POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

TECHNOLOGIE INFORMATYCZNE W ZARZĄDZANIU SYSTEMY WSPOMAGANIA DECYZJI

pod redakcją:
Jana Studzińskiego,
Ludosława Drelichowskiego,
Olgierda Hryniewicza,
Janusza Kacprzyka

TECHNOLOGIE INFORMATYCZNE W ZARZĄDZANIU

SYSTEMY WSPOMAGANIA DECYZJI

Polska Akademia Nauk • Instytut Badań Systemowych

Seria: BADANIA SYSTEMOWE
tom 26

Redaktor naukowy:

Prof. dr hab. Jakub Gutenbaum

Warszawa 2000

TECHNOLOGIE INFORMATYCZNE
W ZARZĄDZANIU
SYSTEMY WSPOMAGANIA DECYZJI

pod redakcją
Jana Studzińskiego, Ludosława Drelichowskiego
Olgierda Hryniewicza i Janusza Kacprzyka

Książka zawiera wybór referatów przedstawionych na konferencji "Komputerowe systemy wielodostepne KSW'2000" w Ciechocinku w 2000 r. Konferencja pod patronatem Komitetu Badań Naukowych została zorganizowana przez Akademię Techniczno-Rolniczą w Bydgoszczy, Instytut Badań Systemowych PAN, Komisję Informatyki PAN - Oddział w Gdańsku oraz Bydgoskie Zakłady Elektromechaniczne "BELAM" S.A. w Bydgoszczy.

ISBN 83-85847-53-7
ISSN 0208-8028

Rozdział 4

Metody i algorytmy obliczeniowe w systemach komputerowych

THEORETICAL FOUNDATIONS OF NEURAL NETWORKS PREDICTION

Maciej Krawczak
Systems Research Institute
Polish Academy of Sciences
University of Computer Technology and Management
Newelska 6, 01-447 Warsaw, Poland
krawczak@ibspan.waw.pl

Most observational disciplines, such as finance, biology, and physics, try to infer properties of an unfamiliar system from the analysis of a measured data of its behavior. There are many mature techniques associated with traditional time series analysis. However, during the last decade, several new and innovative approaches have emerged (such as neural networks (NN) and time-delay embedding), promising insights not available with these standard methods. Unfortunately, the realization of this promise has been difficult. Adequate benchmarks have been lacking, and much of the literature has been fragmentary.

Key words: *dynamic systems, chaos theory, approximation.*

1. Introduction

It is assumed that an interesting system is observed, namely the output of the system is available and observed. The second assumption is such that, generally, the reasons of system performance is unknown - it means the inputs to the system are not available. The aim is to analyze the observed system and to forecast the future behavior of the system. There are many real time series - systems - e.g. financial or in nature, difficult to analyze using, Tong (1990).

There have been developed mathematical methods for analysis of time series - they will be discussed in details. The main difficulties in analyzing the real world time series and related systems are due to presence of nonlinearity and representation of only finite observations - these kind of difficulties justifies the use of empirical as well as nonparametric methods dealt in this work.

For last few years much attention has been paid to the application of the chaos theory for analysis of different time dependent real systems. At the same time much interest has been given to the ability of modern computational tools such as neural networks and genetic programming. It must be mentioned that development as well as various applications of fuzzy logic has found a great interest in artificial intelligence.

The ideas developed in this work are derived from different disciplines, it means the approach to analysis of time series is multidiscipline.

2. Observation of linear systems

For long time in the system theory linear dynamic systems (very often time-invariant) were investigated very deeply. Let us consider a linear time invariant system defined by state-space equations:

$$\dot{x}(t) = Ax(t) + bu(t) \tag{1}$$

$$y(t) = hx(t) + dv(t) \tag{2}$$

The state vector of the system x is of the N dimension; $y(t)$ is the observed output of the system; $u(t)$ is the driving force (control or noise) applied to the system; $v(t)$ is considered as noise; the matrix A and vectors b, h, d are parameters of the system. Having the system parameters as well as the value of the state $x(t)$ it is easy to predict $y(t)$.

Now, let us consider a discrete version of the system (1)-(2), taking samples of the output at times Δt:

$$x(k+1) = \Phi x(k) \tag{3}$$

$$y(k) = hx(k) \tag{4}$$

in these equations $k=n\Delta t$ and $\Phi = \exp(A\Delta t)$. There is a question when and if $y(k)$ can determine corresponding state $x(k)$, in other words there is a question of observability of the system. For linear systems like (3)-(4) we define a so-called observability matrix:

$$\mathrm{O}(h, \Phi) = \begin{bmatrix} h \\ h\Phi \\ \cdot \\ h\Phi^{N-1} \end{bmatrix} \Phi x(k) \tag{5}$$

Putting

$$y(k+n) = h\Phi^{n} x(k) \tag{6}$$

we can describe an observation vector

$$[y(k), y(k+1),..., y(k+N-1)] = O(h,\Phi)x(k) \tag{7}$$

In order to find the state of the system the observability matrix $O(.)$ must be invertible.

The linear autoregression which exactly models the observed sequence has a form

$$y(k) = h\Phi^N O^{-1}[(k-N)\cdots y(k-1) + dv(t)]^T . \tag{8}$$

Now let us consider a nonlinear system, which are much more useful because most of observed time series represent nonlinear dynamic systems. In a similar way we can write a nonlinear system:

$$\dot{x}(t) = F(x(t)) \tag{9}$$

$$y(t) = h(x(t)) \tag{10}$$

where x is N dimensional state of the system, F is a nonlinear transition function, h is an observation function. Applying sampling intervals equal to Δt we can write a delay vector

$$[y(k-1), y(k-2),..., y(k-T)] \tag{11}$$

Takens theorem (1981) allows to apply the observability problem of nonlinear systems. Due to this theorem which states that under very mild conditions, if

$$T > 2D+1 \tag{12}$$

(where D is the fractal dimension), then there exists (for almost all smooth function h) an one-to-one differentiable mapping Ψ between the delay vector and the state vector $x(k)$

$$\Psi(y(k-1), y(k-2),..., y(k-T)) = x(k) \tag{13}$$

In this way we can write an autoregression which models the time series in the form

$$y(k) = h \circ F \circ \Psi(y(k-1), y(k-2),..., y(k-T)). \tag{14}$$

It is worth to notice that equation (13) has a similar form as equation (7). Such a form allows to apply nonlinear autoregressions to model time series.

3. Neural networks prediction

Feedforward neural networks work as an universal approximators. Before using such a network it must be trained. For an input *x(k)* the network response (or output) is given as

$$y(k) = N(w, x(k)) \tag{15}$$

where weights *w* are parameters of the network. During the learning process we try to minimize the average squared error (a learning error)

$$\min_{w} \frac{1}{P} \sum_{p=1}^{P} [d(k) - N(w, x(k))]^2 \tag{16}$$

counting over a training set of examples, where *d(k)* is the desired pattern for each input. Under some stationary and ergodic conditions the learning error (15) converges to an expectation:

$$\begin{aligned} &\lim_{P\to\infty} \min_{w} \frac{1}{P} \sum_{p=1}^{P} [d(k) - N(w, x(k))]^2 \\ &\to E\|D - N(w, X)\|^2. \end{aligned} \tag{17}$$

Here *D* and *X* are considered as random variables while the expectation is taken over the joint probability distribution.

The problem of finding optimal parameters w^* for linear systems is trivial, eg. Shanmugan (1988). For nonlinear systems, the problem is much more complex, especially using neural networks. Namely, we must use the universal approximator properties of feedforward neural networks. The formal explanation of using least square errors for neural network training can be found e.g. in Hecht-Nielsen (1990).

4. Chaotic series

Many interesting systems in the real world are known to be nonlinear or chaotic. Up till now by some mathematical expressions the analysis is constrained to extraction of similar tendencies. For example let us analyze one of the most famous and oldest, perhaps, equation modelling a population growth. The equation is called the logistic equation and is described by the following form (discrete case)

$$x(t+1) = b\, x(t)\,(1 - x(t)), \quad t = 0, \ldots, 300 \tag{18}$$

where b is a real value parameter. The logic equation is drawn in Fig. 1 for the parameter $b = 3.0$, and first 200 points $t = 1,\ldots,200$, and for the initial condition $x(0) = 0.1$,

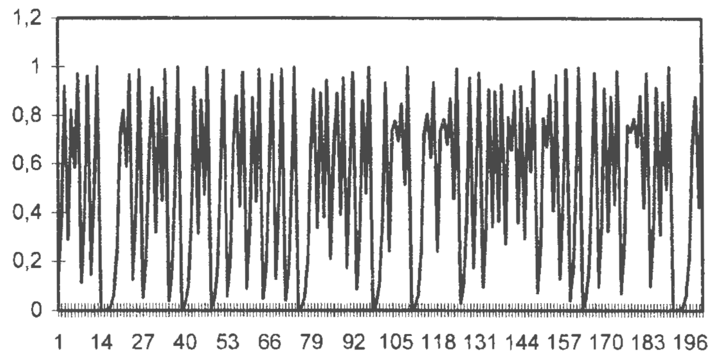

Fig. 1. The first 200 points of the logistic equation.

Another example of chaotic series is generated by the Mackey-Glass equation (1977) of the form:

$$x(t+1) = bx(t) + a\frac{x(t-s)}{1 + x^{c}(t-s)} \tag{19}$$

where the parameters can be stated as follows: $a = 0.2$, $b = 0.9$, $c = 10$, $s = 18$, and the initial conditions are assumed to be $x(0) = x(1) = \ldots = x(18) = 0.7$. Time evolution of the Mackey-Glass equation for the above parameters is shown in the Fig. 2

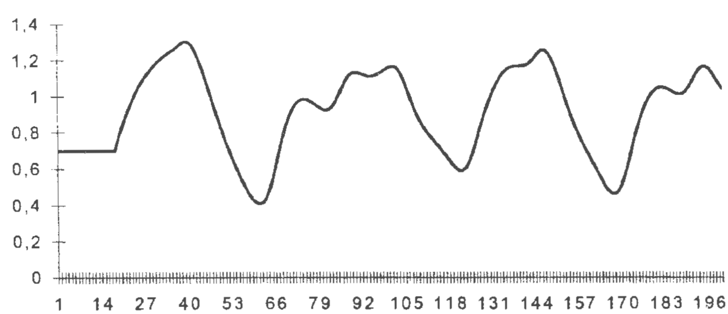

Fig. 2. The first 200 points of the Mackey-Glass equation.

The above time series are shown in order to show that simple nonlinear equations with feedback can cause very complex behavior of the plots. It will be shown little later the attractors of complex time series.

Embedding

There is a pretty simple method for analyzing time series, the method is called the *time series embedding*, Ruelle (1981) and Takens (1981). The approach can be illustrated by plotting pairs of point $x(t)$ and $x(t+1)$ for the considered functions. The case of the logistic equation is shown in the Fig. 3

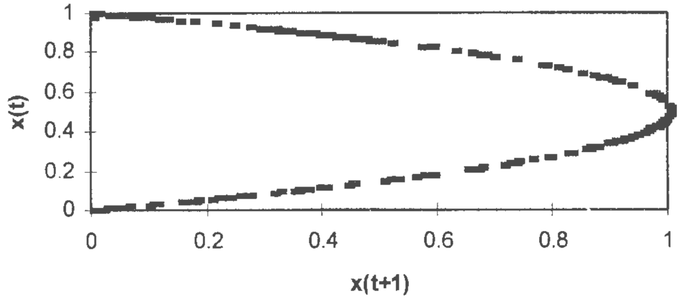

Fig. 3. Embedded logistic function.

In the case of the Mackey-Glass equation the embedding plot is shown in Fig. 4:

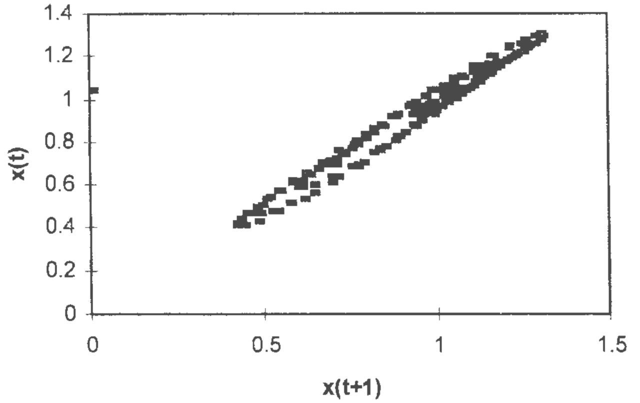

Fig. 4. Embedded Mackey-Glass function.

It should be emphasizes that even plots versus time are very complex time series than the respective embedding plots are rather simple as well as known patterns. The idea is such, namely given the point $x(t)$ it is easy to make a very good estimation of the next point $x(t+1)$ by interpolation. The similar principle can be extended to multiple dimensions:

$$X(t) = x(t), x(d+t), x(2d+t), x(3d+1), \ldots, x(nd+t) \tag{20}$$

where X is the embedded vector, d is the *separation*, and n is the *embedding dimension*.

It has been shown (Takens, 1981) that for a given chaotic series embedded properly there exists a smooth function. This function of embedding can be approximated in various ways, e.g. or by neural networks. Proper parameters that is the dimension parameter d as well as the embedding function must be found empirically. Latter we will discuss difficulties appearing during the modelling process.

5. Elements of the empirical chaos theory

5.1. Lyapunov exponents

The main feature of chaotic systems is their high sensitivity to initial conditions. There is a way to distinguish this feature, namely by calculating the Lyapunov exponents (Wolff, Swift, Swinney, 1985). These exponents indicate whether succeeding points laying on an attractor diverge or converge - with passing time.

The examined trajectories on the attractor are embedded in a space. The divergence between two trajectories can be measured as a difference between two n-tuples. At the beginning there is a need to define the so-called dominant average Lyapunov exponent as:

$$L = \log_2 \frac{\sum_{n-1} \frac{l_{n+1}}{l_n}}{n-1} \tag{21}$$

where n denotes the index of a sample, l is the Euclidean distance between two neighboring trajectories. There is also possible to calculate local Lyapunov exponents, for that we need samples of trajectories of attractor which should be dense. It is obvious that the rate of divergence is not constant at all along the attractor.

It is interesting to allays the Lyapunov exponents. If the Lyapunov exponents are positive it means the system is a chaotic one, while the negative exponents indicate the system behavior is reverting, the value zero of the exponents characterizing cyclic behavior of the systems. For instance, an attractor of a sinusoidal system is a circle.

5.2. Hurst exponent

There is a very important measure of predictability of states of time series. This measure is named the Hurst exponent which is derived by application of so-called R/S analysis. Considering a time series X representing by n points, and choosing a number p which can be taken for convenience as $10 \leq p < n/2$, then the data can be divided into n/p blocks. For any block the average is calculated, then next the maximum range of each block as well as the standard deviation of each block is calculated. The value=range/standard_deviation is calculated for any block and next average of the block is calculated. The average value rs is in relation with the HURST exponent in the following way

$$rs = (p/2)^H \qquad (22)$$

where H is the Hurst exponent.

The Hurst exponent values are between 0 and 1. We can distinguish two ranges of the exponent. A value $0.5 < H < 1$ indicates so called persistent behavior, it means a system can be considered as the values moves to one as a predictable system. While a value $0 < H < 0.5$ indicates probabilistic systems. For $H = 0$ time series must change direction every sample, for $H = 0.5$ time series moves as a random walk, while for $H = 1$ a system is a purely deterministic.

There is a relationship between one definition of the fractal dimension and the Hurst exponent, that is following expression

$$D = 2 - H \ .$$

Analysis and prediction of chaotic time series requires finding the above mentioned parameters - this is another difficult task. We can distinguish two class of methods: the first empirical and the second analytical one. Within the first group we must form a model of the attractor and a supervised learning algorithm of some kind is required (e.g. algorithms for learning feedforward neural networks or genetic algorithm). Within the analytical methods we must base our consideration on the Takens theorem (1981) for determining the upper bounds of an embedding parameter (if we know the fractal dimension of the attractor). There are several methods for deriving a choice of embedding dimension, e.g. Tong (1990), Wolff, Swiftt, Swinney, (1985).

6. Conclusions

In this paper we summarize modern foundations of application neural networks for prediction of time series. Here feedforward neural networks which are universal approximators are used as a tool for modeling unknown nonlinear functions.

We have considered linear systems and associated state observability or autoregression. In a similar way we described nonlinear systems. Next we showed feedforward neural networks as universal approximators. Nonlinear systems, even very simple, can generate in some sense unstable solutions - these kind of systems are called chaotic systems. The main parameters of such systems are described and the role of neural networks as a tool for modelling.

References

Baker, L. and J. P. Gollub (1990): Chaotic Dynamics, an Introduction. Cambridge Univ. Press.

Box, E. P. and G. M. Jenkins (1976): Time series Analysis: Forecasting and Control. Holden-Day.

Bryson, A. and Y. Ho (1975): Applied Optimal Control. Hemisphere Pub. Corp. N-Y.

Cybenko, D. (1989): Approximation by superpositions of sigmoidal function. Mathematics of Control, Signal, and Systems, vol.2, no. 4.

Dreyfus, S. E. (1990): Artificial Neural Networks, Back Propagation, and the Kelley-Bryson Gradient Procedure. Journal of Guidance, vol. 13, no. 5, 926.

Hecht-Nielsen, R. (1990): Neurocomputing. Addison Wesley.

Hertz, J. A., A. Krogh, R.G. Palmer (1990): The Theory of Neural Computation. Addison-Wesley.

Hornik, K., M. Stinchombe and H. White (1989): Multilayer feedforward neural networks are universal approximators. Neural Networks, vol. 2, 359.

Krawczak, M. (2000): Backpropagation versus dynamic programming approach. BULLETIN OF POLISH ACADEMY OF SCIENCES, SERIA TECHNICAL SCIENCES, vol. 48, no. 2.

Mackey, M. and L. Glass (1977):Oscillations and chaos in physiological control systems. Science, 197,287.

Ruelle, D. (1981): Small random perturbations of dynamical systems and the definition of attractors. Commun Math. Phys., 137, 82.

Shanmugan, K. and A. Breipohl (1988): Random Signals: Detection, Estimation, and Data Analysis. John Wiley and Son, N-Y.

Takens, F. (1981): Detecting strange attractors in fluid turbulence,. In: D. Rand and L. S. Young (eds): Dynamical systems and turbulence, Springer, Berlin.

Tong, H. (1990): Non-linear Time Series - A Dynamical System Approach. Oxford Univ. Press.

Weigend, A. and N. Gershenfeld (eds) (1993): Time series prediction: Forecasting the future and understanding the past. Addison-Wesley.

Wolff, A., J. B. Swift, H. L. Swinney and J. A. (1985): Determining Lyapunov exponents from a time series. Physica 16D, 285.

ISSN 0208-8029
ISBN 83-85847-53-7

W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy
prosimy o kontakt z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. 837-35-78 w. 241 e-mail: bibliote@ibspan.waw.pl