Improving EEG Signal Prediction via SSA and Channel Selection

Bahareh Atoufi  
Department of Electrical and Computer Engineering, University of Shahid Beheshti  
atoufi@gmail.com

Ali Zakerolhosseini  
Department of Electrical and Computer Engineering, University of Shahid Beheshti  
a-zaker@sbu.ac.ir

Caro Lucas  
Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, University of Tehran  
lucas@ut.ac.ir

Abstract

Being able to predict the coming seizure can impressively improve the quality of the patients’ lives since they can be warned to avoid doing risky activities via a prediction system. Here, a locally linear neuro fuzzy model is used to predict the EEG time series. Subsequently, this model is utilized in accompany with Singular Spectrum Analysis for prediction. Afterward, an information theoretic criterion is used to select a reliable subset of input variables which contain more information about the target signal. Comparison of three mentioned methods on one hand shows that SSA enables our prediction model to extract the main patterns of the EEG signal and highly improves the prediction accuracy. On the other hand, applying the method of channel selection to the model yields more accurate prediction. It is shown that fusion of some certain signals provides more information about the target and considerably improves the prediction ability.

1. Introduction

Epilepsy is a widespread neurological disease characterized by the unexpected occurrence of seizures. According to the seizure statistics of Epilepsy Foundation of America, 200,000 new cases of epilepsy are diagnosed each year and 10 percent of new patients fail to gain control of seizures despite optimal medical managements. Being able to predict the coming seizure can impressively improve the quality of the patients’ lives since they can be warned to avoid doing risky activities via a prediction system. As the electrical events that produce the symptoms occur in the brain, analyzing the Electroencephalogram (EEG) of the patients guides to identify characterized modifications of the brain activities that precede a seizure. Thus, EEG signal analysis has been proposed for early warning of the epileptic seizure. Several analysis techniques have been examined before such as neural and neuro-fuzzy models [1]. Such models as general function approximators [2] have performed well in the prediction of nonlinear and chaotic time series [3, 4]. However, the nonstationary characteristics of the EEG signal creates some problems. When the prediction horizon increases these models fail in accurate predicting of EEG time series [5] since there is usually a significant change in the frequency before and during the seizure. A frequency analysis such as Singular Spectrum Analysis (SSA) might succeed in dealing with this difficulty since it provides an insight to the unknown dynamics of the signal via extracting the Principal Components (PCs) of the signal. Here, basically, SSA is used in accompany with a locally linear neuro fuzzy model to examine the influence of such analysis. It is shown that SSA can considerably improve the accuracy of EEG signal prediction. Our next experiment includes examining the influence of fusing some signals from different channels to predict the signal of a certain channel. Our experiment shows that fusing signals given by different electrodes can provides an amount of information which is significant in predicting the channel signal since an improvement in prediction is observed after applying the fusion to the input of the predictor model. Our investigation of EEG signals reveals that predicting the signals of some electrodes can be done more precisely by means of some specific electrodes as predictors. Moreover, the number of EEG recording channels is normally high so in order to avoid entrance of redundant data in our
analysis we developed a method for channel selection. This method is based on Information theoretic and analyses the characteristics of channels signals from the nonlinear point of view. Hence, the method of channel selection plays an undeniable role in this part. In this study we aimed at distinguishing the characteristics of different channels' signals, their relationships with the predicted signal and, its reflection in nonlinear statistics applied to EEG signals. The outline of the paper is as follows: section 2 presents an explanation of methods we used for predicting signals. Section 3 looks into multi-channel prediction and describes the channel selection problem from a nonlinear attitude. The nonlinear analysis utilized is Mutual Information (MI) [6, 7] which is briefly described in Section 3. The experiment of developing our selection algorithms and the results are discussed in section 4. This section also includes some discussions about the results of our experiment and some conclusion remarks.

2. Method of prediction: LoLiMoT and SSA

The fundamental technique of locally linear neuro fuzzy model is dividing the input space into small subspaces with fuzzy validity functions. Each produced linear part with its validity function can be described as a fuzzy neuron. Thus, the whole model is a neuro fuzzy network with one hidden layer and a linear neuron in the output layer which simply calculates the weighted sum of the outputs of locally linear models [8].

The tree based methods of training neuro fuzzy networks are appropriate for their simplicity and intuitive constructive algorithm [8]. The Local Linear Model Tree (LoLiMoT) can be considered as a type of Takagi-Sugeno-Kang neuro fuzzy algorithm, which has proven efficient compared to other neuro fuzzy networks in learning nonlinear systems [3, 4, and 8]. LoLiMoT projects a piecewise linear model for which each linear model is welded by a nonlinear function. In other words, LoLiMoT divides the input space into local linear models which has a higher performance and needs lower neuron count compared to normal neural networks [9]. Partitioning of the input space is done via axis-orthogonal splits [2].

Generally, this incremental tree-construction algorithm consists of following iterative steps. The algorithm is initiated with a Locally Linear Model (LLM) with only one neuron. In the second step, the model with the maximum error value is selected for space division. All the divisions of this LLM on input space are constructed and checked. Finally, the best division (with the least estimation error) for the new neuron must be added [8]. Four iterations of LoLiMoT algorithm are illustrated in Figure 1.

![Figure 1. Four iterations of LoLiMoT](image)

As they are not able to distinguish the main patterns of natural time series with quasi periodic patterns such as EEG time series, neuro-fuzzy models fail in long term prediction of such time series [5]. To avoid this difficulty one solution is to use the SSA technique to predict the time series in long horizons. Assuming the time series \(X(t)\) as below, The \(N' p\) trajectory matrix \((D)\) of the time series has the \(M\)-dimensional vectors as its columns, and is obviously a Hankel matrix.

\[
X(t) = (X(t), X(t+1), ... , X(t+p-1)),
\]

\(t=1, ..., N', N' = N \cdot p + 1\) (1)

In the second step, the \(p \times p\) covariance matrix \(C_X\) is calculated as follow,

\[
C_X = \frac{1}{N'} DD^T D
\] (2)

The elements of diagonal matrix \(\Sigma = \text{diag} (\sigma_1, ..., \sigma_M)\) are the singular values of \(D\) and are equal to square roots of the eigen values of \(C_X\). The eigen elements \(\{\lambda_k, \rho_k\}:k = 1, ..., M\) of \(C_X\) are obtained from

\[
C_X \rho_k = \lambda_k \rho_k
\] (3)
Each eigenvalue, $\lambda_k$, estimates the partial variance in the direction of $\rho_k$, and the sum of all eigenvalues equals the total variance of the original time series. In the third step, the time series is projected onto each eigenvector, and yields the corresponding PCs:

$$A_k(t) = \sum_{j=1}^{p} X(t + j - 1) \rho_k(j)$$

(4)

Each of the PCs, being a nonlinear or linear trend, a periodic or quasi-periodic pattern, or a multi-periodic pattern, has narrow band frequency spectra and well-defined characteristics to be estimated. In the fourth step, the time series is reconstructed by combining the associated PCs:

$$R(t) = \frac{1}{p!} \sum_{k=1}^{p} \sum_{j=1}^{p} A_k(t - j + 1) \rho_k(j)$$

(5)

The normalization factor ($P_t$), and the lower ($L_t$) and upper ($U_t$) bounds of reconstruction procedure differ for the center and edges of the time series, and are defined by the following formula:

$$P_t = \left\{ \begin{array}{ll}
1 & 1 \leq t \leq p - 1 \\
1 - \frac{1}{t} & p \leq t \leq N' \\
1 - \frac{1}{N' + 1} - \frac{1}{N + p - 1} & N' + 1 \leq t \leq N
\end{array} \right. \quad (6)$$

SSA describes the main physical phenomena reflected by the data [10]. It also performs a data adaptive filtering in the lag coordinate space of data and yields the principal components (PCs) of time series which have narrow band frequency spectra and obvious temporal patterns [5]. The components obtained from SSA, most of which have linear or simple nonlinear behaviors, are long-term predictable [5]. Therefore, long-term prediction of the original time series is accessible via recombining the predicted components. The PCs related to lower singular values can be omitted in reconstruction stage to obtain adaptive noise cancellation. If all the components are used in reconstructing the time series, no information is lost. Most of the narrow band periodic components can be estimated via simple and optimal linear models, while there are always some more complex patterns which present nonlinear characteristics. Thus, in reconstructing the original time series from the PCs, one should use both linear and nonlinear techniques and also the linearity tests [5].

3. Multi-channel prediction

We use this concept for several channels as the model predictors. Accordingly, multi-channel prediction may lead to more accurate prediction as we have adequate number of input data for obtaining convincing output. It should be considered that fusion provides a large amount of information, full of redundancies, which should be removed. In other words, we should select useful information from the information provided by fusion. This can be performed by selecting most promising input signals and their lagged values among those available.

3.1. Channel Selection: Nonlinear Approach

Normally, several channels are used to collect the EEG data from a patient. Thus there are various signals to fuse. In such a case, providing sufficient information and also avoiding redundancy, is a point that should be considered. This may concern two different approaches (linear and nonlinear) toward selecting channels as predictors. According to [11], applying a linear analysis (Correlation Analysis) on EEG signals is not competent to describe the relations between channels as predictors. Thus, a nonlinear approach is used in order to reveal the special features of some EEG channels which are more trustworthy predictors. In other words, an information theoretic criterion (via utilizing the Mutual Information) is used to select a reliable subset of input variables with the richest information about the output that requires having a reliable prediction.

Nonlinear time series analysis techniques [12] have been developed to analyze and characterize apparently irregular behavior, a distinctive feature, of the EEG. These techniques mainly involve estimation of an effective correlation dimension, entropy related measures, Lyapunov exponent, measures for determinism, similarity, interdependencies, recurrence quantification as well as tests for nonlinearity. During the last decade, a variety of these analyzing techniques have repeatedly been applied to EEG records during physiological and pathological conditions and is shown to offer new information about complex brain dynamics. Nevertheless, nonlinear approaches to the analysis of the brain system have generated new clinical measures as well as new ways of interpreting...
brain electrical function, particularly with regard to epileptic brain states.

3.2. Mutual Information

In probability theory, especially in the information theory, Mutual Information can be used for evaluating the nonlinear dependencies between random variables. Indeed, the MI value between two random variables, such as $X$ and $Y$, can be considered as a measure of amount of knowledge on $Y$ provided by $X$ (or conversely on the amount of knowledge on $X$ provided by $Y$). The MI of two random variables $X$ and $Y$ is defined as:

$$I(X;Y) = H(X) - H(X | Y) = H(Y) - H(Y | X)$$

(7)

Where $H(X)$ and $H(Y)$ are the entropies of $X$ and $Y$, and $H(X|Y)$ is the conditional entropy, and $H(X;Y)$ is the joint entropy of $X$ and $Y$.

$$H(X) = -\int p_x(x) \log p_x(x) dx$$

$$H(Y) = -\int p_y(y) \log p_y(y) dy$$

$$H(X;Y) = -\int \int p_{X,Y}(x,y) \log p_{X,Y}(x,y) dxdy$$

(8)

4. Experiment and results

Currently, Freiburg University presented a unique EEG database with long-term recordings in the Third International Workshop on Epileptic Seizure Prediction. This set of EEG data contains signals recorded using surface electrodes (channels) from three patients during 36 hours. Different number of channels (60, 44 and 22) is used in order to record the data from the patients. The data is sampled at a rate of 512 Hz. A small partition (24 sec.) of the data recorded by 44 channels including preictal and ictal state and related to the patient number two is used in this study.

In order to determine the average predictability horizon of our time series maximum value of Lyapunov Exponent is estimated. The value obtained as the average predictability horizon for this certain part of the data is equal to 90. Thus, the maximum valid prediction can be applied for 90 steps ahead.

In the first phase of our experiment, a neuro-fuzzy network with LoLiMoT learning algorithm is applied to predict the time series. The hidden layer contains three neurons which obtained by trial and error. Initially, a number is chosen for the amount of neurons and then the optimized amount will be defined by comparing the diagrams of training error and testing error. If the error of testing increases, while the training error is decreasing, the number of neurons should be reduced in order to avoid over fitting. Also the Root Mean Square Error (RMSE) is considered as a standard index to compare the accuracy of predictions.

![Singular Spectrum of EEG time series by SSA algorithm](image1)

**Figure 2.** Singular Spectrum of EEG time series by SSA algorithm

![One Step ahead prediction of the EEG signal via (a) LoLiMoT and (b) SSA and LoLiMoT](image2)

**Figure 3.** One Step ahead prediction of the EEG signal via (a) LoLiMoT and (b) SSA and LoLiMoT

In our next experiment, SSA is applied in order to improve the prediction. Initially 30 principal components are extracted from the time series. Figure 2 depicts the singular spectrum of the time series. According to the changes in PC values 10 most important PCs are chosen to be predicted via LoLiMoT and subsequently the predicted components are combined to reconstruct the predicted time series. This
signal is then compared to the reconstructed signal from the original (not predicted) PCs to compute the prediction error. Figure 3(a) and 3(b) show the prediction of time series via LoLiMoT and SSA in accompany with LoLiMoT respectively.

The above experiment is repeated for different prediction horizons to study the influence of SSA during longer horizons. The result shows that SSA improves the prediction accuracy in all valid (less than 90) horizons. This noticeable improvement can be observed in Figure 6.

Next phase of our experiment is motivated to improve the SSA prediction. This improvement is made via fusing different signals from different EEG recording electrodes. In this part our goal is to define the three most efficient channels as predictors. Here the number of channels to be fused is chosen to be equal to three confirming that fusing only a few numbers of channels can improve the result of the prediction considerably and also avoids the redundancy caused by existing additional channels in predictors set. Initially, in order to choose the predictor signals, a method of selecting channels has been developed. The first step is to estimate the MI values between signals. In this part one of the recent estimators based on entropy and estimated from the $k$-nearest neighbor statistics [13], estimates the MI value between two random variables of any dimensional space. The basic idea is to estimate entropy from the average distance to the $k$-nearest neighbors (over all spans of data).

As described in [14] with a small value for $k$, this estimator has a large variance and a small bias, whereas a large value of $k$ leads to a small variance and a large bias [15]. The number of neighbors in our MI estimation algorithm is set to 20. Also the predictors are selected for a window of 50 signal samples. Then SSA is applied on selected predictors to extract 10 PCs from each predictor. Thus 30 PCs are available after applying the SSA to selected predictors. subsequently, 5 lagged values of each PC are chosen to enter the neuro-fuzzy network to predict each PC. In another words, for predicting each component 3 5 predictors are available as we select 3 channels and for each PC of each channel we use 5 lagged values. After predicting each PC via LoLiMoT, predicted PCs are combined to reconstruct the predicted signal. Steps of this algorithm are illustrated in Figure 4.

Figure 5 shows the prediction of the first component via SSA in multi channel and single channel methods. As it is illustrated, the multi channel prediction improves the prediction of first component. Thus, more accurate prediction of original signal is expected.

![Figure 4. The Model of Multi Prediction via SSA](image)

![Figure 5. Comparison of multi channel (upper plot) and single channel (lower plot) prediction methods results, in predicting the first component of the signal to be predicted](image)

Finally, the described experiments are repeated for various horizons (from 1 upto 90) and the prediction errors of three methods are shown in Figure 6. As it is observable, SSA shows an undeniable effect on the accuracy of the prediction. How ever fusing some certain signals with the predictor can improve the results and yields more accurate prediction.
5. References


Figure 6. RMSE of prediction via LoLiMoT, LoLiMoT in accompany with SSA and multi channel prediction via LoLiMoT and SSA.