MA Hong-guang, HAN Chong-zhao

Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction

Abstract A new algorithm is proposed for computing the embedding dimension and delay time in phase space reconstruction. It makes use of the zero of the nonbias multiple autocorrelation function of the chaotic time series to determine the time delay, which efficiently depresses the computing error caused by tracing arbitrarily the slope variation of average displacement (AD) in AD algorithm. Thereafter, by means of the iterative algorithm of multiple autocorrelation and $\Gamma$ test, the near-optimum parameters of embedding dimension and delay time are estimated. This algorithm is provided with a sound theoretic basis, and its computing complexity is relatively lower and not strongly dependent on the data length. The simulated experimental results indicate that the relative error of the correlation dimension of standard chaotic time series is decreased from 4.4% when using conventional algorithm to 1.06% when using this algorithm. The accuracy of invariants in phase space reconstruction is greatly improved.

Keywords phase space reconstruction, embedding dimension, delay time, multiple autocorrelation, $\Gamma$ test

1 Introduction

The characteristics of the strange attractors of a chaotic system can be analyzed by sampling a part of the output chaotic time series of a system. The method that is commonly used is the state space reconstruction in delay coordinate proposed by Packard et al. [1]. It can be proved through Takens’ theorem [2] that the unstable periodic obits (strange attractor) could be recovered properly in an embedding space whenever a suitable embedding dimension $m \geq 2d+1$ ($d$ is the dimension of chaotic system) is detected; that is, the obits in the reconstructed space $R^m$ keeps a differential homeomorphism with the original system.

It is very important to select a suitable pair of embedding dimension $m$ and time delay $\tau$ when performing the phase space reconstruction. The precision of $\tau$ and $m$ is directly related with the accuracy of the invariants of the described characteristics of the strange attractors in phase space reconstruction. For doing this, there are two different points of view. One is that $m$ and $\tau$ are not correlated with each other; that is, $m$ and $\tau$ can be selected independently (Takens has proved that $m$ and $\tau$ are independent in a chaotic time series with infinite length and no noise). Under this golden rule, a commonly used approach, G–P algorithm, for calculating the embedding dimension $m$ was proposed by Albano et al. [3]. For the time delay $\tau$, there are three criterions for its selection: (1) series correlation approaches, such as autocorrelation, mutual information [4], high-order correlations [5], etc., (2) approaches of phase space extension, e.g., fill factor [6], wavering product [7], average displacement (AD; [8]), SVF [9], etc., and (3) multiple autocorrelation and nonbias multiple autocorrelation [10].

The second viewpoint is that $m$ and $\tau$ are closely related because the time series in the real world could not be infinitely long and could hardly avoid being noised. A great deal of experiments indicate that $m$ and $\tau$ tie tightly up with the time window $t_w=(m-1)\tau$ for the reconstruction of the phase space. For a given chaotic time series, $t_w$ is relatively steadfast. An irrelevant partnership between $m$ and $\tau$ will directly impact the equivalence between the original system and the reconstructed phase space. Therefore, the combination approaches for computing $m$ and $\tau$ accordingly come into being, e.g., small-window solution [11], C–C method [12], and automated embedding [13]. Most researchers consider that the second viewpoint is more practical and reasonable than the first one in the engineering practice. Research on the combination algorithm of embedding dimension and delay time will become a hotspot in the category of the chaotic time series analysis.

2 Automated embedding algorithm

This algorithm was proposed by Masayuki Otani and Antonia Jones in October 2000, which is based on the AD method and $\Gamma$ test [14]. By means of this algorithm, a near-optimum embedding dimension and delay time can be
estimated. A brief description about this algorithm is given as follows.

1. Let $X=\{x_i(t)\}, i=1, 2..., N$, be a part of chaotic time series whose evolution through time is described by a $d$-dimension dynamical system. Set an initial value for the embedding dimension; that is, let $m=m_0$. Take the time delay $\tau$ as a variable and let it increase by one for each iteration. At each determined value of $\tau$, reconstruct $X$ into $M=N-(m-1)\tau$ dimensions of vectors $\{x_i\}, i=1, 2..., M, x_i=(x_{i+\tau}, x_{i+2\tau}, ..., x_{i+(m-1)\tau}) x_i \in \mathbb{R}^m$. Then, calculate the AD of the entire vector space using:

$$S(\tau) = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{m-1} [x_{i+j\tau} - x_i]^2,$$

(1)

where $M$ is the number of data points used for the estimation. As the delay time increases from zero, the reconstructed trajectory expands from the diagonal and $S(\tau)$ increases accordingly until it reaches a plateau. With large values of $m$, reconstruction expansion reaches a plateau at a smaller value of the delay time, which maintains the time span as approximately constant. The corresponding value of delay time when $S(\tau)$ gets into saturation is the near-optimum $\tau$ under a certain value of $m$.

2. Take the result of step 1 as a constant and let embedding dimension $m$ become a variable. Estimate the near-optimum $m$ by means of the $\Gamma$ test, which can estimate the best mean-squared output error of a continuous or smooth underlying input/output model without overfitting. That is, suppose that the samples of chaotic time series are generated by a continuous function $f: \mathbb{R}^m \rightarrow \mathbb{R}$, and let $y$ be defined as $y=f(x_1, ..., x_m)+\gamma$, where $\gamma$ represents an indeterminable part, which may be due to noise or lack of functional determination in the input/output relationship. At each given value of $m$, reconstruct $X$ into $M=N-(m-1)\tau$ dimensions of vectors $\{x_i\}$ and construct the input/output pairs $\{\xi_i, y_i\}$ as follows:

$$\xi_i = \{x(i), x((i+1)\tau), \cdots, x((i+m-1)\tau)\}$$

$$y_i = x((i+m)\tau), \quad i = 1, 2, \cdots M (2)$$

Then, find out the $p$th nearest neighbor $\xi_d(N(i, p))$ to $\xi(N(i, p))$ and compute the distances using:

$$dx(h) = \frac{1}{p} \sum_{h=1}^{p} \frac{1}{M} \sum_{i=1}^{M} \|\xi(N(i, p)) - \xi_i\|_2^2$$

$$dy(h) = \frac{1}{p} \sum_{h=1}^{p} \frac{1}{M} \sum_{i=1}^{M} (y(N(i, p)) - y(i))^2$$

(3)

Perform a least-squares fit on the coordinates $(dx, dy)$ to obtain a regression line in the form of $(dy = Adx + \Gamma)$, where $\Gamma$ is the estimated value of $\gamma$.

Increase gradually the value of $m$ by one and repeat steps 1 and 2. The estimated value of $\gamma$ will decrease accordingly.