Physica A 512 (2018) 824–836

Contents lists available at ScienceDirect

Physica A

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

A Method for the computation of entropy in the Recurrence Quantification Analysis of categorical time series

Giuseppe Leonardi *

Paderborn University, Faculty of Cultural Studies, 100 Warburger Str., 33098 Paderborn, Germany
University of Finance and Management, Faculty of Psychology, ul. Pawia 55, 01-030 Warsaw, Poland

HIGHLIGHTS

- In categorical recurrence plots rectangular blocks are the prominent structures.
- The extraction of several recurrence measures could be biased by this fact.
- We propose to change the method of computation of informational entropy.
- Our proposed method seems to be an unbiased estimate of entropy in this case.

ARTICLE INFO

Article history:
Received 15 April 2017
Received in revised form 15 March 2018
Available online xxxx

Keywords:
Recurrence Quantification Analysis
Entropy
Categorical time series
Dynamical measures
Recurrence Plot

ABSTRACT

In this work, I propose a new method for the computation of informational entropy from Recurrence Plots when the analyzed time series are categorical in nature. In such cases, there is typically a simplification in choosing the parameters of the analysis, in the sense that no embedding in multidimensional space is usually assumed and that recurrence is restricted to exact matching (equivalence) of the numerically coded categories. However, such a simplified parameterization brings about some notable changes in the appearance of the obtained Recurrence Plots, which has consequences for the extraction of the standard dynamical measures. Specifically, a categorical Recurrence Plot is often composed of rectangular structures rather than line structures (diagonal and horizontal/vertical), over which the recurrence quantification measures were originally proposed. Starting from this observation, I consider alternative computational procedures to extract a non-biased measure of entropy for the categorical case, showing the viability of such a choice with simulated data

1. Introduction

Recurrence Quantification Analysis (RQA) and its bivariate equivalent, Cross Recurrence Quantification Analysis (CRQA), are becoming very popular methods of analysis in the behavioral sciences since an interest in the dynamical features of the phenomena studied has become more prevalent in this field [1,2]. These methods involve taking as the basic input data one (RQA) or two (CRQA) temporally ordered streams (time series) of an observed behavior and, after the choice of the appropriate parameters for the analysis, building a Recurrence Plot (RP) of these streams by aligning them along the horizontal and vertical axes of the plot. An RP is basically a binary matrix defined as

$$\mathbf{R}_{i,j}(\varepsilon) = \Theta\left(\varepsilon - \left\|x_i - x_j\right\|\right), \qquad i,j = 1,\ldots,N \tag{1}$$

* Correspondence to: University of Finance and Management, Faculty of Psychology, ul. Pawia 55, 01-030 Warsaw, Poland.
E-mail address: g.leonardi@vizja.pl.

https://doi.org/10.1016/j.physa.2018.08.058

where N is the number of measured points x in the time series, i and j indicate the indexes for the horizontal and vertical coordinates of the matrix, ε is a threshold distance (the radius parameter), $\Theta(.)$ is the Heaviside function where $\Theta(x) = 0$ when $x < 0$ and $\Theta(x) = 1$ otherwise, and $\|.\|$ is a norm, most typically the Euclidian norm. $\mathbf{R}$ is then graphically filled with black points $p_{i,j}$ at coordinates i and j when $\mathbf{R}_{i,j} = 1$, under a given parameterization of embedding, delay and threshold (see next section). The general graphical appearance of the RP is visually informative about some global dynamical properties of the time series analyzed. However, some of the structures in the plot are specifically meaningful in terms of such properties and, after the appropriate evaluation, can give rise to a set of recurrence measures, which are the basis of the recurrence *quantification* [3–5] (for a complete introduction and a review of the most significant aspects of RQA see [6]). In the present work, I discuss a possible extension of these recurrence measures for categorical data, sometimes also called symbolic or nominal data.

In what follows, the typical measures that can be extracted from the recurrence plot $\mathbf{R}$ are first briefly reviewed, with a focus on the measure of entropy, which will be discussed in greater detail. Next, I highlight how the application of RQA and CRQA to categorical time series is not only possible but indeed very fruitful, especially in studies of coordination and joint action. This is inferred by the increasing number of studies in this field based on recurrence-based methodologies as well as by the creative research-based innovations brought about into the method itself in recent years (see e.g., [7–10]).

The shift from continuous signals to categorical ones has usually left unchanged the interpretation of the recurrence structures and the computational algorithms of the quantification measures previously introduced, focusing only on the specific parameterization of the categorical case. However, there is a need to discuss and underline the specificity of categorical RPs when contrasted to continuous ones, trying to evaluate the consequences these specificities may have for the RQA measures. I illustrate why this is the case, and I propose a modified computational algorithm for the measure of entropy, tailored for the typical structures of the recurrence points emerging in a categorical RP. Finally, I demonstrate with simulated data how the new measure of entropy seems to better characterize the complexity of the given time series.

2. Recurrence quantification

Recurrence analysis methods arose within the field of mathematics and mathematical physics [11,12]. They are used to characterize the dynamics of the behavior of a dynamical system in terms of recurrences, when we have access to only a limited, usually unidimensional signal originating from that system. In such cases, it has been shown that through a time delayed embedding of the continuous signal, it is possible to appropriately reconstruct a topological equivalent of the original trajectory in the phase space of the system [11]. To do so, we need to set two important parameters in the analysis: the **embedding dimension** (m) and the **delay** (d). The first parameter sets the number of dimensions to which a unidimensional signal is promoted, while the second indicates the sampling distance along that unidimensional signal at which successive embedded dimensions are estimated.

A trajectory that has been reconstructed in an embedded m-dimensional space this way can be concisely represented based on *recurrences* (i.e., the approaching at different times of phase space neighborhoods that had been already visited) in the so called Recurrence Plot (RP) [12]. To determine what counts as a recurrence, we need to set another important parameter, the threshold distance or **radius** (ε), which defines the threshold below which the distance between two trajectories is small enough to be considered recurrent according to a chosen norm $\|.\|$, and hence is represented with a point in the plot (see Eq. (1)).

An RP then represents concisely a dynamical signal graphically through its temporal recurrences (see the right side of Fig. 1 for some examples), but apart from such a visual representation, it is also possible to extract and observe from it some of the most important dynamical features of the system. This last step moved RPs from a qualitative to a quantitative description and has been formalized in a series of seminal studies by Zbilut and Webber [3,4] and by Marwan and colleagues [5,13]. The development of recurrence quantification-based techniques is still ongoing, and new proposals and measures regularly emerge for specific applications.

At the basis of such a step – the quantification of the RPs – is the observation and interpretation of the structures appearing in the RP, the points and particularly the line-structures forming on the plot.

The first basic measure of recurrence quantification is simply the total number of recurrence points in an RP, typically expressed in terms of a proportion – the recurrence rate or percent recurrence – over the total number of possible points in the plot:

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbf{R}_{i,j}(\varepsilon) \tag{2}$$

Obviously, the larger the radius, the more recurrent points will appear in the plot and hence the higher the **Recurrence Rate** (RR) measure.

Important structures in an RP are diagonal lines (parallel to the main diagonal of the RP). A diagonal line of length l can be defined as follows:

$$\left(1 - \mathbf{R}_{i-1,j-1}\right)\left(1 - \mathbf{R}_{i+l,j+l}\right) \prod_{k=0}^{l-1} \mathbf{R}_{i+k,j+k} \equiv 1 \tag{3}$$