Measurement 44 (2011) 653–662

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Use of modified sample entropy measurement to classify ventricular tachycardia and fibrillation

De-Ren Kong [a,b], Hong-Bo Xie [b,*]

[a] School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing, 210094, PR China
[b] School of Electronic and Information Engineering, Jiangsu University, Zhenjiang, 212013, PR China

ARTICLE INFO

Article history:
Received 25 September 2010
Received in revised form 2 December 2010
Accepted 21 December 2010
Available online 30 December 2010

Keywords:
Nonlinear time series
Sample entropy
Complexity
Ventricular fibrillation
Ventricular tachycardia

ABSTRACT

A study of ventricular fibrillation (VF) and ventricular tachycardia (VT) was undertaken using modified sample entropy analysis. Sample entropy (SampEn) is believed to provide the quantitative information about the complexity of experimental data that is corrupted by noise, short in data length. However, the similarity definition of vectors is based on a Heaviside function, of which the boundary is discontinuous and hard, which may cause the problems in the validity and accuracy of SampEn. To deal with the encountered problems, a modified sample entropy (*m*SampEn) based on fuzzy membership function was proposed. Its performances on characterizing VF and VT signals, as well as several simulated time series, demonstrate that *m*SampEn can more efficiently measure the complexity of time series. It is shown that, as a criteria for discriminating between VF and VT, *m*SampEn provides a significantly ($p < 0.0001$) higher (99%) accuracy rate than the standard one.

1. Introduction

The leading cause of death in the industrialized world is cardiovascular disease, among which ventricular fibrillation (VF) is the most frequently seen phenomenon with cardiac patients. High-energy defibrillation is required once a patient was regarded as VF. However, in the daily life, VF often mistakenly diagnosed as ventricular tachycardia (VT), demanding the low-energy cardioversion instead. Therefore, it is crucial to provide an effective and high accurate method to discern between VF and VT [1]. Otherwise, a low-energy cardioversion would be delivered to a patient with VF, which could damage the heart because it is fatal to use the low-energy cardioversion on the heart of a VF patient, hoping that the heart could return to the normal sinus rhythm.

The method earlier used for the classification of VF and VT is based entirely on heart rate. However, the overlap of the heart rate of both arrhythmias of VF and VT may lead to errors. To avoid errors caused by discerning the heart rate, other quantitative analysis techniques for ECG arrhythmia recognition have been proposed previously, such as a sequential hypothesis testing algorithm [1], total least square-based prony modeling algorithm [2], the method based on Lempel–Ziv (L–Z) computation complexity measure [3], qualitative chaos analysis based on symbolic complexity [4,5], correlation dimension method [6,7], Lyapunov exponent method [8]. However, there are some limitations in these methods. Although a normal ECG can be differentiated from an abnormal one using a Lyapunov exponent, VF cannot be separated from VT by this technique [8]. The correlation dimension [6,7] method could be used to measure the degree of complexity of VF and VT, and act as a criterion for discrimination. However, data and computational demands of both methods were so considerable that they were excluded from the clinical use [3]. Zhang et al. [3] proposed a complexity-based method for VF and VT detection, which could achieve an accuracy of

---

[*] Corresponding author. Address: School of Electronic and Information Engineering, Jiangsu University, Xuefu Rd 301#, Zhenjiang 212013, PR China. Tel.: +86 15951288938.

E-mail addresses: derenkong@hotmail.com (D.-R. Kong), xiehb@sjtu.org (H.-B. Xie).

0263-2241/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.measurement.2010.12.004

100% with a 7 s window length, and has a similar result to a sequential hypothesis testing method [1]. It is understood that the same effectiveness of both methods were generated by constructing the binary sequences from raw ECG data and then processing the sequence [3].

To solve the problems of short data length and noisy recording in physiological signals, Pincus [9] developed approximate entropy (ApEn) to measure the system complexity, which is applicable to noisy and short physiological data set. Given N points and tolerance r, ApEn (m, r, N) is approximately equal to the negative average natural logarithm of the conditional probability that two sequences similar for m points within the tolerance remain similar at the next point. Superior to most of nonlinear measures, ApEn has shown potential application to the physiological and clinical signals such as genetic sequences, respiratory patterns, heart rate variability, electrocardiogram, and electroencephalography [10–13].

Nevertheless, ApEn often lacks relative consistency, which induces the difficulty of interpretation of a signal's complexity and deteriorates its differentiation capability. In addition, ApEn suggests more similarity than is present and, thus, is biased. To be free of the bias caused by self-matching, Richman and Moorman [14] developed another related measure of time series' regularity, named sample entropy (SampEn). SampEn has been utilized for the discrimination of VF and VT patterns [15] after it was developed. The results demonstrated SampEn potentially has a clinical value in predicting the onset of spontaneous ventricular tachycardia or fibrillation. Despite its advantage of being less dependent on the length of data set and the relative consistency over a broader range of m, r, and N values, SampEn(m, r, N) is not defined if no template and forward match occurs in the case of small r and N [14]. Moreover, for both ApEn and SampEn, similarity of vectors is based on a Heaviside function. The discontinuity and hard boundary of a Heaviside function may cause a significant change in the SampEn value with a slight change in the tolerance r, which causes the problems in the validity and accuracy of the results [14].

In the present work, we proposed a modified sample entropy (mSampEn) that measures the time series' complexity. Importing the concept of fuzzy sets, vectors' similarity is fuzzily defined in mSampEn on the basis of fuzzy membership function and their shapes. Compared with the standard SampEn, mSampEn showed better relative consistency, more robustness to noise when characterizing signals with different complexities. Moreover, for mSampEn, there is no such restriction of the tolerance r as in the calculation of SampEn. The mSampEn was then applied for the classification of VT and VF patterns and the results of SampEn were also obtained for comparison.

The rest of paper is organized as follows: Section 2 introduces the standard SampEn, proposed mSampEn based on the fuzzy membership function, the test signals, and the VF and VT data sets. Section 3 first makes a widely comparison of mSampEn and SampEn with regard to their ability to capture the degree of time series complexity on the test signals, and then reports the results of applying mSampEn to discriminate VT and VF. Finally, Section 4 includes the discussion and conclusions.

2. Methodology and materials

In this section, we first give a brief overview of the standard SampEn proposed by Richman and Moorman and describe the modified SampEn based on fuzzy membership function in detail [14]. Several benchmark data sets are then introduced to perform the comparison between standard SampEn and mSampEn. Finally, the VT and VF EEG data extracted from MIT database is depicted for the classification purpose.

2.1. Standard sample entropy

For a N points normalized time series $\{u(i) : 1 \leqslant i \leqslant N\}$ with mean zero and unit standard deviation, the following vector sequence can be formed [14].

$$X_i^m = \{u(i), u(i+1), \ldots, u(i+m-1)\} \quad 1 \leqslant i \leqslant N-m+1 \tag{1}$$

Here, X_i^m represents m consecutive u values, commencing with the ith point. The distance d_{ij}^m between X_i^m and X_j^m is defined as

$$d_{ij}^m = d\left[X_i^m, X_j^m\right] = \max_{k \in (0, m-1)} |u(i+k) - u(j+k)| \tag{2}$$

Given tolerance r, and then define

$$B_r^m(i) = \frac{1}{N-m-1} \cdot \sum_{j=1, j \neq i}^{N-m} \Theta(d_{ij}^m - r) \tag{3}$$

$$B_r^m = \frac{1}{N-m} \cdot \sum_{i=1}^{N-m} B_r^m(i) \tag{4}$$

Here, Θ is the Heaviside function

$$\Theta(z) = \begin{cases} 1, & \text{if } z \leqslant 0 \\ 0 & \text{if } z > 0 \end{cases} \tag{5}$$

We then form $(m+1)$-dimensional embedding vectors $X_i^{m+1} = \{u(i), u(i+1), \ldots, u(i+m)\}$ and define A_r^m exactly the same way using X_i^{m+1}. Finally the estimate SampEn of the time series (obtained in the limit of $N \to \infty$) is defined by [14].

$$\mathrm{SampEn}(m, r, N) = -\ln(A_r^m / B_r^m) \tag{6}$$

In the definition of SampEn [14], the similarity of vectors is based on the Heaviside function shown in Eq. (5). The main feature of the Heaviside function is that it provides a step function that converts the input into activity equal to 0 or 1. This function is discontinuous, because there is a "break" in it when its value goes from 0 to 1. This produces a conventional two-state classifier, which judges an input pattern by whether it satisfies the certain precise properties required by membership to a given class. The contributions of all the data points inside the boundary are treated equally, while the data points just outside the boundary are ignored. As a result, the distance d_{ij}^m that is just greater than the tolerance r is not considered in B_r^m or A_r^m and those less than the tolerance are treated equally. As a result, the values of SampEn may be discontinuous and dramatically rise or fall when the tolerance r is slightly changed.