Computers in Biology and Medicine 80 (2017) 137–147

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Binarized cross-approximate entropy in crowdsensing environment

Tamara Skoric[a], Omer Mohamoud[a], Branislav Milovanovic[b,c], Nina Japundzic-Zigon[b], Dragana Bajic[a,*]

[a] University of Novi Sad Faculty of Technical Sciences, Centre of excellence CEVAS, Trg Dositeja Obradovica 6, 21000 Novi Sad, Serbia
[b] University of Belgrade, School of Medicine, Laboratory for Cardiovascular, Pharmacology, Serbia
[c] Bezanijska kosa hospital, Neurocardiology department, 11000 Belgrade, Serbia

ARTICLE INFO

Keywords:
Cross-approximate entropy
Conditional entropy
Differential coding
Crowdsensing
Cardiovascular signals

ABSTRACT

Objectives: Personalised monitoring in health applications has been recognised as part of the mobile crowdsensing concept, where subjects equipped with sensors extract information and share them for personal or common benefit. Limited transmission resources impose the use of local analyses methodology, but this approach is incompatible with analytical tools that require stationary and artefact-free data. This paper proposes a computationally efficient binarised cross-approximate entropy, referred to as (X)*BinEn*, for unsupervised cardiovascular signal processing in environments where energy and processor resources are limited.
Methods: The proposed method is a descendant of the cross-approximate entropy ((X)*ApEn*). It operates on binary, differentially encoded data series split into m-sized vectors. The Hamming distance is used as a distance measure, while a search for similarities is performed on the vector sets. The procedure is tested on rats under shaker and restraint stress, and compared to the existing (X)*ApEn* results.
Results: The number of processing operations is reduced. (X)*BinEn* captures entropy changes in a similar manner to (X)*ApEn*. The coding coarseness yields an adverse effect of reduced sensitivity, but it attenuates parameter inconsistency and binary bias. A special case of (X)*BinEn* is equivalent to Shannon's entropy. A binary conditional entropy for m =1 vectors is embedded into the (X)*BinEn* procedure.
Conclusion: (X)*BinEn* can be applied to a single time series as an auto-entropy method, or to a pair of time series, as a cross-entropy method. Its low processing requirements makes it suitable for mobile, battery operated, self-attached sensing devices, with limited power and processor resources.

1. Introduction

Continuous monitoring of a patient's vital parameters has ceased to be a privilege of medical doctors. The rapid tech logical development of sensors, communications devices and communication protocols, followed by the channel distribution regulations, increased the availability of monitoring equipment and made it accessible to a diverse number of users. Continuous personal monitoring during fitness, sport activities, walking, sleep and work, became part of the modern lifestyle. It has already been recognised as a part of the new mobile crowd-sensing concept, where subjects equipped with adequate sensors extract and share information important for mapping and measuring a diversity of features for personal or common benefit, including those relevant to health [1,2]. This tremendously large number of signals, recorded on a daily basis, constitutes a source of valuable diagnostic and prognostic information waiting to be explored. With proper legislative permissions, it could become part of databases that are available to the research community, but could also be used by health authorities for statistical purposes.

The transmission of recorded signals, however, is subject to bandwidth and energy constraints. As outlined in [1], part of the computing versus energy/bandwidth trade-off would be a local analysis performed using a wearable device. Unfortunately, most of the analytical methods are not compatible with such a concept; they require stationary and artefact-free signals that self-acquisition approaches cannot provide, i.e. devices loosely attached or displaced in the case of a subject undergoing rough movements. An example is cross-approximate entropy (X)*ApEn* [3–5], one of the most valuable methods in biomedical research. It is extensively accepted for assessing the complexity of biomedical signals. For example, the complexity change may indicate an increase of adverse occurrences prior to the alternation of any other parameter. This explains the tremendously

* Corresponding author.
E-mail addresses: ceranic@uns.ac.rs (T. Skoric), hafshook@gmail.com (O. Mohamoud), branislav_milovanovic@vektor.net (B. Milovanovic), nzigon@med.bg.ac.rs (N. Japundzic-Zigon), dragana.bajic@gmail.com, draganab@uns.ac.rs (D. Bajic).

http://dx.doi.org/10.1016/j.compbiomed.2016.11.019
Received 22 August 2016; Received in revised form 23 November 2016; Accepted 30 November 2016
0010-4825/

high number of (X)*ApEn* citations in the open-access literature [6]. It also explains, among others, the number of (X)*ApEn* improvements, alternations, and adjustments to specific applications [7–13].

A reliable (X)*ApEn* estimate requires artefact-free and stationary signals, and its quadratic computational complexity is not fit for a local analysis concept [1]. We propose a robust modification of cross-approximate entropy that ensures a speedy, energy efficient, and unsupervised implementation. It is motivated by a long-known property that infinitely clipped speech signals preserve the intelligibility [14], i.e. the property that a majority of the information content is stored within the difference of the adjacent signal amplitudes. We set a hypothesis that binary, differentially encoded, biomedical signals preserve sufficient information to allow approximate entropy estimation, in spite of the coding coarseness. Differential coding is known to yield stationary signals, whereas the binary representation attenuates the effects of artefacts. The binarisation also allows a considerable reduction of the computational complexity, reducing the cardinality of vector sets required for the entropy estimation procedure.

Binary, differentially encoded signals, coupled with discrete, unconditional Shan n's entropy [16], have already found applicability in numerous, joint symbolic, dynamic (JSD) studies [15].

The aims of the paper are to: 1) introduce and explain binary cross-approximate entropy (X)*BinEn* as a method that considerably increases the computational efficiency of entropy estimation (Section 2.2.), 2) establish a theoretical relationship between (X)*BinEn* and classical Shan n's entropy (Section 2.3.), 3) prove that (X)*BinEn* estimates are comparable to (X)*ApEn* results and identify the technique's limitations (Section 3.2), 4) explore the (X)*BinEn* consistency regarding the the binary bias (Section 3.1) and regarding the entropy parameters known to cause (X)*ApEn* instability (Section 3.3), and 5) append a dynamic complementary measure, based on elements embedded in the (X) *BinEn* procedure (Section 3.4).

2. Materials and methods

2.1. Experimental protocol and signal acquisition

The method is evaluated using signals from laboratory rats exposed to stress. An extensive entropy study has already been performed on these signals, so the referent values to which (X)*BinEn* results can be compared already exist [18]. Physiological aspects of these signals are elaborate and known [17]. The target groups for this method, namely, the mobile crowd-sensing subjects, or individual subjects with wearable sensors, are likely to be healthy. It may be speculated that the stress is the major adverse factor the healthy subjects suffer from.

Signals for this study were derived from blood pressure waveforms, and were recorded from: a) outbred male normotensive Wistar rats (NRM), b) borderline hypertensive rats that were F1 offspring of Wistar dames and spontaneous hypertensive sires (BHR). The mean weight ($\pm$ standard deviation) of each rat was 330 ± 20 g. A pressure sensor with a wireless transmitter (TA11PA-C40, DSI, Transoma Medical, USA) was implanted in the abdominal aorta under combined ketamine and xylazine anesthesia, along with gentamicin, followed by metamizol injections for pain relief. The recording started 10 d after the surgery, to allow post-surgical recovery. The arterial blood pressure (BP) signal was digitised at 1 kHz, and was transmitted to a PC equipped with the Dataquest A.R.T. version 4.0 software for cardiovascular signals analysis. The systolic blood pressure signal (SBP) is defined as the local maximum of the blood pressure waveform, and the pulse interval (PI) signal is estimated as the time interval between successive points of maximum pressure.

During the experiments, the animals were exposed to two types of stresses—shaker stress, with rats positioned on a platform shaking at 200 cycles/min, and restraint stress, with rats placed in supine position in a plexiglass restrainer tube (inner diameter equal to 5.5 cm with holes). Prior to any stress exposure, baseline signals were recorded (BASE). More details about the physiologically relevant design of this experiment can be found in [17].

The number of animals per experimental group was 6 or 7. This number was determined to be satisfactory according to the variability of the parameters in the control group of rats (statistical software 'Power Sample Size Calculation'). All experimental procedures in this study conformed to the European Communities Council Directive of 24 November 1986 (86/609/EEC). The experimental protocol was granted by the School of Medicine, University of Belgrade ethics review board (approval reference number 1306/1–3).

Each one of the original SBP and PI time series is provided with a set of fifty artificially generated control signals that included isodistibutional surrogate time series signals (randomly permuted original signal samples) [19] and pseudo-random time series signals with respective uniform and rmal distributions and with the same mean and standard deviation. It should be ted that pre-processing is necessary for (X)*ApEn* but t for the proposed (X)*BinEn*, and it includes the visual inspection of time series, the elimination of artefacts and slow signal components using a filter designed specifically for cardiovascular signals [20]. A stationarity test [21,22] must also be performed, since standard deviation, required for (X)*ApEn,* can be reliably estimated from stationary signals only. X*ApEn* additionally requires normalised and centralised signals.

The statistical significance of the obtained results was assessed using a repeated measures A VA test, and significance was marked with '*' for $p < 0.05$ and '**' for $p < 0.01$.

2.2. Crowdsensing

The crowdsensing is a novel concept of sharing data collected by sensing devices. Compared to traditional wireless sensor networks, it has a unique advantage of exploiting the worldwide availability of smartphones: it is estimated that more than 80% of world population use a mobile phone, and more than half are smart. Each smartphone has a computing, communication and storage resources of a small computer, and it is provided with a couple of already embedded sensors. A crowdsensing application transforms the phone into a smart data acquisition device, ready to share the information collected from many subjects.

Monitoring the cardiovascular parameters to investigate an ambiental influence can also be a crowdsensing topic, and volunteers for cardiac research are numerous. The most important sensor is single or double channel electrocardiograph (ECG) [23]. In spite of efforts for improvement [24], the data transmission is still the major energy consumer. Available bandwidth would also be compromised if complete acquired data should be transmitted. To save the resources, the processing is performed at the phone and only the final result is transmitted. A simple calculation shows the effects of local analysis: tem minutes – 600 s – ECG recording with 1 kHz sampling frequency and 10 bit A/D convertor yields 6 000 000 bits for transmission, while a single parameter requires 16 or 32 bits.

Unfortunately, one of the major crowdsensing problems is reliability of the recorded data. Unprofessional sensor attachment and moving subject yield n-stationary data full of artifacts that might destroy the validity of the estimated result. An example presented in [25] has shown that eleven (1%) erroneously extracted heart beats lead to five times increased sympathovagal low frequency/high frequency ratio.

An ultimate goal for such a concept is to develop parallel parameter estimation methods that would be a) insensitive to non-stationarities; b) insensitive to reasonable amount of errors and therefore c) implemented without the pre-processing.

2.3. X*BinEn* procedure

The cross-entropy of binary, differentially modulated signals, or (X)