Optimal Knots Allocation in Smoothing Splines using intelligent system. Application in bio-medical signal processing.

O.Valenzuela, M.Pasadas, F. Ortuño, I.Rojas
University of Granada, Spain

Abstract. A novel methodology is presented for optimal placement and selections of knots, for approximating or fitting curves to data (in particular, in this paper ECG will be used), using smoothing splines. Due the relevance of the the placement of the knots in smoothing spline approximation, in order to have an optimize allocation of the number of knots and its position, an Evolutionary Computation Paradigms (ECP) based on a Multi-Objective Genetic Algorithm has been developed, with the main purpose of avoiding the large number of local minima (in terms of approximation error for different system complexity or number of knots) existing in the problem of knots placement. This methodology was successfully applied in bio-medical signal processing. Specifically, good result has been obtained in the approximation of different class of ECG.

1 Introduction

Function approximation and curve fitting is a fundamental problem in almost every scientific field, such as image processing, computer-aided design, computer graphics, geometric modeling in shape design or data visualization and approximation.

In general, traditional methods for function approximation based on polynomial interpolation, have the disadvantage that, when the number of data is large, the interpolant polynomial usually present significant fluctuations in the intervals where no data are available.

The classic solution to this problem is considered as interpolating, spline functions, piecewise polynomial of low degree and good regularity order, ensuring well-behaved approximation error. The use of b-spline for function approximation and curve fitting has the main advantages of stability and simplicity in the calculation, being this methodology frequently used in different application areas [1],[2],[3],[4],[5]

In this paper, a novel methodology is presented for optimal placement and selection of knots, for approximating or fitting curves to data, using smoothing splines. The main purpose of this paper is to solve a well-known problem when design a cubic splines approximation function: the placement of the knots in smoothing spline approximation has an essential and significant effect on the performance of the final approximation (in terms of error measure, usually the mean square error)[2]. The principal application of the presented methodology is the characterization of bio-medical signal, such as the ECG.

2 Cubic spline and function approximation. MOGA optimization.

A spline is a function, typically constructed using low order polynomial functions, joined at breakpoints with certain smoothness conditions. The breakpoints are defined in this context as knots. If n is the order of the spline, in order to ensure the smoothness of the interpolation, typically (n−2) continuity conditions should be fulfilled.

The order n refers to the number of coefficients in the polynomial pieces (n = 2) therefore corresponds with linear splines, whereas cubic splines are of the order n = 4). The order also determines the smoothness of the resulting function approximation as splines fulfilling (n − 2) continuous derivatives. We start with a partition or knot sequence of [a, b] in m subintervals, i.e., an increasing sequence that is uniform or not uniform $\Delta_m = \{a = t_0 < t_1 < \ldots < t_m = b\}$, we define the cubic spline of class C^2 on the partition Δ_m as every function $s : [a,b] \to R$, such that:

i) $s \in C^2[a,b]$ ii) $s|[t_i, t_{i+1}] \in P_3[t_i, t_{i+1}], i = 0,1,...,m-1$

where $P_3[t_i, t_{i+1}]$is the space of all the restrictions of the polynomial functions of a degree less than or equal to three in the interval $[t_i, t_{i+1}]$.

Given t_{-3}, t_{-2}, t_{-1},…, t_{m+1}, t_{m+2}, t_{m+3}, such that $t_{-3} \leq t_{-2} \leq t_{-1}$,…, $t_{m+1} \leq t_{m+2} \leq t_{m+3}$, it is possible to define for each $t \in [a,b]$:

$$B_i^0(t) = \begin{cases} 1, & t_{i-3} \leq x \leq t_{i-2} \\ 0, & otherwise \end{cases}, i = 0,...,m+5-k$$

And $B_i^k(t)$, k = 1, 2, 3, is defined from the recursive relation:

$$B_i^k(t) = \frac{t - t_{i-3}}{t_{i+k-3} - t_{i-3}} B_i^{k-1}(t) + \frac{t_{i+k-2} - t}{t_{i+k-2} - t_{i-2}} B_{i+1}^{k-1}(t), i = 0,1,...,m+5-k$$

The next step, is the optimization of the number of knots and its allocation. To perform this task, a Genetic Algorithm will be used, that simultaneously optimize the number of knots and its position.

First, it is necessary to define several set of data to evaluate the accuracy of the system:

1) Let *nnod* the number of knots to build the Basis of Spline (BS) functions. These knots are denoted by $X^{BS} = [x^{BS}_1,\ldots, x^{BS}_{nnod}]$ and its corresponding output value, i.e., IDS = $[X^{BS}; Y^{BS}]$.
2) Let TDS be a test data set $X^{Test} = [x^{Test}_1,\ldots, x^{Test}_{ntest}]$ to verify the ability of the spline smoothing method and its corresponding output, i.e., TDS =$[X^{Test}; Y^{Test}]$.
3) Finally, let $\overline{Y}^{Test} = \left\lfloor \overline{y}^{Test}_1,\ldots, \overline{y}^{Test}_{ntest} \right\rfloor$ be the output (real numbers) obtained by the spline function on X^{Test}.

In the following figure, a block diagram of the different phases carried out in the MOGA is presented. The use of a MOGA requires the determination of three fundamental issues [11]: solution representation, determination of parameters/factors of the algorithm, and the evaluation function.

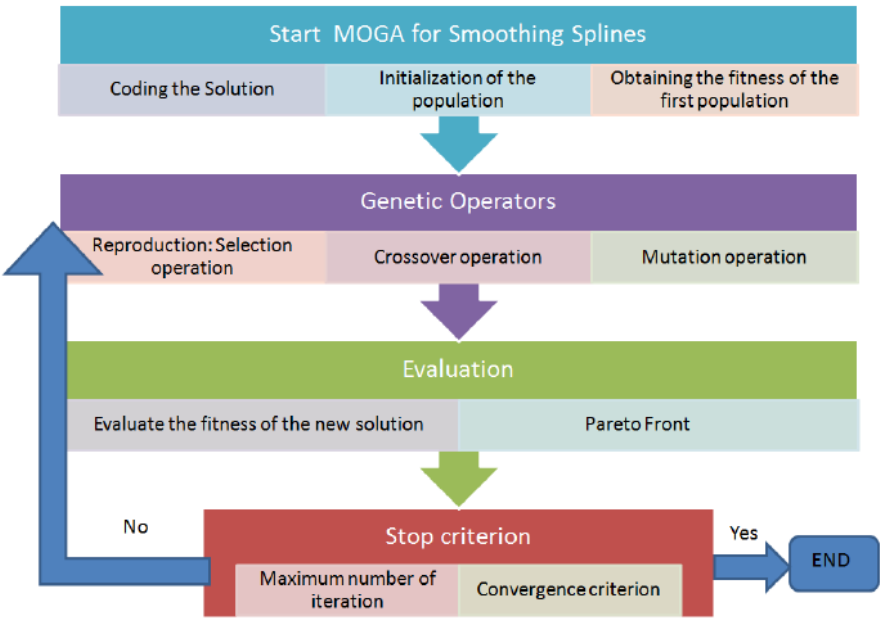

Fig. 1. Block diagram of the different phases carried out in the MOGA

3 Bio-medical signal approximation. ECG of Atrial Fibrillation

Data used in the preparation of this article were obtained from the Physionet web platform (http://www.physionet.org/physiobank/). As pointed out in this page, *"PhysioBank is a large and growing archive of well-characterized digital recordings of physiologic signals and related data for use by the biomedical research community."*

This paper will be focused in one particular pathology: Atrial Fibrillation. The Atrial Fibrillation (AF) is the result of an irregular and repetitive atrial depolarisation. As a result of this the auricle does not contract in a coordinated way, producing so called "F-waves" or fibrillation-waves, that would correspond to the multiple atrial depolarisations (contractions), with disappearing P-waves and disorganised waves appearing in its place. This can be seen in the following figure.

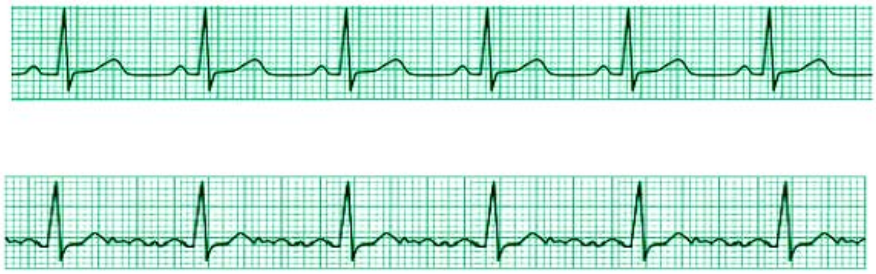

Fig. 2. Example of two superficial ECG's that show the difference between the electrocardiogram of a healthy patient (above), and one of a patient with atrial fibrillation (below). As can be seen there are multiple fibrillate waves in the arrhythmia similar to wave P constantly, throughout the entire ECG.

The atrial fibrillation (AF) is the sustained arrhythmia that is most frequently found in clinical practice, present in 0.4% of the total population [10]. Its frequency increases with age and with the presence of structural cardiopathology. AF is especially prevalent in the elderly, affecting 2-5% of the population older than 60 years and 10 percent of people older than 80 years. It is an important cause of ictus which can be found in about 15% of the patients that suffer from this phenomenon and in 2-8% of the patients with transitory ischemic attacks. The occurrence of ischemic cerebral infarctations in patients with AF non-rheumatic oscillates between 2-5% a year. The recurrences vary in different studies between 2-15% in the first year and approximately five percent in the next year. The most important indication of recurrences is presented by patients with AF and rheumatic heart disease, but the occurrence has been decreasing during the last couple of years. Therefore nowadays the most frequent source of cardioembolism is non-rheumatic AF. AF can be classified as initial or chronic. The risks of sustained AF include stroke and myocardial infarction, caused by the formation of blood clots within stagnant blood volumes in the atrial. The chronic forms of AF can be divided in three groups:

1. Paroxysmal: The episodes generally end spontaneously and usually last less than 48 hours.

2. Permanent: Where the conversion of sinusal rhythm is impossible or where there are quick relapses.

3. Persistent: The AF persists but can be reverted to sinusal rhythm.

The decision to restore the sinusal rhythm or to control the ventricular frequency is of critical importance. In the case of a first episode of AF restoring the sinusal rhythm should always be tried, but in the case of persistent chronic AF it should be attempted to define who benefits from the use of cardioversion and who should be treated with a control of the ventricular frequency and tromboembolic profilaxis.

4 Simulation results for Atrial Fibrillation classification

The first phase was the characterization/approximation of the ECG signal by means of cubic spline. Now, the allocation of the knots and other different features of the signal, will be used as input for a Supports Vector Machine classifier. In this paper, the Challenge 2004 of Physionet was carried out with the characteristics obtained from each author to test the validity of the implementation of their methodology and in order to compare these results with the results obtained by the proposed methodology.

There are two different events:
1. **Event A:** Differentiate between Group N (non-terminating AF, defined as AF that was not observed to have terminated for the duration of the long-term recording, at least an hour following the segment) and Group T (AF that terminates immediately,within one second, after the end of the record).
2.
3. **Event B:** Differentiate between Group S (AF that terminates one minute after the end of the record) and Group T.

Test set A contains 30 records, of which about one-half are from group N, and of which the remainder are from group T. Test set B contains 20 records, 10 from each of groups S and T

	Classification Success (in %)	
	Event A	**Event B**
S.Petrutiu et al. ¡Error! No se encuentra el origen de la referencia.	97	**100**
D.Hayn et al. ¡Error! No se encuentra el origen de la referencia.	93	**80**
F.Cantini et al. ¡Error! No se encuentra el origen de la referencia.	90	**90**
M.Lemay et al ¡Error! No se encuentra el origen de la referencia.	90	**60**
Proposed Methodology	**100**	**97**

Table 1. Comparison of the proposed methodology with the winners of Challenge 2004

5. Conclusion

In this papers an advanced classifier that is able to combine different source of information, such as features from the ECG and also the allocation of the knots of a

spline approximation of a bio-medical signal, is presented. In order to optimize the position of the knots for the cubic-spline approximation, a genetic algorithm is used, that simultaneously optimize the position and number of knots.

Acknowledge: This work has been partially supported by the project: Genil Start-Up Projects For Young Researchers, PYR-2012-8.

References

[1] M. Sarfraz and M. Khan, "Automatic outline capture of arabic fonts," Information Science, vol. 140, pp. 269–281, 2002.

[2] E. Ulker and A. Arslan, "Automatic knot adjustment using an artificial immune system for b-spline curve approximation," Information Sciences, vol. 179, pp. 1483–1494, 2009.

[3] B. Sarkar and C. Menq, "Parameter optimization in approximating curves and surfaces to measurment data," Computer Aided Geometric Design, vol. 8, pp. 267–290, 1991.

[4] M. Sarfraz, "Computer-aided reverse engineering using simulated evolution on nurbs," International Journal of Virtual and Physical Prototyping. Taylor and Francis, vol. 1, no. 4, pp. 243–257, 2006.

[5] Z. Habib, M. Sarfraz, and M. Sakai, "Rational cubic spline interpolation with shape control," Computers and Graphics, vol. 29, pp. 594–605, 2005.

[6] F. Cantini, et al. "Predicting the end of an Atrial Fibrillation Episode: The PhysioNet Challenge", Computers in Cardiology 2004; 31:121-124

[7] M. Lemay et al., "Computers in Cardiology/PhysioNet Challenge 2004: AF Classification Based on Clinical Features", Computers in Cardiology 2004; 31: 669-672

[8] D. Hayn et al., "Automated Prediction of Spontaneous Termination of Atrial Fibrillation from Electrocardiograms", Computers in Cardiology 2004; 31:117-120

[9] S Petrutiu, AV Sahakian, J Ng, S Swiryn, "Fibrillatory Wave Analysis of the Surface ECG to Predict Termination of Atrial Fibrillation: The 2004 Computers in Cardiology/PhysioNet Challenge", Computers in Cardiology 2004; 31, pp.250-261

[10] Chou CC, Chen PS, "New concepts in atrial fibrillation: Mechanism and remodeling" Medical Clinics of North America 92 (1): 53+ JAN 2008

[11] Srinivas M, Patnaik LM "Adaptive probabilities of crossover and mutation in

genetic algorithms" IEEE Trans Syst Man Cybern 24(4):656–666, 1994