Applied Soft Computing 65 (2018) 47–57

Contents lists available at ScienceDirect

Applied Soft Computing

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

Bernstein polynomials for adaptive evolutionary prediction of short-term time series

Kristina Lukoseviciute [a], Rita Baubliene [a], Daniel Howard [b], Minvydas Ragulskis [a,*]

[a] Research Group for Mathematical and Numerical Analysis of Dynamical Systems, Kaunas University of Technology, Studentu 50–147, Kaunas LT–51368, Lithuania
[b] Howard Science Limited, Malvern, UK

ARTICLE INFO

Article history:
Received 5 March 2017
Received in revised form
27 December 2017
Accepted 2 January 2018
Available online 7 January 2018

Keywords:
Bernstein polynomial
Time series prediction
Evolutionary algorithms

ABSTRACT

We introduce a short-term time series prediction model by means of evolutionary algorithms and Bernstein polynomials. This adapts Bernstein-type algebraic skeletons to extrapolate and predict short time series. A mixed smoothing strategy is used to achieve the necessary balance between the roughness of the algebraic prediction and the smoothness of the moving average. Computational experiments with standardized real world time series illustrate the accuracy of this approach to short-term prediction.

1. Introduction

Forecasting is a modelling challenge that relies on a time series analysis. Its aim is to identify a model in time-stamped data presumably generated by some process. Extrapolating by means of this model it makes reliable predictions for unseen data. Recent decades have delivered various models and techniques that are suited to long-term or short-term time series forecasting [1]. Unfortunately, the sheer amount of data needed for training, validating and testing mostly renders long-term time series analysis implausible. Yet, a one-step forward future horizon is adequate for short-term time series forecasting [1] delivering methods which are widely used in high frequency time series analysis with intra-daily data values [2]. Short-term time series predictors are used in finance [3–5]; electricity demand and the associated price forecasting problem [6–8]; wind power; passenger demand [9] and many other industrial applications.

Time series forecasting techniques can be coarsely grouped into classical linear modelling, such as simple exponential smoothing [10], Holt-Winte's methods [11] or Autoregressive Integrated Moving Average (ARIMA) [12], and modern non-linear modelling that is based on soft computing. The latter includes regime-switching models comprising a wide variety of threshold autoregressive models [13–15]: self exciting models [15–17], smooth transition models [18] and continuous-time models [19,20]. Hybrid forecasting methods combine regression, data smoothing, and other techniques to produce forecasts that make up for the comparative deficiencies of individual methods.

A large number of linear and non-linear methods of forecasting appear in the literature, with some methods claiming to do a better job than others under competing assumptions, for example: when given only a short series of input data, or if applied to long-term forecasting [1]. The literature [3–23] covers a wide-variety of techniques that include various flavours of signal processing, support vector machines, ARIMA, Artificial Neural Network (ANN) and Evolutionary Algorithms (EA).

The reader may wonder why there is a continued and strong interest in a plethora of algorithms. The no-free-lunch theorems [24,25] lead to the conclusion that a problem can always be found to defeat any algorithm. Indeed, practical interest in the development of new and hybrid algorithms is warranted because of this reality that no single method will outperform all others in every single situation. At the same time, as Stafford Beer once observed [26], problems of practical interest cannot take an algorithm completely by surprise because the regularities that they comprise are of this world. Real-world problems have neither been designed nor contrived to defeat a popular algorithm. A taxonomy of practical problems, therefore, exists, and it motivates the search for improved algorithms that suit different classes of problems.

In our earlier work [27–29], special EA schemes for the identification of near-optimal algebraic skeleton sequences based on Prony interpolants (represented as linear recurrence sequences (LRS) in

* Corresponding author.
E-mail addresses: kristina.lukoseviciute@ktu.lt (K. Lukoseviciute), rita.palivonaite@ktu.lt (R. Baubliene), dr.daniel.howard@gmail.com (D. Howard), minvydas.ragulskis@ktu.lt (M. Ragulskis).

https://doi.org/10.1016/j.asoc.2018.01.002
1568-4946/

the discrete case) are developed. The high variability of Prony interpolants is managed by means of external [27], internal [28] and mixed smoothing strategies [29]. Such Prony polynomial based techniques are shown to be well applicable for the prediction of stationary short-term time series – but are not well suited for such highly variable time series as Odonovan and Montgome [30].

A natural question arises as to whether some other algebraic interpolants, covering a wider class of functions than the Prony polynomials, might be of benefit to short-term time series forecasting applications. This paper aims to investigate this question with Bernstein polynomials [31] and is structured as follows. Preliminaries on forecasting techniques based on Prony polynomials are given in the second section. Advances over research in [27–29] are presented in the third section. Evolutionary algorithms for the identification of a near-optimal set of corrections are developed in the fourth section, the validation of the model is performed in the fifth section, computational experiments with standard real world time series are presented in the sixth section, and concluding remarks are given in the final section.

2. Preliminaries

The short overview of time series prediction based on Prony polynomials [27–29] in this section, helps to introduce the method that uses Bernstein polynomials.

2.1. The order of a sequence

Let us consider an order n LRS with constant coefficients:

$$x_k = a_{n-1}x_{k-1} + a_{n-2}x_{k-2} + \cdots + a_0 x_{k-n}; \quad k = 0, 1, \ldots; \tag{1}$$

where coefficients $a_j, j = 0, 1, \ldots, n-1$ are constants. The initial conditions $x_k, k = 0, 1, \ldots, n-1$ uniquely determine the evolution of this LRS [32]. The auxiliary polynomial to Eq. (1) reads

$$P(\rho) = \rho^n - a_{n-1}\rho^{n-1} - a_{n-2}\rho^{n-2} - \ldots - a_0, \tag{2}$$

where ρ is the root of the characteristic equation. The LRS takes the form

$$x_j = \mu_1\rho_1^j + \mu_2\rho_2^j + \ldots + \mu_n\rho_n^j \tag{3}$$

if all n roots of Eq. (2) $\rho_1, \rho_2, \ldots, \rho_n$ are distinct and coefficients are determined to fit the initial conditions of the recurrence. If some roots coincide, then the recurrence takes the form:

$$x_j = \sum_{k=1}^{r}\sum_{l=0}^{n_k-1}\mu_{kl}\binom{j}{l}\rho_k^{j-1} \tag{4}$$

where r is the number of distinct roots, n_k is the multiplicity index of the k-th root; $n_1 + n_2 + \cdots + n_k = n$. If the order of LRS is not known in advance, the algorithm for the reconstruction of the model of LRS from a sequence $\left(x_j\right)_{j=0}^{+\infty}$ is more complex. The Hankel transform of $\left(x_j\right)_{j=0}^{+\infty}$ produces the sequence $\left(h_j\right)_{j=0}^{+\infty}$ where $h_j = \det\left(H_j\right)$ and $H_j = (x_{k+l-2})_{1\le k,l\le(j+1)}$ is a Hankel matrix – catalecticant matrix of dimension $(j+1)\times(j+1)$. If there exists an $n \ge 1$ such that $h_n \neq 0$ but $h_k = 0$ for all $k > n$, then $\left(x_j\right)_{j=0}^{+\infty}$ is an LRS and its order is n, and the auxiliary Eq. (2) now reads:

$$\det\begin{bmatrix} x_0 & x_1 & \ldots & x_n \\ x_1 & x_2 & \ldots & x_{n+1} \\ & & \ldots & \\ x_{n-1} & x_n & \ldots & x_{2n-1} \\ 1 & \rho & \ldots & \rho^n \end{bmatrix} = 0. \tag{5}$$

This linear system of algebraic equations has a unique solution because $h_n \neq 0$ [33]. How can one build a model of the process using Eq. (4) if the observed sequence is not an algebraic equation? The idea behind the algebraic prediction technique is based on the identification of the skeleton algebraic sequences and is presented in [27]. Such a concept is based on the assumption that many time series are contaminated with additive noise. This is a strong reason why central to algebraic prediction models is the detection of a base skeleton algebraic sequence in the time series data that removes this additive noise. Further modifications are presented in [28,29].

2.2. Algebraic prediction, external smoothing (APES)

APES is presented in [27]. Let $2n+1$ observations be available for building the model of the process: $(x_k)_{k=0}^{2n}$; where x_{2n} is the value of the observation at the present time. The assumption made that the sequence consists of the addition of noise to some algebraic progression means that the determinant $d_n \neq 0$. The goal now becomes to identify a vector of corrections $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2n})$ such that determinant $\tilde{d}_n$ of differences to these corrections is minimized:

$$\tilde{d}_n = \det\begin{bmatrix} x_0-\varepsilon_0 & x_1-\varepsilon_1 & \cdots & x_n-\varepsilon_n \\ x_1-\varepsilon_1 & x_2-\varepsilon_2 & \cdots & x_{n+1}-\varepsilon_{n+1} \\ & & \cdots & \\ x_n-\varepsilon_n & x_{n+1}-\varepsilon_{n+1} & \cdots & x_{2n}-\varepsilon_{2n} \end{bmatrix} \tag{6}$$

Corrections are identified before any predictions are made with the goal of minimizing any distortions of the original time series and so the fitness function which is the subject of the evolutionary pressure as applied by the EA [27] is given by:

$$F_e(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2n}) = \frac{1}{a|\tilde{d}_n| + \sum_{k=0}^{2n}\lambda_k\left|\varepsilon_k\right|} \tag{7}$$

where the penalty constant, a, is chosen to reflect a desired balance between the magnitude of the determinant and the sum of weighted corrections; λ_k define the tolerance corridor for corrections ε_k. The closer is the element to the last time point, the higher is its weight – and the lower is the variability of its correction.

The evolutionary computation strategy developed in [27] averages 100 reconstructed algebraic skeletons for every single prediction – a single step prediction horizon – to discover a near-optimal vector of corrections.

2.3. Algebraic prediction, internal smoothing (APIS)

An alternative forecasting strategy for short time series in [28] assumes that $2n$ observations are available: $(x_k)_{k=0}^{2n-1}$; and x_{2n-1} is the value of the observation at the current time. Algebraic equation $d_n = 0$ uniquely determines x_{2n}, the element that follows in this sequence. However, such straightforward computations cannot produce satisfactory forecasts. Thus, instead of trying to build such a direct algebraic model into the future, a conciliation between the variability of the skeleton algebraic sequences and the smoothness of the averaged estimates is introduced:

$$F_i(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2n-1}) = \frac{1}{a\sum\lambda_k\left|\varepsilon_k\right| + \left|\tilde{x}_{2n} - \tilde{x}_{2n}\right|}, \tag{8}$$

where $\tilde{x}_{2n}$ is determined from

$$\det\begin{bmatrix} x_0-\varepsilon_0 & x_1-\varepsilon_1 & \cdots & x_n-\varepsilon_n \\ & & \cdots & \\ x_n-\varepsilon_n & x_{n+1}-\varepsilon_{n+1} & \cdots & \tilde{x}_{2n} \end{bmatrix} = 0 \tag{9}$$