Predicting the Future with the Appropriate Embedding Dimension and Time Lag

Georgios Lezos, Monte Tull, Joseph Havlicek, and Jim Sluss

GEORGIOS LEZOS
Graduate Student
School of Electrical & Computer Engineering
University of Oklahoma
313-G Wadsack Dr.
Norman, OK. 73072
PHONE: (405) 325-8620 FAX: (405) 325-7066
E-Mail: gelezos@yahoo.com

MONTE TULL
Associate Professor
School of Electrical & Computer Engineering
University of Oklahoma
414 Carson Engineering Center, 202 West Boyd, Room 219
Norman, OK. 73019-1023
PHONE: (405) 325-4278 FAX: (405) 325-7066
E-Mail: tull@ou.edu

JAMES SLUSS
Associate Professor
School of Electrical & Computer Engineering
University of Oklahoma
414 Carson Engineering Center, 202 West Boyd, Room 219
Norman, OK. 73019-1023
PHONE: (405) 325-4277 FAX: (405) 325-7066
E-Mail: sluss@ou.edu

JOSEPH HAVLICEK
Assistant Professor
School of Electrical & Computer Engineering
University of Oklahoma
414 Carson Engineering Center, 202 West Boyd, Room 219
Norman, OK. 73019-1023
PHONE: (405) 325-4721 FAX: (405) 325-7066
E-Mail: joebob@tobasco.ecn.ou.edu

0-7803-5529-6/99/$10.00

ABSTRACT

Prediction is a typical example of a generalization problem. The goal of prediction is to accurately forecast the short-term evolution of the system based on past information. Neural network and fuzzy logic techniques are used because they both have good generalization capabilities. The embedding dimension (number of inputs) and the time lag selection problem is treated in this paper. It is proposed, that the selection of the appropriate embedding dimension and time lag for the input/output space construction plays an important role in the performance of the above networks. It is shown that the "traditionally accepted" choices for the embedding dimension and time lag are not optimal. The proposed method offers an improvement over the traditionally accepted parameter choices. Different analytical techniques for the determination of these parameters are used, and the results are evaluated.

I. INTRODUCTION

Time series analysis includes three important specific problems: prediction, modeling, and characterizations. Only the prediction problem is treated in this paper. A time series is a sequence of observable quantities $x_1, x_2, x_3, \ldots,$ x_n taken from a system at regular intervals of time. There are different types of time series: linear/nonlinear, stationary/nonstationary or chaotic time series. We are more concerned with nonlinear, chaotic time series because they are more difficult to predict them and are often the case in most real life problems. The goal of prediction is to accurately forecast the short-term evolution of the system based on past information. Prediction is a typical example of a generalization problem if a neural network is used. The network is first trained on a set of input-output pairs. The trained network is tested afterwards with inputs, which are distinct from the inputs used for training it (generalization test). The network is expected to correctly predict the future outputs from the model it has learned during the training session. Neural network and fuzzy logic techniques are used, because they are both capable of mapping an input space to an output space and they have good generalization/interpolation capabilities [1,2]. For the purpose of this paper the Elman recurrent neural network [3] and the Adaptive Network-Based Fuzzy Inference System (ANFIS) [4,5] are used.

II. THE PROPOSED METHOD

A. The input/output construction problem

The first thing one has to deal with when modeling with neural networks and fuzzy logic is the construction of the input/output space. How many inputs are appropriate for the network and what output is the network supposed to give for specific inputs? In addition, how should these input/output data pairs be constructed in order to get the best possible results?

A common method in the literature suggests that, given a time series, a time delay reconstruction of the input/output space should be constructed, in order to view the dynamics [1,6,7]. Thus, if the time series x(n) consists of n-observations, time-lagged vectors of the following form are defined:

$$[x(n),\ x(n+T), x(n+2T), \ldots, x(n+(m-1)T)]$$

where T is the time delay or time-lag and m is the embedding dimension. The problem here is evident: "Given a time series, what is the appropriate embedding dimension and time lag for the input/output space construction?"

In the next section, two different analytical methods are discussed. These methods can give good indications regarding the selection of the time lag T and the embedding dimension m, in order to construct a desirable input/output space.

B. Analytical Methods

Correlation analysis can be used to determine the minimum embedding dimension m for a time series [6]. Correlation analysis calculates the correlation dimension as a function of the embedding dimension by using the correlation integral [8,9]. The correlation integral is the probability that a pair of points in the attractor are within a distance R of one another. The number of points is counted in the following manner. Whenever the correlation dimension saturates, the attractor is unfolded for the specific embedding dimension (m). The implication is that, if one were to develop a model to simulate the behavior of a time series and to predict its behavior, a maximum of m-independent variables would suffice [6].

Hurst analysis can be used for studying the cyclical behavior of time series [10,11]. Two important items of information can be determined from Hurst analysis: the Hurst coefficient and the average cycle length of the time series. These cycles are good indication for determining the time horizon one has to consider in order to predict the next points in time. If, for example, a time series has an average cycle length of 10-days it won't be reasonable to predict the 11-th day based on the 10-day history because there is a trend change at day 10. Rather, it will be more reasonable to predict on the 10-th day from the last 9-day history because within the 10-day cycle the trend is still persistent. Thus the selection of the embedding dimension (m) and time lag (T) in the input/output construction should exploit the average cycle length. In other words, the empirical formula: m*T = average cycle length can be considered for the determination of the above parameters.

C. Simulation results

For simulation studies of the techniques described above, two different popular time series are used, the Mackey-Glass time series [12] and the monthly sun spot data [13]. To allow comparison with earlier work, the simulation settings are as close as possible to those reported in [4, 14]. One step prediction as well as multiple step prediction with both the Elman recurrent neural network and the ANFIS network is performed. Different values for the embedding dimension and time lag are used, and the non dimensional error index [14] is recorded in order to check if the previous analytical methods can give good indications for choosing the appropriate input/output space. For example, when predicting the Mackey-Glass time series with the ANFIS network (one step prediction) the results are shown on Table 1. The first 500 input/output data pairs are used as the training data set, and the remaining 500 pairs are used for prediction purposes (one step prediction).

	ANFIS (One step prediction)	
Time lag (T)	Non dimensional error index m =5	Non dimensional error index m = 6
1	0.0025	0.0021
2	0.0144	0.0086
3	0.0244	0.0070
4	0.0171	0.0013
5	0.0039	0.0013
6	0.0070	0.0025
7	0.0104	0.0022
8	0.0119	0.0016
9	0.0143	0.0018
10	0.0068	0.0027
11	0.0146	0.0047
12	0.0183	0.0042
13	0.0158	0.0029
14	0.0154	0.0040
15	0.0141	0.0043
16	0.0173	0.0049
17	0.0165	0.0053

Table 1 Predicting the Mackey-Glass with the ANFIS using different parameters.

In order to show that better prediction results can be achieved with different values of embedding dimension and time lag, numerous simulations were carried out. ANFIS was trained for 500 epochs each time and the non-dimensional error index was recorded for an embedding dimension of five and six, with different time lags in the range 1....17. Because the number of linear and nonlinear parameters grows exponentially with the number of inputs, ANFIS cannot be tested with an embedding dimension of 10. In fact, for an embedding dimension greater than six, ANFIS becomes very inefficient. Therefore ANFIS was tested with m = 5 and 6.

The following useful conclusions can be derived from the above results:

1. The traditional selection of m = 5 and T = 6 is not the best choice for the embedding dimension and the time lag respectively (non-dimensional error index = 0.0070).

2. In order to exploit the 50-step average cycle, m = 5 and T =10 length (5*10 = 50) were chosen. The NDEI is 0.0068, which is better than the traditional selection, though there is not a great improvement (3%).

3. Very good predictions can be achieved when T = 1. That is because one step prediction is being applied, and it appears that time lag 1 training vectors work well in one step prediction. In other words, since the network is fed with the last updated points at each time, when only the next value is to be predicted, time lag 1 seems to work well for the training.

4. Similar conclusions can be derived for m = 6. When T = 8 (m*T = 48), the NDEI = 0.0016 which gives a 77% improvement over the traditional error.

5. Surprisingly, when m, T are chosen to exploit half the average cycle length the smallest errors can be achieved. For m = 5, T = 5 (m*T = 25), the NDEI = 0.0039. For m = 6 and T = 4 (m*T = 24), the best prediction with NDEI = 0.0013 (81% improvement over the traditional error) is achieved. The network seems to work better if the input/output training set is constructed with a time horizon equal to half the average cycle length. This is because there is a major change around the middle of the 50-step average cycle. The network prefers to be trained based on this 25-step trend change rather than the 50-cycle length.

Multiple step prediction is usually required in real life prediction problems. In this kind of prediction, one step ahead is initially predicted, and afterwards the first prediction is fed back in order to predict the value two steps ahead. By feeding back the last prediction, multiple step prediction can be achieved. Most forecasting techniques do not perform very well in multiple step prediction [4, 14-16]. Thus, in order to check if the selection of initial parameters is important, multiple step prediction simulations are applied.

Correlation analysis can give a strong indication of the chaotic nature of the Mackey-Glass time series. This indication for chaotic behavior implies that the time series is predictable strictly in the short term. A 13-step prediction was attempted, and simulation tests were conducted to investigate ANFIS behavior. The training data set x(101)-x(600) was used again, and the next 13 values were predicted. The network was trained for 100 epochs and for embedding dimensions $m \in [3,6]$ and time lags $T \in [1,15]$. The non-dimensional error index (NDEI) was used as a performance measure. The results are shown in Table 2. The following conclusion can be derived from these results:

1. When m and T are close to half the average cycle length (25) very good predictions can be achieved. For example, when m = 4, T = 6 (m*T = 24), the NDEI = 0.1259. For m = 5, T = 5 (m*T = 25), the NDEI = 0.0148. The reasons are the same as those stated earlier.

	ANFIS (Multiple Step Prediction)			
Time Lag	NDEI (m= 3)	NDEI (m= 4)	NDEI (m= 5)	NDEI (m= 6)
1	2.0264	0.4928	0.3802	0.6513
2	2.2706	1.2661	2.1860	1.7981
3	3.7459	4.0591	0.3377	0.0663
4	5.6565	2.4305	0.0916	0.0030
5	6.0270	0.4479	0.0148	0.0030
6	1.4305	0.1259	0.0176	0.0093
7	0.8650	0.1396	0.0392	0.0054
8	0.6552	0.3990	0.0920	0.0022
9	0.4450	0.5098	0.0179	0.0018
10	0.2474	0.3895	0.0325	0.0187
11	0.2701	0.4750	0.0115	0.0256
12	0.4485	0.2553	0.0683	0.0067
13	0.7143	0.0607	0.0637	0.0041
14	1.0229	0.0939	0.1399	0.0050
15	1.4517	0.2823	0.0278	0.0061

Table 2 Error index for 13-step prediction, with different parameters

2. Better results can be achieved when the time horizon is chosen around 50 (average cycle length). For example, for m = 5 and T = 11 (m*T = 55), NDEI = 0.0115. The 25-step trend change appears to be not as important in multiple step prediction.

3. The choice of T = 1 does not work well here as in the case of one step prediction. Predicting several steps ahead is not the same as predicting one step at the time. The reason is that the prediction one step ahead is more correlated than the prediction two steps ahead, and so on. When predicting 13 steps ahead, knowledge of initial conditions is lost. Constructing the appropriate input/output space seems to be helpful, since better predictions can be achieved for a longer future time horizon. Thus, choosing the appropriate initial parameters is more crucial in multiple step prediction, than in one step prediction.

4. The best predictions can be achieved when m = 6 and T = 9 (NDEI = 0.0018). It can be seen that m*T = 54 is very close to the traditional 50-step average cycle length.

Similar results were achieved when predicting with the Elman neural network (one step and multiple step prediction), and in the case of the sun spot data [17].

III. CONCLUSIONS

The selection of the appropriate embedding dimension and time lag for the input/output space construction plays an important role in the performance of the networks that have been used for time series prediction. The proposed method offers an improvement over the "traditionally accepted" choices. The analytical methods described in Section A are helpful for determining the appropriate embedding dimension and time lag for a given time series.

When performing one step prediction, choosing a time lag of one is generally the best choice for the input/output vector construction. However, in multiple step prediction, the suitable embedding dimension m and time lag T are very important factors. The empirical formula:

$$m * T = \text{average cycle length}$$

can be considered for the determination of the above parameters. This paper has shown that better performance is obtained with the appropriate values of the embedding dimension and time lag. An embedding dimension greater than the appropriate one does not dramatically improve the prediction results.

This paper demonstrated that good one step prediction is obtained when the time horizon extends to half the average cycle length. This was because of the dynamics of the time series that were used. Both have a potential trend change at about half of their period. However, this is not the case when performing multiple step prediction. The reason is that the prediction one step ahead is more correlated than the prediction two steps ahead, and so on. When predicting multiple steps ahead, the knowledge of initial conditions is lost. Constructing the appropriate input/output data pairs seems to work better, since better predictions can be achieved.

REFERENCES

[1] Haykin, S., *Neural networks: A comprehensive foundation*, Prentice Hall 1994.

[2] Kosko, B., *Fuzzy engineering*, Prentice-Hall, 1997.

[3] Elman, J. L., "Finding structure in time," *Cognitive Science*, Vol.14 (2), pp.179-211, 1990.

[4] Jang, J.-S.R., Sun C.-T., Mizutani E., *Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence*, Prentice Hall, 1997.

[5] Jang, J.-S.R., "ANFIS: Adaptive-network-based Fuzzy Inference System," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No.3, May/June, 1993.

[6] Deboeck, G.J., *Trading on the edge: Neural, genetic, and fuzzy systems for chaotic financial markets*, Wiley & Sons, 1994.

[7] Haykin, S., *Neural networks: A comprehensive foundation*, 2nd edition, Prentice Hall 1999.

[8] Packard, N.H., Crutchfield, J.P., Farmer, J.D., and Shaw, R.S., "Geometry from a time series," *Physical Review Letters*, Vol. 45, No. 9, September, 1980.

[9] Takens, F., "Detecting strange attractors in fluid turbulence," *Lecture Notes in Mathematics*, Vol. 898, pp. 366-381, 1981.

[10] Hurst, H.E., Black, R.P., and Simaika, Y.M., *Long-term storage: An experimental study*, London Constable, 1965.

[11] Hurst, H.E., "Long-term storage capacity of reservoirs", *Transactions of the American Society of Civil Engineers*, Vol. 116, p. 770, 1951.

[12] Mackey, M., and Glass, L., "Oscillation and chaos in physiological control systems," *Science*, Vol. 197, pp. 287-289, July, 1977.

[13] Sun Spot Index Data Center (SIDC), http://www.oma.be/KSB-ORB/SIDC/, Observatory Meteorology Aeronomy, Royal Observatory of Belgium, Brussels.

[14] Crowder, R.S.III, "Predicting the Mackey-Glass timeseries with the cascade-correlation learning," *Proceedings of the 1990 Connectionist models summer school*, pp.117-123, 1990.

[15] Lapedes, A.S., and Farber, R., "Nonlinear signal processing using neural networks: Prediction and system modeling," *Technical Report*, LA-UR-87-2662, Los Alamos, National Laboratory, Los Alamos, New Mexico, 875 45, 1987.

[16] Matsuba, I., Masui, H., and Hebishima, S., "Optimizing multilayer neural networks using fractal dimensions of time series data," *International Joint Conference on Neural Networks*, pp. 583-588, Baltimore, Maryland, June 7-11, 1992.

[17] Lezos, G., "Neural network & fuzzy logic techniques for time series forecasting", *Graduate thesis research*, University of Oklahoma, Electrical and computer engineering department, Dec. 1998.