



ELSEVIER

Contents lists available at ScienceDirect

Neurocomputing

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

Non uniform Embedding based on Relevance Analysis with reduced computational complexity: Application to the detection of pathologies from biosignal recordings



Jorge A. Gómez-García^{a,*}, Juan I. Godino-Llorente^a, Germán Castellanos-Dominguez^b

^a Bioengineering and Optoelectronics (ByO) Group, Universidad Politécnica de Madrid, Ctra. Valencia, km. 7, Madrid 28031, Spain

^b Signal Processing and Recognition (SPR) Group, Universidad Nacional de Colombia, km. 7 vía al Magdalena, Manizales, Colombia

ARTICLE INFO

Article history:

Received 23 February 2012

Received in revised form

26 November 2012

Accepted 14 January 2013

Available online 18 October 2013

Keywords:

Nonlinear dynamics

Non uniform embedding

Time-delay embedding

ABSTRACT

Nonlinear analysis tools for studying and characterizing the dynamics of physiological signals have gained popularity, mainly because tracking sudden alterations of the inherent complexity of biological processes might be an indicator of altered physiological states.

Typically, in order to perform an analysis with such tools, the physiological variables that describe the biological process under study are used to reconstruct the underlying dynamics of the biological processes. For that goal, a procedure called time-delay or uniform embedding is usually employed. Nonetheless, there is evidence of its inability for dealing with non-stationary signals, as those recorded from many physiological processes.

To handle with such a drawback, this paper evaluates the utility of non-conventional time series reconstruction procedures based on non uniform embedding, applying them to automatic pattern recognition tasks. The paper compares a state of the art non uniform approach with a novel scheme which fuses embedding and feature selection at once, searching for better reconstructions of the dynamics of the system. Moreover, results are also compared with two classic uniform embedding techniques. Thus, the goal is comparing uniform and non uniform reconstruction techniques, including the one proposed in this work, for pattern recognition in biomedical signal processing tasks. Once the state space is reconstructed, the scheme followed characterizes with three classic nonlinear dynamic features (Largest Lyapunov Exponent, Correlation Dimension and Recurrence Period Density Entropy), while classification is carried out by means of a simple k -nn classifier. In order to test its generalization capabilities, the approach was tested with three different physiological databases (Speech Pathologies, Epilepsy and Heart Murmurs).

In terms of the accuracy obtained to automatically detect the presence of pathologies, and for the three types of biosignals analyzed, the non uniform techniques used in this work lightly outperformed the results obtained using the uniform methods, suggesting their usefulness to characterize non-stationary biomedical signals in pattern recognition applications. On the other hand, in view of the results obtained and its low computational load, the proposed technique suggests its applicability for the applications under study.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The automatic detection of pathologies is an increasingly important issue that has gained popularity in the last few years. The main goal in this field is to develop computer-aided diagnostic systems, enabling an objective assessment, hence helping diagnosis, reducing the evaluation time, and improving the clinical treatment given to each patient [1].

These automatic detection systems require signals recorded from the biological processes under study, which are further analyzed to make decisions about the state of the patient. For this purpose, the analysis by means of nonlinear dynamic tools has become useful for understanding and characterizing biological processes, among other reasons, due to the high nonlinearity, complexity, and non-stationary behavior of biological signals [2], as well as the known relationship between pathophysiological dysfunction and changes in nonlinear dynamic features [3].

In this regard, the nonlinear analysis of time series depends on the reconstruction of the dynamical behavior of the underlying system, so that the evolution of the system is represented in a m -dimensional space. The most common technique for this purpose is based on the Takens' Time-Delay Embedding Theorem [4], which

* Corresponding author. Tel.: +34 913367829.

E-mail addresses: jorge.gomez.garcia@upm.es (J.A. Gómez-García), ignacio.godino@upm.es (J.I. Godino-Llorente), cgcastellanosd@unal.edu.co (G. Castellanos-Dominguez).

requires the computation of two parameters: the embedding dimension m and the time lag τ , chosen to optimize the spread of the time series without confusing the underlying dynamics in the m -dimensional space.

Takens' theorem provides a sufficient condition for assessing the embedding dimension m , suggesting it to be twice plus one larger than the actual dimension of the system. Since that a priori knowledge of the system's properties is not achievable in most of the cases, other approaches are preferred for estimating the embedding dimension, for example, the coherence of embedded data points [5]. In this respect, one of the most popular methods is the *False Nearest Neighbors* (FNN) method, aiming to search for points in time series being neighbors in the embedding space, but that should not be, since their future temporal evolution is too different [6].

On the other hand, and since the Takens' theorem is stated for ideal time series, for which an infinite amount of data is available and no noise is present, the theorem is silent on how to estimate the time delay τ , because in the aforementioned ideal scenario any τ value might be valid. However, in real world time series, the choice of the time delay should be carried out thoroughly to assure a proper reconstruction. At this point, several methods have been proposed, the *Average Mutual Information* (AMI) criterion [7] being one of the most prominent. Despite the AMI and FNN criteria are the most utilized in the state of the art, some works have stated that instead of focusing on searching τ and m separately, it might be more effective to consider directly the embedding window as it relates them both [8–11].

Although the uniform embedding technique has been widely used for reconstructing attractors providing successful results in several pattern recognition tasks, some issues still arise regarding the analysis of non-stationary biosignals with multiple periodicities. In fact, when choosing τ , a short time lag might be optimal for high frequency components, whereas a long lag might be so for low frequency components and modulations. Therefore, a compromise remains inadequate for both time-scales [12]. This fact is crucial for the characterization of complex and non-stationary signals, such as those produced by biological processes. Bearing that in mind, it has been argued that the *non uniform embedding* [12] might deal with such issues, by replacing the single time delay τ with a lag vector $\vec{\tau}$.

There are some approaches for searching an optimum lag vector $\vec{\tau}$ in a non uniform embedding reconstruction, though they have been employed mainly for modeling or forecasting purposes. For instance, an early approach is based on the *Minimum Description Length* (MDL) principle [12], which combines embedding and modeling into a single procedure, on which the quality of the embedding is related to the quality of the whole model. Another approach grounded on the False Nearest Neighbors criterion is also presented in [13], which employs the ratio of false neighbors for increasingly adding time delays until that ratio drops to zero. In another approach [14], a feed-forward neural network trained by structural learning is suggested for finding the optimal set $\vec{\tau}$. In [15], two statistics are introduced for evaluating the quality of a non uniform embedding reconstruction. The usage of geometrical criterion for selecting the time lag vector is also described in [16], which aims to spread the attractor in the state space as much as possible, such that the magnitude of the attractor's spreading in the state space is a measure of the embedding quality. This approach can be improved using evolutionary algorithms as discussed in [17]. Lastly, an objective function guiding the search for an optimal state-space reconstruction, in terms of noise amplification and minimization of the complexity of reconstruction, is given in [18].

Nevertheless, all the above referred approaches provide either no clue of how to find the best delay vector, or have a high

computational burden (as in most of cases they need to model, to optimize several parameters, or to employ greedy search algorithms), making them unsuitable for pattern recognition tasks.

To overcome the required high computational costs, providing a proper non uniform embedding within the framework of the nonlinear characterization of biosignals, this work proposes a novel method that attempts, at the same time, minimizing the redundancy and maximizing the relevance as part of the feature selection-like schema. To test the feasibility of the proposed schema, several experiments are performed using different biosignal databases. For characterization purposes, the following nonlinear features have been used: the Correlation Dimension (d_2), the Largest Lyapunov Exponent (Λ) [6], and Recurrence Period Density Entropy (ρ) [19], while for classification, a simple k -nn classifier is employed. It must be remarked that the main concern of this paper is not to improve the accuracies found in the state of the art for the aforementioned applications (Speech Pathologies, Epilepsy and Heart Murmurs detection), but to investigate the suitability of the non uniform embedding state space reconstruction technique developed for pattern recognition tasks, comparing with the classic uniform and other non uniform embedding techniques.

This paper is organized as follows: Section 2 presents theoretical aspects of the embedding procedures, as well as the Non Uniform embedding proposal. Section 3 presents the methodology and describes the databases used in this study. Section 4 presents the experiments and the results. Finally, Section 5 presents the discussions, and Section 6 the conclusions of this work.

2. Theoretical background

Fig. 1 depicts an outline of the system that has been developed to automatically detect pathologies. The system characterizes different types of biosignals by using nonlinear features extracted from a previous reconstruction of the state space using both uniform and non uniform techniques. The different stages of the process are described next.

2.1. State space reconstruction

Roughly speaking, the state space has been reconstructed either using uniform or non uniform techniques. Two uniform embedding methods have been used throughout this work, termed as Uniform 1 and Uniform 2; and two non uniform methods have also been used, termed as Non Uniform 1 and Non Uniform 2. The following presents an introduction to these four methods.

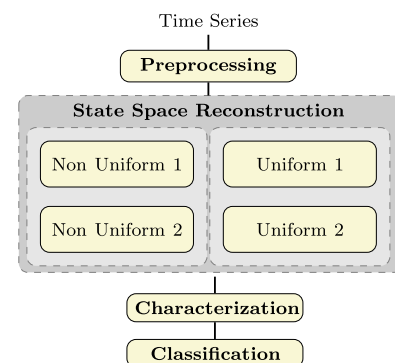


Fig. 1. Diagram of an automatic pathology detection system based on Uniform and Non uniform embeddings.

Download English Version:

<https://daneshyari.com/en/article/406562>

Download Persian Version:

<https://daneshyari.com/article/406562>

[Daneshyari.com](https://daneshyari.com)