A Genetic Algorithm Approach For Pattern Recognition In Biomedical Signals

Kleydis SUÁREZ[1], Jesús SILVA[2], Mohamed NAJIM[3]

[1,2] GBBA.Edif. Electrónica, Of. 115. Universidad Simón Bolívar, Sartenejas, Venezuela

[3] Equipe Signal et Image. ENSEIRB.BP 99,F-33 402 Talence Cedex

email: [1] ksuarez@usb.ve, [2] jgsilva@usb.ve, [3] najim@tsi.u-bordeaux.fr

Résumé - Dans cette communication, nous présentons le problème de détection automatique d'ondes sur un enregistrement non stationnaire, à partir d'une approximation polynomiale fondée sur des algorithmes génétiques. Cette approximation est estimée en minimisant une fonction d'erreur non linéaire, correspondant à l'équation du cercle. Nous appliquons cette approche pour détecter les fuseaux de sommeil dans les électroencéphalogrammes (EEG), l'onde R et les battements ectopiques dans les électrocardiogrammes (ECG).

Abstract – *This article deals with the problem of on-line wave detection in non-stationary signals, through a polynomial approximation based on genetics algorithms. This approximation is estimated by minimising a nonlinear error function, described by the circle equation. We apply this approach to detect sleep spindle waveforms in electroencephalograms (EEG) and R wave and ectopics beats in electrocardiograms (ECG). The proposed method makes it possible to recognise time-variant waveform with a simple decisional parameters.*

1. Introduction

A great deal of attention has been paid to on-line wave detection problem in non-stationary signals. Several approaches were developed in the literature, particularly in biomedical applications. When using over-parameterised models such as those proposed in [1] and [2], a priori knowledge of the signal is necessary to complete the detection. As nonlinear models are used, a set of meaningful parameters is necessary for typical classification and/or discrimination task. In addition, baseline, artefacts and noisy signals require signal pre-processing. Automatic detection cannot therefore be completed due to the high computational cost of the methods.

To discern these drawbacks, we propose a new approach based on nonlinear approximations, operating in two steps:

1. the pattern to be detected is approximated by a polynomial function, which is estimated by using Genetic Algorithms (GAs). The model parameters, named «descriptors», are estimated by minimising a nonlinear distance function. This distance is derived from the Hausdorff's neighbouring [3] and corresponds to the circle equation.
2. The detection is completed by comparing the polynomial function and the analysed signal.

The remainder of the paper is organised as follows: In section II, the theoretical basis dealing with the minimisation of the quadratic nonlinear distance are introduced. More particularly, we define the polynomial approximation, the proposed distance function and present the approach to estimate the model parameters. In section III, we provide further details about the two steps of the nonlinear approximation proposed. In section IV, presented method is used to detect patterns in sleep electroencephalogram (EEG) registers [4] and high-resolution electrocardiogram (HRECG) database [5]. Comments and conclusions are given in the last section.

2. Theoretical Basis

Consider a functional relation, which defines a transformation $T(.)$ of a segmented pattern x. This transformation provides a set of values y, by means of a metric function or distance $d(x, y)$. This function is a notion of «proximity» between the signal x and its transformation y. Minimising this distance provides the best functional transformation $T_o(x)$.

Assume that $T_o(\ )$ is applied to the analysed signal x_a, which has the same length as x.

The level of «neighbourhood» between x and x_a can be evaluated through a comparative analysis between their error functions. The analysed signal and the pattern are associated providing $d(x_a, y_a)$ and $d(x, y)$ are similar. Detection can then be carried out.

2.1 Definition of the signal transformation

Consider the observation vector $\underline{x} = [x_1,..., x_M]$ associated to the «*neighbouring function*», $\{y_i = T(x_i)\}_{i=1,...,M} \in \Re$. Our purpose is to determine the transformation T.

We propose a polynomial approximation of $\underline{y} = [y_1,..., y_M]$:

$$y_i = T(x_i) = \sum_{n=0}^{p} a_n x_i^n = \underline{a}_n^T x_i. \qquad (1)$$

where $\underline{a}_n$ denotes the weight coefficient vector or descriptor:

$$\underline{a}_n = [a_0, a_1,..., a_p]^T. \qquad (2)$$

The descriptor is estimated by minimising the distance function.

However, choosing this transformation for $\underline{x}$ and another vector built with $\{x_i\}_{i=1, M}$ provide the same minimal error. For this reason, we must introduce the sequential information by modelling the pattern, as follows:

$$y_i = \sum_{n=0}^{p} a_n (ix_i)^n \ . \tag{3}$$

2.2 Distance function proposed

Let E be a metric space with a defined distance function $d(.,.)$, and S a sub-set of E. Searching an approximation of an element $y \in E$ for an element $g \in S$, consists in determining g such that [6]:

$$d(y, g) = \min_{h \in S} (d(y, h)) \ . \tag{4}$$

If this element exists, g will be is the best approximation of y in S, within the meaning of the distance $d(.,.)$.

There are several distances proposed in metric space, essentially in the construction and systematic treatment of the geometries [7,8].

Hausdorff's neighbourhood [3,8,9] defines the most basic concept of «proximity» between functions: «In $\Re$, the set J=]a,b[of real numbers such that $a < x < b$ is an open part (open interval). This is an open neighbourhood of all points of J. If it closes this part, we obtain the closed intervals K=[a,b]. So, J is the interior of K...» for every $S \subset E$, and a scalar $r > 0$, an open neighbourhood of S, $N_r(S)$, is defined as [8]:

$$N_r(S) = \{b : dist(a, b) < r, a \in S\} \ . \tag{5}$$

For any two subsets A and B from space S, the Hausdorff's distance, $h(A,B)$, is therefore defined in terms of the neighbourhood [7] as follows:

$$h(A, B) = \inf\{r : A \subset N_r(B) \ \& \ B \subset N_r(A)\}, \tag{6}$$

where $inf\{\Phi\}$ denotes the greatest lower bound of a set Φ.

This definition generates balls in $\Re^2$ spaces. These are convex compact curves, which satisfy the properties of distance function in metric space [9]:

9   Positiveness: $dist(A,B) \geq 0$. If $dist(A,B)=0$, then $A=B$

9   Symmetry: $dist(A,B) = dist(B,A)$.

9   Triangle inequality: $dist(A,C) \geq dist(A,B)+dist(B,C)$

Therefore, we propose a metric space S with a distance function $e(\underline{x}, \underline{y}): SxS \rightarrow [0, \infty[$, expressed by circle equation with ratio r, like a convex compact curve distance [10]. This means :

$$e_r(x, y) = \sum_{1 \leq i \leq M} (x_i^2 + y_i^2 - r^2)^2 \ . \tag{7}$$

This nonlinear error function limits the signals energies into a compact space. Module of the vector, shaped between the points x and y, circumscribes a surface of regular geometry, i.e. a circle (Figure 1).

To guarantee the convergence during the parameter estimation process, the pattern signal must be normalised in the interval [-r, r].
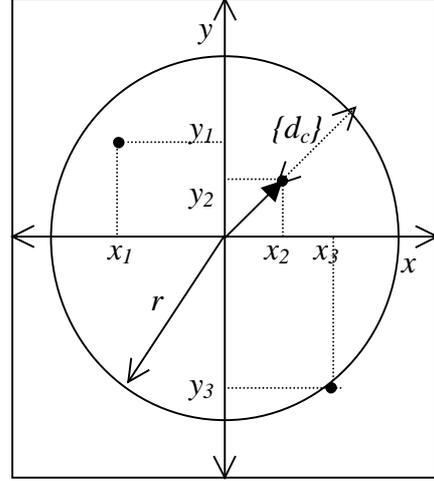
FIG. 1 : Geometric interpretations of circle of radius r used as convex compact curve distance, $\{d_c\}$.

2.3 Method of parameter estimation

We choose a non-probabilistic method to estimate the descriptor. It is the genetic algorithms (GAs). The best descriptor minimises the quadratic nonlinear error established in equation (7).

GAs are a robust search technique inspired in Darwin's theory, which solves the multidimensional optimisation problems by *implicit parallelism* [11]. GAs operate in four steps:

(a) Random population generation, in which a set of specimens is created initially. Each specimen encodes a weight coefficient vector with random polynomial order.

(b) «Fitness» evaluation, which defines the specimen capacity for detection.

(c) Iterative generation of new specimens. It includes parent selection, crossover, mutation, accepting and replacement into population

(d) The iteration process is carried out as long as the best fitness is greater than 90%.

For a pattern $\underline{x}$ with length L, the fitness is defined by:

$$fitness = \left(1 \Big/ 1 + \sum_{i=1}^{L} e_r(x_i, y_i)\right) \times 100 \ \% \ . \tag{8}$$

The GAs setting parameters include [11] the mutation rate, the population size, the parent selection mechanism as roulette wheel, rank, steady state or elitism; and the crossover method.

3. Methodology used

This approach operates in two steps.

3.1 Descriptor estimation

In equation (7), we take $r = 1$ to define the error criterion as the unitary circle. We therefore normalise the pattern signals to [-1,1] interval to assure the estimation convergence. In addition, normalized pattern signal is modulated in time using equation (3).

The genetic algorithms (GAs) search the best specimen by minimising error criterion. This is the descriptor associated to a fitness value greater than 90%. The setting of GAs is: 0.2% of rate mutation, population size of 60 specimens, parent selection for elitism and single point crossover scheme.

3.2 Pattern recognition

The signal register is analysed in a 1-sampled moving window, which has the same length L as the pattern signal. We calculate the fitness during each analysed window using expressions (1), (7) and (8). Each fitness value is assigned to the central point of the window.

Detection signal is obtained from the fitness analysis. A pulse of high fitness (> 90%) makes it possible to detect the pattern in the analysed signal.
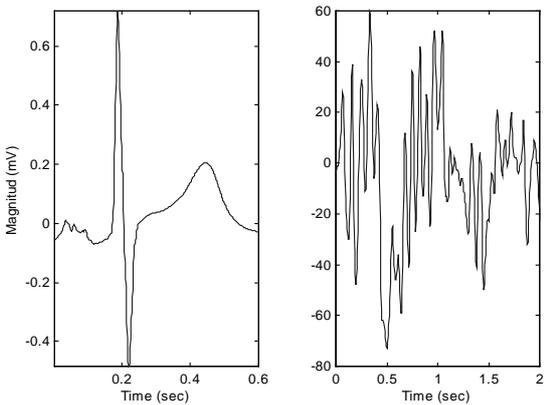
FIG. 2 : Averaged beat (left) from ECG and Spindle (right) from EEG.

4. Examples

We apply this approach in ten ECG registering and three sleep-EEG signals, to recognise the waveforms given in figure 2. These are:

… Average beats from high resolution ECG, acquired using orthogonal electrode leads X, Y, Z.

… Sleep spindle waveforms from 3mn-EEG registers obtained with 10-20 international electrode systems.

4.1 Application to electrocardiography

In ECG, R wave correspond to ventricular contraction events [4]. Measurement of R-R intervals make it possible to investigate diagnosis issue, as the Heart Rate Variability (HRV).

In this approach, R wave is detected when the pulses in fitness plot are greater to 90% and occurs a maximum variation into PE. Analysing moving window and pattern signal haves the same length: 600 samples. Figures 3 and 4 show the results of this example.
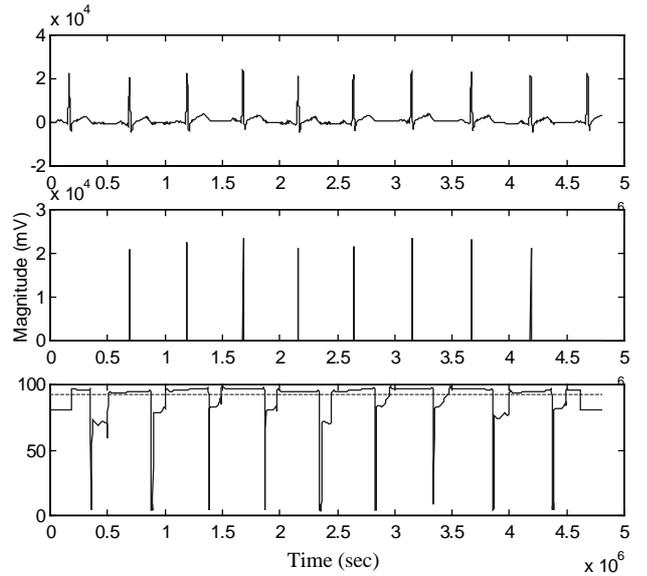
FIG. 3 : R wave detection using the 8-th order descriptor: (a) ECG register from healthy person (upper); (b) Detection signals of R wave (centre). (c) Fitness function evaluated via unitary (lower). Line represents the level of maximum fitness.
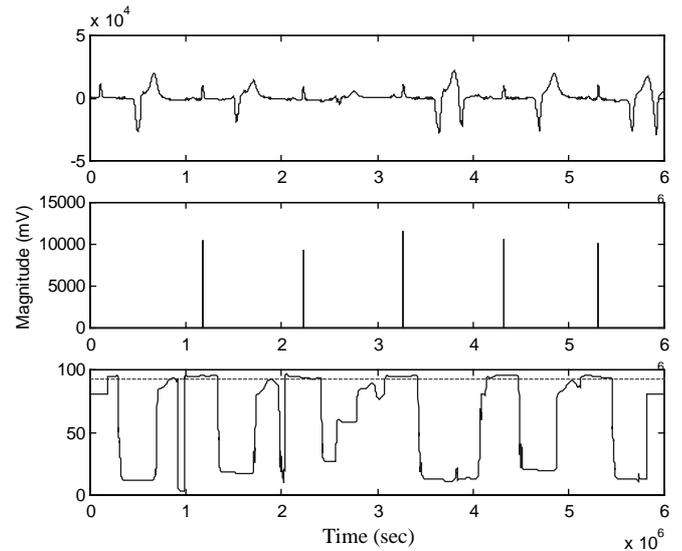
FIG. 4 : R wave detection using the 8-th order descriptor: (a) ECG register from unhealthy person (upper); (b) Detection signals of R wave (centre). (c) Fitness function evaluated via unitary (lower). Line represents the level of maximum fitness. level of maximum fitness.

4.2 Application to electroencephalography

There are three phases in sleep EEG signals: the waking (W), quiet sleep (QS) and rapid eye movement (REM). In addition, there are four stages in the QS phase. Superposed bursts or noise-free waveforms with periodic, quasiperiodic and complex oscillations characterize the stage 2 of QS phase. These waves specifically are named *spindles* and *K-complexes*.

The spindles have a frequency band from 11 to 115 Hz, duration from 0,5 to 1,5 s and amplitude between 15 to 25 μV [3].

The detection is derived from fitness plot selecting the pulses with duration between 0.5 and 1.5 seconds (duration of spindles); and with fitness > 95%.
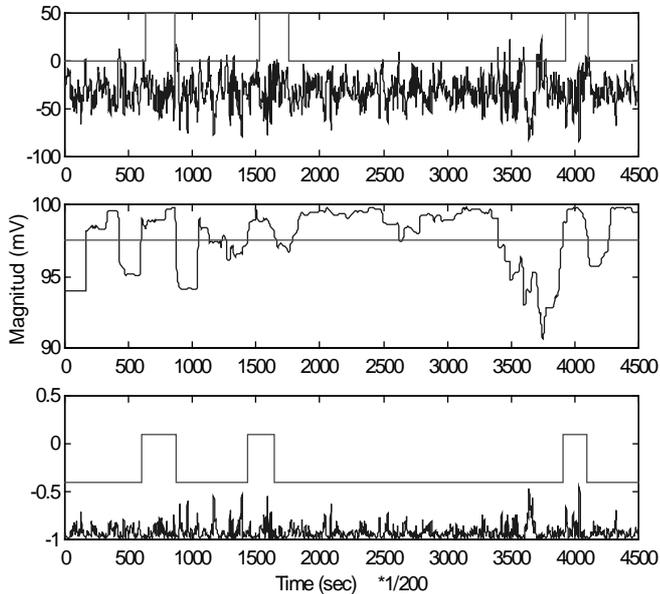
FIG. 5 : Spindle detection in EEG using a 11-th order descriptor: (a) EEG register (upper) with the annotations of spindle segments (pulses); (b) Fitness function (centre). Line represents the level of maximum fitness; (c) Polynomial evaluation (lowest) and Detection signal (pulses).

5. Discussion and conclusion

The transformation proposed in this approach is a nonlinear approximation of a neighbouring function from pattern signal. The approximation response follows the pattern without performing a complete matching. A comparison of the distance functions evaluate the «nonlinear proximity» between the pattern and the registering. In this approach, detection is based on response model (polynomial evaluation) and description capacity (fitness function). The combination of this information offers a most robust detection process. segments during recognition process. Baseline and time-varying features of signal do not affect the detection capacity of this method.

Once the descriptor have been estimated, computational cost of detection involves trivial process, as a polynomial evaluation and a small number of discrimination tasks. These are an improvement for on-line applications.

In addition, the parameter estimation based on GAs does not involve statistical approximation and can estimate the polynomial order by minimising the error criterion.

References

[1] Corrales C., Auñon J. "Nonlinear system identification and over-parameterisation effects in multisensory evoked potentials studies". *IEEE Trans. on Biome. Eng.* 2000,vol. 47(4): pp.472-485.

[2] Schimming T., Orgorzalek M., Dedieu H., "A nonlinear dynamical model for compression and detection of ECG signals". *Proceeding EUSIPCO-98*, Rhodes, Greece; September, 1998

[3] Edgar G.A. " Measure topology and fractal geometry". *Springer* (3er edition). UTM, 1995.

[4] Berthoumieu Y., Boudart E., Najim M. " High resolution spectral characterization of sleep spindles. *14th annual Inter. Conf. Of IEEE Eng. In Med. And Biol. Society*, 1995 , No. 11.

[5] Mora F., Passarriello G., Gomis P. "Señales electrocardiográficas de alta resolución en Chagas: El proyecto SEARCH". *Acta Científica Venezolana*, 1999: vol. 50, pp.187-194.

[6] Marinov P., Approximate determination of the generalized polynomial of the best on-side Hausdorff approximation. *Mathematica Balkanika, New series*; vol. 3,1989. Fasc. 1: pp.97-105

[7] Stanislawa K., Adam L., Mariusz S.. "Metric space". *Journal of Formalized Math.*, vol. 2, 1990.

[8] Ahmed M., Eskicioglu, Fisher P. " Image quality measures and their performance". *IEEE Trans. on Communication*. 1995, vol. 43(12); pp. 2959-2965.

[9] Serra J. "Image Analysis and mathematical morphology". *Academic Press*, 1993; pp. 72-75.

[10] J. Silva. Electrocardiographic signal analysis using genetic algorithms (GAs). *Computers in Cardiology: IEEE Computer Society Press*, 1996; pp. 445-8.

[11] Bäck T. Evolutionary algorithms in theory and practice. *Oxford University Press*, 1996: pp. 106-131.