Biomatemática ***30*** *(2020), 177–186* ***ISSN 1679-365X***
Uma Publicação do Grupo de Biomatemática IMECC – UNICAMP

Novel method based in fuzzy clustering for EEG signal analysis

Ana Maria A. Bertone [1],
Faculdade de Matemática, UFU, 38.400-902, Uberlândia, MG.

Jefferson B. Martins [2],
Fac.Engenharia Elétrica, UFU, 38.400-902, Uberlândia, MG

A novel fuzzy identification method is apply to a **E**lectro**E**ncephalo**G**ram, EEG for short, data obtained in the Bio-medics laboratory of Electrical Engineering faculty of the Federal University of Uberlândia to obtain useful information for the specialists. The technique is based in a fuzzy clustering combined with a Takagi-Sugeno fuzzy inference to obtain extracted simplifyed function, that has the same qualification in terms of Higuchi fractal dimension and the sample entropy parameters. The Hilbert transform is apply to this simplify signal that provides amplitude modulation bandwidth, the associated instantaneous frequency, and the reconstruction of the signal through the Hilbert spectrum, representing the energy-frequency, for further analysis by the specialists.

Fuzzy clustering; Hilbert transform; EEG analysis.

1. Introduction

The signals from an EEG for short, are well known for being highly subjective and may appear at random in the time scale. These are the main reasons for extracting and analyzing EEG signal parameters using computers as more efficient in diagnostics. With this motivation, nonlinear methods have been proposed to extract parameters for analysis and classification of time series signals. The Higuchi fractal dimension (HFD) parameter Higuchi (1990)

[1]amabertone@ufu.br
[2]jefferson@iftm.edu.br

measures the complexity, and the chaotic nature of the time series signals. The sample entropy (SampEn) is a statistical parameter to quantify its predictability Richman and Moorman (2000).

New methods have been proposed for the analysis and simplification of the original time series captured, for instance, by the electroencephalography procedure. Among others, the **E**mpirical **M**ode **D**ecomposition (EMD), developed by Huang et al. Huang et al. (2008), has been applied to nonlinear and non stationary signal analysis. The EMD method objective to break down the signal in **I**ntrinsic **M**ode **F**unctions (IMF), without leaving the time domain.

A novel method is proposed in this study with the objective of extracting a single IMF function from the **O**riginal **T**imes **S**eries (OTS). This function, that we denote as ESF (extracted simplifyed function), matches the HFD and the SampEn parameters of the OTS, with a small absolute difference error as a validation of the method. A fuzzy clustering based in Gusftason and Kessel algorithm Gustafson and Kessel (1979) is applied to the time series interpreted as a time-output system, that divides the temporal universe in subsets containing the highest degree of memberships information, obtained from the clustering procedure. From there, and using the Hilbert transform (HT), the amplitude modulation bandwidth (BAM) and frequency modulation bandwidth (BFM) are extracted from analytic representation of the single IMF. The spectrum and instantaneous frequency are other two crucial information extracted for further analysis.

2. Mathematical background

The method proposed in this work is based in a fuzzy identification technique that includes fuzzy clustering and a Takagi-Sugeno (TS) fuzzy inference. In general, clustering is a classification non supervised of data, forming groups of elements called clusters. The objective of this technique is the organization of the patterns represented by vector or points in the multidimensional space, according with a similarity measure. The fuzzy clustering allows that the elements of the data belong to different groups simultaneously with different membership degrees. Finally, it is obtain a number c of fuzzy sets. There are many techniques refer to the fuzzy clustering. In this study is used the well-known due to Gustafson and Kessel (GK) Gustafson and Kessel (1979).

Describing the mathematically process, we consider a time series type of vector

$$z_j = (t_j, o_j) \in \mathbb{R}^2, j = 1, \dots N,$$

where t_j is the time and o_j represents the output of the system. In the first step of clustering through the GK algorithm, it is consider the following elements:

- The Euclidean distance as a measure of similarity;
- A set, $Vini$, of center of cluster which are chosen from the set $\{z_j\}_{j=1}^{N}$ in a way that the centers are well distributed among the data.

With those elements as a starting point, the next step of the clustering process is to find the entries of the matrix

$$U = \begin{bmatrix} \mu_{11} & \dots & \mu_{1k} & \dots & \mu_{1N} \\ \mu_{21} & \dots & \mu_{2k} & \dots & \mu_{2N} \\ \vdots & \vdots & \ddots & \vdots & \\ \mu_{c1} & \dots & \mu_{ck} & \dots & \mu_{cN} \end{bmatrix}.$$

For this end, it is formulate an optimization problem given by

$$\text{minimize} \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{i,j}^{m} d_{i,j}^{2} \tag{2.1}$$

where v_i is an element of $Vini$ for $i = 1, \dots c$,

$$d_{i,j}^2 = d^2(z_j, v_i) = (z_j - v_i) I (z_j - v_i)^T, \ i = 1, \dots, c, \ j = 1, \dots, N, \tag{2.2}$$

and I is the identity matrix. The optimization problem has the following constrains:

$$\begin{aligned} & 0 \le \mu_{i,j} \le 1, i = 1, \dots, c, \ j = 1, \dots, N \ \text{(degree of membership)}; \\ & \sum_{j=1}^{N} \mu_{i,j} > 0, \ i = 1, \dots, c \ \text{(non emptiness of the cluster)}; \\ & \sum_{i=1}^{c} \mu_{i,j} = 1, \ j = 1, ..., N \ \text{(total degree of membership)}, \end{aligned} \tag{2.3}$$

where m is known as the fuzzy parameter of the clustering Abonyi et al. (2002). The number m determines how much fuzziness is wanted for the clustering. It has been taken for this study $m = 2$.

Once the optimization problem (2.1) is solved using the Lagrange Multipliers, it is obtained the membership function μ_i for each cluster, which, at each element of the data z_k, is given by

$$\mu_i(z_k) = \left(\frac{d_{i,k}^2}{\sum\limits_{j=1}^{c} d_{j,k}^2} \right)^{-\frac{1}{m}}, \; i = 1, \dots, c, \; k = 1, \dots, N. \tag{2.4}$$

As a convenient choice, the new center of clusters are chosen as the center of mass of the membership function μ_{ij}^m given by

$$v_i = \frac{\sum\limits_{j=1}^{N} \mu_{ij}^m z_j}{\sum\limits_{j=1}^{N} \mu_{ij}^m}. \tag{2.5}$$

Therefore, it has been constructed a set $V = \{v_1, v_2, ..., v_c\}$, $v_i \in \mathbb{R}^3$, that represents the new centers of the clusters. Next, considering for each $i = 1, \dots, c$ the matrix A_i, $2 \times N$ and B_i, $N \times 2$, given by

$$A_i = \frac{\mu_{ij}^m}{\sum\limits_{j=1}^{N} \mu_{ij}^m} (z_j - v_i)^T, \qquad B_i = (z_j - v_i)$$

we define the matrix 2×2 given by $F_i = A_i \cdot B_i$, where $\cdot$ is operator product of matrix. The F_i entries represent the variance of the memberships of the elements of the cluster i through μ_{ij}^m. This matrix is known as the covariance matrix of the cluster i. Finally, it is defined the matrix that induces the new distance in the GK process, defined by

$$M_i = \det(F_i)^{\frac{1}{2}} F_i^{-1}, \tag{2.6}$$

that determines the new geometry features of the cluster with center v_i. From the matrix M_i, it is updated the distance measure defined as

$$d_{i,j}^2 = (z_j - v_i) M^i (z_j - v_i)^T, \; i = 1, \dots, c, \; j = 1, \dots, N. \tag{2.7}$$

The procedure can be repeat, ending when an established criteria of stop, s, is set, and the error

$$\text{error} = \max \left\| U_{ij}^k - U_{ij}^{k+1} \right\|,$$

is lesser than s, where U_{ij}^k represents the entries of the memberships degrees of the data z_j to the cluster i at the step k of the GK algorithm.

After fuzzy clustering process is finished, the TS fuzzy inference is apply following the classic steps illustrated in Figure 1, when the defuzzificated output is obtained.

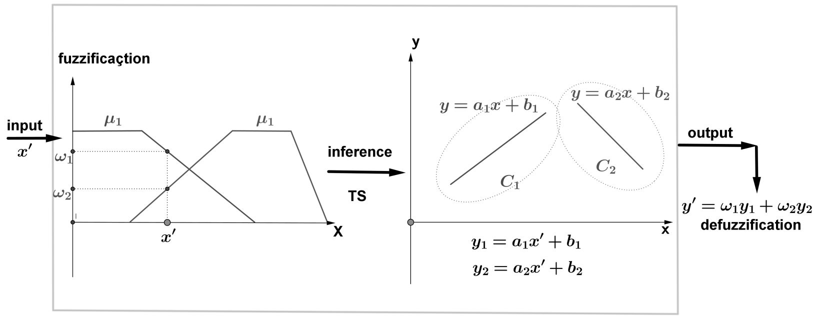

Figure 1: Figures for TS fuzzy inference.

The fuzzy identification is apply to the EEG data obtained in the Bio medics laboratory - FEELT -UFU to obtain useful information for the specialists analysis. From the identification process, a ESF is obtained with a $c = 20$ number of clusters followed by the application of the Hilbert transform. Indeed, in digital signal processing, it is often needed to look at the relationships between the real part and imaginary part of a complex signal. The main feature of this transform is that provides an analytic function as output result. By definition, given a signal, $s(t)$, its Hilbert transform, $\hat{s}(t)$, is defined by

$$\hat{s}(t) = H[s(t)] = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{s(\tau)}{1-\tau} de\tau, \tag{2.8}$$

when the integral exists. In other words, $\hat{s}(t)$ is the convolution $\frac{1}{\pi t} * s(t)$. Note that that in equation 2.8 the independent variable does not change through the transformation, what makes $\hat{s}(t)$ also a time dependent function. It is normally not possible to calculate the Hilbert transform as an ordinary improper integral because of the possible singularity at $\tau = t$. The integral is to be considered as a Cauchy principal value defined as

$$\lim_{\varepsilon\to 0^+}\left(\int\limits_a^{b-\varepsilon} s(t)\ dt + \int\limits_{b+\varepsilon}^{c} s(t)dt\right),$$

where b is a point at which the behavior o f the signal $s(t)$ verifies

$$\int\limits_a^b s(t)\ dt = \pm\infty \text{ for all } a < b, \text{ and } \int\limits_b^c s(t)\ dt = \mp\infty \text{ for all } c > b.$$

The analytic signal $z(t)$ is the complex function created by taking a $s(t)$ and its Hilbert transform $\hat{s}(t)$ as follows:

$$z(t) = s(t) + i\hat{s}(t) = A(t)\exp(i\Theta(t)), \tag{2.9}$$

where $A(t)$ is instantaneous amplitude (envelope) of the analytic signal given by

$$A(t) = \sqrt{s^2(t) + \hat{s}^2(t)},$$

and $\Theta(t)$ is the instantaneous phase of the analytic signal given by

$$\Theta(t) = \arctan\left(\frac{\hat{s}(t)}{s(t)}\right).$$

This allows the calculation of the envelope and the phase of $s(t)$ and the two functions are then plotted in polar form, that results in a waveform display. Thus the resulting magnitude versus angle plot is used for further analysis. At the same time that because using a sampled-data waveform sampled at a fixed frequency (usually 360 Hz), the time information is still implicitly available to the specialist.

All these features of the Hilbert Transform of the ESF obtained by fuzzy identification are shown in the following section.

3. Methodology

In this section we show the results of the numerical simulations of the method described in the previous section. Once the signal was collected in the laboratory, is submitted to the a pre-analysis using the parameters HFD and SampEn parameters to confirm that the EGG signal comes from a normal subject. The results of the parameters was

After this pre-analisys, a fuzzy identification with $c = 20$, was developed through an own application. As a result it is obtain an ESF signal on which we applied the spectral analysis. Before this analysis, we compared the parameters of normality between the two signals, the OTC and ESF, which results are shown in Table 1. We remark that the OTC and ESF had practically the same performance in the parameter tests with an difference, Diff, described in the third column of Table 1. This means that the ESF signal represents well the OTC signal in terms of these parameters.

Table 1: Informative table of the results in simulations.

	HFD	SampEn	Diff
Original time series	1.41617	1.41616	0.00001
Simplify time series	0.6215	0.6217	0.0002

At this stage of the study, we use the Matlab software to obtain the Hilbert transform of the ESF obtained, the amplitude modulation bandwidth, the associated instantaneous frequency, the Hilbert spectrum, and polar plot of the signal, for further analysis of the specialists.

As a result, in Figure 2 the EEG signal capture in the laboratory together with the ESF obtained by the fuzzy identification method are shown. The choose of the number of cluster is empirically done, according with the opinion of the specialists.

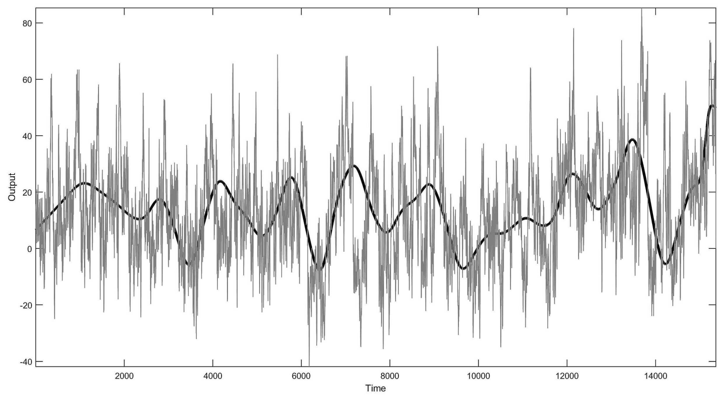

Figure 2: The plot on time of the OTC and the ESF.

Figure 3 presents the instantaneous frequency. This is the main guide to the decision of the clusters' number, due that the specialists agree that the frequency shown is corresponding to a normal EGG.

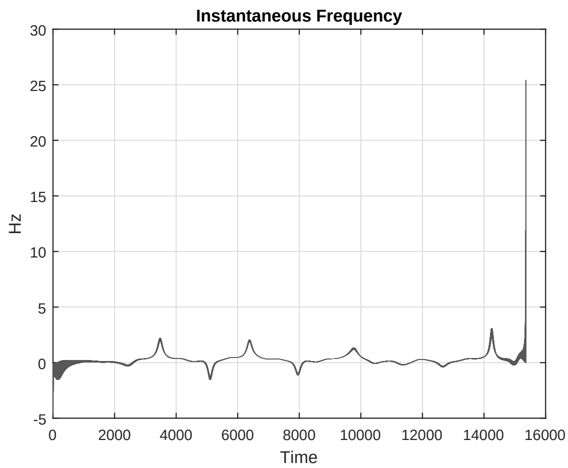

Figure 3: The plot on time the ESF's instantaneous frequency.

Next Figure 4 shows the ESF's Hilbert spectrum, which appears as the expected result form the specialists' analysis.

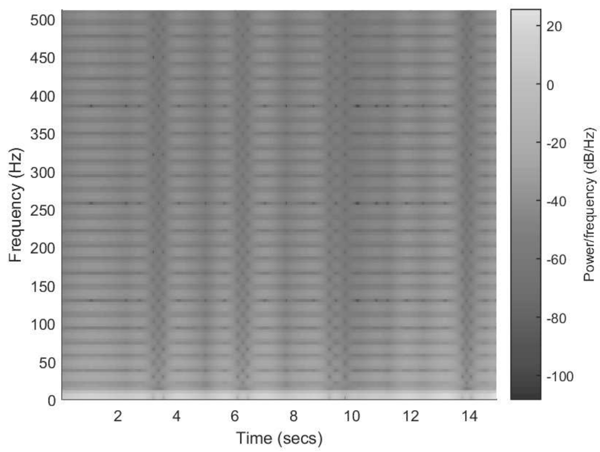

Figure 4: The ESF's Hilbert spectrum.

Finally, Figure 5 shows the ESF's polar plot of the signal described in equation 2.9.

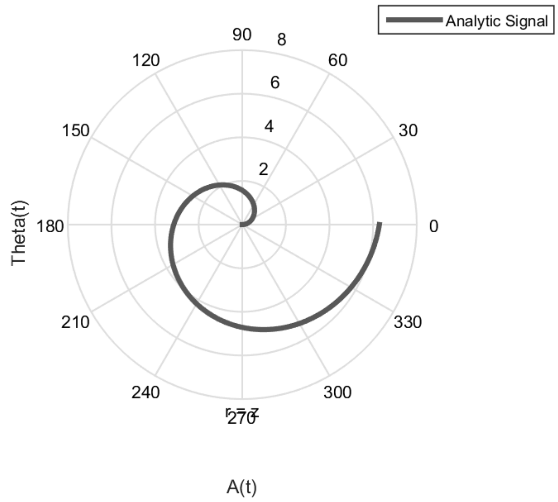

Figure 5: The ESF's polar plot.

4. Conclusion

This study propose a new method for the preparation analysis of an electroencephalogram (EEG), which is based in the Hilbert transform apply to a simplify signal returned from a fuzzy identification system. This technique presents as an appropriate approach to represents nonstationary and nonlinear EEG signals into extracted simplifyed that turns an analytic function through the Hilbert transform. It is well-known that analytic signals facilitated to extract amplitude modulation bandwidth and frequency modulation bandwidth. The features, amplitude modulation bandwidth and frequency modulation bandwidth of EEG signals, have been found useful to categorize, for instance, sleep and wake EEG signals. Future research may consist to use the same approach for EEG signal classification in different data with diverse brain conditions.

References

Abonyi, J., Babuska, R., and Szeifert, F. (2002). Modified gath-geva fuzzy clustering for identification of takagi-sugeno fuzzy models. *IEEE Trans. on Systems, Man, and Cybernetics*, 5(1):612–621.

Gustafson, D. E. and Kessel, W. C. (1979). Fuzzy clustering with fuzzy covariance matrix. In *Proceedings of the 18th IEEE Conference on Decision & Control, Dec. 12-14*, Fort Lauderdale/FL, USA. IEEE.

Higuchi, T. (1990). Relationship between the fractal dimension and the power law index for a time series: A numerical investigation. *Physica D*, 46:254–264.

Huang, Y., Schmitt, F., Lu, Z., and Liu, Y. (2008). An amplitude-frequency study of turbulent scaling intermittency using hilbert spectral analysis. *Europhys. Lett.*, 84:4–10.

Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology. Heart and Circulatory Physiology*, 6:2039–2049.