ARTICLE IN PRESS

Physica A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physica A

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

Spatial-dependence recurrence sample entropy

Tuan D. Pham [a,*], Hong Yan [b]

[a] *Department of Biomedical Engineering, Linköping University, 58183 Linköping, Sweden*
[b] *Department of Electronic Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong*

HIGHLIGHTS

- Quantifying irregularity with sample entropy is based only on the distance measure.
- Sequential ordering is an important criterion for computing sample similarity.
- The new method is introduced to consider both sources of information.

ARTICLE INFO

Article history:
Received 7 August 2017
Received in revised form 8 October 2017
Available online xxxx

Keywords:
Time series
Irregularity
Sample entropy
Recurrence plot
Binary-level co-occurrence matrix
Spatial dependence

ABSTRACT

Measuring complexity in terms of the predictability of time series is a major area of research in science and engineering, and its applications are spreading throughout many scientific disciplines, where the analysis of physiological signals is perhaps the most widely reported in literature. Sample entropy is a popular measure for quantifying signal irregularity. However, the sample entropy does not take sequential information, which is inherently useful, into its calculation of sample similarity. Here, we develop a method that is based on the mathematical principle of the sample entropy and enables the capture of sequential information of a time series in the context of spatial dependence provided by the binary-level co-occurrence matrix of a recurrence plot. Experimental results on time-series data of the Lorenz system, physiological signals of gait maturation in healthy children, and gait dynamics in Huntington's disease show the potential of the proposed method.

1. Introduction

In the combination of statistics and information theory, the approximate entropy, denoted as ApEn, was developed for quantifying the amount of irregularity or predictability of fluctuations in time series data [1,2]. The sample entropy [3], denoted as SampEn, was then introduced as a modified algorithm of ApEn that removes the bias in counting self-matching patterns included in the ApEn. Both ApEn and SampEn, and their modified versions have been increasingly found useful in many applications, particularly in the analysis of physiological time series. Some recently published works include applications in network theory [4], analyses of heart rate variability and systolic blood pressure variability [5], postural analysis [6], and analysis of traffic signals [7]. However, several applications and studies are reportedly in favor of SampEn [8–10].

Given its popularity as a useful measure for quantifying irregularity of time series, SampEn has several technical shortcomings. Many attempts have been made to improve SampEn mainly to reduce its sensitivity to the selection of values for its model parameters [11–13]. One of the most recent efforts has tried to modify SampEn to overcome the limitation

* Corresponding author.
E-mail addresses: tuan.pham@liu.se (T.D. Pham), h.yan@cityu.edu.hk (H. Yan).

https://doi.org/10.1016/j.physa.2017.12.015
0378-4371/

Please cite this article in press as: T.D. Pham, H. Yan, Spatial-dependence recurrence sample entropy, Physica A (2017), https://doi.org/10.1016/j.physa.2017.12.015.

about the relationship between the parameters and the length of time series [7]. However, there is little effort in developing the formulation of SampEn that can also capture the sequential ordering of the time series, except for the indirect case of using multiple scales of time series data [10,14–16]. In this paper, we introduce a method for quantifying irregularity in time series based on the formulation of SampEn that computes the probability of sample similarity by incorporating sources of information from the distance measure of sample points of a time-series and the spatial orientation of the binary-level co-occurrence matrix of its recurrence plot, where the latter information is a spatial representation of the sequential ordering.

The rest of the paper is organized as follows. Section 2 briefly presents the mathematical formulations of the sample entropy, recurrence plots, and binary-level co-occurrence matrix, which are used as the basis for developing the framework of the proposed spatial-dependence recurrence sample entropy. Experimental results and discussion about the proposed method are presented in Section 3, which includes the testing and comparison of proposed method and the sample entropy using three datasets: the time series of the Lorenz system and their surrogate time series, the complex physiological signals of gait maturation in children, and gait dynamics in Huntington's disease obtained from publicly accessible PhysioNet databases.

2. Methods

2.1. Sample entropy

The sample entropy (SampEn) [3] is a measure of irregularity in time series. The formulation of SampEn is briefly described as follows. Consider a time series X of length N taken at regular intervals: $X = (x_1, x_2, \dots, x_N)$, and a given embedding dimension m, a set of newly reconstructed time series from X, denoted as Y^m, can be established as $Y^m = (y_1^m, y_2^m, \dots, y_{N-m+1}^m)$, where $y_i^m = (x_i, x_{i+1}, \dots, x_{i+m-1})$, $i = 1, 2, \dots, N-m+1$. The probability of vector y_i^m being similar to vectors y_j^m is computed as $(N-m-1)^{-1}$ times the number of vectors y_j^m within a similarity tolerance of y_i^m, where self-matches are excluded, and mathematically expressed as follows

$$B_i^m(r) = \frac{1}{N-m-1} \sum_{j=1}^{N-m} H[d(y_i^m, y_j^m)],\ i \neq j, \tag{1}$$

where r is a real positive value for the similarity tolerance, and $H(d(y_i^m, y_j^m))$ is the Heaviside function, defined as

$$H[d(y_i^m, y_j^m)] = \begin{cases} 1 \ : d(y_i^m, y_j^m) \leq r \\ 0 \ : d(y_i^m, y_j^m) > r \end{cases} \tag{2}$$

The distance between the two vectors is obtained by using the Chebyshev distance or the L_∞ metric, where the distance between two vectors is the largest of their differences along any coordinate dimension and mathematically expressed as

$$d(y_i^m, y_j^m) = \max_k(|x_{i+k-1} - x_{j+k-1}|),\ k = 1, 2, \dots, m. \tag{3}$$

The probability of pairs of vectors or data points of length m having the Chebyshev distance $\leq r$, denoted as $B^m(r)$, is expressed as

$$B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r). \tag{4}$$

Similarly, $A_i^m(r)$ is defined as $(N-m-1)^{-1}$ times the number of vectors $y_j{}^{m+1}$ within a similarity tolerance of $y_i{}^{m+1}$, where $j = 1, \dots, N-m, j \neq i$, and setting

$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r). \tag{5}$$

Finally, SampEn is calculated as

$$\text{SampEn}(m, r, N) = -\log\left[\frac{A^m(r)}{B^m(r)}\right], \tag{6}$$

where $A^m(r) \leq B^m(r)$, which is imposed by the Chebyshev distance.

2.2. Recurrence plots

In nonlinear dynamics and chaos theory, a recurrence plot (RP) [17] is a visualization method that shows the times at which a phase-space trajectory approximately revisits the same area in the phase space. Let $X = \{x\}$ be a set of phase-space states, in which x_i is the ith state of a dynamical system in m-dimensional space. An RP is constructed as an $N \times N$ matrix in which an element (i, j), $i = 1, \dots, N, j = 1, \dots, N$, is represented with a black dot if x_i and x_j are considered to be closed to

Please cite this article in press as: T.D. Pham, H. Yan, Spatial-dependence recurrence sample entropy, Physica A (2017), https://doi.org/10.1016/j.physa.2017.12.015.