Neurocomputing 127 (2014) 161–171

Contents lists available at ScienceDirect

Neurocomputing

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

Short-term time series algebraic forecasting with internal smoothing

Rita Palivonaite [a,1], Minvydas Ragulskis [b,*]

[a] Research Group for Mathematical and Numerical Analysis of Dynamical Systems, Kaunas University of Technology, Studentu 50-325, Kaunas LT-51368, Lithuania

[b] Research Group for Mathematical and Numerical Analysis of Dynamical Systems, Kaunas University of Technology, Studentu 50-222, Kaunas LT-51368, Lithuania

ARTICLE INFO

Article history:
Received 30 April 2012
Received in revised form
15 April 2013
Accepted 17 August 2013
Communicated by Bijaya Ketan Panigrahi
Available online 17 October 2013

Keywords:
Hankel matrix
Time series forecasting
Algebraic sequence
Evolutionary algorithms

ABSTRACT

A new algebraic forecasting method with internal smoothing is proposed for short-term time series prediction. The concept of the *H*-rank of a sequence is exploited for the detection of a base algebraic fragment of the time series. Evolutionary algorithms are exploited for the identification of the set of corrections which are used to perturb the original time series. The proposed forecasting method is constructed to find a near-optimal balance between the variability of algebraic predictors and the smoothness of averaging methods. Numerical experiments with an artificially generated and real-world time series are used to illustrate the potential of the proposed method.

1. Introduction

Time series prediction is a challenging problem in many fields of science and engineering. Conditionally, time series prediction could be classified into long-term time series prediction techniques and short-term time series prediction techniques. In general, the object of long-term time series prediction techniques is to build a model of the process and then use this model to extrapolate past behavior into relatively long future horizons. The main objective of short-term time series prediction techniques remains the same—to build the model of the process and to project this model into the future. Unfortunately, the applicability of techniques for building the model of a long time series for a short time series is often impossible simply due to the fact that the amount of available data is too small. On the other hand, one step-forward future horizon (sometimes accompanied by the estimates of local minimums and local maximums) is a sufficient requirement for a short-term time series predictor.

Short-term time series forecasting procedures include different techniques and models. The use of general exponential smoothing to develop an adaptive short-term forecasting system based on observed values of integrated hourly demand is explored in [1]. A self-organizing fuzzy neural network is used to predict a real-time short-peak and average load in [2]. Day-ahead prices are forecasted with genetic-algorithm-optimized Support Vector Machines in [3]. Artificial neural networks (ANN) are used for day-ahead price forecasting in restructured power systems in [4]. A fuzzy logic based inference system is used for the prediction of short-term power prices in [5]. A hybrid approach based on ANN and genetic algorithms is used for detecting temporal patterns in stock markets in [6]. Short-time traffic flow prediction based on chaotic time series theory in developed in [7]. The radial basis function ANN with a nonlinear time-varying evolution particle swarm optimization (PSO) algorithm are used to forecast one-day ahead and five-days ahead of a practical power system in [8]. PSO algorithms are employed to adjust supervised training of adaptive ANN in short-term hourly load forecasting in [9]. A new class of moving filtering techniques and of adaptive prediction models that are specifically designed to deal with runtime and short-term forecast of time series which originate from monitors of system resources of Internet based servers is developed in [10]. A generalized regression neural network based on principal components analysis is used for electricity price forecasting in [11]. A technique based on Self-Organizing Map neural network and Support Vector Machine models are proposed for predicting day-ahead electricity prices in [12].

An algebraic prediction technique based on the Hankel rank for the identification of the skeleton algebraic sequences in short-term time series is developed in [13]. Such an approach is used to extract information about the algebraic model of the process and then to use this model to extrapolate past behavior into future. It has been demonstrated in [13] that such algebraic predictor can be effectively used for the estimation of local minimums and

* Corresponding author. Tel.: +370 69822456; fax: +370 37330446.
E-mail addresses: rita.palivonaite@ktu.lt (R. Palivonaite), minvydas.ragulskis@ktu.lt (M. Ragulskis).
URL: http://www.personalas.ktu.lt/~mragul (M. Ragulskis).
[1] Tel.: +37067287255.

0925-2312/$ - see front matter
http://dx.doi.org/10.1016/j.neucom.2013.08.025

maximums in day-ahead forecasting applications. It is agreeable that no single method will outperform all others in all situations. It is shown in [13] that the proposed predictor is outperformed (for some real-world time series) by such simple standard techniques as the moving average or the exponential smoothing method—if only the averaged prediction errors are considered.

The main objective of this paper is to enhance the algebraic predictor proposed in [13] by modifying the procedure for the identification of the skeleton algebraic sequences. Our goal is to employ the procedure of internal smoothing which should enable reaching a healthy balance between excellent variability of skeleton algebraic sequences and valuable properties of predictors based on the moving averaging method. The goal is to develop such a predictor which could produce reliable forecasts for very short time series—in situations when the available data is not enough for such predictors as ARIMA or short term time series nonlinear forecasting methods such as neural networks or support vector machines.

This paper is organized as follows. The concept of the Hankel rank is reviewed in Section 2; the forecasting strategy is developed in Section 3; parameter selection for evolutionary algorithms is done in Section 4; computational experiments are discussed in Section 5 and concluding remarks are given in the last section.

2. Preliminaries

Let a sequence of real numbers is given:

$$(x_0, x_1, x_2, \ldots) := (x_k; k \in Z_0). \tag{1}$$

The Hankel matrix (the catelecticant matrix with constant skew diagonals) $H^{(n)}$ constructed from the elements of this sequence is defined as follows:

$$H^{(n)} := \begin{bmatrix} x_0 & x_1 & \cdots & x_{n-1} \\ x_1 & x_2 & \cdots & x_n \\ & & \cdots & \\ x_{n-1} & x_n & \cdots & x_{2n-2} \end{bmatrix} \tag{2}$$

where the index n denotes the order of the square matrix. The determinant of the Hankel matrix is denoted by $d^{(n)} = \det H^{(n)}$; $n \geq 1$. The rank of the sequence $(x_k; k \in Z_0)$ is such natural number $m = Hr(x_k; k \in Z_0)$ that satisfies the following condition [13]:

$$d^{(m+k)} = 0 \tag{3}$$

for all $k \in N$; but $d^{(m)} \neq 0$. Let us assume that the rank of the sequence is $Hr(x_k; k \in Z_0) = m$; $m < +\infty$. Then the following equality holds true [13]:

$$x_n = \sum_{k=1}^{r} \sum_{l=0}^{n_k-1} \mu_{kl} \binom{n}{l} \rho_k^{n-l}; n = j, j+1, j+2, \ldots \tag{4}$$

where characteristic roots $\rho_k \in C$; $k = 1, 2, \ldots, r$ can be determined from the characteristic equation

$$\begin{vmatrix} x_0 & x_1 & \cdots & x_m \\ x_1 & x_2 & \cdots & x_{m+1} \\ & & \cdots & \\ x_{m-1} & x_m & \cdots & x_{2m-1} \\ 1 & \rho & \cdots & \rho^m \end{vmatrix} = 0; \tag{5}$$

The recurrence indexes of these roots n_k $(n_k \in N)$ satisfy the equality $n_1 + n_2 + \cdots + n_r = m$; coefficients $\mu_{kl} \in C$; $k = 1, 2, \ldots, r$; $l = 0, 1, \ldots, n_k - 1$ can be determined from a system of linear algebraic equations which can be formed from the systems of equalities in Eq. (4).

The set of characteristic roots $P(x_0, x_1, \ldots, x_{2m-1}) = \{\rho_k \in C; k = 1, 2, \ldots, r; \quad n_1 + n_2 + \cdots + n_r = m; \quad m \geq 1\}$ is associated to the finite sequence $x_0, x_1, \ldots, x_{2m-1}$ which is denoted as the base fragment of the algebraic progression [13].

3. The forecasting strategy

Let us assume that $2n$ observations are available for building a model of the process and then using this model to extrapolate the past behavior into the future:

$$x_0, x_1, x_2, \ldots, x_{2n-1}; \tag{6}$$

where x_{2n-1} is the value of the observation at the present moment. Let us assume that the sequence $(x_k; k \in Z_0)$ is an algebraic progression and its H-rank is equal to n. Then it is possible to determine the next element of the sequence x_{2n} from the following equality:

$$\det{}^{(n+1)} = \det \begin{bmatrix} x_0 & x_1 & \cdots & x_n \\ x_1 & x_2 & \cdots & x_{n+1} \\ & & \cdots & \\ x_n & x_{n+1} & \cdots & x_{2n} \end{bmatrix} = 0; \tag{7}$$

where the only unknown is x_{2n} (the horizon of the prediction is equal to 1). Unfortunately, real world series are usually contaminated with more or less noise. Thus, such a straightforward assumption that the sequence $(x_k; k \in Z_0)$ is an algebraic progression does not hold in practice (a random sequence does not have a rank).

3.1. The identification of a skeleton algebraic sequence

In the previous paper [13] we have used an evolutionary strategy to identify the algebraic sequence by removing the additive noise. In fact, we were using $2n+1$ observations to build a model of the process. The idea was based on the assumption that the sequence in Eq. (1) is produced by adding noise to some unknown algebraic progression (the H-rank of that algebraic progression is assumed to be equal to n). In other words, we made a proposition that the sequence

$$\tilde{x}_k = x_k - \varepsilon_k; k = 0, 1, 2, \ldots \tag{8}$$

is an algebraic progression and this sequence is some sort of a skeleton sequence determining the global dynamics of the time series. Then, according to Eq. (3),

$$\det(\tilde{H}^{(n+1)}) = 0 \tag{9}$$

where

$$\tilde{H}^{(n+1)} = \begin{bmatrix} \tilde{x}_0 & \tilde{x}_1 & \cdots & \tilde{x}_n \\ \tilde{x}_1 & \tilde{x}_2 & \cdots & \tilde{x}_{n+1} \\ & & \cdots & \\ \tilde{x}_n & \tilde{x}_{n+1} & \cdots & \tilde{x}_{2n} \end{bmatrix}. \tag{10}$$

Corrections ε_k; $k = 0, 1, 2, \ldots, 2n$ had to be identified before any predictions could be made. Since the goal was to minimize any distortions of the original time series, the fitness function for the set of corrections $\{\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2n}\}$ was introduced in [13]:

$$F(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{2n}) = \frac{1}{a\left|\det\left(\tilde{H}^{(n+1)}\right)\right| + \sum_{k=0}^{2n} \lambda_k |\varepsilon_k|}; a > 0; \tag{11}$$

where

$$\lambda_k = \frac{\exp(b(k+1))}{\sum_{j=0}^{2n} \exp(b(j+1))}; k = 0, 1, \ldots, 2n; b > 0. \tag{12}$$