COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 99 (2010) 124–132

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Optimal parameters study for sample entropy-based atrial fibrillation organization analysis

Raúl Alcaraz[a,*], *Daniel Abásolo*[b], *Roberto Hornero*[b], *José J. Rieta*[c]

[a] *Innovation in Bioengineering Research Group, University of Castilla-La Mancha, Spain*
[b] *Biomedical Engineering Group, University of Valladolid, Spain*
[c] *Biomedical Synergy, Electronic Engineering Department, Universidad Politécnica de Valencia, Spain*

ARTICLE INFO

Article history:
Received 1 April 2009
Received in revised form
25 February 2010
Accepted 28 February 2010

Keywords:
Atrial fibrillation
Electrocardiogram
Main atrial wave
Sample entropy
Parameters selection
Nonlinear analysis

ABSTRACT

Sample entropy (SampEn) is a nonlinear regularity index that requires the *a priori* selection of three parameters: the length of the sequences to be compared, m, the patterns similarity tolerance, r, and the number of samples under analysis, N. Appropriate values for m, r and N have been recommended and widely used in the literature for the application of SampEn to some physiological time series, such as heart rate, hormonal data, etc. However, no guidelines exist for the selection of that values in other cases. Therefore, an optimal parameters study should be required for the application of SampEn to not previously analyzed biomedical signals. In the present work, a thorough analysis on the optimal values for m, r and N is presented within the context of atrial fibrillation (AF) organization estimation, computed from surface electrocardiogram recordings. Recently, the evaluation of AF organization through SampEn, has revealed clinically useful information that could be used for a better treatment of this arrhythmia. The present study analyzed optimal SampEn parameter values within two different scenarios of AF organization estimation, such as the prediction of paroxysmal AF termination and the electrical cardioversion outcome in persistent AF. As a result, interesting recommendations about the selection of m, r and N, together with the relationship between N and the sampling rate (f_s) were obtained. More precisely, (i) the proportion between N and f_s should be higher than 1 s and $f_s \geq 256$ Hz, (ii) overlapping between adjacent N-length windows does not improve AF organization estimation with respect to the analysis of non-overlapping windows, and (iii) values of m and r maximizing successful classification for the analyzed AF databases should be considered within a range wider than the proposed in the literature for heart rate analysis, i.e. $m = 1$ and $m = 2$ and r between 0.1 and 0.25 times the standard deviation of the data.

1. Introduction

The application of nonlinear regularity metrics to physiological signals is a valuable tool because "hidden information" related to underlying mechanisms can be obtained [1,2]. To this respect, the employment of sample entropy (SampEn) to estimate non-invasively atrial fibrillation (AF) organization has revealed clinically useful information, which could be used for a better treatment of the arrhythmia [3–5]. The orga-

* *Corresponding author at*: E. U. Politécnica, Campus Universitario, 16071 Cuenca, Spain. Tel.: +34 969 179 100x4847; fax: +34 969 179 119.
E-mail address: raul.alcaraz@uclm.es (R. Alcaraz).
0169-2607/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2010.02.009

nization estimation of AF is a relevant aspect to improve its comprehension, since its mechanisms are still unexplained [6] despite of affecting up to 1% of the general population [7,8].

Given a time series with N data points, the *a priori* selection of two unknown parameters, m and r, is required to compute SampEn [2]. The parameter m determines the length of the sequences to be compared and r is the tolerance for accepting similar patterns between two segments. Although these parameters are critical in determining the outcome of SampEn, no guidelines exist for optimizing their values. Typically recommended m and r values are $m=1$ and $m=2$ and r between 0.1 and 0.25 times the standard deviation (SD) of the data [9]. This recommendation is largely based on the application of approximate entropy (ApEn) to relatively slow dynamic signals such as heart rate [1,9,10] and hormone secretion data [11]. Given that SampEn is a modified version of ApEn to solve its shortcomings, such as bias, relative inconsistency and dependence on the sample length [2], these values are also applicable to SampEn.

However, a recent work has demonstrated that the typically recommended values for ApEn are not always appropriate for fast dynamic signals [12]. As a consequence, since only few values, within the range suggested in the literature [9], were tested in the previous works where SampEn was applied to AF organization estimation [3–5], the main goal of the present study is to carry out an in depth analysis on SampEn parameters able to achieve optimized classification of AF events which are directly dependent on AF organization. Thus, several combinations of m and r, over a range wider than the one typically recommended, together with the number of analyzed samples, N, and the overlapping effect between adjacent N-length windows have been analyzed. Additionally, since in AF organization estimation, SampEn is applied to a time series that depends on the original ECG sampling rate (f_s), the relationship between N and f_s is also addressed.

Due to the lack of a method able to generate AF signals with *a priori* controlled organization, the use of simulated signals is not possible. As a consequence, real AF signals with organization-dependent events were selected. Two different scenarios, such as the prediction of paroxysmal AF termination and the electrical cardioversion (ECV) outcome in persistent AF, in which organization plays an important role, as corroborated by invasive recordings [13], were analyzed.

2. Materials

In the present work, two databases were used with the aim of providing general recommendations, for optimal SampEn computation, that are applicable to a wide range of AF studies in which organization has to be estimated, like ablation guiding or drug effects. Firstly, a set of paroxysmal (i.e., spontaneously terminating) AF recordings were analyzed to predict the termination of the arrhythmia and, secondly, a set of persistent AF recordings (i.e., requiring external intervention for termination) were studied to predict ECV outcome. The recordings belonging to each data set were sampled at different rates, as will be described in the next subsections.

2.1. Paroxysmal AF database

Fifty Holter recordings of 30 s in length and two leads (II and V1) available in Physionet [14] were analyzed. The database included non-terminating AF episodes (group N), which were observed to continue in AF for, at least, one hour following the end of the excerpt, and AF episodes terminating immediately after the end of the extracted segment (group T). These signals were digitized at a sampling rate of 128 Hz and 16-bit resolution. Nevertheless, they were upsampled to 1024 Hz in order to allow better alignment for QRST complex subtraction, such as Bollmann et al. suggested [13]. This step is necessary to extract the atrial activity (AA) from surface ECGs, see Section 3.1. A cubic splines interpolation method was used because it provided the best resolution (lower than 1 ms) in the R peak detection in comparison with other methods analyzed for the same purpose [15]. After the AA extraction, the residual signal was downsampled back to 128 Hz.

2.2. Persistent AF database

Sixty-three patients (20 men and 43 women, mean age 73.35 ± 9.02 years) with persistent AF lasting more than 30 days, undergoing ECV were followed during 4 weeks. A standard 12-lead ECG was acquired for each patient during the whole procedure and a segment of 30 s in length was extracted from each recording for the analysis. All the signals were digitized at a sampling rate of 1024 Hz and 16-bit resolution.

After the ECV, 22 patients (34.93%) maintained normal sinus rhythm (NSR) during the first month. On the contrary, in 31 patients (49.20%), NSR duration was below 1 month and, in the remaining 10 (15.87%), AF recurred immediately after ECV. These 41 patients constituted the group of AF recurrence. All the patients were in drug treatment with amiodarone. The median arrhythmia duration was 10.58 months (range 1–47.22), echocardiography demonstrated a mean left atrium diameter of 45.82 ± 6.93 mm and 20.63% of the patients presented underlying heart disease. No significative differences were found in the aforementioned clinical parameters between the patients who maintained NSR and relapsed to AF.

2.3. Data preprocessing

In both databases, lead V_1 was chosen for the analysis because previous works have shown that AA is prevalent in this lead [16]. The recordings were preprocessed in order to improve later analysis. Firstly, baseline wander was removed making use of bidirectional high pass filtering with 0.5 Hz cut-off frequency [17]. Secondly, high frequency noise was reduced with an eight order bidirectional IIR Chebyshev low pass filtering, whose cut-off frequency was 70 Hz [18]. Finally, powerline interference was removed through adaptive notch filtering, which preserves the ECG spectral information [19].