Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 1539–1547
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

# Local prediction of non-linear time series using support vector regression

K.W. Lau, Q.H. Wu*

*Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool L69 3GJ, UK*

Received 8 September 2005; received in revised form 24 August 2007; accepted 29 August 2007

**Abstract**

Prediction on complex time series has received much attention during the last decade. This paper reviews least square and radial basis function based predictors and proposes a support vector regression (SVR) based local predictor to improve phase space prediction of chaotic time series by combining the strength of SVR and the reconstruction properties of chaotic dynamics. The proposed method is applied to Hénon map and Lorenz flow with and without additive noise, and also to Sunspots time series. The method provides a relatively better long term prediction performance in comparison with the others.


*Keywords:* Time series analysis; Local prediction; Support vector regression; Radial basis function; Least square; Delay coordinates; State space reconstruction

## 1. Introduction

The study of prediction has been influenced, for a long time, by statistical methods such as the ARMA model [1] for linear stationary time series. Recent developments in non-linear and/or non-stationary time series analysis can be found in Refs. [2,3]. Another approach to predicting time series using neural networks has been investigated [4]. All of these methods are known as global time series prediction in which only one function is engaged for all available data. Local prediction uses more than one function to fit the data. This approach is also known as lazy learning [5].

Support vector regression (SVR) [6,7] and radial basis function (RBF) networks [8,9] are two different approaches to non-linear regression problems, but both methods are known as single layer networks. The relationship between the SVR and RBF networks can be found in Ref. [7]. SVR has been applied to such as drug discovery [10], Travel time prediction [11] and computational vision [12]. Recently, SVR has been used for predicting chaotic time series [13].

The RBF has not only the ability of approximating scattered data without using any mesh. Therefore it is a good solution for the multivariate interpolation problems. It is also a method of turning an ill-posed problem into a well posed problem by regularization [14–16]. RBF has been applied to such as 3D object recognition [17] and facial expressions [18].

A chaotic attractor is obtained by measuring a chaotic time series. The properties of the chaotic attractor can be retained through a reconstruction procedure. This procedure is known as the delay coordinate embedding [19] resulting in a reconstructed state space which contains a reconstructed chaotic attractor preserving both geometrical and dynamical properties of the original chaotic attractor.

In this paper we propose the SVR as a local predictor. The local predictor chooses a set of nearest neighbours which evolves similarly in the reconstructed chaotic attractor. This approach to predicting the chaotic time series has interested many researchers such as Farmer and Sidorowich [20], Casdagli [21] and Sauer [22]. Our approach is different from the work done by Mukherjee et al. [13] and Casdagli [21]. We combine the strength of SVR and local predictor to achieve a better prediction result. The resulting predictor is referred as SVR based local predictor (SVRLP).

The proposed algorithm is applied to two benchmark problems of chaotic time series, known as the noisy Hénon time series [23] and the noisy Lorenz time series [24], respectively. The benchmark problems are mainly concerned with chaotic dynamics which is difficult to predict. Then the proposed algorithm is further applied to Sunspots series. Through the

* Corresponding author. Tel.: +44 151 7944535; fax: +44 151 7944540.
*E-mail address:* q.h.wu@liv.ac.uk (Q.H. Wu).

0031-3203/$30.00 
doi:10.1016/j.patcog.2007.08.013

simulation study presented in this paper, based on these three benchmark problems, it is demonstrated that the prediction performance of SVRLP is better than RBF based local predictor (RBFLP) in most situations.

In this paper Section 2 describes the delay coordinate embedding methodology which shows an attractor can be reconstructed from an univariate time series. Section 3 reviews the existing method of local prediction. The SVR is introduced in Section 4. The comparison of the prediction performances of SVRLP, RBFLP and least square local predictor (LSLP) is given in Section 5. The conclusion is given in Section 6.

## 2. Delay coordinate embedding

Takens embedding theorem [19] provides theoretical foundation for the analysis of time series generated by non-linear deterministic dynamical systems. Later Sauer [25] shows a phase space can be reconstructed from an univariate chaotic time series. Let an univariate time series $\{x_i\}_{i=1}^{l}$, where $l$ is the length of the time series, is generated from in a $D$-dimension chaotic attractor, a phase space $\mathbf{R}^D$ of the attractor can be reconstructed by using delay coordinate defined as

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

where $d$ is known as the embedding dimension of reconstructed phase space, $m$ is the delay constant and $\top$ denotes the vector transpose.

The delay coordinate method is currently the most widely used choice for chaotic time series analysis. It has one good property that the signal to noise ratio on each component is equal. For a noise free infinite long time series, the embedding dimension is the smallest integer greater than $2D$, i.e. $d > 2D$, and the delay constant $m$ is almost arbitrary. In practice, both the dimension of the attractor and the delay constant are unknown and must be estimated from the time series.

### 2.1. Estimate the embedding dimension

The correlation dimension method [26] is introduced to estimate the dimension of a dynamical system from the given time series. It can be determined from the correlation integral which is defined as follows:

$$C(r) = \lim_{\hat{l}\to\infty} \frac{1}{\hat{l}^2} \sum_{i,j=1}^{\hat{l}} \theta(r - \|\mathbf{x}_i - \mathbf{x}_j\|), \tag{2}$$

where $\hat{l} = l - (d-1)m$ and $\theta(x)$ is the Heaviside step function:

$$\theta(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x > 0, \end{cases} \tag{3}$$

and the correlation dimension $d_2$ is given by

$$d_2 = \lim_{r\to 0} \frac{\log(C(r))}{\log r}. \tag{4}$$

Obviously, the correlation dimension depends on the size of $r$. When $r$ is small, the behaviour of the correlation dimension is dominated by the characteristics of the noise, which has infinite dimension [27]. Let $d_2(r_L, r_U)$ denote a range of correlation dimensions corresponding to a range of radii $\{r_L, r_U\}$. In practice, the correlation dimension is chosen such that $d_2(r_L, r_U)$ is a constant over a number of embedding dimensions exceeding $[2d_2(r_L, r_U) + 1]$ [28].

### 2.2. Estimate the delay constant

In order to use the delay coordinate embedding, it is necessary to estimate the delay constant $m$. If $m$ is too small, each coordinate is almost the same and the trajectories of the reconstructed space are squeezed along an identity line; this phenomenon is known as redundancy. If $m$ is too large, in the presence of chaos and noise, the dynamics at any one time become effectively and causally disconnected from the dynamics at a later time, so that even simple geometric objects look extremely complicated; this phenomenon is known as irrelevance.

A reasonable choice of $m$ is the first minimum of the autocorrelation function such that each coordinate is linearly independent. However, it is not sufficient for the non-linearity of the chaotic time series. Mutual information measures both linear and non-linear dependence of each coordinate. Therefore the first local minimum of mutual information provides a better criterion for estimating $m$. The mutual information from a time series has been proposed by Fraser et al. [29]. It is defined as

$$I_d(m) = dH_0 - H_d(m), \tag{5}$$

where

$$H_d(m) = \frac{1}{\hat{l}} \sum_{i=1}^{\hat{l}} \ln P_r(\mathbf{x}_i(m)). \tag{6}$$

However, Fraser's algorithm is difficult to implement and has only been applied to a two-dimensional case. Another algorithm has been proposed by Liebert et al. [30]. The first local minimum of mutual information $I_d(m)$ is equal to the first local minimum of $\log C_1^d$, where $C_1^d$ is defined as the $\lim_{q\to 1} C_q^d$ and $C_q^d$ is defined as

$$C_q^d = \left( \frac{1}{\hat{l}} \sum_{i}^{\hat{l}} P_r^{q-1}(\mathbf{x}_i(m)) \right)^{1/(q-1)}, \tag{7}$$

where $P_r(\mathbf{x}_i(m))$ is defined as

$$P_r(\mathbf{x}_i(m)) = \frac{1}{\hat{l}} \sum_{j=1}^{\hat{l}} \theta(r - \|\mathbf{x}_i(m) - \mathbf{x}_j(m)\|), \tag{8}$$

where $r$ is a constant.

## 3. Existing local prediction methods

The local prediction method relies on a set of nearest neighbours which evolves similarly in the reconstructed chaotic