Biomedical Signal Processing and Control 31 (2017) 217–230

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Classification of cardiac arrhythmias based on alphabet entropy of heart rate variability time series

Alan Jovic [a,*], Franjo Jovic [b]

[a] University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, HR-10000 Zagreb, Croatia
[b] Faculty of Electrical Engineering, Josip Juraj Strossmayer University of Osijek, Kneza Trpimira 2B, HR-31000 Osijek, Croatia

ARTICLE INFO

Article history:
Received 2 November 2015
Received in revised form 23 March 2016
Accepted 18 August 2016

Keywords:
Nonlinear dynamics
Entropy
Classification
Feature selection
Cardiac arrhythmia

ABSTRACT

Background: Symbolic dynamics' methods provide a description of time series variability that allows inference of new predictive markers. Classification of disorders using symbolic dynamics is accomplished through the use of nonlinear entropies, measured upon encoded series.
Method: This work applies a recently developed symbolic dynamics method, alphabet entropy (AlphEn) to heart rate variability (HRV) analysis in order to improve automatic classification of cardiac arrhythmias. Experiments are conducted on PhysioNet MIT-BIH Arrhythmia Database. The approach is experimentally compared with other HRV linear and nonlinear feature combinations established in literature. AlphEn is experimentally compared with other common nonlinear entropies: Shannon's entropy, approximate entropy, sample entropy, etc. Feature selection using symmetrical uncertainty is used for discovering relevant AlphEn features and random forest algorithm is used for arrhythmia classification.
Results: The best classification result obtained for six heart rhythms on 20 s segments is achieved for AlphEn no-change threshold $\theta = 100$ ms. AlphEn features improved mean sensitivity of other feature combinations by 2% on average, with the best results achieved: SENS: 91.2%, SPEC: 97.1%, AUC: 99.0%. AlphEn may be used efficiently by adding top 10 ranking features, obtained with symmetrical uncertainty, to other established combinations. AlphEn provides the best incremental result to linear feature combination with respect to the inspected entropies.
Conclusions: AlphEn improves the results of established HRV feature combinations on the problem of automatic cardiac arrhythmia classification. The method enables the extraction of a number of potentially significant, domain-oriented features. It can be used as an accurate first-hand screening for arrhythmia problems.

1. Introduction

One of the most studied, and not yet fully understood, biomedical signals is the heart rhythm. The analysis of the fluctuations of the heart inter-beat intervals (RR-intervals) is known as heart rate variability (HRV) analysis. It is an important prognostic marker of cardiac health [1]. Research shows that heart rhythm of healthy subjects displays interesting short- and long-scale complex fluctuations. HRV decreases with the occurrence of cardiac disorders and aging [2,3]. Through the analysis of HRV, the researchers, among other objectives, aim at classifying heart rhythm patterns [3,4], obtaining accurate cardiac disorder models [5,6], and predicting the onset of cardiac disorders [7].

Disorder modeling using HRV usually includes feature extraction. Biomedical time series variability features in general can be classified into: statistical, geometric, informational, energetic, and invariant [8]. Additionally, several transformations of the heart rate series are possible, with the most common ones being frequency domain transformations: discrete Fourier transform [9], bispectrum calculation [10], higher-order spectrum calculation [11], discrete cosine transform [12], and Hilbert-Huang transform [13]. Other transformations are also possible, such as wavelet transform [14] and symbolic dynamics [15]. The purpose of these transformations is to adapt the time series into a more suitable form that will enable the calculation of some highly accurate domain-oriented modeling features.

In this work, the focus is on symbolic dynamics transformation methods. These methods were first introduced into heart rhythm analysis by Voss et al. [16]. There are several known approaches to symbolic dynamics analysis of HRV [17,18]. Two phases can be

* Corresponding author.
E-mail addresses: alan.jovic@fer.hr (A. Jovic), fjovic@inet.hr (F. Jovic).

http://dx.doi.org/10.1016/j.bspc.2016.08.010
1746-8094/

perceived in the approaches. In the first phase, the original signal is transformed into a new series by discretizing the numerical values into a set of categories, with the potential to obtain more relevant information. This phase is termed symbolization and leads to an overall loss of signal information [19]. It can be achieved by: (1) calculating first derivatives of the series values and retaining only binary information (heart beat shortening or prolongation) [20], (2) taking several categories between minimum and maximum value of the signal (coarse-graining) [17], or (3) taking several categories based on the distance from the series' mean value [19,21]. The second phase includes analyzing the newly encoded sequence using complexity measures such as approximate entropy [20,22], Lempel-Ziv complexity [23], Rényi's entropy [21], or permutation entropy [24], with the aim of quantifying the characteristics of the observed signal.

This work presents the application of a recently developed symbolic dynamics method, appropriately named alphabet entropy [25], to the analysis of HRV. Alphabet entropy preserves both qualitative and quantitative information about the biomedical time series' dynamics on short time scales. It starts by transforming the original signal into a sequence of alphabet letters. The method proposes 27 possible letters of the alphabet, which are encoded by observing four-by-four consecutive measurements and by analyzing their possible changes (prolongation, shortening, or no-change). The method then proceeds to calculate the expanded version of properly adapted Carnap's one-dimensional entropy for each quadruple of measurements. Our previous research showed that the method is highly sensitive to significant changes in the signals, but due to a threshold parameter for no-change, it is mostly insensitive to noise [25]. Also, previous results indicated that common nonlinear measures improve the results of automatic classification of cardiac disorders in comparison with standard linear time and frequency measures [26,27], which supports the use of nonlinear entropy measures in feature combinations.

In order to test the capabilities of the method, the aim of this work is to apply it to automatic classification of cardiac arrhythmias. While classification of arrhythmias based on the whole ECG can provide us with highly accurate results [28,29], classification based on HRV is more challenging. When using HRV alone, accuracy of the constructed models is high only in the cases where the arrhythmias are markedly different, as shown by Asl et al. [3] and Yaghouby et al. [30].

However, the classification of significantly different arrhythmias has limited applicability in practice, because there are several types of cardiac arrhythmias that are not noticeably different if HRV information is used alone (e.g. atrial vs. ventricular arrhythmias), but are quite common in patients [27]. The arrhythmias ought to be classified as accurately as possible in order to enable early detection of potentially significant cardiac disorders.

Therefore, this work focuses on the application of alphabet entropy to classification of several, not easily discernible cardiac arrhythmias. The problems are analyzed on a set of records taken from MIT-BIH Arrhythmia Database from PhysioNet web portal [31].

The major goal of this paper is to discover whether alphabet entropy can increase the accuracy of models of linear and nonlinear HRV feature combinations on the problem. This could be clinically relevant, because more accurate models would enable more reliable detection of arrhythmias, which would lead to earlier treatment for patients. The secondary goal is to compare alphabet entropy with other commonly used entropy features in order to establish its relevance.

The paper is structured into the following sections. In Section 2, a brief overview of symbolic dynamics methods is given in order to provide a rationale for exploring and utilizing an additional symbolic dynamics method. Section 3 describes the alphabet entropy method and its resulting features. Considerations of applying alphabet entropy to HRV series is given in Section 4. Evaluation methodology is presented in Section 5 and results are shown in Section 6. The results are discussed and conclusions are drawn in Sections 7 and 8, respectively.

2. Symbolic dynamics methods overview

Symbolic dynamics methods have been used extensively in biomedical time series analysis: HRV [18,20,32,33], ECG [23,34], EEG [35,36], and joint time series analysis [15]. Most researchers perform binary encoding of the original series [18,32,34], although ternary encoding (with adjustable threshold) can also be applied, particularly for those signals where the lack of significant change is relevant for classification [33]. Coding to a richer alphabet has occasionally been explored [19]. Several complexity features of such encoded sequences are usually analyzed, such as:

1. Entropy and information-based features calculated from short-term encoded sequences' distributions [18,33,34].
2. Other complexity features, e.g. Lempel-Ziv complexity [35] and detrended fluctuation analysis (DFA) [15].
3. Similarity measures between encoded sequences [32]

The commonly used topological entropy measures include: Shannon's entropy (ShEn) [37], correlation entropy (CorrEn) [38], and Rényi's entropy (RenEn) [21,39]. These methods quantify the information in a topological setting where there is no reference to movement of the system's trajectory over time. The entropy methods that do consider the movement of the system trajectory over time are based on approximations of Kolmogorov's entropy for a smaller number of measurements [40]. These include: approximate entropy (ApEn) [22], sample entropy (SampEn) [41], corrected conditional Shannon entropy (CCShEn) [37], and more recently, fuzzy approximate entropy (FuzzyApEn) [42]. These entropy methods proved to be successful for the majority of biomedical time series [35,43], although theoretically, they still require the analysis of several hundred measurements, at least [44].

A different topological approach to measuring system entropy was proposed by Carnap in 1977, as presented in the work of Pudmetzky [45]. The idea was that, instead of dividing a d-dimensional phase space R^d into a number of cells (bins) with a fixed volume v^d, one should divide the space into cells that are attributed to each measurement point. The division of the phase space into such environments is called Voronoi tessellation [46]. The entropy measure that Carnap defined on the Voronoi diagram of an arbitrary finite dimension d is known as Carnap's entropy (CarnEn).

The issues with some of the existing symbolic dynamics' methods, including CarnEn, can be summarized as:

1. Separation of the encoding process method (qualitative) and the complexity evaluation method (quantitative) [18,19,32].

2. Lack of assurance that the complexity evaluation method will provide different results for two sets of different sequences, because of different binning options. This limits the precision of such methods [24,45].

3. Effective application of common entropy methods depends on a large number of available measurements [22,44].

The methodology that we use, based on alphabet entropy, is derived from CarnEn and explained in Section 3. Our approach is novel and timely, because it solves all the three aforementioned issues related to efficient and accurate quantification of time series changes.