A MUTUAL INFORMATION-BASED METHOD FOR THE ESTIMATION OF THE DIMENSION OF CHAOTIC DYNAMICAL SYSTEMS USING NEURAL NETWORKS

Christos Chatzinakos[1], *Constantinos Tsouros*[4]

Aristotle University of Thessaloniki
Polytechnic School
University Campus, Zip Code 541 24

Nikos Kofidis[2], *Athanasios Margaris*[3]

University of Macedonia
Department of Applied Informatics
Egnatia 156, Zip Code 540 06

ABSTRACT

In this paper, a method of estimating the dimension of dynamical systems from a time series, using neural networks, is examined. It is based (a) on the hypothesis that a member of a time series can be optimally expressed as a deterministic function of the d past series values (where d is the dimension of the system), and (b) on the observation that neural networks'learning ability is improved rapidly when the appropriate amount of information is provided to a neural structure which is as complex as needed. To estimate the dimension of a dynamical system, neural networks are trained to learn the component of the attractor expressed by a reconstructed vector in a suitable phase space whose embedding dimension m, has been estimated using the mutual information method. More specifically, the information supplied to the networks is represented by vectors consisting of the m past values of the time series, where m varies from 1 to $D + 2$, D being a pre-estimation for the maximum value of the embedding dimension of the system. The current method proposes that when m meets the dimension d of the dynamical system, the neural model of the attractor remarkably improves its learning ability, minimizing locally the RMS error of the training set. The logistic and the Henon map as well as the Lorenz and the Rosler attractors expressed as systems of difference equations, were examined to test the validity of the method.

Index Terms— chaotic systems, neural networks, embedding dimension

1. INTRODUCTION AND REVIEW OF PREVIOUS WORK

Chaotic analysis and neural networks' modeling can be very helpful tools in understanding both chaotic systems and neural networks. Much of the recent research work has been orientated in the common field of chaos and neural networks.

Lapedes.A & Farber R. [1] constructed back propagation neural networks to predict chaotic time series, using the appropriate number of past system values to feed the input module.

SF Masri. AG Chassiakos and TK Caughey [2] have investigated the ability of back propagation networks with static or dynamic neurons to learn non-linear dynamical systems. Small changes of parameter values were allowed, while experimental measurements were used to provide the network with the displacement and velocity values of the dynamical system.

George J. Mpitsos and Robert M. Burton. Jr [3] studied in detail the effect of input signal dynamics to the neural network learning process, by modeling the logistic map for $\lambda = 3.95$, random noise, and sine functions, with back propagation neural networks, while Ramazan Gencay [4] constructed feed forward networks to predict non linear time series produced by the Henon map application when noise is present.

I-Cheng Yeh [5] constructed back propagation neural networks with added extended layer and auxiliary output neurons to model the Henon and Ikeda maps. Kugiumztis D. [6] has used neural networks for the attractor state space reconstruction in order to predict chaotic series. Renals Steve and Rohwer Richard [7] investigated the dynamics of discrete neural networks consisting of N sigmoid nodes fully connected via non-symmetric (in general) weight matrix. John F. Kolen and Jordan B. Pollack [8] have experimented on simple back propagation networks, to indicate sensitivity of convergence on initial choice weights, learning rate, and momentum. Han L.J Van der Maas Paul F.M.J Verschure and Peter C.M Molenaar [9] have explored chaotic behavior of the sum of the weights of a 3-cell autoassosiator, trained by the use of back propagation and Hebian rules. Thomas B. Kepler, Sumeet Datt, Robert B. Meyer and L.F Abbot [10] implemented a four-node neural network circuit in order to explore possible pathways to chaotic behavior. As shown period doubling, intermitency, and quasi-periodic pathways are all possible roots to chaos for neural networks.

Paul F.M.J Verschure [11] has taken the advantage of neural networks' chaotic behavior to create a chaos-based learning algorithm.

Francois Chapeau-Blondeau and Gilbert Chauvet [12] studied the behavior of two and three node neural networks with full or partial delay.

K.Aixara,T.Takabe and M.Toyoda [13] proposed a neuron model with chaotic dynamics, based on the Nagumo-Sato model.

E. K. Blum and Xin Wang [14] explored the dynamics of small neural networks of the sigmoidal type, using the corresponding difference equations, under the consideration of time discreteness and synchronization.

G. Randons, H. G. Schuster and D. Werner [15] considered the iterational procedure of weight updating, with learning rate as the varying parameter. This resulted in fractal measures (Cantor sets) for the invariant distributions of the weights P(w).

2. THE PROPOSED DIMENSION ESTIMATION ALGORITHM

The current work is based on two essential ideas, the former concerning the mathematical formulation of a dynamical system and the latter concerning some features of the learning ability of neural networks.

The hypothesis concerning dynamical systems stands that a member of a time series can be optimally expressed as a function of the d past series values, where d is the dimension of the system, obeying an underlying iterative mechanism. It can be proven, that under certain constraints, a system of m difference equations with m variables $\{x_1, x_2, \ldots, x_m\}$, can be replaced by one or more alternative systems of m equations containing one variable's m past values $\{x_{1,n-1}, x_{1,n-2}, \cdots, x_{1,n-m}\}$. Provided that any system of differential equations can be transformed, within some accuracy level, to a system of difference equations, there ought to be an underlying iterative mechanism governing any deterministic time series. Taken's theorem of phase space reconstruction [16] sets the bounds of this relationship between members of time series, arguing on the ability of the resulting dynamical system to describe dynamical properties of the original one. On the other hand, if none of the above holds for a given time series, there is no simple neural structure (as presented in the following) to model it properly. In this case the method should be abandoned.

The second idea, concerning neural networks, focuses on the observed property of neural networks to respond optimally when the necessary amount of information is provided to their input. However, it is well known that the determination of both the necessary information and the optimum level of complexity it is not always possible. To overcome this limitation, reliable and accurate models of the dynamical systems under consideration are constructed, by using the appropriate value of the embedding dimension, based on the fact, that different neural models of the same attractor should result to distinguishable network performances. When a neural network is provided with the necessary information, its learning ability increases rapidly. This is indicated by a great slope of the curve of RMS error of the training set, leading to its local minimum. The information provided to the network is represented by an m-dimensional vector consisting of m past values of the time series. The parameter m - known as network order - starts at the value $m = 1$ and gradually increases with a step $s = 1$. Provided that a member of a time series can be optimally expressed by the d past values, d being the dimension of the system, it can be considered that the learning starting point of the network (the local minimum of the RMS) appears when m takes a value m_j which is very close to d. The value m_j reveals the dimension of the system if no significant improvement occurs with further increase of m. If further improvement of the networks' performance can be observed (correspondingly the slope of the RMS error with respect to m is comparable to the previous slopes) m_j may not be a reliable estimation of the dimension. In this case the investigation goes on until an m-value indicates the starting point of a relative stabilization (probably slight improvement or worsening) of the networks' performance. It should be pointed out that the networks are trained K-times ($K = 10$ in this work) and the average performance, represented by the average of the RMS error of the training set, is considered. Training the networks as many times as possible is necessary since the resulting RMS error is sensitive to the initial values of the weight vector [17].

Typical results of the application of the previously described algorithm can be found in [18]. In that project, the time delay used in the phase reconstruction process had been selected in a random way and without applying any of the available time delay estimation methods. In this work, an improvement of the above algorithm is the estimation of the time delay by using the mutual information method [19][20]. More specifically, the algorithm considers m networks as models of a dynamical system instead of one parametric network with varying number of input neurons. The networks are set to an increasing order, with respect to m, and are trained. The number of input neurons of the network that shows remarkable learning ability, compared to the previous with lower m, reveals the dimension of the system under consideration. In a more mathematical form, the proposed algorithm includes the following steps:

1. The appropriate time delay value is estimated by using the mutual information method.

2. For $L = 1$ to K

   (a) Train m networks $N_1, N_2, \ldots, N_m$. The network N_j has an input layer of j neurons fed with j past values of the time series.

   (b) For each network N_j, keep the resulting $RMSL_j$ error of the training set after n epochs of training.

   (c) Stop after K cycles.

3. Compute the average value, $Av(RMS)$, of $RMSL_j$ with respect to L. Construct a diagram of the $Av(RMS)$ with respect to m. The dimension of the modelled system is indicated by the last minimum of the $Av(RMS)$ which is reached through a significant negative slope of the curve. In the next section the mechanism of selecting the right dimension is clarified through the study of certain cases.

As pointed out earlier in this section, the number of cycles K is kept to 10 and the maximum value of m is considered to be $m = D + 2$, where D is a pre-estimated upper level of the embedding dimension of the system.

3. EXPERIMENTAL RESULTS

The proposed algorithm for estimating the embedding dimension has been applied to a set of known chaotic systems which in short, are the following:

The logistic map: The Logistic map is an iterative model expressed by the equation

$$x_{n+1} = \lambda x_n(1 - x_n) \quad (1)$$

It was studied for $\lambda = 3.93$, a parameter value that produces a member of the family which lies in the middle between the interior and the boundary crises. The expected value of the dominant Lyapunov exponent is 0.8 bits/iteration for this member.

The Henon map: The Henon map is expressed by the difference equation

$$x_{n+1} = 1 - \alpha x_n^2 + \beta x_{n-1} \quad (2)$$

The above form of the system is equivalent to the original one expressed in a two-dimensional phase space. The elimination of the y_n variable by using the double recursion of x_n makes the neural simulation of the system easier and more effective. The parameter values are set to the values $\alpha = 1.4$ and $\beta = 0.3$.

The Lorenz attractor: The Lorenz attractor is expressed by a system of three differential equations, which in this case is transformed to a system of three difference equations

$$\begin{aligned} x_{n+1} &= x_n + \Delta t(sy_n - sx_n) \\ y_{n+1} &= y_n + \Delta t(rx_n - y_n - z_n x_n) \\ z_{n+1} &= z_n + \Delta t(x_n y_n - \beta z_n) \end{aligned} \quad (3)$$

The Rössler attractor: As in the previous case, the original system is transformed to the system of difference equations

$$\begin{aligned} x_{n+1} &= x_n + \Delta t(-x_n - z_n) \\ y_{n+1} &= y_n + \Delta t(y_n - \alpha x_n) \\ z_{n+1} &= z_n + \Delta t(\beta + (x_n - \gamma) z_n) \end{aligned} \quad (4)$$

The value of Δt is set to 0.06 while the parameter values are $\alpha = 0.1, \beta = 0.1$ and $\gamma = 18$. The resulting system is different from the original one. The expected value of the dominant Lyapunov exponent, estimated by the Wolf's algorithm [21], is between $0.01 - 0.015$.

The parameters of training and testing the neural network models are shown below:

- Type of network: backpropagation network with one hidden layer
- Learning rate $\eta = 0.1$
- Momentum $m = 0.85$
- Range of initial weight space $[-0.1, +0.1]$
- Data set used for training: 500 up to 100 pairs of input-output vectors
- Epochs (for batch training): $M = 10$
- Maximum number of epochs: $M_{max} = 20000$

The epoch is 1 to 15, because the results are same we put one characteristic value epochs=10. Besides has happened proof from 1 to 15 the recognition of the determinism. The time series has 1000 data from each network The information has direct affinity with probability to come over a fact, as much as bigger is the firstly probability to happen a fact as small scale as is the magnitude information that results by the fact. The mutual average information is the formation that we take of the section two or more sets (manifolds). This method has object to release the data from each relation and not only from linear or higher degree. For that reason we choose one time delay in order the mutual average information that we will take from the data before the time delay and the data after the time delay to be the least possible. This method is the best, he only disadvantage is that needs too much data. The optimumum time delay value used in the reconstruction stage was estimated using the mutual information method and the estimated value was equal to $\tau_\ell = 10$ for the logistic map, $\tau_h = 13$ for the Henon map, $\tau_{\ell x} = 11$ and $\tau_{\ell y} = \tau_{\ell z} = 10$ for the x, y and z components of the Lorentz map respectively, and $\tau_{rx} = 23$, $\tau_{ry} = 26$ for the x and y components of the Rössler map. The experimental results for all the map types presented above are shown in Tables 1-4. In those tables, the parameter N the number of input neurons of the neural network, while, the parameter RMS represents the mean square error of the network training. The variation of the RMS error with the number of input neurons for the logistic and the Henon map is shown in Figure 1, while, similar diagrams can be constructed for the remaining chaotic systems.

Fron the above results it is clear that the proper embedding dimension for the phase space reconstruction stage - this dimension is associated with the minimum RMS trainig error - is equal to $d_\ell = 3$ for the logistic map, $d_H = 2$ for the Henon

Fig. 1. The variation of the RMS training error with the number of input neurons for the logistic and the Henon map

Table 1. Results for the logistic map

N	1	2	3	4	5
RMS	0.0229	0.0221	0.0109	0.0205	0.0203

Table 2. Results for the Henon map

N	1	2	3	4
RMS	0.0959	0.0183	0.0980	0.0205
N	5	6	7	
RMS	0.0201	0.0287	0.0295	

Table 3. Results for the Lorenz attractor

Results for the X data series					
N	1	2	3	4	5
RMS	0.0293	0.0196	0.0094	0.0102	0.0111
N	6	7	8	9	9
RMS	0.0125	0.0137	0.0203	0.0209	
Results for the Y data series					
N	1	2	3	4	5
RMS	0.0341	0.0201	0.0099	0.0204	0.0223
N	6	7	8	9	9
RMS	0.0234	0.0243	0.0241	0.0251	
Results for the Z data series					
N	1	2	3	4	5
RMS	0.0332	0.0065	0.0101	0.0103	0.0122
N	6	7	8	9	9
RMS	0.0141	0.0148	0.0188	0.0185	

map, $d_{Lx} = d_{Ly} = 3, d_{Lz} = 2$ for the X, Y and Z time series of the Lorenz attractor, and $d_{rx} = d_{rz} = 3, d_{ry} = 4$ for the X,Y and Z time series of the Rössler attractor, in complete accordance with the values found in the literature.

4. CONCLUSIONS AND FUTURE WORK

The objective of this research was the investigation of the ability of artificial neural networks to estimate the proper embedding dimension of chaotic dynamical systems using a pre-estimated time delay value. The results showed that neural networks can be used for this task, since the estimated embedding dimension values are identical with those estimated by using other methods. Future work includes the application of this method to noisy time series emerging from known chaotic systems or real time measurements to explore the way it behaves and the accuracy of the results emerged from it.

5. REFERENCES

[1] Lapedes A., and Farber R, "Nonlinear signal processing using neural networks: Predictive and system modeling", LA-UR-87-2662, Los Alamos National Laboratory Technical Report

[2] SF Masri, A G Chassiakos, and T.K.Caughey, "Structure-unknown non-linear dynamic systems: identification through neural networks", Smart Mater. Struct. 1 pp 45-56, 2002

[3] Mpitsos J George, and Burton M. Robert Jr, "Convergence and Divergence in neural networks: Processing of Chaos and Biological Analogy", Neural networks, Vol. 5, pp 605-625, 1992.

[4] Ramazan Gencay, "Nonlinear Prediction of Noisy Time Series with Feedforward Networks". Physics Letters A 187 pp 397-403, 1994

[5] I-Cheng Yeh, "Modeling chaotic Dynamical Systems Using Extented Neuron-Networks. Neural", Parallel & Scientific Computations, Vol 5, pp 429-438.

[6] Kugiumtzis D. "State Space Reconstruction In the Prediction of Chaotic Time Series with Neural Networks", Proceedings of the Norwegian Neural Network Symposium, 1994.

[7] Renals Steve, and Rohwer Richard, "A Study of Network Dynamics", Journal of Statistical Physics, Vol 58, Nos.5/6, 1990

[8] Kolen F.John and Jordan Pollack B. "Back Propagation is Sensitive to Initial Conditions" Complex systems 4, pp. 269-280, 1990.

[9] Van der Maas L.J Han, Verschure F.M.J Paul, and Molenaar C.M Peter, " Note on Chaotic Behavior in Simple Neural Networks", Neural Networks, Vol.3, pp 119-122, 1990

[10] Kepler B.Thomas, Sumeet Datt, Robert B. Meyer, and L.F Abbot, "Chaos in a neural network circuit", Physica D46, pp. 449-457, 1990.

[11] Verschure F.M.J Paul, "Chaos-based learning", Complex Systems, Vol.5, pp. 359-370, 1991.

[12] Chapeau-Blondeau Francois and Chauvet Gilbert, "Stable, Oscillatory and Chaotic Regimes in the Dynamics of Small Neural Networks With Delay", Neural neyworks, Vol 5, pp 735-743,1992.

Table 4. Results for the Rössler attractor

Results for the X data series					
N	1	2	3	4	5
RMS	0.0197	0.0185	0.0115	0.0128	0.0148
N	6	7	8	9	9
RMS	0.0145	0.0207	0.0140	0.0173	
Results for the Y data series					
N	1	2	3	4	5
RMS	0.0194	0.0209	0.0061	0.0047	0.0051
N	6	7	8	9	9
RMS	0.0055	0.0060	0.0067	0.0069	
Results for the Z data series					
N	1	2	3	4	5
RMS	0.0327	0.0298	0.0299	0.0298	0.0297
N	6	7	8	9	9
RMS	0.0297	0.0293	0.0304	0.0322	

[13] Aixara K., Takabe T, and Toyoda M., "Chaotic Neural Networks", Physics Letters A, Vol.144, No. 6,7. 1990.

[14] Blum K.E and Wang Xin "Stability of Fixed Points and Periodic Orbits and Bifurcations in Analog Neural Networks", Neural Networks, Vol 5, pp. 577-587, 1992.

[15] Randons, H. G. Schuster, and D. Werner, "Fractal Measures and Diffusion as Results of Learning in Neural Networks", Physics Letters A, Vol. 174, pp. 293-297, 1993

[16] F.Takens, "Detecting strange attractors in turbulence", Dynamical Systems and Turbulence, Lecture Notes in Mathematics 898, Springer-Verlag, pp. 366-81, 1980

[17] Kofidis N., Roumeliotis M, and Adamopoulos M., "Chaotic Properties Of Neural Networks: Chaotic Response of RMS Series and effects on the Absolute Error Distribution, "International Journal of Pure and Applied Mathematics", Vol.2, No 2, pp135-153, 2002.

[18] Kofidis N., Kotsialos E., Margaris A., and Roumeliotis M.: "Estimation of the Dimension of Chaotic Dynamical Systems Using Artificial Neural Networks", Proceedings of 1st International Conference "From Scientific Computation to Computational Engineering", 1st IC-SCCE, Athens 8-10 September, pp. 545-555, 2004.

[19] Fraser A., and Swinney H, "Independent coordinates for strange attractors from mutual information", Physical Review A 33, pp. 1134-1140, 1986

[20] Martinerie J. M., Albano A.M., Mees A. I., Rapp P. E, "Mutual information, strange attractor, and the optimal estimation of dimension", Physical Review A 45, pp. 7058-7064, 1992

[21] Wolf A., Swift B.J., Swinney L.H., and Vastano A.J., "Determining Lyapunov Exponents from a Time Series", Physica 16D, pp. 285-317, 1985