Digital Signal Processing 25 (2014) 114–122

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Partially adapting multi-step local linear prediction ☆

Zhengmin Zhao

Key Laboratory of Modern Acoustics and Institute of Acoustics, Nanjing University, #22 Hankou Rd., Nanjing 210093, China

ARTICLE INFO

Article history:
Available online 25 October 2013

Keywords:
Locally linear prediction
Multi-step
Partially adaptive
Prediction precision

ABSTRACT

An adaptive multi-step local linear prediction method is derived for nonlinear time series by partially updating the predictor with an implicit vector inner product involved in every prediction. The predictor mainly consists of two parts, the basic vector which is calculated at one step based on the nearest neighbors of the current state point, and the updating vector which ensures the validity of the multi-step prediction. Analyses on the precision and computational complexity are made with close comparisons with other prediction methods based on the local linear model. Straightforward performance comparisons made with the classical Lorenz series show that while preserving an admirable precision, the proposed method reduces the computational complexity, thus rendering it more applicable of real time signal processing than the others based on the local linear model. Experiments on voiced sounds are also made and the results demonstrate that, as compared with the traditional linear prediction (LP) method, a remarkable performance gain in precision is achieved with the proposed method, while the computations it costs are similar to that consumed by the LP method.

1. Introduction

Nonlinear time series occur in many fields of science and engineering and a lot of nonlinear prediction methods have been proposed. The methods based on the intrinsic nonlinear models, including the neural network models and support vector machine ones, such as the radial basis function (RBF) neural network [1], finite impulse response (FIR) neural network [2] and support vector regression (SVR) [3,4], have been extensively studied. As expected, these methods can give better performance in terms of prediction precision, but the intensive computation in solving the relevant optimization problems in these methods makes them unrealistic in practical realtime application. Another kind of the nonlinear prediction model often used is that based on the Volterra series [5,6]. A second order Volterra series based method can improve the reconstructed speech quality up to 2 dB overall [6]. Nevertheless, this method consumes more computational resources, even if the fast recursive algorithm is exploited [7,8].

On the other hand, approaches based on the theory of nonlinear dynamics were also attempted for some kinds of nonlinear time series. Among them are those that are based on the locally linear prediction (LLP) model which is first proposed by Farmer and Sidorowich [9] for predicting chaotic time series, and further developed by many others [10,11]. They were regarded as being most promising because of the relative simplicity in implementation as compared to the above-mentioned nonlinear methods [12,13]. But so far their application to realtime processing was still impeded largely by the prohibitive computational complexity of constructing the predictor [14].

To overcome this problem, an iterative LLP (iLLP) was suggested in [10]. In fact, it introduces nothing new but simply predicts a frame of time series iteratively with the same one-step predictor determined by the LLP analysis at the first step of the frame, so as to save the computational cost for multi-step prediction. In this way, the prediction error will grow rapidly as the prediction step advances [1,10], especially when its Lyapunov exponent is large. Thus the maximum predictable term is limited.

As a compromise between the LLP and iLLP, another modified LLP was proposed in [14] to reduce the computation by periodically updating the predictor every more than one step, but its performance is still limited in balancing the computational complexity and the accuracy of prediction.

In this paper, the mathematics involved in the kernel algorithm for multi-step prediction based on the local linear model is optimized, and a method is derived by partially updating the predictor with an implicit vector inner product involved in every prediction. The predictor consists of the basic vector and the updating vector which ensures the validity of the multi-step prediction. It is estimated at one step based on a group of nearest neighbors of the current state point. And then in the following steps, the predictor is partially updated while the prediction is made by an inner product. Thus it adapts to step advances, and this prediction method is referred to as the *partially adapting multi-step Local Linear*

☆ This work is supported by the National Science Foundation of China under Grant Nos. 10574070 and 10874085. It is also supported by the Natural Science Fund for Colleges and Universities in Jiangsu Province under Grant No. 08KJD510002.

E-mail address: if@sohu.com.

1051-2004/$ – see front matter
http://dx.doi.org/10.1016/j.dsp.2013.10.009

Prediction (paLLP). Our focus is improving the computational efficiency of nonlinear prediction while keeping an acceptable accuracy which is similar to that achieved by the one-step LLP method.

The paper is organized as follows. Section 2 formulates the theoretical aspects of the proposed method (Section 2.1), and presents a practical scheme for the paLLP in the form of pseudo code (Section 2.2). Analyses as well as some remarks are given on the prediction precision and computational complexity by comparisons with the relevant nonlinear prediction methods based on the local linear model (Sections 3.2, 3.3 and 3.4). The performance of the paLLP is evaluated by applying it to the typical nonlinear time series generated by the classic Lorenz system, along with close comparisons with other LLP-based methods, showing the superiority of the proposed method to the other two LLP-based predictions in terms of precision and to the LLP in reduction of computation cost (Section 4.1). After the applicability of the paLLP to speech processing, especially to a kind of the Analysis-by-Synthesis (A-B-S) codec, is briefly analyzed (Section 4.2.1), human voiced sounds are taken as examples to validate the paLLP for practical signals (Sections 4.2.2 and 4.2.3). Except the comparisons with the other LLP-based methods, comparisons with the traditional LP are also made and the results show that maybe the paLLP is a potential candidate for the speech signal processing. The work is summarized and further discussion is given in the concluding section (Section 5).

2. Partially adapting multi-step local linear prediction

2.1. Formulation of algorithm

Consider an observed scalar time series x_i, $i = 0, 1, 2, \ldots$, sampled from a continuous signal at an interval τ_s. By applying Takens' theorem [15,16], we can reconstruct the dynamics of the time series in d-dimensional embedding space $\mathbf{S}$ with a series of vectors

$$\mathbf{x}_i = (x_i, x_{i-m}, \ldots, x_{i-(d-1)m})^T, \quad i = (d-1)m, (d-1)m+1, \ldots, \tag{1}$$

where d is the embedding dimension, m is the delay time in unit of τ_s, and the superscript "T" denotes transpose of a vector or a matrix.

In the d-dimensional reconstructed phase space, the dynamics of the system can be described by an iterative function

$$\mathbf{x}_{n+1} = \mathbf{f}(\mathbf{x}_n) \tag{2}$$

with $\mathbf{f} : \mathbb{R}^d \mapsto \mathbb{R}^d$ being a nonlinear vector function. Applying the iteration k times gives

$$\mathbf{x}_{n+k} = \underbrace{\mathbf{f}(\mathbf{f}(\ldots(\mathbf{f}}_{k}(\mathbf{x}_n))\ldots)) \equiv \mathbf{f}^k(\mathbf{x}_n), \tag{3}$$

in particular

$$x_{n+k} = f^k(\mathbf{x}_n) \tag{4}$$

with $f^k : \mathbb{R}^d \mapsto \mathbb{R}$ also being a nonlinear mapping function.

Thus, the future $\mathbf{x}_{n+k}$ could be predicted, provided that $\mathbf{f}$ be known. Unfortunately, the exact form of $\mathbf{f}$ is unknown in general for complex systems, such as human voicing one. However, as long as $\mathbf{x}_{n+k}$ is sufficiently close to the vector $\mathbf{x}_n$ in the reconstructed phase space, $\mathbf{f}^k(\mathbf{x}_n)$ can be locally linearized via its expansion in the Taylor series about $\mathbf{x}_n$ around a point $\mathbf{p}_{n,k}$ near to $\mathbf{x}_n$ [17], and its approximation can be written as

$$\begin{aligned}\mathbf{x}_{n+k} &= \mathbf{f}^k(\mathbf{p}_{n,k}) + \left.\frac{\partial \mathbf{f}^k}{\partial \mathbf{x}^T}\right|_{\mathbf{x}=\mathbf{p}_{n,k}} (\mathbf{x}_n - \mathbf{p}_{n,k}) + \boldsymbol{\epsilon}_{n,k} \\ &= \mathbf{q}_{n,k} + \mathbf{Q}_{n,k}\mathbf{x}_n + \boldsymbol{\epsilon}_{n,k},\end{aligned} \tag{5}$$

where $\mathbf{q}_{n,k} = \mathbf{f}^k(\mathbf{p}_{n,k}) - \left.\frac{\partial \mathbf{f}^k}{\partial \mathbf{x}^T}\right|_{\mathbf{x}=\mathbf{p}_{n,k}} \mathbf{p}_{n,k} \in \mathbb{R}^d$, $\mathbf{Q}_{n,k} = \left.\frac{\partial \mathbf{f}^k}{\partial \mathbf{x}^T}\right|_{\mathbf{x}=\mathbf{p}_{n,k}} \in \mathbb{R}^{d\times d}$ and $\boldsymbol{\epsilon}_{n,k} \in \mathbb{R}^d$ stands for the higher-order-term (h.o.t.) of $(\mathbf{x}_n - \mathbf{p}_{n,k})$. Let $\mathbf{C}_{n,k} = (\mathbf{Q}_{n,k}, \mathbf{q}_{n,k}) \in \mathbb{R}^{d\times(d+1)}$, $\mathbf{z}_n = (\mathbf{x}_n^T, 1)^T \in \mathbb{R}^{d+1}$, then we have

$$\mathbf{x}_{n+k} = \mathbf{C}_{n,k}\mathbf{z}_n + \boldsymbol{\epsilon}_{n,k}, \tag{6}$$

in particular,

$$x_{n+k} = \mathbf{z}_n^T \mathbf{c}_{n,k} + \epsilon_{n,k}, \tag{7}$$

where $\mathbf{c}_{n,k} = (c_{n,k,0}, c_{n,k,1}, \ldots, c_{n,k,d})^T \in \mathbb{R}^{d+1}$ is the first column of $\mathbf{C}_{n,k}^T$, $\epsilon_{n,k} \in \mathbb{R}$ stands for one-dimensional h.o.t. of $(\mathbf{x}_n - \mathbf{p}_{n,k})$. Thus the x_{n+k} can be predicted with the vector $\mathbf{z}_n$ and the estimated coefficient vector $\hat{\mathbf{c}}_{n,k}$ as

$$\hat{x}_{n+k} = \mathbf{z}_n^T \hat{\mathbf{c}}_{n,k}, \tag{8}$$

which gives the local linear prediction of x_{n+k} from the so-called *current state vector* $\mathbf{x}_n$ based on the local linear model (7). The prediction error $e_{n,k}$ is

$$e_{n,k} = x_{n+k} - \hat{x}_{n+k} = x_{n+k} - \mathbf{z}_n^T \hat{\mathbf{c}}_{n,k}, \tag{9}$$

which includes the $\epsilon_{n,k}$, the error caused by the estimation error of $\hat{\mathbf{c}}_{n,k}$ and noise in x_{n+k}.

The problem, predicting x_{n+k}, thus is reduced to estimate the coefficient vector $\mathbf{c}_{n,k}$. To do so, a group of K $(K > d+1)$ nearest neighbors of $\mathbf{x}_n$ are searched from the past state vectors $\mathbf{x}_i$, $(d-1)m \leqslant i < n$, and K pairs $(\mathbf{x}_{n_j}, x_{n_j+k})$ $(j = 0, \ldots, K-1)$ are formed, where n_j indicates the jth nearest neighbor of the $\mathbf{x}_n$. After $k-1$ steps, the x_{n_j+1}, which corresponds to $\mathbf{x}_{n_j}$, evolves into x_{n_j+k}. With the K nearest neighbors $\mathbf{x}_{n_j}$ and the corresponding x_{n_j+k}, we form matrix

$$\mathbf{Z}_n = (\mathbf{z}_{n_0}, \mathbf{z}_{n_1}, \ldots, \mathbf{z}_{n_{K-1}})^T \in \mathbb{R}^{K\times(d+1)}, \tag{10}$$

and vector

$$\mathbf{y}_{n,k} = (x_{n_0+k}, x_{n_1+k}, \ldots, x_{n_{K-1}+k})^T \in \mathbb{R}^K. \tag{11}$$

The best estimation of $\mathbf{c}_{n,k}$ in terms of minimizing square prediction errors therefore becomes an optimization problem

$$\min_{\mathbf{c}_{n,k}} \left\| \mathbf{W}_n(\mathbf{y}_{n,k} - \mathbf{Z}_n\mathbf{c}_{n,k}) \right\|_2, \tag{12}$$

where $\mathbf{W}_n = \mathrm{diag}(w_0, w_1, \ldots, w_{K-1}) \in \mathbb{R}^{K\times K}$ is an appropriate weighting matrix, which is used to indicate the confidence of every point pair for the predictor estimation. Generally, the more $\mathbf{x}_{n_j}$ close to the $\mathbf{x}_n$, the higher the confidence of the point pair $(\mathbf{x}_{n_j}, x_{n_j+k})$ is, and the most similar $\mathbf{x}_{n_j}$ get the most weight. Thus the weighting factor w_j is set to be a function of $\|\mathbf{x}_n - \mathbf{x}_{n_j}\|_2$, $\|\cdot\|_2$ is L_2 norm in $\mathbf{S}$. For example [14],

$$w_j = \|\mathbf{x}_n - \mathbf{x}_{n_j}\|_2^{-1}. \tag{13}$$

The Least-Mean-Square (LMS) solution to (12) turns out to be

$$\hat{\mathbf{c}}_{n,k} = \mathbf{Z}_n^{\dagger}\mathbf{W}_n\mathbf{y}_{n,k}, \tag{14}$$

where

$$\mathbf{Z}_n^{\dagger} = \left(\mathbf{Z}_n^T\mathbf{W}_n^T\mathbf{W}_n\mathbf{Z}_n\right)^{-1}\mathbf{Z}_n^T\mathbf{W}_n^T \tag{15}$$

denotes the Moore–Penrose generalized inverse of $\mathbf{W}_n\mathbf{Z}_n$.

Substituting (14) into (8), we finally obtain an explicit formula for optimized local linear prediction

$$\hat{x}_{n+k} = \mathbf{z}_n^T\hat{\mathbf{c}}_{n,k} = \mathbf{z}_n^T\mathbf{Z}_n^{\dagger}\mathbf{W}_n\mathbf{y}_{n,k} = \mathbf{h}_n\mathbf{y}_{n,k} \tag{16}$$