ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 139-144

Error Measure for Identifying Chaotic Attractors

Tomas Hrycej

Daimler-Benz AG, Research Center Ulm
Postfach 2360, D-89013 Ulm, Germany

Abstract

Behavior of chaotic systems cannot be exactly forecast for all state variables by identified models since inaccuracies of model parameters lead to exponential forecast errors. However, under certain conditions a model can be identified that possesses the same strange attractor. This can be reached by using an error measure based on error volume evaluation, instead of additive squared error. In this way, the forecast horizon can be substantially extended.

1 Introduction

The behavior of chaotic systems cannot be exactly forecast by identified models since a deviation in model parameters leads to exponential forecast error (see, e.g. Nicolis and Prigogine [4]). This results from the fact that a small deviation in the initial state $\vec{x}_0$ of a chaotic system leads to an exponentially growing deviation in the state $\vec{x}_t$ for $t \rightarrow \infty$. A small deviation in model parameters leads, for an identical initial state, to a small deviation in the state $\vec{x}_1$ and has thus the same consequences as a small deviation in the initial state.

However, *dissipative systems* are not completely unpredictable. By their definition, for dissipative systems, the volume enclosing a family of trajectories in the state space shrinks with time (e.g., Schuster [5]). As a consequence, even a chaotic dissipative system is attracted into a region of dimension smaller than the complete state space, into the so called strange attractor. This attractor is typically a set that is "thin" in some directions and "thick" in others. This property of dissipative systems makes a certain forecast of the system behavior possible. We cannot know where the system is exactly, but we can try to (A) *identify the attractor*, that is, look for a model whose attractor has similar topology to that of the real system, and (B) *determine the position along "thin", predictable dimensions.* Since practically all technical, and many natural, systems are dissipative, solving this problem would mean a substantial shift of our ability to forecast the behavior of nonlinear systems.

The present work shows a possibility to address this problem with neural networks, using an appropriate error criterion for learning agorithm.

2 Problems with conventional least squares

To formalize the identification problem, let us denote the mapping representing the system (in the state space representation) as

$$\mathbf{z}_{t+1} = f(\mathbf{z}_t, \mathbf{u}_t) \tag{1}$$

with $\mathbf{z}_t$ being the system state and $\mathbf{u}_t$ the input into the system at time t. This mapping corresponds to the neural network (or, generally, any mathematical model) by which the system is identified.

The identification task can be formulated as looking for the model which minimizes the deviation between the model forecast for the i-th state variable z_i and measurement from the real system d_i. Usual approach is to take square deviations of single-step forecasts and sum them up over all state variables

$$\sum_i (z_i - d_i)^2 \tag{2}$$

This criterion is used in classical (e.g., Åström and Wittenmark [1]) and neural (e.g., Nguyen and Widrow [3], Werbos [6]) system identification.

Minimizing (2) is a successful approach to linear system identification. For nonlinear chaotic systems, the problem with all *additive* least squares criteria is that the *inherently necessary error* (given by the unpredictability of the exact position *within* the strange attractor) in dimensions where the attractor is "thick" may prevent minimization of the error in other, "thin", dimensions. The reason for this is that even the smallest, casual improvement along the former dimension (caused, e.g. by a minor local adaptation to the concrete measurements) may be numerically larger than a substantial improvement along the latter. This suggests the use of *nonadditive* minimizing criteria.

3 Error volume criterion

Dissipative systems are characterized by the contraction of initially n-dimensional phase space to a (possibly chaotic) attractor of dimension lower than n. The intuition behind the following developments is that *although the forecast error cannot be reduced down to zero, it can be reduced to fill a volume of dimension lower than n.*

This is why an optimization criterion appropriate for identifying chaotic dissipative systems must have "volume-like" properties. A large deviation in one

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 139-144

direction can be offset by a small deviation in another direction, in a multiplicative way. Furthermore, these directions need not coincide with individual state variables - "flat" dimension of a strange attractor may be oriented arbitrarily. This excludes the use of a simple multiplicative criterion of the form $\prod_i (z_i - d_i)^2$.

The most appropriate volume measure is the determinant of the matrix of errors $z_{ij} - d_{ij}$, $i, j = 1, ..., n$ of n multi-step forecasts of system trajectories with slightly perturbed initial conditions, with i being the state variable index and j the perturbation index

$$V = det\left[z_{ij} - d_{ij}\right] = det\left[e_{ij}\right] \tag{3}$$

If the perturbations are linearly independent, they span, together with their origin, the nonperturbed initial state, a n-dimensional convex set defined by the coordinate origin and the n forecast errors of perturbed trajectories. The determinant represents the volume of this set. It is zero if the vectors $\vec{e}_j$ are all in a hyperplane. This is, in turn, the case if all the model forecasts and the real plant states differ only in directions constituting a hyperplane, and do not differ in directions from the subspace orthogonal to this hyperplane - which is all what we can expect in a chaotic attractor.

Criterion (3) is applicable only if measurement trajectories can be sorted to extract a sufficient number of n-tuples of their segments starting at n sufficiently close points.

If this is not feasible, the next best criterion is

$$V = det\left[z_{ij} - d_i\right] \tag{4}$$

representing the deviations of perturbed forecasts from a *single (nonperturbed) trajectory* of the real plant. What is actually perturbed is the initial states of the model trajectories. Each model trajectory starts at a point $\vec{z}_0 = \vec{d}_0 + \delta\vec{e}_j$, with $\vec{e}_j$ being unit vectors and δ a small number.

If (4) converges to zero with increasing distance from the forecast initial point $\vec{z}_0$, the model possesses an attractor of dimension lower than n. If this attractor is a strange attractor of the same dimension as that of the real chaotic system attractor, their shapes can be made identical by training on an extensive training set containing sufficiently many trajectories through the attractor. Insufficient correspondence between the model attractor and that of real system would cause their local interpolating hyperplanes to be different. This, in turn, would inevitably lead to a volume criterion not converging to zero for some forecasts.

However, criterion (4) may have a spurious minimum corresponding to the situation in which the model has an attractor of a lower dimension than that of the real plant - typically, a point attractor while the real plant has a strange attractor. This can be prevented by extending (4) to

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 139-144

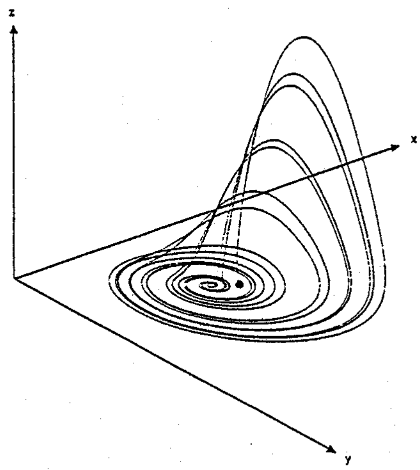

Figure 1: Roessler attractor

$$V = det\,[z_{ij} - d_i] + \sum_j \min_i \|z_{ij} - d_i\| \tag{5}$$

requiring that forecast errors are small at least for some variables.

Criterion (3) or (5) is to be minimized by some training algorithm. The horizon for multi-step forecasts is to be sufficiently long for convergence to the strange attractor, but not excessively long, for information loss by chaotic behavior not to be complete.

4 Computational experiments

The concepts of Section 3 are particularly illustrative if applied to the Roessler model described by the equations

$$\begin{aligned} \dot{x} &= y - z \\ \dot{y} &= x + ay \\ \dot{z} &= bx - cz + xz \end{aligned} \tag{6}$$

with parameter values $a = 0.38, b = 0.3, c = 4.82$. For this parameter combination, the Roessler model has a strange attractor. Moreover, with this parameter combination, the system is in the proximity of a homoclinic point (Nicolis and Prigogine [4]), in which all chaotic trajectories pass through an infinitesimally small bottleneck at the proximity of origin (Fig. 1). This configuration makes the forecast particularly difficult. The results received with neural network based multi-step identification procedure (Hrycej [2]) are given in Fig. 2. Fig. 2A shows the forecast of variable Z for 1000 time periods of length 100 ms with additive multi-step criterion. (the measurements in bold line). Fig. 2B shows the same

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 139-144

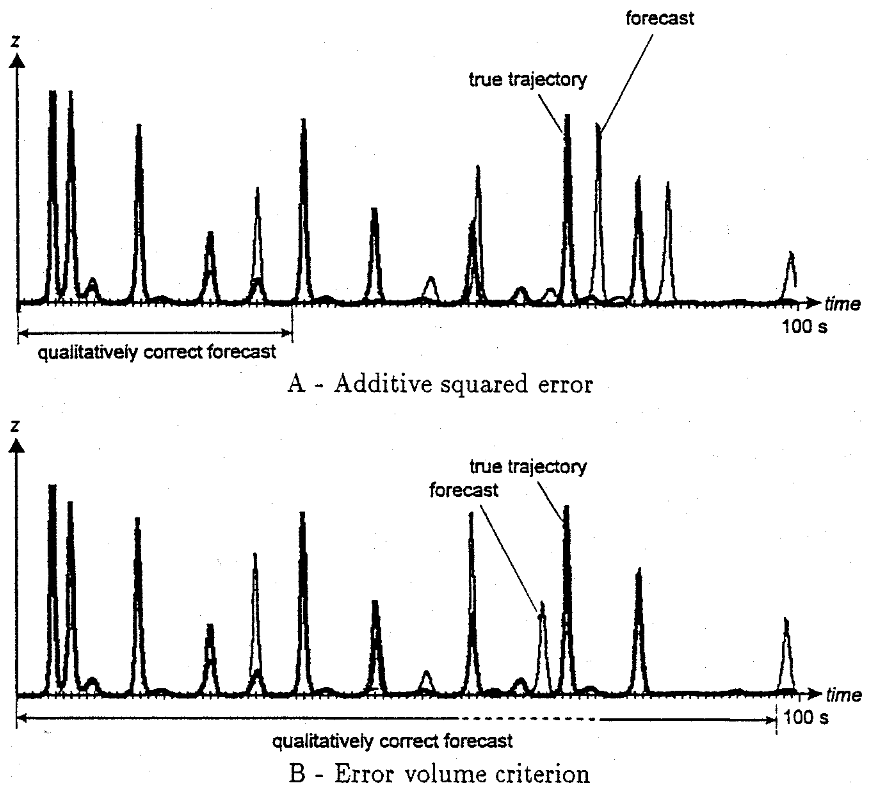

Figure 2: Roessler system identification

for criterion of type (4). Obviously, some parts of the attractor are flat in the Z-direction while others consist of peaks of chaotic height. A successful forecast should identify the flat parts of the trajectory with a large precision while the height of the peaks remains, after reaching a certain time horizon, unpredictable.

It can be seen that while additive criterion model "loses the contact with reality" after about 350 steps (site I). Error volume criterion model loses the contact only temporarily, rejoining the measurement after about 380 steps (site J) - in other words, the model fails to forecast only the peak between sites I and J. The forecast deteriorates only after about 600 steps (site K), but once more correctly identifying the peak at the 800-step (site L).

To measure the quality of the identification of attractor itself, viewed as a subset of the state space, the state space has been divided into 18018 cells, and the visits of the real system and of the model in individual cells have been counted. For each cell, the minimum of both counts has been determined. The sum of all minima, if related to the total count of iterations is a simple measure of correspondence of both attractors: the sum is equal to the iteration count if

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 139-144

both attractors are identical within the given observation grain, and is zero if the attractors are completely disjoint. In other words, it is a measure of intersection of both attractors. For additive squared error, the intersection has been only 63.87 %, while for error volume criterion, it has been 83.20 %.

5 Limitations

The arguments of Section 3 are based on the assumption of the local approximation of the attractor by a hyperplane. Of course, strange attractors are no hyperplanes. The approximation is good only as long as there are several regions that are approximately planar, such as at the bottom of the Roessler attractor of Fig. 1. Then, the "thick" dimension will be able to be forecast. However, the divergence between forecast and real state can go so far that both are in different such regions. In this case, the chaos will be complete, and even the error volume criterion cannot be expected to help. However, it is then still possible to model the attractor, that is, to look for a model whose strange attractor (as a state space subset) is a good approximation of that of the real system. A (frequently satisfied) precondition for this is that the system is structurally stable, i.e., that a small perturbation in model parameters does not change the qualitative type of the attractor. Nevertheless, the horizon of possible forecast will be substantially extended by the use of the error volume criterion.

References

[1] Åström, K.J. and Wittenmark, B., *Adaptive Control*, Addison-Wesley: Reading, MA, 1989.

[2] Hrycej, T., Plant Identification Using Temporal Sequences, *Proc. International Conference on Artificial Neural Networks '94*, Sorrento, Italy, 1994.

[3] Nguyen, D. and Widrow, B., The truck backer-upper: An example of self-learning in neural networks, *Proc. International Joint Conference on Neural Networks*, Volume 2, Washington DC, 1989, 357-363.

[4] Nicolis, G. and Prigogine, I., *Exploring Complexity*, W. H. Freeman and Company, New York, 1989.

[5] Schuster, H. G., *Deterministic Chaos*, VCH Verlagsgesellschaft, Weinheim, Germany, 1988.

[6] Werbos, P., Backpropagation and neurocontrol: a review and prospectus, *Proc. International Joint Conference on Neural Networks*, Volume 1, Washington DC, 1989, 209-216.