Physica A 405 (2014) 63–69

Contents lists available at ScienceDirect

Physica A

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

A parametric, information-theory model for predictions in time series

M.T. Martín [a], A. Plastino [b,*], V. Vampa [c], G. Judge [d]

[a] Facultad de Ciencias Exactas and Conicet, Universidad Nacional de La Plata, Argentina
[b] IFLP-CCT-CONICET, Universidad Nacional de La Plata, C. C. 727, 1900 La Plata, Argentina
[c] Facultad de Ingeniería, Universidad Nacional de La Plata, Argentina
[d] Giannini Hall, University of California at Berkeley, CA, United States

HIGHLIGHTS

- Economic, chaotic time-series are analyzed with information theory methods.
- The maximum entropy principle is appealed to.
- Predictions concerning the Dow-Jones series are made.

ARTICLE INFO

Article history:
Received 29 October 2013
Received in revised form 20 February 2014
Available online 12 March 2014

Keywords:
Time series
Information theory
Parametric Inference

ABSTRACT

In this work, a method based on information theory is developed to make predictions from a sample of nonlinear time series data. Numerical examples are given to illustrate the effectiveness of the proposed method.

1. Introduction

One of the most important aspects of nonlinear dynamics is concerned with time series analysis and how to predict the future behavior of a system. Economic-behavior systems, like physical and biological systems, are stochastic in nature. Thus the predictability of a dynamic system's behavior may best be considered in a random outcome context and probability distribution functions may be used to measure the statistical nature of the system. System predictability may then be studied quite naturally by information theoretic methods-functionals, since the focus is random in nature, and the functionals may be interpreted in terms of uncertainty and by measures of the difference between the statistical distributions. Thus, an information theoretic basis is provided for unlocking the dynamic content of nonlinear time series data and using this information to predict the future behavior of the system.

2. Method

Given a signal $\mathbf{x}$ from a dynamical system $D : \mathbb{R}^S \to \mathbb{R}^S$, the corresponding time series consists of a sequence of measurements $\{v(t_n), n = 1, \ldots, N\}$ on a system considered to be in a state described by $\mathbf{x}(t_n) \in \mathbb{R}^S$ at discrete times

* Corresponding author. Tel.: +54 22145239995; fax: +54 2214523995.
E-mail addresses: plastino@fisica.unlp.edu.ar, angeloplastino@gmail.com (A. Plastino).

http://dx.doi.org/10.1016/j.physa.2014.02.055
0378-4371/

t_n, where N is the length of the time series. It is known (see Refs. [1,2]) that for $T \in \mathbb{R}$, $T > 0$, there exists a functional form of the type,

$$v(t+T) = F(\mathbf{v}(t)), \tag{1}$$

where,

$$\mathbf{v}(t) = [v_1(t), v_2(t), \ldots, v_d(t)], \tag{2}$$

and $v_i(t) = v(t-(i-1)\Delta)$, for $i = 1, \ldots, d$, where Δ is the time lag and d is the embedding dimension of the reconstruction. In this paper d is determined from the data itself, by the method of false nearest neighbors.

We consider (as in Ref. [2]) a particular representation for the mapping function of Eq. (1), expressing it, using Einstein's summation notation, as an expansion of the form

$$F^*(\mathbf{v}(t)) = a_0 + a_{i_1} v_{i_1} + a_{i_1 i_2} v_{i_1} v_{i_2} + a_{i_1 i_2 i_3} v_{i_1} v_{i_2} v_{i_3} + \cdots + a_{i_1 i_2 \ldots i_{np}} v_{i_1} v_{i_2} \ldots v_{i_{np}}, \tag{3}$$

where $1 \le i_k \le d$ and np is the polynomial degree chosen to expand the mapping F^*. The number of parameters in Eq. (3) corresponding to k terms (the degree), is the combination with repetitions,

$$\binom{d}{k}^* = \frac{(d+k-1)!}{k!(d-1)!}. \tag{4}$$

The length of the vector of parameters, **a** is

$$N_c = \sum_{k=1}^{np} \binom{d}{k}^*. \tag{5}$$

As an information recovery criterion we determine the vector **a** by using the maximum entropy principle (MEP) [3]. Our objective is a model that attains high predictive ability. Computations are made on the basis of a specific information supply, given by M points of the series

$$\{\mathbf{v}(t_n), v(t_n+T)\}, \quad n = 1, \ldots, M. \tag{6}$$

Given the data set in Eq. (6), the parametric mapping in Eq. (3) will be determined by satisfying following condition:

$$v(t_n+T) = F^*(\mathbf{v}(t_n)) \quad n = 1, \ldots, M. \tag{7}$$

In this way, a rectangular system of equations is obtained,

$$W \cdot \mathbf{a} = (v(t_n+T))_{n=1,\ldots,M}, \tag{8}$$

where W is a matrix of size $M \times Nc$, M is the length of the information-points in Eq. (6), and Nc is the number of parameters of the model (cf. Eq. (3)). Using the Moore–Penrose pseudo-inverse of the matrix W [4], the solution is

$$\mathbf{a} = P_{MP}(W) \cdot (v(t_n+T))_{n=1,\ldots,M}, \tag{9}$$

where $P_{MP}(W) = (W^T * W)^{-1} * W^T$. Thus, the most probable configuration is the one that is linked to the mean value of the probability distribution, associated to the pseudo-inverse matrix of W.

Once the pertinent parameters are determined, they are used to predict MP new series' values,

$$(\widehat{v}(t_n+T))_{n=1,\ldots,MP} = \widehat{W} \cdot \mathbf{a}, \tag{10}$$

where $\widehat{W}$ is a matrix of size $MP \times Nc$. MP is such that $MP - M$ new series' values are predicted.

3. Applications

In order to illustrate the performance of the method we now discuss two specific time series predictions, with reference to possible chaotic systems. In particular, we deal with (i) the well-known Logistic system and (ii) an actual, empirical time series.