Detection of Event Related Potentials Using Biologically Inspired Networks

Ali Motie Nasrabadi
Biomedical Engineering
Shahed University
Tehran, Iran
nasrabadi@shahed.ac.ir

Neda Afzalian
Neuroscience
Institute for Studies in Theoretical Physics and Mathematics
Tehran, Iran

Elahe’ Yargholi
Biomedical Engineering
Islamic Azad University, Science and Research Branch
Tehran, Iran

Abstract—The present research was proposed to classify biosignals based on chaotic models. Recurrent networks, capable of describing data variation by the means of the interaction between internal layer neurons, were designed. The result demonstrated remarkable stability against external disturbance and the ability for extraction of the system original dynamics. Also a reduction of precision was shown in detection of synchronic regions through the data filtering process. The method dependence on structure not on frequency may explain why this phenomenon happens.

Biosignal; recurrent network; chaotic model; qualitative resonance

I. INTRODUCTION

Classification is a major problem in machine learning. Modelling the data diversity of one class through classification is the most demanding task, because it requires the other classes data non-interference [1]. The methods of signal recognition usually describe the data diversity of one class as a kind of external noise. So, such diversities can be defined by the means of random process like: filtered white noise, hidden Markov model or differential equations [2]-[4]. With regard to the existing interaction of all body organs with each other and with the environment and also the existence of chaotic dynamics all over the body in both microscopic (neurons performance) and macroscopic view (for instance cardiac performance or brain performance during sleep), it has been concluded that the data variation of a signal source is of a chaotic nature and it is not randomly produced [5], [6]. The chaotic order has the capability of producing ambiguous data, so the chaotic model can be used to model the data diversity of each class. So far, various methods have been applied in this area. In the following, several chaotic classifiers will be introduced.

Chaotic neural networks with negative feedback put similar data in the same clusters. Such a network does not converge at local minimums and it is this network superiority over the other neural networks [7].

Chaotic map clustering is another method which does not require the prior knowledge of data for classification. It assigns a one dimensional chaotic map to each data category.

The mutual information of maps are considered as the similarity index of data used for data classification.

The other approach places data on stable fixed points and so does the classification [8]-[11].

KIII neural network is based on the imitation of human olfactory system [12]-[13]. The record of information depends on chaotic bases of attraction. The method enjoys high potentials in the case of EEG signals recognition.

Presentation of signals variations by the means of non-linear models is another method. Each chaotic system is capable of producing all trajectory family, being different with each other, while having some features in common. Having designed the model, classification, using synchronization error can be performed.

The present study sought for a biologically inspired network. To classify the biosignals, the method made an attempt towards biosignals modification. It is necessary to mention that qualitative resonance is the index of classification.

II. MATERIAL

The EEG signals used in this research were adopted from BCI-2005 match and they were divided into two categories: including P300 (target), not including P300 (non-target). EEG was recorded through standard 64-channel protocol [14].

III. METHODS

A. Model Description

Baire introduced a biologically inspired model (Fig. 1.).

The internal state is:

\[ x_k = [x_{1,k}, x_{2,k}, ..., x_{N,k}]^T, \]

\[ x_{n,k} \] denotes the state value at node number \( n = [1, ..., N] \) and time \( k = [1, ..., K]. \)

The state equation for this can be written:

\[ x_{k+1} = \alpha x_k + \beta \tanh(w_{int} x_k + w_{in} u_k + w_{back} y_k + B), \]

\[ y_k = \tanh(w_{out} x_k). \]  

(1)

\( y_k, u_k \) are respectively output and input of the model.

Cognitive function of the brain is the result of connection and interaction of vast areas of brain cortex and several parameters such as entropy can be used in order to measure the statistical connection of brain elements. Entropy decreases in both cases either the whole members of a system function alike or completely independently. In a case of the existence of
a structural connection, on the other hand, entropy increases. So, based on this principle, to maximize entropy, the network parameters are determined.

![General topology of the biologically inspired model](Image)

Figure 1. General topology of the biologically inspired model. Solid arrows indicate internal connections held in $w_{int}$, dashed arrows feedback connections in $w_{back}$, dotted arrows input connections held in $w_{in}$, and dash dotted arrows output connections held in $w_{out}$. [15]

Design Parameters

In fact, any term in the state equation, except $w_{out}$, can be seen as a parameter with which the properties of the system can be controlled.

The memory of the nodes. The memory of the nodes is controlled by $\alpha$ and $\beta$. $\beta$ scales the strength of the network connections, whereas $\alpha$ controls the speed of the exponential decrease of a node’s state in the absence of an input signal. The necessary condition to have a asymptotically unique behavior is:

$$[\alpha + \beta \lambda_{\text{max}}] < 1$$

(2)

where $\lambda_{\text{max}}$ is the maximal eigenvalue of the matrix $w_{int}$.

The connection matrices $w_{in}$ is not of major interest for system identification, the resulting system should exhibit sustained oscillations, without need for external excitation. The feedback connections are drawn from a uniform distribution between -1 and 1. The properties of $w_{int}$ are discussed later.

Identification Process

Consider a time series $y_{train}$ should be identified. At first we note that if the system would have been identified correctly already, it would be possible to observe $y_{train}$ on the output node. Now suppose the output node would be already, it would be possible to observe $y_{train}$ on the output node. Now suppose the output node would be.

The entropy of the nodes. The entropy of the nodes is needed to measure the quality of inversion, the entropy quantifier proposed by Tononi is employed [17]. As the weights are obtained by the use of (3), the network will always be stable.

Consequently, the parameters of the network should be set values, so that a high entropy results.

Optimizing the system’s entropy

The range of parameters can be set to such extents that an exact high entropy is attained and be therefore convenient choices for network generation. In the following the dependence of entropy on several parameters is analyzed.

Nonuniform bias. The set of bias values $b$ are indicated with matrix B controlled by (1). Jaeger [18] proposed to put a nonuniform bias on the nodes. Apart from entropy dependence on bias, it also depends on feedback weights and nodes interactions. Thus entropy is analyzed in function of bias, supposing that nodes are not connected to each other. The mean value of 50 entropy realizations (50 different distributions of bias and feedback) for each bias value was obtained. (Fig. 2.)

In the network having nodes interacted with each other, to analyze the effect of bias, each of supposed parameters were changed and mean value of 100 realizations is plotted (Fig. 3.).

Dimension. In this section dimension ranges with high entropy is selected. For this all the other parameters remain fixed (connection probability 5%, bias 2% and maximal eigenvalue 0.95). Based on Fig 4., 15 nodes with 0.05 connection probability will produce the maximal entropy.

Connection probability and memory. It has been mentioned that memory comes to one part from the memory in the nodes and from connections held in $w_{int}$. As a measure for memory simply the maximal eigenvalue of the matrix $A = \alpha I + \beta w_{int}$ can be taken. First, $w_{int}$ with 0.05 connection probability is established. Then the matrix is scaled so that its maximal eigenvalue is scaled to 0.95. The next step continues with entropy analysis in function of $\alpha$ and $\beta$ (Fig. 5.). The effect of (2) on network behaviour is observable. $\alpha$ and $\beta$ were
selected to such extents to meet the condition of (1) ($\alpha=0.604, \beta=0.44$).

The feedback loop is broken up and as entry to the model a sum of the external forcing signal and the output system is taken, this leads to the modified state equations:

$$x_{k+1} = \alpha x_k + \beta \tanh \left( w_{int} x_k + w_{in} u_k + w_{back} \left( (1-\rho) y_k + \rho y_{class,k} \right) + B \right)$$

$$y_k = \tanh (w_{out} x_k)$$  \hspace{2cm} (4)

Designing multiple network

Each class of data enjoys the existence of variant dynamics. So, to increase the classifier efficiency, the idea of multiple network was proposed and the clustering methods are applied to extract similar dynamics for the data of each class. The $\rho$s were regulated in $[0-1]$ to have the maximal model error for the data of other clusters. In the step of recognition, the test data were applied to each hybrid network. The error values of any model were normalized and the test data were assigned to the model, having the minimal error.

C. Efficiency Improvement and Other Channel’s Information

A method should be applied for omission of background and undesired information. If it is possible to filter the effects of occipital channels, an accurate calculation is attainable. To do so, the information of occipital channels was assumed as disturbance and with the aim of adaptive filters, the omission was accomplished. However the filtering outcome represents the omission of main signal and information and the classification precision was remarkably decreased. It probably happened because all recorded channels are related to each other, where as the dependence of those channels was ignored, in the omission process. Furthermore some records of different parts of brain are so synchronous that supposing them as disturbance is not correct.
Modeling can be based on channels in which P300 occurs more powerfully. So, regarding the separation and summation potentials of brain in data processing [19] following the method proposed by Carmeli, a multichannel method was adopted [20], [21]. The method first used synchronous channels at P300 occurrence (C1, CPz, Cz, Fz, FCz, FC3, and FC4). Then with regard to suggestions some other channels (Pz, C4, Cz, C3, Fz, Po8 and Po1) were used. For implementation of method two different networks were studied.

Multichannel model 1

The model consists of C input and C output. The nodes of internal layer map a pattern of all inputs:

\[ x_{k+1} = \alpha x_k + \beta \tanh \left( \sum_{j=1}^{C} w_{\text{int},j} x_k + \sum_{j=1}^{C} w_{\text{int},j} u_{k,j} \right) + \sum_{j=1}^{C} w_{\text{back},j} y_{k,j} + B \]

\[ y_{k,j} = \tanh (w_{\text{out}} x_k) \]  

(5)

For classification, the network was stimulated by data of 7 channels:

\[ x_{k+1} = \alpha x_k + \beta \tanh \left( \sum_{j=1}^{C} w_{\text{int},j} x_k + \sum_{j=1}^{C} w_{\text{int},j} u_{k,j} \right) + \sum_{j=1}^{C} w_{\text{back},j} y_{k,j} + \rho y_{\text{class},k,j} + B \]  

(6)

\( \rho \) was regulated to have the minimal error value between Cz channel and its input.

Multichannel model 2

Structure of multichannel model 2 is shown in Fig.13.

For each channel data, an individual network was designed. The changes of entropy in function of bias were analyzed for each channel data individually.

Fig.14. shows the type of data does not affect the entropy or internal layer weights. So, the parameters of just one channel were determined and they can be generalized to the others.

For classification, the network was stimulated by data of 7 channels:

\[ x_{k+1} = \alpha x_k + \beta \tanh \left( \sum_{j=1}^{C} w_{\text{int},j} x_k + \sum_{j=1}^{C} w_{\text{int},j} u_{k,j} \right) + \sum_{j=1}^{C} w_{\text{back},j} y_{k,j} + \rho y_{\text{class},k,j} + B \]  

(6)

\( \rho \) was regulated to have the minimal error value between Cz channel and its input.

![Figure 9. Performance of multiple network.](image)

![Figure 10. Structure of multi channel network 1.](image)

![Figure 11. Entropy changes versus connection probability of neurons.](image)

![Figure 12. Entropy changes versus bias value.](image)

![Figure 13. Structure of multi channel network 2.](image)

![Figure 14. Entropy changes with networks of one channel.](image)
In the discussed model, state variables of internal layer nodes have no effect on each other, so, just using the data of their own output channel, they can be calculated.

\[ x_{k+1,j} = \alpha x_{k,j} + \beta \tanh\left( w_{\text{lin},j} x_{k,j} + \sum_{j=1}^{C} w_{\text{out},j} u_{k,j} + \Sigma_{j=1}^{C} w_{\text{back},j} y_{k,j} + B \right) \]  

(7)

In this structure, output layer weights contain the information of the other channels. As a result, the effect of other channels information is reflected through output layer. There are amplification of similar weights of other channels information of the other channels. As a result, the effect of their own output channel, they can be calculated. Nodes have no effect on each other, so, just using the data of the same as that in model 1.

IV. RESULTS

The method was employed for P300 component recognition. To classify target and non-target signal of BCI data, first the data of two classes were partitioned into train, validation and test sets using k-fold method. Training signal was produced putting 50 records of two classes after each other. The average error of validation data sets of two classes were analyzed, to choose the corresponding \( \rho \). Eventually, \( \rho = 0.3 \) was selected. The results demonstrated that the network can not accurately distinguish the real dynamics. This leads to the inaccuracy of the \( \rho \) computation as well. Test data were also applied to the network (Table I).

<table>
<thead>
<tr>
<th>Stimulate</th>
<th>Target</th>
<th>Non-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>56.23%</td>
<td>43.77%</td>
</tr>
<tr>
<td>Non-target</td>
<td>44.69%</td>
<td>57.31%</td>
</tr>
</tbody>
</table>

TABLE I
CLASSIFICATION RESULT FOR TARGET AND NONTARGET DATA WITH 50 TRAINING SAMPLES

It was decided to increase the training signal length, aimed at the efficiency improvement. The whole target data set (namely 170×40%) and same amount of non-target data set were used in the process of identification (Table II).

<table>
<thead>
<tr>
<th>Stimulate</th>
<th>Target</th>
<th>Non-Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>60.32%</td>
<td>39.68%</td>
</tr>
<tr>
<td>Non-target</td>
<td>34.27%</td>
<td>65.73%</td>
</tr>
</tbody>
</table>

TABLE II
CLASSIFICATION RESULT FOR TARGET AND NONTARGET DATA WITH 68 TRAINING SAMPLES

Due to the limited number of target data set (170 samples), it was impossible to have more increase in the length of target training series. However, the increase of non-target samples (850 samples) brought a higher degree of accuracy (Table III).

<table>
<thead>
<tr>
<th>Stimulate</th>
<th>Target</th>
<th>Non-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>51.73%</td>
<td>94.13%</td>
</tr>
<tr>
<td>Non-target</td>
<td>71.11%</td>
<td>70.37%</td>
</tr>
</tbody>
</table>

TABLE III
CLASSIFICATION RESULT FOR SYNCHRONOUS TARGET AND NONTARGET DATA WITH 340 TRAINING SAMPLES

Target data model was recognized insufficient in classification process, so the other model, using non-target data was applied (Table IV).

<table>
<thead>
<tr>
<th>Stimulate</th>
<th>Target</th>
<th>Non-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>49%</td>
<td>94.74%</td>
</tr>
<tr>
<td>Non-target</td>
<td>46%</td>
<td>70.37%</td>
</tr>
</tbody>
</table>

TABLE IV
CLASSIFICATION RESULT FOR TARGET AND NONTARGET DATA WITH NONTARGET MODEL

The insufficiency of model suggests that 600ms data length may not demonstrate true dynamics of system. Thus, applying the period in which P300 take place, model includes just part (not the whole) of the dynamic changes-related information. It is essential to eliminate the linear part of data while modeling is in progress. Moreover, P300 increases the recorded signal amplitude. A short window length can omit linear part meanwhile the information obtained through the increase of signal amplitude is also missed. Furthermore, undesired information is possible to be magnified during data normalization process. Thus, for classification, total length of data (1 second, including 240 samples) was used (Table V).

<table>
<thead>
<tr>
<th>Stimulate</th>
<th>Target</th>
<th>Non-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>51.73%</td>
<td>94.13%</td>
</tr>
<tr>
<td>Non-target</td>
<td>71.11%</td>
<td>70.37%</td>
</tr>
</tbody>
</table>

TABLE V
CLASSIFICATION RESULT FOR TARGET AND NONTARGET DATA WITH ONE SECOND LENGTH

The comparison of results in Tables VI and VII, indicates the classification efficiency improvement due to application of total length of data. Regarding the method sensitivity to delay, the data, using the fractal dimension computation according to the occurrence time of P300, had the same phase. To train and test the network, the phased data were used. Implementation of the preprocessing caused the efficiency of the classification to increase (Table VI).

<table>
<thead>
<tr>
<th>Stimulate</th>
<th>Target</th>
<th>Non-target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>51.73%</td>
<td>94.13%</td>
</tr>
<tr>
<td>Non-target</td>
<td>71.11%</td>
<td>70.37%</td>
</tr>
</tbody>
</table>

TABLE VI
CLASSIFICATION RESULT FOR SYNCHRONOUS TARGET AND NONTARGET DATA

The study shows that application of \( (C_1, CP_z, C_z, F_z, FC_2, FC, \text{and } FC_z) \) channels in comparison with \( (P_4, C_4, C_z, FC_z, PO_8, \text{and } PO_1) \) channels resulted in higher accuracy. Furthermore, the accuracy obtained through multichannel network is not as much as that of single channel networks.

www.MatlabSite.com | متلب سایت
Proceeding of ICCE2010, Isfahan University of Technology, 11-13 May 2010

11
TABLE VII
CLASSIFICATION RESULT OF MODEL 1 WITH THE USE OF (C1, CPz, Cz, Fz, FC2, FC1, and FCz) CHANNELS

<table>
<thead>
<tr>
<th></th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target data</td>
<td>87.514</td>
</tr>
<tr>
<td>Non-target data</td>
<td>89.34</td>
</tr>
<tr>
<td>Total accuracy</td>
<td>88.427</td>
</tr>
</tbody>
</table>

TABLE VIII
CLASSIFICATION RESULT OF MODEL 1 WITH THE USE OF (Pz, Cz, C4, Cz, C1, CPz, Po, and Po) CHANNELS

<table>
<thead>
<tr>
<th></th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target data</td>
<td>41.27</td>
</tr>
<tr>
<td>Non-target data</td>
<td>86.34</td>
</tr>
<tr>
<td>Total accuracy</td>
<td>63.785</td>
</tr>
</tbody>
</table>

TABLE IX
CLASSIFICATION RESULT OF MODEL 2 WITH THE USE OF (C1, CPz, Cz, Fz, FC2, FC1, and FCz) CHANNELS

<table>
<thead>
<tr>
<th></th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target data</td>
<td>71.381</td>
</tr>
<tr>
<td>Non-target data</td>
<td>82.33</td>
</tr>
<tr>
<td>Total accuracy</td>
<td>76.85</td>
</tr>
</tbody>
</table>

V. CONCLUSION

The present research was proposed to classify biosignals based on chaotic models. A reduction of precision was shown in detection of synchronous regions through the data filtering process. The method dependence on structure not on frequency may explain why this phenomenon happens.

Another approach was using chaotic feature for target data to have the same phase on occurring time of P300. The results revealed the criteria of the applied method were appropriate to extract the occurrence time of P300 and it improved the accuracy of classification. Recurrent networks, capable of describing data variation by the means of the interaction between internal layer neurons, were designed. The result demonstrated remarkable stability against external disturbance and the ability for extraction of the system original dynamics. Designing individual models for each subcluster led to having higher levels of accuracy in the case of classification. The application of this method on EEG signal showed that its efficiency depends on the length of input data. The length of target data set was limited, so the model could not describe the signal features. As a result the classification was performed just by use of nontarget data. With regard to the fact that different variables are involved in the brain system activities, it can be concluded just the signals obtained through one channel cannot provide us with the whole system information. To extract the common information, the regions active in the process of recognition, were studied. The implementation of filtering methods, because of the dependency, existing between channels, caused the information to be wiped out. So, a desirable result was not attained.

The lack of enough precision of computing methods, such as entropy, … may explain the reasons for the errors occurred in these methods. Another cause of error is that to compute the parameters in these methods, some assumption are used which may not be true in those systems (e.g. the assumption of being ergodic can be referred to). Also, the application of the averages can ruin a major part of related information.

REFERENCES