A Survey of Multi-Channel Prediction of EEG Signal in Different EEG States: Normal, Pre-Seizure, and Seizure

Bahareh-Atoufi[1,2], Caro-Lucas[1], Ali-Zakerolhosseini[2]
[1] University of Tehran,
[2] University of Shahid Beheshti
Tehran, Iran
e-mail: atoufi@gmail.com

ABSTRACT
Epilepsy is a brain disorder that causes people to have recurring seizures. The ability to predict epileptic seizures not only would enable the use of novel approaches to seizure control but also can impressively improve the quality of the patients' lives. As the electrical events that produce the symptoms occur in the brain, analyzing the Electroencephalogram (EEG) of the patient guides to identify characterized modifications of the brain activities that precede a seizure. In this study a neuro fuzzy model is utilized to predict the EEG signal. Then this method is improved using multi channel prediction. Multi channel prediction is using several channels as predictors in the predicting system. Predictor signals must be the most relevant ones to predict the target signal. This means they should contain the most amount of information about the target. In order to find such predictors, a novel strategy of selection is presented in this paper which utilizes the nonlinear dynamics of the EEG signal. Also, the power of our selection method and the influence of multi channel prediction are evaluated by comparing the results of multi channel and single channel predictions.

Keywords
EEG signal prediction, LoLiMoT, multi channel prediction, Mutual Information.

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.

In order to analyze the EEG signal, 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]. The use of neural networks in engineering applications has increased dramatically over the last few years. Computational neural networks offer exciting advantages, such as learning, adaptation, fault- tolerance, parallelism and generalization [5] while fuzzy logic performs an inference mechanism under cognitive uncertainty [6]. To enable a system to deal with cognitive uncertainties in a manner more like humans, one may incorporate the concept of fuzzy logic into neural networks. The resulting hybrid system is called neuro fuzzy network. In hybrid form they i.e. fuzzy logic and neural networks, can provide a perfect platform to take into account changing knowledge [7].

Since there is usually a significant change in statistical properties of EEG signal [8, 9], especially before and during the seizures, we attempt to investigate the predicting ability of neuro fuzzy models in different states of the EEG signal. Normally, epileptic EEG signals contain normal state and abnormal state which is related to the epileptic seizures. However, in some cases, pre-seizure state can be found via applying an unsupervised fuzzy clustering algorithm to the EEG signal [10]. The initial objective of this paper is to illustrate the changes in neuro fuzzy models ability of predicting EEG signals during mentioned states.

As our second and main objective we use information fusion in order to improve the mentioned model's EEG signal prediction accuracy. As it will be described, this goal is obtained by multi-channel prediction of the signal. The main challenge in multi-channel prediction is selecting the channels which their signals should be used to construct the predictor.

In the field of EEG analysis the search for the hidden information predictive of an impending seizure has a long history. Nonlinear time series analysis techniques [11] have been developed to analyze and characterize apparently irregular behavior distinctive feature of the EEG. These techniques mainly involve estimates of an effective correlation dimension, entropy related measures, Lyapunov exponents, measures for determinism, similarity, interdependencies, recurrence quantification as well as tests for nonlinearity. Applying such nonlinear methods such as Mutual Information (MI) [12] concept may be useful to find the most relevant channels for multi-channel prediction of the EEG signal since it directly leads us to the channels which contain higher amount of information about the signal to be predicted. As more information means more predicting ability, we can get the best predictor channels through this method. Multi-channel prediction of EEG is then studied in different states of the signal in order to illustrate its influence on neuro fuzzy prediction in each state.
This paper is organized as follows: Section 2 illustrates the neuro fuzzy model used for prediction in our experiment. In Section 3, we describe the MI concept briefly and then present

our channel selection algorithm. Finally, Section 4 discusses the experimental results.

2. PREDICTION: LOLIMOT

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 [13].

The tree based methods of training neuro fuzzy networks are appropriate for their simplicity and intuitive constructive algorithm [13]. 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 13]. 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 [14]. 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 [13]. Four iterations of LoLiMoT algorithm are illustrated in Fig. 1.

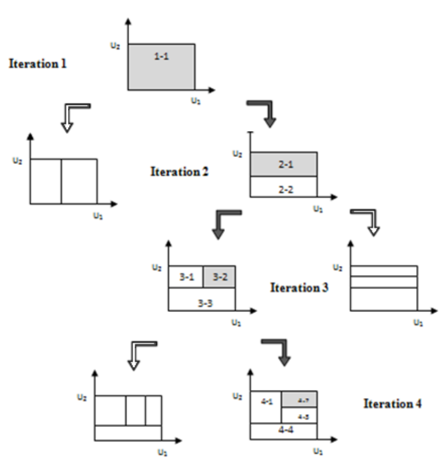


Figure 1. Four iterations of LoLiMoT

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

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, i.e. linear and nonlinear, toward selecting channels as predictors. According to [15], applying a linear analysis, i.e. 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 [11] 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, MI 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:

$$\begin{aligned} I(X;Y) &= H(X) - H(X \mid Y) = H(Y) - H(Y \mid X) \\ &= H(X) + H(Y) - H(X;Y) \end{aligned} \qquad (1)$$

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.

$$H(X) = -\int_X p_X(x) \log p_X(x) dx$$

$$H(Y) = -\int_Y p_Y(y) \log p_Y(y) dy$$

$$H(X;Y) = -\int_Y \int_X p_{X,Y}(x,y) \log p_{X,Y}(x,y) dx dy \qquad (2)$$

3.3. Selection Algorithm

In this section a new algorithm for channel selection using mutual information is presented. In the process of selecting channels, we would like to reduce the number of selected channels by excluding irrelevant or redundant channels among all EEG recording channels. This technique is very similar to what ***Battiti*** named as a "feature reduction" problem [16]:

> Given an initial set of n features, find the subset with $k < n$ features that is "maximally informative" about the class.

From the information theory, mutual information between two random variables measures the amount of common information contained in these variables [17]. The problem of selecting predictor channels (signals) with a high amount of information about the channel whose signal is to be predicted (target) can be solved by computing the mutual information between predictor channels and target channel. If the mutual information between these signals could be exactly obtained, the above problem could be reformulated as follows:

> Given an initial set F of all channels, find the subset $S \subset F$ that minimizes $H(C|S)$, i.e., that maximizes the mutual information $I(C; S)$.

Here C is the target signal.

The major deficiency of this strategy is that the result would be a set of selected predictors which contain high degree of redundancy. Since the MI between each of them and the target is high, it is very probable that we chose very similar signals, i.e. they have a lot in common. In order to solve this difficulty our algorithm selects the channels with the most amount of information about the target but with the least information about each other, i.e. the MI value between the predictors is low.

Here our strategy of selecting initial channel is 'generate and test' strategy. Initially, for each channel f and target channel C, $H(C|f)$ is computed. Then channel with the most H value is selected as the first channel. As a matter of course, this is optimal and it is feasible as the number of channels is limited.

From now on, we try to find a channel say f_j which maximizes the statement below:

$$I(C|f_j) - \Sigma I(f_j|S) \tag{3}$$

S is the set of channels which are already selected. Table 1 briefly illustrates our strategy of selection.

Table1. Algorithm Of selection.

1. (Initialization) set $F \leftarrow$ all channels, $S \leftarrow$ empty.
2. (Computation of MI with the target) $\forall f \in F$,
   Compute $I(C; f)$.
3. (Selection of the first channel) finding the channels that maximizes $I(C; f)$, omit f from F and add it to S.
4. (Greedy selection) repeat until desired number of channels are selected.
   4.1. (Compute the joint MI between variables) $\forall f \in F$, Compute $I(C; f)$ and $I(f; S)$.
   4.2. (Selection of the next channel) choose the channel $f \in F$, that maximizes $I(C; f)$ and minimizes $I(f; S)$, *i.e.* Maximize $I(C; f) - \beta \sum I(f; F)$. Omit f from F and add it to S.
5. Output the set S containing the selected channels.

Here β is a redundancy variable which is used to consider the redundancy between selected predictors. If $\beta = 0$, mutual information between predictors is not taken into consideration and the algorithm selects the channels in the order of the mutual information between predictors and the target signal, so the redundancy between predictors is never reflected. As β grows, the mutual information values between predictor channels influence the selection procedure much and the redundancy is reduced. But in the case of too large β, the algorithm only considers the relation between predictors and cannot reflect the predictor-target relation.

4. EXPERIMENT AND THE RESULT

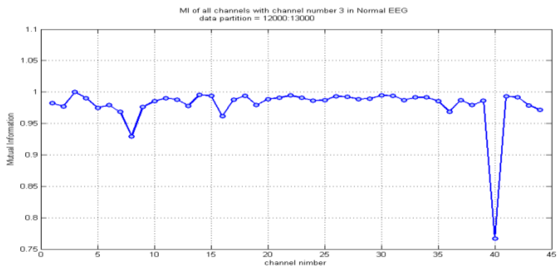
Currently, Freiburg University presented a unique EEG database with long-term recordings in the Third International Workshop on Epileptic Seizure Prediction. This set of EEG data contains signals recorded using surface electrodes (channels) from three patients during 36 hours. Different number of channels (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. 100 second of the signals recorded by 44 channels including normal, pre-seizure, and seizure state and related to the patient number two is used in this study.

The data is first divided into three segments. The first segment contains normal EEG signal, the second one is related to the pre-seizure state and the last one only includes seizure state. Here the neuro fuzzy model described before is applied to predict the EEG signal in three different segments of the signal, and then the mentioned algorithm is applied to each segment of the signal separately in order to select the channels for multi channel prediction. The influence of multi channel prediction on prediction accuracy is then studied in each part.

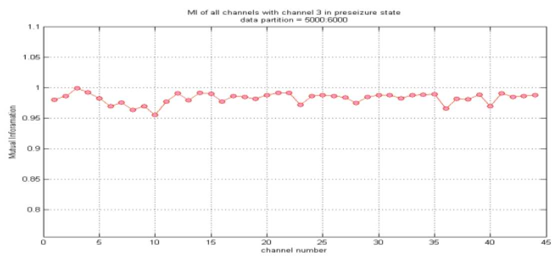
In the first phase of our experiment, a neuro-fuzzy network with LoLiMoT learning algorithm is applied to predict the time series. The hidden layer contains three neurons which obtained by trial and error. Initially, a number is chosen for the amount of neurons and then the optimized amount will be defined 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.

For single channel prediction, we used the lagged values of the signal to predict its own future values. The result of single channel prediction of channel number three is illustrated in the second column of figure 3. As it is shown in the first row, the amplitude of the signal is not very high in the normal segment. In pre-seizure segment which is shown in second row, the amplitude grows in comparison with the normal segment, however a quasi-periodic behavior can be observed. Therefore, predicting the signal is possible in relatively longer horizons for this segment. The third row shows the seizure segment which contains higher amplitude and the signal behavior seems to be quasi-periodic.

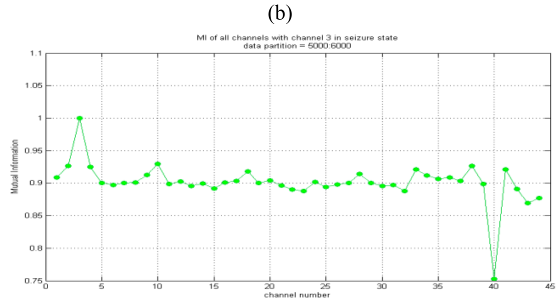
In our next experiment we predicted each segment by means of three different channels in order to improve the prediction accuracy. In this part our goal is to define the three most efficient channels as predictors. Here the number of channels to be fused is chosen to be three confirming that fusing only 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. In order to choose the predictor signals, our method of selecting channels is applied. 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 [18], 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).



(a)



(b)



(c)

Figure2. MI between channel three and all channels in normal (a), pre-seizure (b), and seizure (c) segments.

As described in [19] 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 [20]. The number of neighbors in our MI estimation algorithm is set to 20.

As figure 2 shows, the range of MI values between the channels and the target signal varies between the segments. Especially, in pre-seizure segment (Figure.2.b) the values of $I(f_i;C)$ are very close to each other. This shows that in pre-seizure state signals recorded by all channels have a lot in common. In such a case selecting the proper channels have less influence on multi-channel prediction; however, the influence of multi channel prediction is reduced slightly in pre-seizure segment. In contrast, in seizure segment (Figure.2.c) the MI values are reduced. This means that during the seizure the signals recorded by different channels are more different in comparison with the pre-seizure state. This behavior affects the channel selection result and also the influence of multi-channel prediction on improving the prediction result.

Also, the result of channel selection to predict channel number three is obviously different for different segments. This is caused by variation of MI values in different segments. Table 2 shows the channels selected in each segment as best predictors of channel number three.

Note that the channel to be predicted is always one of the channels used for multi channel prediction. This is because the target signal is usually the one which contains the most information about its future. There for, we assume that it is always exists in the set of predictor signals by default.

Table2. Channels selected in different states via presented algorithm to predict channel number three.

State	**Data Segment** (start from sample# : ends at sample#)	**Selected Channels**
Normal	(1 : 14000)	21, 3, 2
Pre-seizure	(1361441 : 1373001)	1, 3, 4
Seizure	(1373001 : 1387000)	6, 3, 42

However, there are some other channels containing extra information about the future of the target signal and using them in training phase of predicting yields more accurate result.

Our next and the most important experience is applying the channels selected by our algorithm in multi-channel prediction of the target signal and comparing the result of each segment with the one in single prediction results.

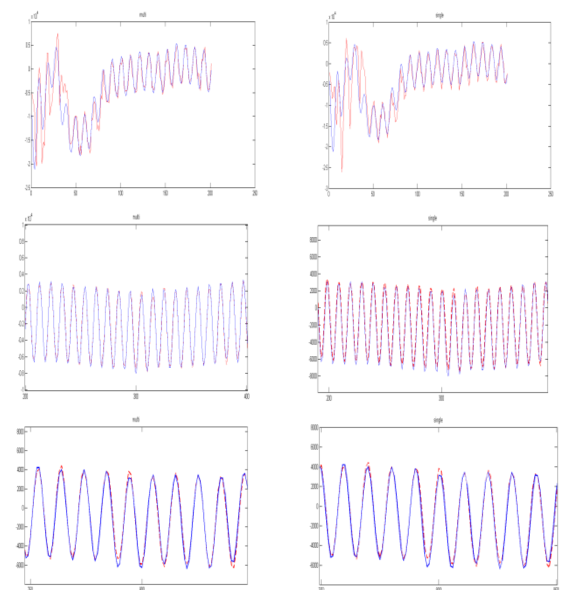


Figure3. Original and predicted signals in normal state (first row), pre-seizure state (second row), and seizure state (last row) via multi channel (first column) and single channel (second column) prediction.

Here, the main goal is to study the influence of multi-channel prediction on improving the prediction accuracy in different states of the EEG signal. To this end, some lagged values of selected channels are used to predict the target in each segment. Figure 3 shows the result of multi channel prediction in its first column and one can compare them to the single channel prediction result which is shown in second column.

As it is illustrated in Figure 3, multi-channel prediction improves the prediction accuracy considerably in pre-seizure state but, the improvement in normal and seizure segments is not noticeable. For a closer look we need to check the error of each prediction. The RMSE values corresponding to the predictions of Figure 3 are shown in Table 3. In fact, multi-channel prediction has improved the accuracy in all states but this improvement is not the same for segments. As it is shown, the most influence was in the pre-seizure state .This is obvious due to the MI values between the channels and channel number three in this state (Figure2.b). The MI values in pre-seizure state are very close to each other and their values are more, i.e. in comparison with the normal and seizure states. More MI with the target means more information about the target signal and the channels ability of predicting the target increases by increasing their MI value.

Therefore, for a state in which the signals have more predicting ability, the influence of multi-channel prediction is the most among different states.

Table3. RMSE of predicting channel number three in different states via multi channel and single channel prediction.

State	Multi channel	Single channel
Normal	141.2165	143.5915
Pre-seizure	20.4504	28.5085
Seizure	21.3076	21.3526

REFERENCES

[1] S. Haykin, "Neural networks: a comprehensive foundation", *Macmillan*, *New York*, 1994.

[2] O. Nelles, "Nonlinear system identification", *Springer, Berlin Heidelberg New York*, 2001.

[3] B. Lillekjendlie, D. Kugiumtzis, N. Christophersen, "Chaotic time series, part II: system identification and prediction", *Model Identif Control*, pp. 225–243, 1994.

[4] A. Cichoki, R. Chichester, "Neural networks for optimization and signal processing", *Wiley, New York*, 1993.

[5] P. D. Wasserman, "Neural Computing: Theory and Practice", *Van nostrand Reinhold, New York*, 1989.

[6] L. A. Zadeh, "Fuzzy logic", *IEEE Computer*, pp. 83-92, 1988.

[7] S. V. Barai, R. S. Nair, "neuro-fuzzy models for constructability analysis", *Journal of Information Technology in Construction*, pp. 65-73, 2004.

[8] B. Boashash, M. Mesbah, and P. Colditz, "Time frequency detection of EEG abnormalities", *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference, B. Boashash, Ed.*, Elsevier, Oxford, pp. 663-670, 2003.

[9] R.B. Pachori and P. Sircar, "EEG signal analysis using FB expansion and second-order linear TVAR process", *Signal Processing*, pp. 415-420, 2008.

[10] A. B. Geva and D. H. Kerem, "Forecasting Generalized Epileptic Seizures from the EEG Signal by Wavelet Analysis and Dynamic Unsupervised Fuzzy Clustering", *IEEE Ttransactions on biomedical engineering,* pp. 1205-1216, 1998.

[11] H. Kantz and T. Schreiber, "Nonlinear Time Series Analysis", *Cambridge, UK: Cambridge Univ. Press,* 1997.

[12] T. Cover, J. Thomas, "Elements of Information Theory", *John Wiley*, 1990.

[13] A. Pedram, M.R. Jamali, T. Pedram, S.M. Fakhraie, and C. Lucas, "Local Linear Model Tree (LOLIMOT) Reconfigurable Parallel Hardware", *International Symposium on Parallel Computing in Electrical Engineering*, pp.198 – 201, 2006.

[14] O. Nelles, "Nonlinear system identification with local linear neuro fuzzy models", *PhD Thesis, TU Darmstadt, Shaker Verlag, Aachen, Germany,* 1999.

[15] B. Atoufi, C. Lucas, A. Kalhor, M.M.R Yousefi, "Channel Selection in EEG Prediction: Linear and Nonlinear Approach", *Lecture Notes in Management Science*, pp. 320-327, 2008.

[16] R. Battiti, "Using mutual information for selecting features in supervised neural net learning", *IEEE Transaction Neural Networks*, pp. 537 – 550, 1994.

[17] T.M. Cover, and J.A. Thomas, "Elements of Information Theory", *John Wiley & Sons*, 1991.

[18] A. Kraskov, H. Stögbauer, P. Grassberger, "Estimating Mutual Information", *Physical Review*, 2004.

[19] A.H. Vahabie, M.M.R Yousefi, B.N. Araabi, C. Lucas, S. Barghinia, P. Ansarimehr, "Mutual Information Based Input Selection in Neuro-Fuzzy Modeling for Short Term Load Forecasting of Iran National Power System", *IEEE International Conference on Control and Automation*, China, pp. 2710–2715, 2007.

[20] F. Rossi, A. Lendasse, D. François, V. Wertz, M. Verleysen, "Mutual Information for the Selection of Relevant Variables in Spectrometric Nonlinear Modeling", *Chemom. Intell. Lab. Syst*., pp.215–226, 2006.