Modelling Nonstationary Dynamics

M.I. Széliga, P.F. Verdes, P.M. Granitto, H.A. Ceccatto [1]

Instituto de Física Rosario, Consejo Nacional de Investigaciones Científicas y Técnicas, Boulevard 27 de Febrero 210 Bis, 2000 Rosario, República Argentina

Abstract

We incorporate the use of validation data to cope with noisy records in a neural network-based method for modelling the dynamics of slowly-changing nonstationary systems. As a byproduct, we obtain a precise criterion to find the optimal value of a required internal hyperparameter. Testing these ideas on a controlled problem shows that the resulting algorithm is able to outperform previous methods in the literature, allowing a more accurate modelling of nonstationary dynamics.

Key words: Dynamical Systems, Nonstationary Time Series, Neural Networks
PACS: 05.45.Tp, 95.75.Wx

1 Introduction

Most real-world time series have some degree of nonstationarity due to external perturbations and/or changes in the internal parameters of the observed system. Because of this, in the last years an increasing effort is being devoted to devise nonlinear methods for nonstationary time series analysis [1,2].

In this work we focus on the accurate modelling of slowly-changing dynamical systems. In particular, we will use artificial neural networks (ANNs) to learn the dynamics and, at the same time, reconstructing the profile of the driving parameter responsible for the nonstationary behavior. This last problem has been already discussed in [1–3], and a preliminary investigation of the ideas explored here have been reported in [4]. The present work refines the proposed method by incorporating the use of validation points, set aside from the learning data, which is required for studying noisy time series. We assess the performance of the algorithm by applying it to synthetic data and comparing with the results in [3,4].

* Corresponding author.

2 Learning Driving Forces

Consider an observational record $D = \{x_t, t = 1, N\}$, which can be modelled in a d-dimensional pseudo-phase space according to $x_{t+1} = f(\boldsymbol{x}_t, \alpha_t) + \varepsilon_t$. Here $\boldsymbol{x}_t = (x_t, x_{t-1}, ..., x_{t-d+1})$, time t is measured in units of the lag τ between observations, and ε_t is some residual additive noise of zero mean. The parameter α_t accounts for the effects of either a slow external perturbation acting on the system or internal parameters or degrees of freedom varying on large time scales $T >> \tau$ not modelled by f. We want to reconstruct both the intrinsic dynamical function f and the nonstationary signal α from the available data D. To this end, in [4] we proposed the following algorithm:

(1) Train a feedforward ANN with architecture $(d + 1) : h : 1$, minimizing with respect to weights and biases $\boldsymbol{w}$ the normalized error

$$E_D = \frac{1}{N_p \sigma_D^2} \sum_{t=1}^{N_p} (x_{t+1} - o(\boldsymbol{x}_t, \alpha_t = 0; \boldsymbol{w}))^2 \tag{1}$$

Here $o(\bullet; \boldsymbol{w})$ is the network's output, $N_p = N - d$ the number of training patterns and σ_D^2 the data variance. In this step the input α_t corresponding to the extra neuron in the input layer is set to 0.

(2) Switch on the nonstationary perturbation α, and retrain the network minimizing the error $E = E_D + \lambda E_\alpha$ with respect to $\boldsymbol{w}$ and the unknown α_t's. Here

$$E_\alpha = \frac{1}{N_p \sigma_D^2} \sum_{t=1}^{N_p} (\alpha_t - \alpha_{t-1})^2 \tag{2}$$

is a smoothing error term introduced to penalize sudden α_t variations (we normalize it using the data variance σ_D^2). The hyperparameter λ sets the appropriate scale between E_D and E_α, as discussed below.

(3) Repeat the last step for different values of λ, and plot the E_α's obtained at the minimum of E_D in each case as a function of this relative scale. The presence of a plateau in this curve will indicate a region of weak sensitivity of α with respect to λ. In [4] we proposed and empirically proved that the curves α_t vs. t in this region are good reconstructions of the driving force.

The algorithm above described has two major drawbacks: i) For noisy time series the expected plateau is much reduced, and there is always the possibility of overfitting the ANN to the noisy dynamics. ii) The criterion for determining λ (step 3) is not very precise. Although there are small changes in α for different λ's in the region of interest, the algorithm does not indicate the optimal value λ_{opt} that leads to the best reconstruction. To cope with these problems, we propose the following refinement of the ideas discussed in [4]:

1) Split at random the data set D in training T and validation V sets, and perform step 1 above stopping the ANN training process at the minimum of the *validation* error $E_V^{\min}$ ("early-stopping" criterion). 2) Perform step 2 above, training the ANN by minimizing $E = E_T + \lambda E_\alpha$ and stopping at the new $E_V^{\min}$. 3) Repeat this last step for different values of λ, and find $\lambda_{\rm opt} = \mathrm{argmin}_\lambda E_V^{\min}(\lambda)$. This value should lead to the best reconstruction of α and, consequently, to the best modelling of the nonstationary dynamics.

There is a question in point 2 above that needs clarification: What α values should we input to the ANN for predicting points in the validation set V? Since we assumed that the driving signal was slow in comparison with the lag τ between observations, for every $x_i \in V$ it is reasonable to take $\alpha_i = (\alpha_{i+1} + \alpha_{i-1})/2$, where $x_{i\pm1}$ are the nearest neighbor points (in time) that belong to the training set T.

The intuition behind the modified algorithm is the following: i) for λ too small one will obtain rough parameter α_t variations and, consequently, large E_α, very small E_T and fairly large E_V values. That is, the α_t's will be highly correlated to the residual errors in step 1, being tuned to produce outputs of the ANN that exactly match the targets in the training set and, consequently, producing wrong inputs for the validation points. ii) For λ too large E_α will be almost zero, α_t will be nearly constant and, consequently, E_T and E_V will not vary much from the values obtained in step 1. iii) For intermediate values of λ there should be the expected plateau in E_α as before, but E_V must point to the optimal λ and the best α reconstruction. To check this, in the next section we perform a study along these lines of the controlled problem investigated in [3,4], which will also serve as a comparison test for the proposal in this work.

3 Forced Logistic Map

We consider the logistic map with additive noise: $y_{t+1} = r y_t(1 - y_t)$, $x_{t+1} = y_{t+1} + \varepsilon_{t+1}$, where ε_t is Gaussian noise with a small noise-to-signal ratio of 10^{-3}. Like in [3,4], the parameter r was slowly driven according to the law $r_t = r_0 + A\cos[(2(t/T)\exp(t/T)]$, with $r_0 = 3.8$, $A = 0.045$ and $T = 50$ (see Fig. 1), keeping the map in the interesting chaotic regime. Since here we know the exact $\alpha_t \equiv r_t$, we can assess the efficacy of the algorithm by monitoring

$$E_{\rm rec} = \frac{1}{N_p \sigma_\alpha^2} \sum_{t=1}^{N_p} (\alpha_t - r_t)^2. \tag{3}$$

We considered only $N = 100$ iterates in the data set D, taking 80 points for training and 20 points for validation purposes. We used simple 2:3:1 archi-

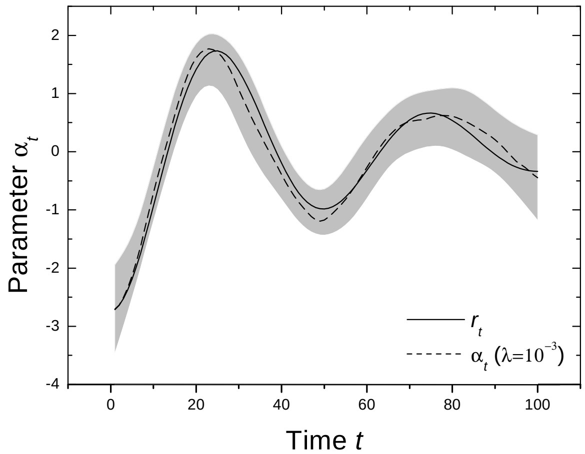

Fig. 1. Driving parameter r_t for the logistic map (full line) and its best reconstruction α_t for $\lambda = 10^{-3}$ (dashed line). The gray area indicates the $\pm 1.96\sigma$ dispersion of the best 50 independent runs of our algorithm.

tectures for the ANN and the standard backpropagation rule for error minimization. In Fig. 2 we plot the different errors' behaviors has a function of λ. For comparison, we also show the curves obtained by stopping the ANN training at the minimum of E_T, like in [4], where no validation set was used. The plotted values correspond to an average over the best 50 networks (having the smallest validation errors) in 100 different experiments, starting from the random partition of D in sets T and V (this filter outs the possibility of badly trained ANN due to random fluctuations). Let's discuss the different errors in detail: i) For both stopping criteria, $E_{\rm rec}$ –only available in this controlled case– shows a minimum at $\lambda_{\rm opt} = 10^{-3}$. However, the reconstruction based on E_V gives $E_{\rm rec} \simeq 0.08$, which is nearly half the value corresponding to the use of E_T ($E_{\rm rec} \simeq 0.15$). Moreover, the modelling accuracy improves almost 40%, from $E_V \simeq 2.2 \times 10^{-5}$ to $E_V \simeq 1.4 \times 10^{-5}$. On the other hand, there are fairly good α reconstructions having $E_{\rm rec} < 0.2$ for $10^{-4} \lesssim \lambda \lesssim 1$. ii) In agreement with this, E_α has the expected plateau in approximately the same region, as advocated in [4]. Note, however, that the criterion used there could not identify $\lambda_{\rm opt} = 10^{-3}$ (one might be tempted to choose instead $\lambda_{\rm opt} = 1$ or 10). iii) The behavior of E_V as a function of λ points clearly to the correct $\lambda_{\rm opt}$. Furthermore, due to the small level of noise in this problem, E_T also shows a minimum at this value. However, for higher noise levels overfitting problems will force the use of a validation set V, which will lead to a larger gain with respect to the algorithm proposed in [4]. Finally, in Fig. 1 we show the mean α curve obtained at $\lambda_{\rm opt} = 10^{-3}$ by averaging the best 50 reconstructions. For this curve $E_{\rm rec} \simeq 0.029$, which should be compared with $E_{\rm rec} \simeq 0.035$ obtained with the method in [4], and $E_{\rm rec} \simeq 0.045$ obtained with the independent algorithm proposed in [3].

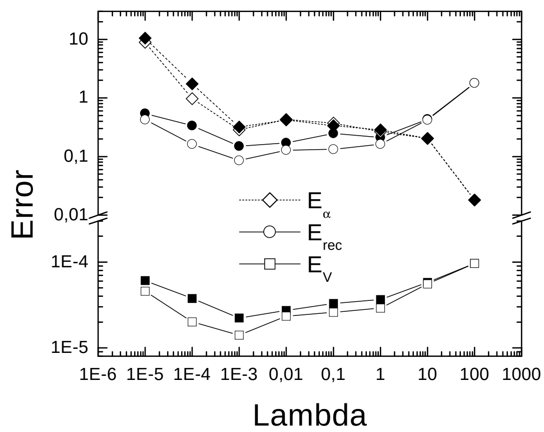

Fig. 2. Different errors at the minimum of E_V (open symbols) and E_T (full symbols) as a function of λ.

4 Summary and Conclusions

We discussed a refinement of a previously-proposed ANN method for the accurate modelling of nonstationary dynamical systems. The algorithm incorporates the use of validation data to treat noisy records and, as a byproduct, leads to a precise criterion for determining the optimal value of an internal hyperparameter. Using synthetic data from a forced logistic map, we showed that this modified algorithm is able to outperform similar methods in the literature, allowing a more accurate modelling of the nonstationary dynamics. In principle, the algorithm here proposed could be extended to trace the simultaneous variation of several parameters, something that has not been fully explored in the literature. Furthermore, these ideas can also be applied in general regression problems where the underlying phenomenon is changing in time. Work in these directions is in progress.

References

[1] M. Casdagli, Physica D 108, 12 (1997), and references therein.

[2] T. Schreiber, Phys. Rep. 308, 1 (1999).

[3] P.F. Verdes, P.M. Granitto, H.D. Navone and H.A. Ceccatto, Phys. Rev. Lett. 87, 124101 (2001).

[4] M.I. Széliga, P.F. Verdes, P.M. Granitto and H.A. Ceccatto, Proceedings of the VII Brazilian Symposium on Artificial Neural Networks (SBRN) (Porto de Galinhas, Recife, November 11-14, 2002), pp. 101-105.