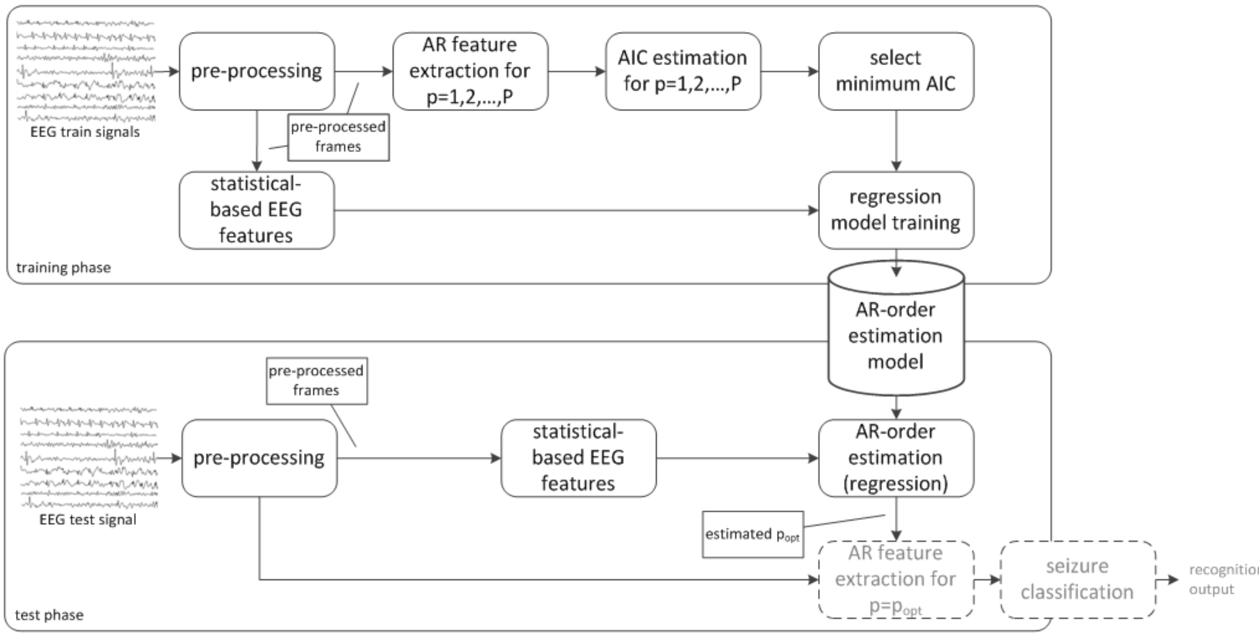

Fig.1. Proposed Architecture for the estimation of the AR order. The estimated optimal order is denoted as $\boldsymbol{p_{opt}}$.

the proposed methodology for AR-order estimation is presented. Section IV provides details about the evaluation data and the experimental protocol followed and presents the achieved results. Finally in Section V we conclude this work.

II. Background

The AR model of a single-channel EEG signal is defined as follows [25]:

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

where $a_k, k = 1,2, \ldots p,$ are the linear model parameters, p is the model order, n denotes the discrete sample time, and $\varepsilon(n)$ is white noise with zero mean and unity variance. Intuitively, the AR modeling method, expresses the signal with lagged terms of itself. In particular, the current value of the signal amplitude is given by the sum of the amplitudes of the previous samples and the estimation error.

The parameter p is the model order and is usually defined through trial and error. However, several criteria has been suggested in the literature to determine the optimum model order including Akaike information criterion (AIC), final prediction error (FPE) and CAT function [11]. When the model degree to be used is lower than the square root of the number of samples the AIC criterion is used to determine the optimum model order.

For an N-length data sequence, the optimum AR model order p is obtained by minimizing the AIC information criterion [21]:

$$AIC(p) = N ln(\rho_p) + 2p. \qquad (2)$$

where ρ_p is the model error variance defined as

$$\rho_p = \sum_{n=0}^{N-1} |e_p(n)|^2 \,. \qquad (3)$$

which can be determined using

$$e_p(n) = \sum_{k=0}^{p} a_p(k)\varepsilon(n-k). \qquad (4)$$

However, the computation of the AIC criterion for each candidate AR order requires the computation of the error of the model (and consequently the computation of the AR coefficients) for each AR order, which is time consuming and computationally inefficient. In the following section, we present a computationally efficient method for the automatic estimation of the optimal AR order avoiding the computation of the AIC criterion for each candidate AR order value.

III. Automatic Estimation of the Autoregressive Model Order

In order to avoid the computation of the AIC criterion for each candidate AR order, we propose the estimation of the optimal (or close to optimal) AR order using regression analysis on statistical measures extracted from the EEG signal samples under consideration. The block diagram of the proposed methodology is illustrated in Figure 1. Short time analysis is performed in single channel EEG data and regression models for the estimation of the optimum AR order are built.

During the training phase, a set of training data including EEG epochs from both background and ictal periods of the signal are used to build the regression models. Specifically, the single channel EEG data are initially frame blocked to epochs of constant length with constant time shift and without time overlap between successive epochs (frames). For each frame the AR coefficients of order 1 up to p are computed and based on the minimization of the AIC value one optimal value of the AR-order is assigned to each frame.

Afterwards, from each preprocessed frame statistical features based on the frame's samples are estimated. The extracted statistical features include the minimum value, maximum value, mean, variance, standard deviation, percentiles (25%, 50%-median and 75%), interquartile range, mean absolute deviation and range of the samples. The estimated (according to the AIC criterion) optimal order value is appended to the extracted statistical features and the corresponding feature vectors are used to train an AR-order estimation model using a regression algorithm.

During the test phase for each unknown test frame a statistical based feature vector identical to the training phase is computed. The test statistical-based feature vectors are processed by the AR-order model estimator, which using the regression algorithm provides an estimation of the optimal AR order, denoted as p_{opt}. Finally, for each test frame AR coefficients of order p_{opt} are computed in order to proceed to seizure classification. The last two steps are not in the case of the current study and thus are denoted in dashed lines. Note that the estimation of the optimal AR order is performed per frame since the EEG signal is a non-stationary signal.

The proposed architecture allows the estimation of optimal AR-order without the need for iterative and exhaustive computations, but using a regression model based on a bootstrap training set.

IV. Experimental Setup and Results

The previously described methodology for the estimation of AR order was evaluated on multi-parametric EEG recordings contained within the CHB-MIT database [6], which can be downloaded from the PhysioNet website http://physionet.org/physiobank/database/chbmit/. This database, which was collected at the Children's Hospital Boston, consists of EEG recordings from pediatric subjects with intractable seizures. Subjects were monitored for up to several days following withdrawal of anti-seizure medication in order to characterize their seizures and assess their candidacy for surgical intervention. For each clinical seizure, an expert indicated the earliest EEG change associated with the seizure. The data was segmented into one hour long records. The data set consists of 916 hours of continuous scalp EEG with sampling frequency 256 Hz. All data were stored in EDF formatted files. For the experiments we use healthy (control) and seizure epochs from 10 subjects of the CHB-MIT data base. The selected EEG channel was the CZ-PZ.

In a first step, the signal from the CZ-PZ channel was frame blocked to epochs of one second length, without time-overlap between successive epochs. Approximately the same number of normal and ictal epochs were isolated. Each epoch was parameterized using the statistical features described in Section III. Also, for each epoch the optimum AR order was calculated by minimizing the AIC criterion. Specifically, for AR order 1 to $p = 100$ the AIC criterion was computed. The AR order that corresponded to the minimum AIC value was selected as the optimum AR order. The optimal AR order of each epoch (frame) was used as a class label for the evaluation of the regression model.

TABLE I
AR order estimation performance in terms of MAE for several regression algorithms

Model	Mean Absolute Error
Additive Regression	4.311
Decision Table	4.4423
Linear regression	4.5077
M5P	4.295
M5Rules	4.2803
MLP	5.1033
REP Tree	4.2732
SMO Reg (Polykernel)	**4.2047**
SMO Reg (RBF kernel)	4.4715

TABLE II
AR order estimation performance in terms of MAE for several regression algorithms using contextual epochs

Model	Mean Absolute Error
Additive Regression	4.2887
Decision Table	4.4416
Linear regression	4.3472
M5P	4.1612
M5Rules	4.1953
MLP	4.6474
REP Tree	4.1166
SMO Reg (Polykernel)	**4.0739**
SMO Reg (RBF kernel)	4.2082

The computed feature vectors, were used to train regression models. In order to evaluate the ability of the statistical features to predict the optimum AR order, we examined several regression algorithms as implemented by the WEKA machine learning toolkit software [20]. The evaluated regression algorithms are: Additive Regression, Decision Table, Linear Regression, M5P, M5Rules, MLP, REP-Tree, SMO-Regression with polynomial kernel and RBF kernel. The free parameters of each regression algorithm were empirically selected.

For the evaluation we followed a 10-fold cross-validation protocol, in order to avoid overlap between training and test subsets. Table 1 shows the AR-order estimation performance in terms of mean absolute error (MAE), i.e. the absolute difference between the "ground truth" AR-order and the predicted one. As "ground truth" here we consider the optimal AR-order selected during the training phase, as described in Section 3.

As can be seen in Table 1, the overall lowest error of the proposed methodology is 4.2047 for the SMO-Reg regression algorithm with polynomial kernel. However, all algorithms present mean absolute error of AR-order that ranges from 4.2047 for the SMO-Reg algorithm to 5.1033 for the MLP algorithm with mean value approximately 4.4. These MAE are relatively low comparing to the range of AR-order values that have been found in the literature [12],[16],[27].

In a second step, we repeated the evaluation using as features not only statistics extracted from the epoch under consideration (current epoch), but also statistics extracted from the preceding and following epoch (current and two contextual epochs). The results are tabulated in Table 2. Once again SMO-Reg with polynomial kernel achieved the lowest mean absolute error. When features from both current and contextual epochs are used, all algorithms present a slightly lower prediction error (4.3 in average) compared to the initial case where features only from the current epoch were used. However, the improvement cannot be considered significant enough to change the selection of the order, comparing to the non-contextual case.

V. Conclusion

In this paper, we investigated the problem of estimation of the optimum AR order from EEG signals in an efficient way using statistical features and regression analysis. Examination of several regression algorithms showed the feasibility of our estimation methodology with low mean absolute error. The proposed methodology was evaluated with EEG data from 10 subjects and achieved AR-order estimation performance with mean absolute error approximately 4 units. The method has been tested using statistical-based features extracted from the EEG epochs under consideration as well as features extracted from both the current and the two adjacent epochs (preceding and following epochs).

Since preliminary analysis have shown that the accuracy of seizure detection can be significantly improved if the model order is optimized, we deem that the use of such a methodology for the estimation of the AR-order per epoch for each EEG channel separately can offer more robust fine-tuning of seizure classification schemes and improve their performance especially in real-time applications where time-consuming setups are not feasible.

References

[1] E. Niedermeyer and F. L. da Silva, "Electroencephalography: Basic Principles, Clinical Applications, and Related Fields," Lippincott Williams and Wilkins, 2005.

[2] K.S. Ng, H.J. Yang and S.H. Kim, "Hidden pattern discovery on event related potential EEG signals, vol. 97, Biosystems," 2009, pp. 15-27.

[3] S.A. Mousavi, M.R.H. Arshad, H.H. Mohamed and S.A. Alomari, "An efficient method for minin Event-Related Potential Patterns," vol. 8, International Journal of Computer Science Issues, 2011.

[4] G. Buzsaki, "Rhythms of the brain," Oxford University Press, 2006.

[5] B. Swartz and E. Goldensohn, "Timeline of the history of EEG and associated fields," vol. 106 Electroencephalography and clinical Neurophysiology, 1998, pp.173–176.

[6] A.H. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment," PhD Thesis, Harvard-MIT Div. of Health Sciences and Technology, 2009.

[7] J. Corsini, L. Shoker, S. Sanei and G. Alarcon, "Epileptic seizure predictability from scalp EEG incorporating constrained blind source separation," vol. 53, IEEE Trans. Biomed Eng.,2006, pp. 790-9.

[8] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S.T. Treves and J. Guttag, "Patient-specific seizure onset detection," vol. 5, Epilepsy Behavior, 2004, pp.483-98.

[9] R. S. Fisher, W. V. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee and J. Engel, "Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) ," vol. 46, Epilepsia, 2005, pp. 470-472.

[10] I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, M. Richardson, V. Megalooikonomou, "Seizure detection using EEG and ECG signals for computer-based monitoring, analysis and management of epileptic patients", Expert Systems with Applications, Vol. 42(6), pp. 3227–3233.

[11] B.R. Greene, G.B. Boylan, R.B. Reilly, P. de Chazal and S. Connolly, 2007. "Combination of EEG and ECG for improved automatic neonatal seizure detection", vol. 118 Clin Neurophysiol.,2007,pp. 1348-59.

[12] M.K. Kiymik, A. Subasi and H.R. Ozcalhk, "Neural Networks With Periodogram and Autoregressive Spectral Analysis Methods in Detection of Epileptic Seizure", vol. 28 Journal of Medical Systems, 2004.

[13] G. R. Minasyan, J.B. Chatten, M.J. Chatten and R.N. Harner, "Patient-Specific Early Seizure from Scalp EEG", vol. 27, Clinical Neurophysiology, 2010, pp. 163-178.

[14] H.R. Mohseni, A. Maghsoudi and M.B. Shamsollahim, "Seizure detection in EEG signals: a comparison of different approaches", In Conf. Proc IEEE Eng Med Biol Soc., 2006, pp. 6724-7.

[15] S. Nasehi and H. Pourghassem, "Seizure Detection Algorithms Based on Analysis of EEG and ECG Signals: a Survey", vol. 44, Neurophysiology, 2012, pp. 174-186.

[16] A.F. Rabbi and R.F.Rezai, "A fuzzy logic system for seizure onset detection in intracranial EEG", vol. 2012, Computational Intelligence and Neuroscience, 2012.

[17] A. T. Tzallas, P. S. Karvelis, C. D. Katsis, D. I. Fotiadis, S. Giannopoulos and S. Konitsiotis, "A method for classification of transient events in EEG recordings: application to epilepsy diagnosis," vol. 45, Methods Inf Med, 2006, pp. 610-621.

[18] M. Valderrama, S. Nikolopoulos, C. Adam, V. Navarro and M. Le Van Quyen. "Patient-specific seizure prediction using a multi-feature and multi-modal EEG-ECG classification", vol. 29, XII Med. Conf. on Medical and Biological Engineering and Computing, 2010, pp. 77-80.

[19] V. Lawhern, W. D. Hairston, K. McDowell, M. Westerfield, and K. Robbins, "Detection and classification of subject-generated artifacts in EEG signals using autoregressive models", Journal of Neuroscience Methods, Vol. 208 (2), pp. 181–189.

[20] H.R. Mohseni, A. Maghsoudi, M. B. Shamsollahi, "Seizure detection in EEG signals: a comparison of different approaches", In Proc. of the IEEE EMBS'06, pp. 6724-6727.

[21] S. Tong and N.V. Thakor, "Quantitive EEG Analysis methods and clinical applications," Engineering in Medicine and Biology, 2009.

[22] J.D. Markel, A.H.Gray, "Linear prediction of speech", Springer-Verlag, Berlin, 1976.

[23] I. Mporas, V. Tsirka, E.I. Zacharaki, M. Koutroumanidis, V. Megalooikonomou, "Online Seizure Detection from EEG and ECG signals for Monitoring of Epileptic Patients", Artificial Intelligence: Methods and Applications, LNCS, Vol. 8445, 2014, pp. 442-447.

[24] ARMOR project. Web-site: http:// www.armor-project.eu/.

[25] J.L. Semmlow and B. Griffel, "Biosignal and medical image processing", MATLAB-Based Applications, New York: Marcel Dekker, 2004.

[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA Data Mining Software: An Update," vol. 11, SIGKDD Explorations, 2009.

[27] D. Gorur, U. Halici, H. Aydin, G. Ongun and K. Levlebicioglu, "Sleep spindles detection using autoregressive modeling", ICANN 2003 and ICONIP 2003, LNCS, vol. 2714, Springer, Heidelberg, 2003