Selection of Embedding Dimension and Delay Time in Phase Space Reconstruction

Ma Hongguang, Han Chongzhao,Li Xihai; Wang Guohua, Xu Jianfeng, Zhu Xiaofei
Xi'an Jiaotong University, School of Electronic & Information Engineering, Xi'an 710049, P.R. China; Research Institute of High Technology, Xi'an 710025, P.R. China
FAX. 0086- 29-82668775, E-mail: mhg_xian@163.com, http://www.xjtu.edu.cn

Abstract
Embedding techniques provide a powerful advance in the development of experiment chaos. However there seems no universal method to find the best set of parameters to use. In this paper, we analyze the drawback of an algorithm of automatic embedding dimension and time delay presented in Reference[1](Massayuki Otani and Antonia Jones, Oct. 2000), and propose a new approach for computing the embedding dimension and delay time based on the multiple autocorrelation and Γ-test. This approach is provided with a sound theoretic basis, and its computing complexity is relatively lower and not strongly depended on the data length. The experimental results indicate that a near optimum embedding dimension and delay time can be estimated by using this approach, and the accuracy of invariants in phase space reconstruction is efficiently improved.
Keywords: Phase Space Reconstruction, Embedding Dimension, Delay Time, Multiple Autocorrelation, Γ-test.

1 Introduction

The characteristics of strange attractors of a chaotic system can be analyzed by sampling a part of the output chaotic time series of system. The method in common use is the state space reconstruction in delay coordinate proposed by Packad[2]. It can be proved by Takens' theorem[3] that the unstable periodic obits (strange attractor) could be recovered properly in an embedding space whenever a suitable embedding dimension $m \geqslant 2d+1$, (d is the dimension of chaotic system) were found out, i.e. 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 τ when performing the phase space reconstruction. For doing this there are two different points of view: one is that m and τ are not correlated with each other, i.e. m and τ can be selected independently (Takens has proved that m and τ are independent in a chaotic time series with infinite length and noiseless). Under this golden rule, a commonly used approach, G-P algorithm for calculating the embedding dimension m was proposed by Grassberger and Procacia[4]. For the time delay τ, there are three criterions to select it: 1. Series correlation approaches, such as Autocorrelation [4], Mutual Information [5], and High-order Correlations [6], etc. 2. Approaches of phase space extension, e.g. Fill Factor [7], Wavering Product [8], Average Displacement [9] and SVF [10], etc. 3. Multiple Autocorrelation and Non-bias Multiple Autocorrelation [11].

The second viewpoint is that m and τ are closely related, because the time series in the real world could not be the infinite long, and hardly avoid being noised. A great deal of experiments indicate that m and τ tie tightly up with the time window $t_w = (m-1)\tau$ for the reconstruction of phase space. For a given chaotic time series, t_w is relatively steadfast. An irrelevant partnership of m and τ will directly impact the equivalence between the original system and the reconstructed phase space. Therefore, the combination approaches for computing m and τ are accordingly come into being, e.g. small-window solution [12], C-C method [13] and automated embedding [1]. We consider that the second viewpoint is more practical and reasonable than the first one in the engineering practice. The 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 Oct. 2000, which is based on the Average Displacement Method (AD) and Γ-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 $\boldsymbol{X}=\{x_i(t)\}$, 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, i.e. let $m = m_0$. Take the time delay τ as a variable and let it increase by one for each iteration. At each determinate value of τ, reconstruct X into $M=N-(m-1)\tau$ dimensions of vectors $\{\boldsymbol{x}_i\}$, i=1,2, ...M, $\boldsymbol{x}_i = (x_i, x_{i+1}, \ldots, x_{i+(m-1)\tau})$ $\boldsymbol{x}_i \in R^m$. Then calculate the average displacement of entire vector space by using formula (1).

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

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 smaller value of the delay time, which maintains the time span approximately constant. The corresponding value of delay time when $S(\tau)$ gets in saturation is the near optimum τ under the certain value of m.

2. Take the result of step 1 as a constant and let embedding dimension m is a variable. Estimate the near optimum m by means of Γ-test, which can estimate the best mean squared output error of a continuous or smooth underlying input/output model without overfitting, i.e. suppose the samples of chaotic time series are generated by a continuous function f: $R^m \rightarrow R$, and let y be defined as $y = f(x_1, \cdots x_m) + \gamma$. Where γ 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:

$$\xi_i = \{x(i), x((i+1)\tau), \cdots, x((i+m-1)\tau\} \qquad (2)$$

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

Then find out the pth nearest neighbour $\xi_i(N(i,p))$ to ξ_i (p_{max}=20～50) and compute the distances by means of the formula 3.

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

$$dy(h) = \frac{1}{p}\sum_{h=1}^{P}\frac{1}{2M}\sum_{i=1}^{M}(y(N(i,p)) - y(i))^2 \qquad (3)$$

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

Increase the value m by one gradually and repeat steps 1 and 2. The estimated value of γ will decrease accordingly until it is much closed to zero. At this moment, the values of m and τ are the near optimum embedding dimension and time delay for the given chaotic time series. By chance if the estimated value of γ is not close to zero, the data set is non-deterministic; therefore we cannot hope to reconstruct the attractor accurately. This may happen if the SNR is lower, or the choice of time delay is poor.

The experimental results indicate that this algorithm is very efficient for the continuous chaotic time series. But the computing accuracy of this algorithm is tightly depended on that of AD algorithm. The average displacements of Lorenz and Rossler flows are depicted in fig.1 and fig.2. It can be seen clearly that the time delay is decreasing with the increasing of embedding dimension, and also there are some waviness when the waveshapes get into saturation.

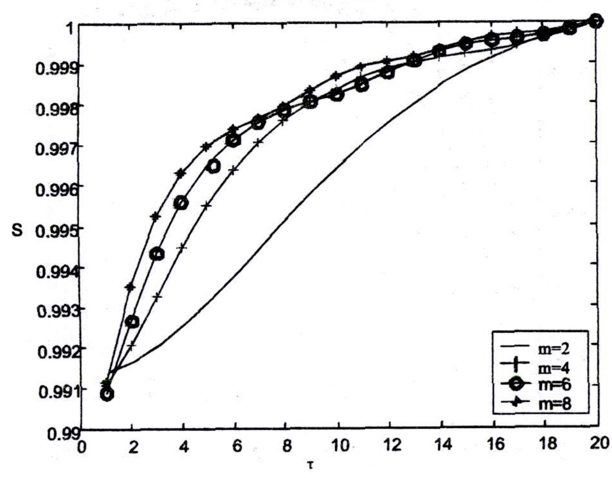

Figure 1: Average Displacement of Lorenz Flow.

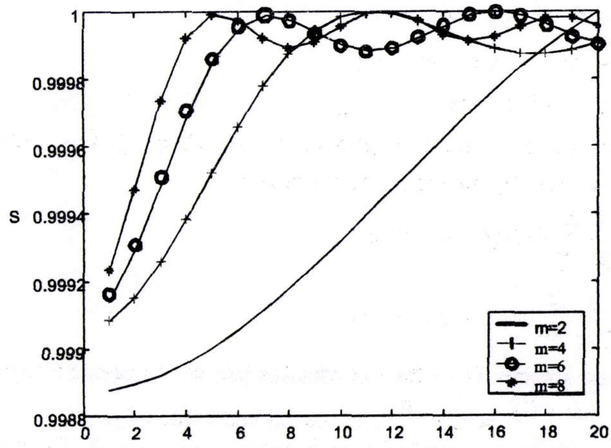

Figure 2: Average Displacement of Rossler Flow.

However, this algorithm cannot directly process the discrete chaotic time series, such as Henon, Logistic and Quadratic, etc. The major causation is that the sampling spacing of the discrete chaotic time series is "too large" that make the relativity between the data change so swiftly, and it seems that those maps behave like the random series. Hence, the discrete chaotic time series must be interpolated before processing. Fig.3 and 4 depict the average displacements of Henon and Quadratic maps after the interpolation with spline function. The data are the 10 times more than that of the originals.

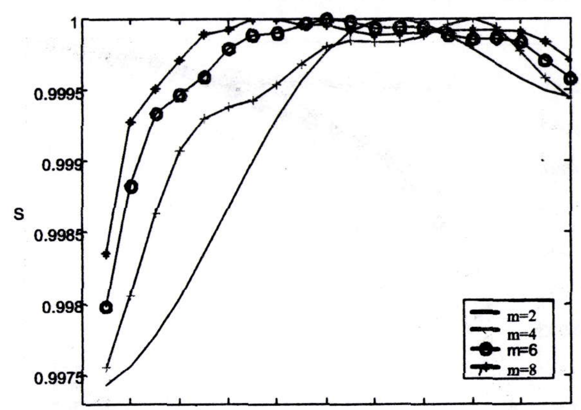

Figure 3: Average Displacement of Henon Map.

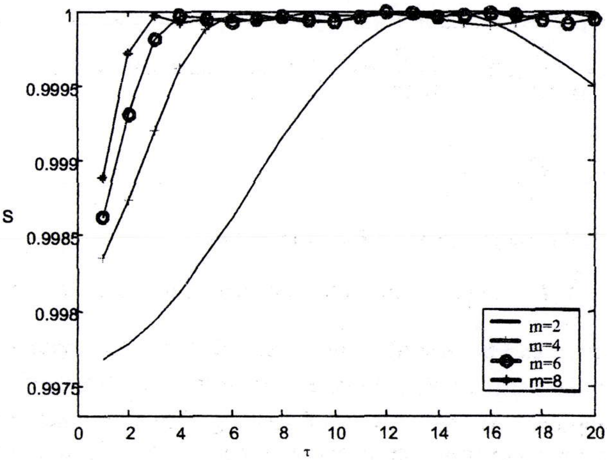

Figure 4: Average Displacement of Quadratic Map.

The average displacement algorithm is a geometry-based approach that can overcome the drawbacks of the autocorrelation-based methods, since the autocorrelation can ensure x_i and $x_{i+\tau}$, $x_{i+\tau}$ and $x_{i+2\tau}$ are not correlated respectively, but it cannot guarantee that x_i and $x_{i+2\tau}$are not correlated, too. Therefore, the autocorrelation-based method cannot be generalized in the high-order dimensions. So the AD algorithm looks like a suitable approach for the high-order system. In practice, the sloping variation of statistic S(τ) should be measured to figure out the corresponding delay time, usually we take the time point at the slope decreases to 40% of its initial value as the near optimum time delay. But from Figures 1 and 2 we can see that there intermix some wobbles in the entire variation of S(τ). Thereby, using the changing slope to determine the time delay sometimes will introduce a non-ignored error, and this error will influence the computing accuracy of embedding dimension in Γ-test. Hence, a modification should be done for the algorithm of time delay.

3 Multiple Autocorrelation Approach[11]

The multiple autocorrelation approach is derived from autocorrelation and average displacement. From formula 1 we can rewrite the statistic S(τ) of the chaotic time series $\{x_i\}$ in m dimension as follows:

$$S_m^2(\tau) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{m-1}(x(i+j\tau)-x(i))^2 \tag{4}$$

Extend the left part of formula 4 and ignore the errors caused by the border data. Consider that $E = \frac{1}{M}\sum_{i=1}^{M} x(i)^2 = \frac{1}{M}\sum_{i=1}^{M} x(i+j\tau)^2$ is a constant within 1≤j≤m-1, we can get:

$$S_m^2(\tau) = 2(m-1)E - 2\sum_{j=1}^{m-1} R_{xx}(j\tau) \tag{5}$$

Where, $R_{xx}(j\tau)$ is the autocorrelation function of $\{x_i\}$.

Define $R_{xx}^m(\tau) = \sum_{j=1}^{m-1} R_{xx}(j\tau)$, the multiple autocorrelation approach for the series $\{x_i\}$ in m dimension space can be described like that: select the corresponding time as the time delay τ when the value of $R_{xx}^m(\tau)$ decreasing to the $1\text{-}e^{-1}$ times of its initial value. Obviously, this approach is the ecdysis of AD algorithm. It inherits the geometric property of AD in the reconstruction of phase space. Meanwhile, it can be regarded as the extension of autocorrelation approach in the high-order dimensions. It overcomes the drawback of the autocorrelation, i.e. the multiple autocorrelation not only guarantees that x_i and $x_{i+\tau}$, $x_{i+\tau}$ and $x_{i+2\tau}$ are not correlated with each other respectively, but also ensures that x_i and $x_{i+2\tau}$ are not correlated. Therefore, the multiple autocorrelation has a sound theoretic basis.

Finally, the algorithm we adopt to replace the AD algorithm is the "Non-bias Multiple Autocorrelation":

$$C_{xx}^m(\tau) = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{m-1}(x(i)-\bar{x})(x(i+j\tau)-\bar{x}) = R_{xx}^m(\tau)-(m-1)(\bar{x})^2 \tag{6}$$

Where, $\bar{x}$ is the mean value of $\{x_i\}$. So employing the non-bias multiple autocorrelation for $\{x_i\}$ to select a near optimum time delay τ in m dimension of phase space is to choose the corresponding time when $C_{xx}^m(\tau)$ goes to zero at first time. The strongpoint of this approach is that it is endow with the merit of AD algorithm but gets rid of its drawback. The mathematic expression is sententious and easy to computation.

In order to validate the accuracy of the improved approach, we took Henon map and Lorenz flow as examples to reconstruct them with AD and non-bias multiple autocorrelation plus Γ-test respectively. Thereinto, the data of Henon map has been interpolated 10 times with spline function and then took out 500 data to be in for experiment, for Lorenz flow, we firstly generated 10,000 data and then chosen 1,000 points between 5,000 and 6,000 for experiment. Then calculate the correlation dimensions of them and made a comparison with their nominal values [15] to figure out the errors. The experimental results are shown in table 1.

Table 1. Experimental Results

Modle	Sample Period	AD+Γ-test			C_{xx}+Γ-test			Nominal Value
		Embedding Dimension Time Delay	Correlation Dimension	Error	Embedding Dimension Time Delay	Correlation Dimension	Error	
Henon (a=1.4,b=0.3)	0.1	m=3 τ=0.8	1.3158	0.0558	m=3 τ=0.7	1.2734	0.0134	1.26
Lorenz (a=10,b=8/3, γ=28)	0.01	m=5 τ=0.35	2.0772	0.0172	m=5 τ=0.25	2.0539	0.0061	2.06

4 Conclusion

We have described an efficient method for choosing a pair of delay time and embedding dimension which facilitates an accurate reconstruction of the high dimensional dynamics. This technique is based on the non-bias multiple autocorrelation and Γ-test methods, the combination of which is computationally inexpensive. The choices of delay time and embedding dimension are important, as a good choice can reduce both the amount of data required and the effect of noise. Throughout our experiments, we have consistently found that the delay time and embedding dimension are tightly correlated. Choosing a near optimum pair of them can effectively describe the strange attractors in a nonlinear chaotic system. Since the embedding techniques are widely employed to model a physical system in cases where the mathematical description is unknown, such an automated reconstruction has a wide applicability.

References

[1]Otani Masayuki, Jones Antonia (2000). Automated Embedding and Creep Phenomenon in Chaotic Time Series, http://www.cs.cf.ac.uk.

[2]Packard N.H., Crutchfield,J.P. Farmer J.D. and Shaw R.S. (1980). Geometry from a time series, Phys. Rev. Lett. , 45, pp.712-716.

[3]Takens F. (1980). Detecting strange attractor in turbulence. Lecture notes in Mathematics, 898, pp.361-381.

[4]Albano A.M., et al.,(1988) SVD and Grassberger-Procaccia algorithm. Phy. Rev. A, 38, pp.3017-3026.

[5]Fraser A.M.(1989). Information and entropy in strange attractors. IEEE Tr. On IT, Mar. 35(2), pp.245-262.

[6] Albano A.M., et al.(1991). Using high-order correlations to define an embedding window. Physica D,54, pp.85-97

[7] Pfister Buzug T.(1992). Optimal delay time and embedding dimension for delay-

time coordinates by analysis of the global static and local dynamical behavior of strange attractors. Phy. Rev. A, 45, pp.7073-7084.

[8] Pfister Buzug T.(1992). Comparison of algorithms calculating optimal embedding parameters for delay time coordinates. Physica D, 58, pp.127-137.

[9] Rossenstein M. T., Colins J. J., Luca C.J.De (1994). Reconstruction Expansion as a Geometry-based Framework for Choosing proper Delay Times, Physica D 73, pp. 82-98.

[10] Kember G., Fowler A.C.(1993). A correlation function for choosing time delays in phase portrait reconstructions. Phy. Lett. A, 179, pp.72-80.

[11]林嘉宇，王跃科，黄芝平，沈振康。(1999) 语音信号相空间重构中时间延迟的选择—复自相关法，信号处理，Vol. 15, pp.220-225.

(Lin Jiayu, Wang Yueke, Huang Zhiping and Shen Zhenkang.(1999) Selection of Proper Time –Delay in Phase Space Reconstruction of Speech Signals, Signal Processing, Vol.15,pp.220-225)

[12]Kugiumtzis D. (1996). State Space Reconstruction Parameters in the Analysis of Chaotic Time Series—the Role of the Time Window Length. Physica D, 95, pp.13-28.

[13] Kim H. S., Eykholt R., Salas J. D.(1999). Nonlinear dynamics, delay times, and embedding windows, Physica D 127, pp. 48-60.

[14]Stefansson A., Koncar N., Jones A. J. (1997). A Note on the Gamma test, Neural Computing & Applications, 5, pp.131-133.

[15]Kugiumtzis D.,Lillekjendlie B.,Christophersen N.(1994).Chaotic time series Part I:Estimation of some invariant properties in state space, Modelling Identification and Control, 15(4), pp. 205-224.