A Neural-Based Method for Choosing Embedding Dimension in Chaotic Time Series Analysis

Sepideh J. Rastin and Mohammad Bagher Menhaj

Summary. This paper introduces applying a neural-based method for determining minimum embedding dimension for chaotic time series analysis. Many methods have been proposed on selecting optimal values for delay embedding parameters. Some frequently used methods are investigated and practically implemented, and then by using artificial neural networks (ANN) as one of components of the computational intelligence (CI) an approach was proposed to determine the minimum embedding dimension. This approach benefits from the multilayer feedforward neural networks ability in function approximation. The advantage of this method is that it gives a global nonlinear model for the system that can be used for many purposes such as prediction, noise reduction and control. Based on the achieved neural model an indirect algorithm for maximal Lyapunov estimation was suggested.

Key words: Neural networks, Chaos, Embedding, Time series.

1 Introduction

Analysis of time series derived from successive measurements of the underlying system is the most straightforward way to understand the nature of the underlying system. *Chaotic systems* that show extremely complex behavior and amazing structures are of the great interest for researchers because the time series data arise from such systems seem to be originated from the intrinsically random phenomena, but they come from deterministic nonlinear dynamical systems [3]. When we are encountered with a nonlinear system which behaves chaotically in some parts of its parameter space, linear data analysis fails and despite of the determinism leads to the false conclusion that the system is stochastic [1, 3]. This was a strong reason for developing some nonlinear techniques to uncover the deterministic structures. Chaotic behavior has appeared in economics, astrophysics, meteorology, biology, chemical processes and so many other real life events [1, 3].

The nonlinear time series methods studied here are based on the theory of dynamical systems which are defined by an m-dimensional map or an m-dimensional flow in the forms presented in (1) and (2):

$$x_{n+1} = F(x_n) \tag{1}$$

$$\dot{x}(t) = f(x(t)) \tag{2}$$

Moreover, we are interested in dissipative systems for which the volume is contracted by the time evolution if the phase space is finite dimensional [2]. For such systems, a set of initial conditions of positive measure will be attracted to some invariant subspace of phase space called the *attractor* after some transient time. The set of initial conditions leading to the same nontransient behavior is referred to the *basin* of attraction [1].

Since the dynamics of such systems are defined in some phase space, it is natural to reconstruct the phase space of the investigated system from the observations taken from system's output. There are two fundamental methods for phase space reconstruction, Delay coordinates and Derivative coordinates [3]. The last is not suitable for experimental data, because Derivatives are susceptible to noise [1, 3]. Thus the Delay reconstruction is considered for practical aspects. This method has two parameters, the delay embedding and the delay time. The appropriate adjustment of these parameters is important in practice [1, 3–5, 12]. This paper discusses various conventional methods to select these parameter in an optimal manner, and then introduces a method for determining minimum embedding dimension estimation. This method utilizes the artificial neural networks (ANN) as one of the elements of computational intelligence (CI) and has called the *predictive* method. Traditional methods for embedding dimension estimation are usually exploiting from the fact that the determinism should not be violated and the invariants should not be changed due to the reconstruction process. Besides satisfying these conditions, the proposed method processes more flexibly resulting in a global nonlinear model for the underlying system.

The rest of the paper organized as follows. In Section 2 we first review the eminent features of chaotic systems, then investigate some nonlinear tools in order to distinguish chaotic time series from the others via quantifying these characteristics. To do so, we present the Lyapunov exponents in Section 3. The natural instability of a chaotic systems manifest itself in positive maximal Lyapunov exponent. A robust direct algorithm was described to measure this nonlinear statistics. Section 4 is devoted to the phase space reconstruction from the given time series and its related theorems. Sections 5 and 6 present various routines for time lag selection and embedding dimension determination for delay reconstruction. Section 7 was dedicated to neural based predictive approach and its procedure. Based on this proposed approach, an algorithm was suggested to estimate the maximal Lyapunov exponent in Section 8. This algorithm reduces the computation complexity with respect to the algorithm described in Section 3. The given methods are practically implemented and applied to the measured data of Colpitts chaotic oscillator. Simulation results are presented in Section 9. Finally Section 10 concludes the paper.

2 Characteristics of Chaotic Systems

The first key feature of chaotic systems is their determinism. Chaos theory says that the random variables are not the only possible sources of irregularity [1]. Irregularity can be seen in deterministic nonlinear dynamical systems' outputs in some part of their parameter space. Interesting attractors can occur in such deterministic systems.

The hallmark of chaos is the exponential divergence of nearby trajectories due to the instability of solutions. This property has been referred to *sensitive dependence on initial conditions*, and makes the system unpredictable in spite of the deterministic evolution. For dissipative systems exponential separation happens in the stretching directions. Other directions are so much contracted such that the dissipation condition satisfies [2].

This dynamical aspect of chaos has its corresponding side in the geometry of the attractor [1]. The nonlinearity, the dissipation and the invariance of the attractor together with the exponential divergence cause the attractor folded in the phase space and mapped to itself. This process leads to some kind of self-similarity known as *statistical* type [5]. If a piece of a strange attractor is enlarged, it will resemble itself. Due to this reason the attractors of chaotic systems have been called strange. Strange attractors show globally bounded but locally instable behavior.

The last key property is related to the power spectrum of these systems. Although their power spectra still may contain peaks, a noisy background of broadband spectrum is present [1,3]. We cannot use this feature to distinguish a noisy quasiperiodic signal from a chaotic one.

In order to verify the chaos, we can define some criteria for investigating these properties.

3 Maximal Lyapunov Exponent

According to the sensitive dependence on initial conditions in chaotic systems, an initial infinitesimal perturbation will typically grow exponentially; the averaged exponent of this growth rate is called the Lyapunov exponent that quantifies the strength of chaos [1, 15].

The number of definable Lyapunov exponents is equal to the phase space dimensions [1,15]. Such Lyapunov spectrum is denoted by $(\lambda_1, \ldots, \lambda_m)$, where subscript m denotes the phase space dimension. The maximal Lyapunov exponent, λ, is the most important element of the spectrum because it has a dominant behavior, and its positiveness is a signature of exponential divergence of nearby trajectories. The quantity λ is defined by the following equation [3]:

$$\lambda = \lim_{\Delta n \longrightarrow \infty} \lim_{\delta 0 \longrightarrow \infty} \frac{1}{\Delta n} \ln[\frac{\delta_{\Delta n}}{\delta_0}] \tag{3}$$

where δ_0 is the initial distance between two points in phase space and $\delta_{\Delta n}$ is the distance between two trajectories deriving from these points at time n. If δ_0 is finite rather than infinitesimal, $\delta_{\Delta n}$ cannot get larger than the diameter of the attractor [2].

It should be noted that the Lyapunov spectrum and thus the maximal exponent are characteristic exponents for the system because they are invariant under smooth transformations [1, 2, 8]. Oseledec (1967) studied on invariant probability measures, and proved that the Lyapunov spectrum has the property of ergodicity by means of his "Multiplicative Ergodic Theorem" [2, 16].

In accorded with the importance of maximal Lyapunov exponent, it is necessary to estimate it from a given data series. Various methods are classified in direct and indirect approaches. Here, a direct approach is presented [7]:

1. Choose a point y_{n_0} in the m-dimensional phase space.
2. Find all of its neighbors with distance smaller than ϵ.
3. Compute the average over the distances of all neighbors to the reference part of the trajectory as a function of relative time.
4. Repeat the above steps for many values of n_0.

Finally (4), has to be computed

$$S(\Delta n) = \frac{1}{N} \sum_{n_0=1}^{N} \ln\left(\frac{1}{|N_\epsilon(y_{n_0})|} \sum_{y_n \in N_\epsilon} |s_{n_0+\Delta n} - s_{n+\Delta n}|\right) \quad (4)$$

where reference points y_{n_0} are phase space vectors and N_ϵ is the neighborhood of y_{n_0} with ϵ radius. If for some ranges of Δn and for some choices of m and ϵ, the function $S(\Delta n)$ exhibits a linear increase, its slope would be an estimation of the maximal Lyapunov exponent. This method was implemented and applied to the time series of 2,000 data points obtained from observable variable x of Henon chaotic map given with (5),

$$\begin{aligned} x_{n+1} &= 1 - a{x_n}^2 + y_n \\ y_{n+1} &= b x_n \end{aligned} \quad (5)$$

for values $a = 1.4, b = 0.3$ the system is chaotic [2]. We computed (4) for different values of ϵ and m. The results for $\epsilon = 0.2$ and $m = 1, \ldots, 5$ are plotted in Fig. 1. The parameter ϵ was choose so that the 500 reference points have enough neighbors and the distances do not saturate for small Δn. Using linear regression for the linear parts of the curves, we can determine λ with a tolerence of .01 as $\lambda = 0.41 \pm 0.01$.

4 Phase Space Reconstruction, Embedding Theorems

Since we do not confront with a phase space object but a time series, we have to convert it into state vectors such that the invariant characteristics of the

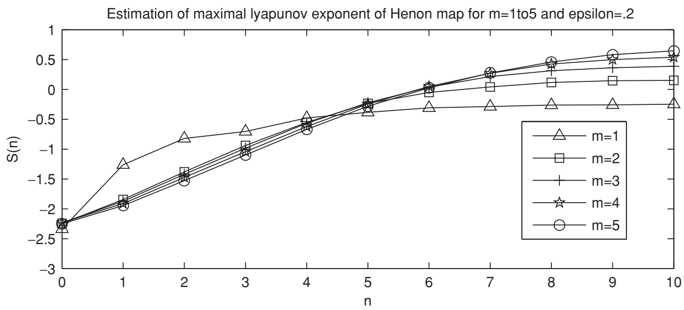

Fig. 1. Maximal Lyapunov exponent estimation for the time series of 2,000 data points obtained from observable variable x of Henon map given with (5) and choosing $\epsilon = 0.2$ and $m = 1, \ldots, 5$

original unknown attractors are preserved. Thus, an embedding of a compact smooth manifold A into R^m should be a map G that is a one to one and immersion on A [1].

As we mentioned before one of the reconstruction method is the delay coordinates established upon the Takens' *delay embedding theorem* [1, 3, 8, 9]. Let us denote the measurement function with s. A sequence of scalar measurements taken at multiples of a fixed sampling time can be shown as:

$$s_n = s(x(n\Delta t)) \tag{6}$$

then the delay reconstruction is formed by the vectors y_n,

$$y_n = (s_{n-(m-1)k}, ..., s_n)) \tag{7}$$

In the above, $k\Delta t$ is referred to the lag or delay time denoted by τ and m is the dimension of delay reconstruction. Takens proved that for an infinite noise free data series, a delay map of dimension $m \geq 2D + 1$ is an embedding of a D-dimensional compact manifold, i.e., it is a deffeomorphism [1, 9]. This theorem was generalized by Saur et al. called the *fractal delay embedding prevalence theorem*. They replaced the condition $m \geq 2D + 1$ with $m \geq 2D_f + 1$, where D_f denotes the Capacity (Box Counting Dimension) of the attractor [11]. Moreover, it has been shown that an embedding dimension $m > D_f$ suffices [1, 8, 10].

The delay reconstruction is consisted of two parameters adjustment: the embedding dimension m and the lag τ. It is easy to show that in reality with a finite number of noisy data, the estimates of the invariants depend on both m and τ [3]. Therefore, their optimal selection is of practical importance.

5 Choosing the Delay Time

Except for the fact that in *fractal delay embedding prevalence theorem* certain values for Δt and τ are not allowed to be chosen, these values are not the topic of embedding theorem under its ideal conditions [1, 3]. Different τs result in diffeomorphically equivalent attractors. If τ is selected small compared to the time scales of the system, components of the delay vectors are strongly correlated. In such cases, all reconstructed vectors are collected around the bisectrix of R^m, unless m is very large [1, 12]. This situation gets better when τ is increased. In these cases, the attractor unfolds and its structure becomes visible on larger scales. If τ is increased to very large amounts, the successive elements get independent which may lead to self-intersection in reconstructed trajectories [8]. The most conventional method is the minimization of the redundancy of the coordinates of the reconstructed space [8]. To do so, we can choose the time at which the autocorrelation function reaches $1/e = 1/2.7183$ as the lag time. Since the autocorrelation is a linear statistical quantity, it is more sophisticated to choose the time corresponding to the first minimum of the mutual information function as the delay time. The mutual information for time delay τ is:

$$I(\tau) = \sum_{i,j} p_{ij} \ln p_{ij}(\tau) - 2\sum_{i} p_i \ln p_i \tag{8}$$

where p_i is the probability to find a time series in the ith bin of the histogram created for the probability distributions of the data p_{ij} is the joint probability. Note that there is no assurance that $I(\tau)$ has an apparent minimum [1, 3].

6 Embedding Dimension Estimation

Embedding theorem says that the choice of m needs a priori knowledge of D_f of the original attractor which is unrealistic for experimental data [1, 3, 10]. Thus several methods have been proposed on embedding dimension estimation [3, 4, 13]. The main classical method can be classified into three types [4].

The first method is the computation of some invariant quantity like the maximal Lyapunov exponent while increasing the parameter m from low values to high values. When the estimated value for the invariant stops changing, the adequate m is achieved. This method is very data intensive and time consuming [4].

The second method is the singular value decomposition based approach. This method is very subjective and also the resultant reconstruction is not always optimal [3].

The last conventional method is a geometrical approach based on finding false nearest neighbors [4, 6]. As m increases in the reconstruction of a data series, the attractor unfolds and when it gets completely unfolded, a trajectory will never cross itself. The method of false nearest neighbors (FNN) recognizes that where the trajectory has some self-intersections, two neighboring points actually will be far away in the true embedding space. Based on this approach, Kennel (1992) was proposed an algorithm to determine the minimum m [6]. This algorithm was subjective in determining whether a neighbor is false.

To avoid this problem, Cao (1997) introduced a modified version of the Kennel algorithm that has been presented below [4]:
Let

$$E1(m) = \frac{E(m+1)}{E(m)} \tag{9}$$

with

$$E(m) = \frac{1}{N - m\tau} \sum_{t=0}^{N-m\tau-1} \frac{\|y_{m+1}(t) - y_{m+1}^{NN}(t)\|}{\|y_m(t) - y_m^{NN}(t)\|} \tag{10}$$

and

$$\|y_m(t) - y_m^{NN}(t)\| = \max_{0 \le j \le m-1} |s(t + j\tau) - s^{NN}(t + j\tau)| \tag{11}$$

where N is the length of the data series and m, τ denote the embedding dimension and the lag, respectively. The superscript NN means the nearest neighbor to the other vector as defined by the metric of (11). The optimal embedding dimension is given by the value of m where $E1(m)$ stops changing. Cao also proposed a related method to distinguish deterministic signals from the stochastic ones for practical conditions. He defined

$$E2(m) = \frac{E^*(m+1)}{E^*(m)} \tag{12}$$

where

$$E^*(m) = \frac{1}{N - m\tau} \sum_{t=0}^{N-m\tau-1} |s(t + m\tau) - s^{NN}(t + m\tau)| \tag{13}$$

for random data, E2(m) will be equal to one for any m. However, for deterministic data the values of E2 (m) will not equal to 1 for any m.

7 Predictive Method for Minimum Embedding Dimension Estimation

In this section, we propose a neuro based method for minimum embedding dimension estimation. This method benefits from the multilayer feedforward neural networks ability in function approximation. It has been shown that a three layered feedforward net with sigmoid functions in hidden layer and a linear function in output layer is able to approximate all of the squared integrable function with any approximation order, provided that there are enough number of neurons in the hidden layer. This fact has been called the *universal function approximation theorem* [14]. However, this theorem has some practical limitations.

For using this method, assume that the given system can be observed through the measurement function

$$y = h(\underline{x}), \underline{x} \in R^k \tag{14}$$

where R^k denotes the original phase space. Let

$$F_m : R \longrightarrow R^M \tag{15}$$

where F_m is the delay map given below:

$$\underline{Y}_m(n) = F_m(\underline{x}) = [y(n), y(n-\tau), ..., y(n-(m-1)\tau)]^T \tag{16}$$

Furthermore, the time evolution of the dynamics of the underlying system can be described by a deterministic map like F

$$\underline{x}(n) = F(\underline{x}(n-\tau)) \tag{17}$$

We want to find m_E such that

$$\underline{Y}_{m_E}(n) \approx \underline{x}(n) \tag{18}$$

This implies that the reconstruction attractor approximates the original one such that the time evolution from $\underline{Y}_{m_E}(n)$ to $\underline{Y}_{m_E}(n+1)$ follows the time evolution from $\underline{x}(n)$ to $\underline{x}(n+1)$ in original attraction. From (16) we have

$$\underline{x}(n) = F_m^{-1}(\underline{Y}_m(n)) \tag{19}$$

Besides, we have

$$\begin{aligned} y(n) &= h(\underline{x}(n)) \\ &= h \circ F(\underline{x}(n-\tau) \\ &= g(\underline{x}(n-\tau)) \end{aligned} \tag{20}$$

Combination of (19) and (20) yields

$$\begin{aligned} y(n) &= g \circ F_m^{-1}(\underline{Y}_m(n-\tau)) \\ &= q(\underline{Y}_m(n-\tau)) \\ &= q([y(n-\tau), \ldots, y(n-m)\tau)]) \end{aligned} \tag{21}$$

therefore $y(n)$ can be approximated in the form of:

$$y(n) = \hat{q}(\underline{Y}_m(n-\tau)) \tag{22}$$

Function $\hat{q}$ as an approximation of q can be obtained using a feedforward net with *error backpropagation* (BP) training algorithm. To do so, the net architecture is made of the input layer, the hidden layer and the output layer. The input layer consists of m units, and the elements of delay vectors are distributed to the neurons. In order to determine $m = m_E$:

1. Start from $m = 1$.
2. Train the net and apply the test set to the trained net to obtain the $\hat{y}(n)$. Compare $\hat{y}(n)$ with $y(n)$. Compute the prediction error, $e = \hat{y} - y$.
3. Put $m = m + 1$ and compute e again.
4. The routine will be finished when root mean squared prediction error, $e(rms)$, has no remarkable changes as m.

The value of m for which the $e(rms)$ begins to be constant is equal to the minimum embedding dimension m_E. This approach is illustrated in Fig. 2.

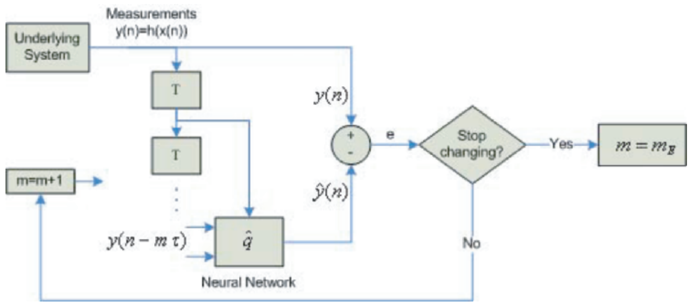

Fig. 2. The predictive approach for minimum embedding dimension estimation

8 Indirect Method for Maximal Lyapunov Estimation

The predictive method results in determination of m_E and also gives a neural model. Based on this model, an indirect algorithm for maximal Lyapunov exponent estimation is suggested. This approach is summerized below:

1. Select an arbitrary vector $\underline{Y}_m(n)$ in the delay reconstructed space. Then, use the neural model to obtain $y(n+\tau)$.
2. Compute $\tilde{\underline{Y}}_m(n) = \underline{Y}_m(n) + \underline{\epsilon}_0$, where ϵ_0 is a perturbation vector in the form of $(\epsilon_0, 0, \cdots, 0)$ with a small ϵ_0 tending to zero. Then, apply the neural model to compute the $\tilde{y}(n+\tau)$.
3. Compute $S(\Delta n\tau)$ given by (23). Plot the graph $S(\Delta n\tau)$ versus $\Delta n\tau$ and compute its slope which gives an estimation of the quantity λ.

$$S(\Delta n\tau) = \frac{1}{N} \sum_{n_0=1}^{N-m\tau} \{ln(|\tilde{y}_m(n+\Delta n\tau) - y_m(n+\Delta n\tau)|) \tag{23}$$

This approach requires less computations than that of the direct one, because there is no need of neighbor searching.

9 Simulation Results

The methods presented for choosing optimal values for m and τ were practically implemented and applied to the experimental data derived from the Colpitts chaotic oscillator [17]. A schematic of Colpitts circuit is given in Fig. 3. The given time series was made of 6,000 measured points. As mentioned before, mutual information is a powerful technique for selecting τ, the mutual information function was computed for lag from 0 to 32 units. As Fig. 4 shows, this function has an obvious minimum. Thus, we choose the corresponding time of this minimum for $\tau = 4$.

The next step is to determine m. We first computed the maximal Lyapunov exponent by the presented direct method for m = 1 to m = 6. The results are

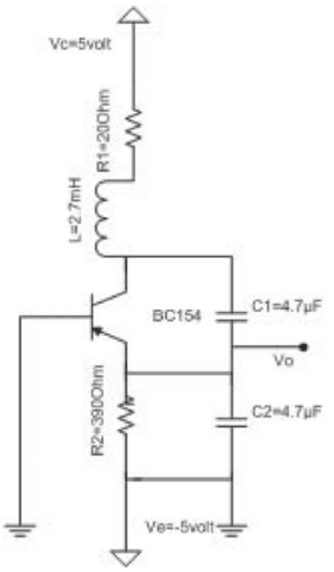

Fig. 3. The Colpitts oscillator

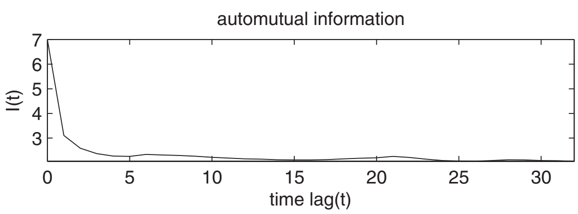

Fig. 4. Computed mutual information function for Colpitts time series

exhibited in Fig. 5. The estimated λ is independent of m for $m \geq 3$. The violations of linear growth in small scales may be due to the measurement noise or due to the lack of neighboring points. Then, we applied the Cao geometrical algorithm to the data series. The maximum value for m was set to 8. Figure 6 shows the resultant E1(m) and E2(m). Form the figure, it can be inferred that $m = 3$ will give an appropriate selection for embedding dimension. Moreover, one can figure out that the given time series is not stochastic.

As the last approach, the predictive method was implemented. A feedforward net was composed of three layers. Having tested different numbers for neurons, we put 8 units in the hidden layer. The given time series was divided into the training and test sets. We applied 5,500 data for training and 500 data for testing. The BP learning algorithm with the mean of squared errors as the index function was used for the training procedure. The value of index

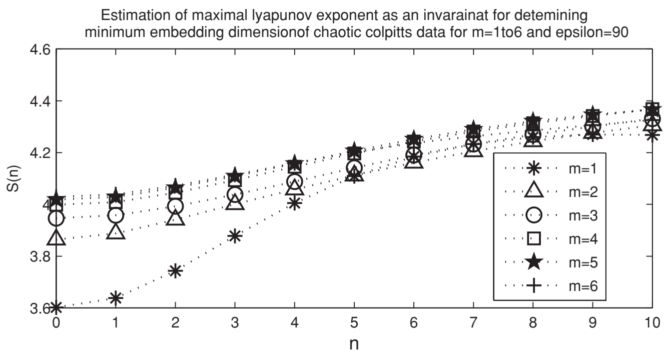

Fig. 5. Computing the maximal Lyapunov exponent as an invariant for estimating m

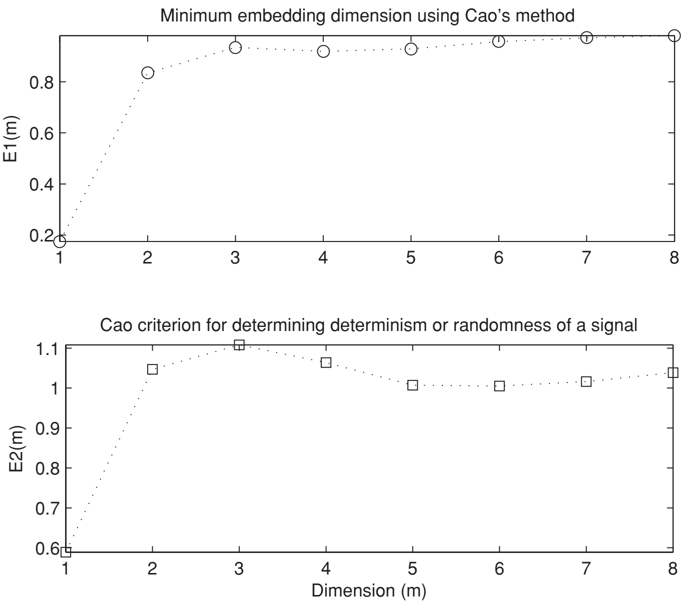

Fig. 6. Applying Cao geometrical approach for choosing m with $\tau = 4$. (**a**) E1(m) (**b**) E2(m)

function to stop the iteration was set to 10^{-8}. The learning curves of the utilized nets with m-inputs is presented in Fig. 7.

As we can see, in cases $m = 1, 2$ the learning curves did not reach to the desired value, but by increasing m to $m \geq 3$. The curves have come to the desired predetermined value. Besides, from Fig. 8 it is obvious that for $m \geq 3$ the prediction error has not remarkable changes. Thus we choose $m = 3$ again. The reconstructed attractor with $m = 3$ and $\tau = 4$ is shown in Fig. 9.

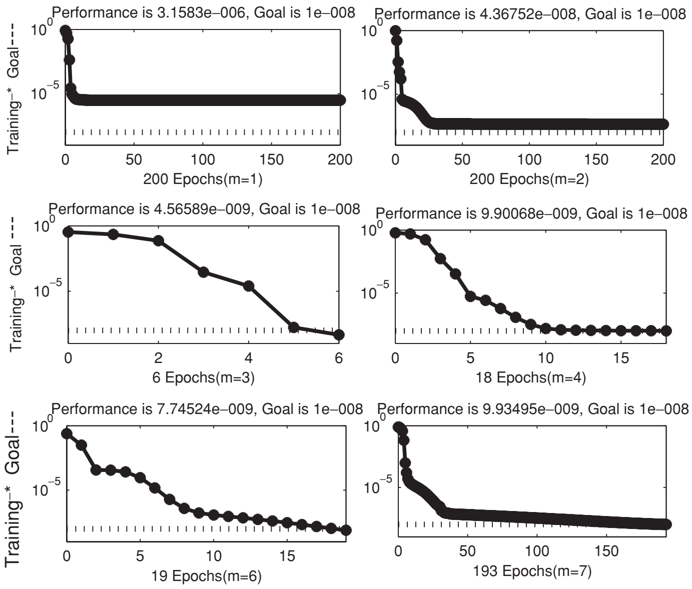

Fig. 7. Learning curves corresponding to m-inputs neural nets

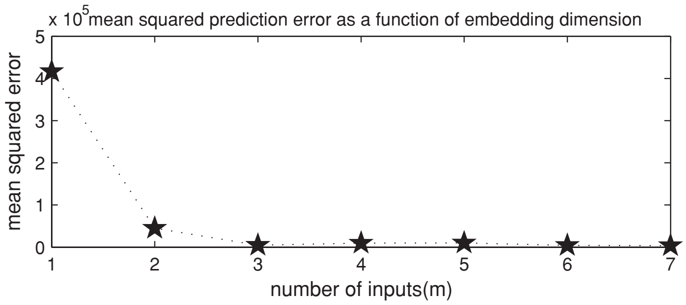

Fig. 8. Mean of squared prediction error of each used neural network as a function of m

10 Conclusion

Characteristics of chaotic time series were first studied. In order to investigate these properties, some nonlinear analysis tools were reviewed. Since the reconstruction of phase space is the basis of all nonlinear time series analysis, delay reconstruction and its parameter adjustments were studied. In addition to classical method for choosing these parameters a neuro based method known as predictive was presented. All methods were practically implemented and applied to the experimental data of Colpitts chaotic oscillator. Among methods described in the paper, the Cao geometrical approach and our suggested method were very promising. Cao algorithm is better than the other

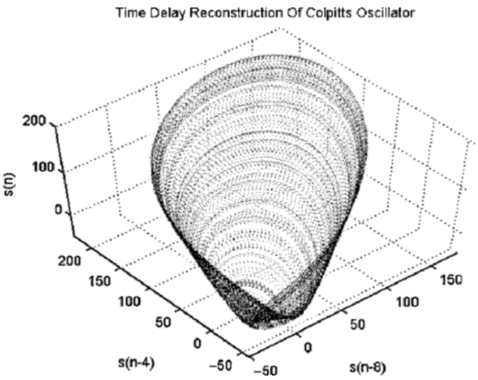

Fig. 9. The reconstructed attractor with $m = 3$ and $\tau = 4$

traditional methods, because it is not subjective and data intensive. The presented method gives a global nonlinear model for underlying system. This model can be used for many purposes such as prediction, noise reduction and control. Besides, based on the achieved neural model an indirect algorithm for maximal Lyapunov exponent estimation was suggested. This algorithm reduces remarkably the computational complexity.

References

1. H. Kantz, and T. Schreiber, Nonlinear time series analysis, Cambridge University Press, 1997
2. J.P. Eckmann, and D. Ruelle, Ergodic Theory of Chaos and Strange Attractors, Rev. Mod. Phys. 57(3), Part I, 617, 1985
3. D. Kugiumtzis, B. Lillekjendlie, and N. Christophersen, Chaotic time series. Part I. Estimation of invariant properties in state space, Model. Identification Control, 15(4), 205–224, 1994
4. L. Cao, Practical Method for Determining the Minimum Embedding Dimension of a Scalar Time Series, Physica D, 110, 43–50, 1997
5. J. Theiler, Estimating Fractal Dimension, Opt. Soc. Am. A, 6(7), 1055, 1990
6. M.B. Kennel, R. Brown, and H.D.I. Abarbanel, Determining Embedding Dimension for Phase Space Reconstruction Using a Geometric Construction, Phys. Rev. A, 45, 3403–3411, 1992
7. H. Kantz, A Robust Method to Estimate the Maximal Lyapunov Exponent of a Time Series, Phys. Lett. A, 185, 77, 1994
8. U. Parlitz, Nonlinear Time-Series Analysis, Nonlinear Modeling, Advanced Black-Box Techniques, Kluwer, Boston, 209–239, 1998
9. F. Takens, Determining Strange Attractors In Turbulence, Lecture Notes In Math., vol. 898, Springer, New York, 1981
10. T. Sauer, and J.A. Yorke, How Many Delay Coordinates Do You Need?, Int. J. Bifurcation Chaos, 3, 737–744, 1993
11. T. Sauer, J.A. Yorke, and M. Casdagli, Embedology, J. Stat. Phys., 65, 579–616, 1991

12. M. Casdagli, S. Eubank, J.D. Farmer, and J. Gibson, State Space Reconstruction in the Presense of Noise, Physica, D51, 52, 1991
13. Z. Aleksic, Estimating the Embedding Dimension, Physica D52, 352–368, 1991
14. K.M. Hornik, M. Stinchcombe, and H. White, Mutilayer Feedforwrd Networks are Universal Approximators, Neural Networks, 2, 1989
15. T. Schreiber, Interdisciplinary Application of Nonlinear Time Series Methods, Phys. Rep., 308, 1–64, 1999
16. V.I. Oseledec, A Multiplicative Ergodic Theorem, Lyapunov Charactristic Numbers for Dynamical Systems, Tran. Moscow Math. Soc., 19, 197, 1968
17. M.P. Kennedy, Chaos in the Colpitts Oscillator, IEEE Tran. Circ. Syst., 41(11), 771–774, 1994