EEG signal classification by improved MLPs with new target vectors

José Ricardo Gonçalves Manzan¹, Shigueo Nomura¹, and João Batista Destro Filho¹
¹Faculty of Electrical Engineering, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brasil

Abstract - This paper proposes the use of new target vectors for MLP learning in EEG signal classification. A large Euclidean distance provided by orthogonal bipolar vectors as new target ones is explored to improve the learning and generalization abilities of MLPs. The data set consisted of EEG signals captured from normal individuals and individuals under brain-death protocol. Experimental results are related to MLP performance comparison by training the networks with three types of target vectors (conventional, orthogonal bipolar and non-orthogonal ones). We have concluded that the use of orthogonal bipolar vectors as target ones has contributed to improve the MLP performance on tasks for EEG signal classification.

Keywords: EEG signal brain-death, pattern classification, multilayer perceptron, target vectors, orthogonal bipolar vectors.

1 Introduction

Artificial neural networks (ANN) is a science field that has emerged as a set of powerful tools capable of solving engineering problems that previously could not be solved. One of those tools is multilayer perceptron (MLP) that has been applied to function approximation, time series prediction, pattern recognition [1], sound signal processing [2], and biomedical signal processing [3] problems.

Related works are presented in Section 2. Section 3 presents a motivation for the work. Hypothesis to be solved by this work is described in Section 4. The different types of target vectors are defined in Section 5. In Section 6, we can verify the experimental procedure. Section 7 presents experimental results. Section 8 performs the discussion and Section 9 presents the conclusion.

2 Related Works

Researches on pattern recognition mainly description and classification are fundamental to engineering and science. Several techniques such as statistical approach, theoretical decision and syntactic approach have been adopted [11]. In recent years, ANN techniques represented by MLPs have been widely used because of promising results. One of the advantages of using ANN is the ability for training in a supervised or unsupervised form.

The traditional approaches on artificial intelligence use the sequential processing. On the other hand, MLP models use a learning mode with parallel and distributed processing. His principles, architectures and training have inspired in biological neurons and training takes place by means of examples. Trial and error strategies contribute to the ability to differentiate patterns. MLP has a similar behavior when a large number of neurons send excitatory or inhibitory signals to other neurons composing the network.

Several researchers [4] [5] [6] have focused on improving ANN performances. Some proposed strategies are regarded to input pattern improvement, MLP architecture optimization, and learning algorithm enhancement.

However, strategies related to the use of different target vectors rather than conventional ones are not so usual. For instance, Dietterich has published a work based on error-correcting output codes [7].

EEG signal analysis is very important for public health applications in the context of Brazilian health system, for which costless, non-invasive and simple diagnostic examinations such as electroencephalography (EEG) are needed [8]. In this case, brain death protocols (BDP) require complementary examinations for medicines taking a decision about the therapeutic procedure to be performed on patients at the Intensive Care Unit [9]. Very few works of the literature are devoted to the application of EEG to BDP [10], but none of them, to our knowledge, make use of MLP.

3 Motivation

The biological cognition has abilities to recognize and distinguish patterns, even if they have a high degree of degradation on their features [11] [12] [13]. In the case of ANN, the appropriate choice of parameters can provide good training. In this case the network can get an excellent power of generalization. This is good for constructing a model with high flexibility to properly recognize too degraded patterns. Moreover the choice of parameters may be inappropriate and it can cause the overtraining problem of networks. With the
overtraining, the network loses the generalization ability to correctly classify those degraded patterns.

Several proposals in order to improve the ability to recognize degraded patterns have been carried out. In most cases they have focused on how to treat input vectors [14]. The previous work [15] shows satisfactory results in using improved target vectors for degraded pattern recognition. This work shows effects of adopting orthogonal bipolar vectors as targets on improving the MLP performance to classify the EEG signals.

In addition, for the application discussed in this paper, neural nets are very important, since noise disturbs recorded data, and since too many clinical details influences the standard visual analysis of EEG by neurologists. Generally, decisions on BDP are taken based just on this visual analysis [16].

4 Hypothesis

In case of conventional bipolar vectors (CBV), the inner product between two of them is not null. On the other hand, orthogonal bipolar vectors (OBV) always have null inner product between them. Also, the similarity between two OBVs is lower than that corresponding similarity between two CBVs. Furthermore, the orthogonality between two OBVs leads to the largest Euclidean distance as well as possible. The hypothesis is that larger Euclidean distance and lower similarity of OBVs can affect on the MLP performance improvement to recognize degraded patterns.

This paper presents experimental results of use the OBVs and CBVs in recognition of EEG signals of normal patients and individuals under the brain-death protocol. We have realized that the results are consistent with the hypothesis.

5 Representation of Vectors

5.1 Orthogonal Bipolar Vectors

Equations (1) and (2) represent two possible target vectors, the equation (3) represents the inner product and equation (4) the Euclidean distance.

\[
\overrightarrow{V} = \begin{pmatrix} v_1, v_2, \ldots, v_n \end{pmatrix}
\]

\[
\overrightarrow{W} = \begin{pmatrix} w_1, w_2, \ldots, w_n \end{pmatrix}
\]

\[
\overrightarrow{V} \cdot \overrightarrow{W}^T = v_1 \cdot w_1 + v_2 \cdot w_2 + \ldots + v_n \cdot w_n
\]

\[
d_{\overrightarrow{V}, \overrightarrow{W}} = \sqrt{(w_1 - v_1)^2 + (w_2 - v_2)^2 + \ldots + (w_n - v_n)^2}
\]

Consider the case where \(\overrightarrow{V}\) and \(\overrightarrow{W}\) are orthogonal with length \(n\). There will be \(n/2\) components whose product is positive and \(n/2\) components whose product is negative. Positive product of components is related to the ones which the terms have the same signal. These terms do not affect on the result of the Euclidean distance given by equation (4). On the other hand, for the terms with opposite signals, the square of their difference is 4. The squares of differences contribute into the Euclidean distance resolution. Therefore, if we have larger number (\(n\)) of components then we have larger Euclidean distances. Equations (5) and (6) represent examples of OBVs. The inner product of those OBVs is given by equation (7). The OBVs can be generated by implementing the algorithm as described in [17].

\[
\overrightarrow{V} \overset{\text{def}}{=} (1,1,1,1, -1, -1, -1, -1)
\]

\[
\overrightarrow{W} \overset{\text{def}}{=} (1,1, -1, -1,1,1, -1, -1)
\]

\[
\overrightarrow{V} \cdot \overrightarrow{W} = 1 \cdot 1 + 1 \cdot 1 + (-1) \cdot (-1) + (-1) \cdot 1 + (-1) \cdot 1 + (-1) \cdot (-1) = 0
\]

5.2 Conventional Bipolar Vector (CBV)

In case of conventional bipolar vector (CBV), the component corresponding to the pattern \(i\) represented by vector \(\overrightarrow{U}\) at the position \(i\) values 1. All the other components value (−1) as represented by equation (8).

\[
\overrightarrow{U} \overset{\text{def}}{=} (-1, -1, \ldots, 1, \ldots, -1)
\]

If \(\overrightarrow{U}_1\) and \(\overrightarrow{U}_2\) are conventional then the terms of equation (4) are null except for two terms corresponding to the positive component of the vector \(\overrightarrow{U}\) given by equation (8). So, the Euclidean distance for CBVs is smaller than the distance for OBVs. Consider the example of two CBVs, \(\overrightarrow{U}_1\) and \(\overrightarrow{U}_2\) that represent two patterns defined by equations (9) and (10) from a set of four patterns to be classified.

\[
\overrightarrow{U}_1 = (1, -1, -1, -1)
\]

\[
\overrightarrow{U}_2 = (-1, -1, -1, -1)
\]

The Euclidean distance between \(\overrightarrow{U}_1\) and \(\overrightarrow{U}_2\) is calculated by equations (11) and (12). Those equations show
that only two terms are non nulls. Therefore, the Euclidean distance value between two CBVs is always $2\sqrt{2}$.

$$d_{\overrightarrow{U_1}, \overrightarrow{U_2}} = \left(1 - (-1)^2 + [(-1) - (-1)]^2 \right)^{1/2}$$

$$d_{\overrightarrow{U_1}, \overrightarrow{U_2}} = \sqrt{[2]^2 + [2]^2 + 0^2 + 0^2} = 2\sqrt{2}$$

5.3 Non-Orthogonal Bipolar Vector (NOV)

The non-orthogonal bipolar vectors (NOVs) are similar to CBVs. However, they have larger lengths and they are equal to OBVs. They consist of a unitary component at the position “i” to represent the $i^{th}$ pattern and other components are equal to “– 1” as given by equation (13).

$$\overrightarrow{T} = (-1, -1, ..., 1, ..., -1, -1)$$

Let $\overrightarrow{T_1}$ and $\overrightarrow{T_2}$ as two NOVs with $n$ components represented by equations (14) e (15). The Euclidean distance between $\overrightarrow{T_1}$ and $\overrightarrow{T_2}$ is calculated by equations (16) and (17).

In a similar way to CBVs, the Euclidean distance between NOVs is equal to $2\sqrt{2}$.

$$\overrightarrow{T_1} = (1, -1, -1, ..., 1)_{n \ text{ \ components}}$$

$$\overrightarrow{T_2} = (-1, 1, -1, ..., -1)_{n \ text{ \ components}}$$

$$d_{\overrightarrow{T_1}, \overrightarrow{T_2}} = \frac{\left[1 - (-1)^2 + [(-1) - (-1)]^2 \right]^{1/2} + \left[(-1) - (-1)]^2 + \left[(-1) - (-1)]^2 \right.}{n \ text{ \ terms}}$$

$$d_{\overrightarrow{T_1}, \overrightarrow{T_2}} = \sqrt{[2]^2 + [2]^2 + 0^2 + 0^2} = 2\sqrt{2}$$

NOVs can be defined as an extended version of CBVs. For this reason, NOVs can have arbitrary lengths. Then, one can generate NOVs and OBVs with the same lengths. This allows a fair comparison of MLP performances, since OBVs have different lengths compared to CBVs.

6 Experimental Procedure

6.1 Experimental Data

Adult normal individuals and patients were both recorded at Uberlandia University Hospital (HCU), by using a Linx BNT EEG amplificator, and the standard 10-20 electrode placement. For each person, a 20-minute recording was performed, then a neurologist randomly selected 15 segments without any kind of artifact, by means of visual analysis. Notice that one of these segments is composed of data of all 20 channels or electrodes during a one-second duration window. All these procedures have been fully authorized by the local Ethics Committee, through protocol 369/11.

“Normal individuals” translates into people aged 18-60 years old, men or women, without any previous neuropathology and not taking absolutely no neurological drug during lifetime. Adult patients under BDP at our local Intensive Care Unit have been selected, so that no effects of analgesics were involved. For patients, a retrospective data collection was performed, considering the period February 2009 - February 2012. Based on this, five patients have been randomly chosen for the MLP processing.

Input data to the MLP are prepared as follows. First, for each normal individual or patient, consider the 15 segments extracted from the EEG recording by the neurologist. These 15 segments were normalized according to (20), which takes into account the average (18) and the standard deviation (19). Then we used Fast Fourier Transform (21) for the estimation of the twenty median frequencies of each segment, wherein each median frequency is tied to one single channel.

$$\overrightarrow{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \overrightarrow{x})^2}$$

$$X_i = \frac{(x_i - \overrightarrow{x})}{s}$$

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y)e^{-j2\pi(xu/M + yv/N)}$$

Wherein:

* $n$ is a number of samples.
- $x_i$ is the ith sample value.
- $x$ is the mean of data.
- $s$ is standard deviation of data.
- $X_i$ is ith sample normalized.
- $M$ is number of samples of variable 1;
- $N$ is number of samples of variable 2;
- $f(x,y)$ is discretized continuous function in M samples along the x-axis and N samples along the y-axis;
- $F(u,v)$ is value of fourier discrete transform.

Finally, we calculated the mean and standard deviation of the median frequencies of each channel, considering all the fifteen segments. The feature vector for the MLP neural network training finally consisted of 40 values. They were ordered according to the order of the channels, the first of which is the average and the second is the standard deviation.

The current amount of experimental data is small due to the serious difficulties for collecting and registering this kind of data in hospital. Furthermore, various strict conditions for validating the data have limited the amount of experimental data.

Given this fact, sample combinations with the available data set were performed. In each combination were used 2 normal individuals and 2 patients under BDP for training and 3 normal individuals and 2 patients under BDP to test the performance of MLPs. Therefore, six combinations according to Table 1 were obtained.

### Table 1: Combinations Used In The Experiment

<table>
<thead>
<tr>
<th>Combination</th>
<th>Training data</th>
<th>Test data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients under BDP</td>
<td>Normal individuals</td>
</tr>
<tr>
<td>1</td>
<td>1 and 2</td>
<td>1 and 2</td>
</tr>
<tr>
<td>2</td>
<td>1 and 3</td>
<td>1 and 3</td>
</tr>
<tr>
<td>3</td>
<td>1 and 4</td>
<td>1 and 5</td>
</tr>
<tr>
<td>4</td>
<td>2 and 3</td>
<td>2 and 3</td>
</tr>
<tr>
<td>5</td>
<td>2 and 4</td>
<td>2 and 5</td>
</tr>
<tr>
<td>6</td>
<td>3 and 4</td>
<td>3 and 5</td>
</tr>
</tbody>
</table>

### 6.2 MLP topologies

Five topologies for MLP training according to the target vector length were adopted. They have been defined by the combination (number of input neurons x number of output neurons) as follows: 40 x 50 x 2; 40 x 50 x 4; 40 x 50 x 8; 40 x 50 x 16 and 40 x 50 x 32.

The number of 2 units in the output layer was defined by number of diagnoses, being normal and brain death. In case of using OBVs as target vectors, since OBVs have been generated with the lengths as 2, 4, 8, 16, 32, 64, 128 and so on, OBVs with lengths of 4, 8, 16 and 32 units were chosen in the output layer. In parallel to the OBVs, NOVs with lengths of 4, 8, 16, and 32 units have been adopted in the experiments.

### 6.3 MLP training stage

The experimental simulations were performed using the traindx toolbox of Matlab software version R2008. The traindx toolbox adopts the momentum and adaptive learning rate for MLP learning. For this reason, we can get a more rapid convergence assuring consistent results. Furthermore, in order to get fair experimental results, we have adopted MLP topologies that allow the use of same initial synaptic weights for training. This procedure was done to assure a fair comparison between different target vectors.

The adopted initial learning rate was 0.1, since traindx works with adaptive learning rates. For each training epoch is calculated an error that is compared to the tolerance (stopping criterion for training).

The equations for calculating the mean squared error to be used in the stopping criterion are as follows:

$$E_p = \frac{1}{2} \sum_{j=1}^{N_p} (d_j - y_j)^2$$  \hspace{0.5cm} (22)

$$E_m = \frac{1}{N} \sum_{p=1}^{N} E_p$$  \hspace{0.5cm} (23)

Where:
- $E_p$ is the squared of a pattern $p$;
- $N_p$ is the number of output neurons;
- $d_j$ is the desired output for neuron $j$;
- $y_j$ is the net output for neuron $j$;
- $E_m$ is the mean squared error of all patterns for each epoch;
- $N$ is the number of patterns;

The training is finalized when the stopping condition $E_m < \varepsilon$ is satisfied. The value of $\varepsilon$ is the tolerance for error during the training process.

The simulations were performed on a computer with INTEL(R) CORE i5-2410TM, 2.30GHz processor and 4 GB RAM.
6.4 Experimental Results

The graph of Figure 1 compare the performance of CBV(2), OBV(4) and NOV(4). Figure 2 show that compare their performance of CBV(2), OBV(8) and NOV(8). The comparison of CBV(2), OBV(16) and NOV(16) is represented in the graph of Figure 3. Finally the Figure 4 shows the graph the comparison of CBV(2), OBV(32) and NOV(32).

The best performance is 93.33% with the use of OBV(4) and OBV(8) and 86.7% with the use of CBV(2). In all the cases, the use of OBVs has lead to more consistent results. Furthermore, the results with the use of OBVs show good performances at large values of tolerance for error in training. The performance of CBV(2) and NOVs only achieves good performance with small tolerance values. The overtraining problem has been detected with few epochs of training by using CBVs and NOVs as targets.

Previously, the OBV experimental results for application to the digits extracted from license plate degraded images [15] have presented an increase of 8% on the MLP performance. All these results have strengthened the viability of using OBVs as target vectors for MLP in pattern recognition.

7 Discussion

The results have shown that the use of OBVs as targets improves the performance on recognizing biomedical patterns represented by EEG signals in all the cases.

Besides obtaining better recognition rates, the use of OBVs has contributed into reducing the computational load. With the use of OBVs, the number of epochs for training the ANN is smaller than the case with the use of other vectors.

This is quite evident when comparing CBV(2) and NOV (32) with OBV (32). For a tolerance of 2.5E-01, the recognition rate for OBV (32) is 90% while for CBV(2) is 76.67% and for NOV (32) is 60%.
8 Conclusion

This work presented the experimental results by using MLP models for classifying EEG signals of normal individuals and individuals under brain-death protocol.

The experimental results using real input data from a hospital have shown the advantages of adopting orthogonal bipolar vectors as targets for MLP learning improvement.

We also have concluded that the use of orthogonal bipolar vectors provides a better separation of pattern features due to larger Euclidean distance between these vectors.

So, the results can confirm the hypothesis of our work suggesting orthogonal bipolar vectors as new expectation values for an improved MLP learning in EEG signal classification.

9 References


