ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 311-316

A multi-step-ahead prediction method based on local dynamic properties

Gianluca Bontempi, Mauro Birattari

Iridia - Université Libre de Bruxelles
Av. F. Roosevelt 50, CP 194/6
1050 Bruxelles - Belgium
http://iridia.ulb.ac.be
e-mail: {gbonte,mbiro}@ulb.ac.be

Abstract. The task of forecasting a time series over a long horizon is commonly tackled by iterating one-step-ahead predictors. Despite the popularity that this approach gained in the prediction community, its design is still plagued by a number of important unresolved issues, the most important being the accumulation of prediction errors. We introduce a local method to learn one-step-ahead predictors with the aim of reducing the propagation of errors during the iteration. For each prediction, our method selects the structure of the local approximator using, in a local version, well-known results of dynamic system theory Experimental results on two time series from the Santa Fe competition show that the technique is competitive with state-of-the-art forecasting methods.

1. Introduction

The paper presents a local method for long horizon forecasting based on the iteration of a one-step-ahead predictor. Local approaches do not estimate a global model of the dynamic system underlying the time series but defer the processing of data until a prediction is explicitly requested [6]. A database of observed values is stored in memory, the dataset is embedded in a state space and the prediction is derived from an interpolation based on a neighborhood of the current state (*locally weighted regression*). A key issue in local learning is *model selection*, that is the procedure which aims to select the local model structure (e.g. size of the neighborhood, degree of the local fitting) which is expected to have the best prediction accuracy given a set of historical observations. This issue is still more relevant if we want to avoid the accumulation of prediction errors during the iteration of a one-step-ahead predictor.

In previous works [4, 3], the authors proposed the adoption of linear statistical procedures (e.g. the PRESS leave-one-out statistic) to assess different local configurations and to select the most accurate one. Here, we use a local description of the dynamic system underlying the time series in order to solve the model selection problem. The idea is to replace an assessment criterion based

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 311-316

on cross-validation with a criterion based on the local topological properties of the system.

In particular, we consider here the notion of *conservative/dissipative* dynamical system [7]. This notion which is global for a linear system can have a local interpretation in a nonlinear system. When the volume of a local region surrounding a trajectory remains constant during the evolution of the dynamical system, the system can be said *locally conservative*. On the contrary, when the volume surrounding a trajectory contracts, it can be defined as *locally dissipative*.

A local model of a dynamical system returns, together with a prediction, also a description of the local behavior of the system, i.e. whether the system is dissipative or conservative. As a consequence, it is possible to check if the prediction returned by the local approximator is consistent with the expected behavior. The main idea of the paper is to use, as assessment criterion of a given local model, the consistency of the model prediction with the expected change of volume of the local area surrounding the current trajectory.

The idea of using dynamic parameters (e.g. the Lyapunov exponents) to support time series forecasting is not new in literature [5]. What is innovative here is the adoption of a theoretical result of system theory to implement the model selection step of a local learning algorithm.

The experimental session will show that this local iterated method is competitive with state-of-the-art multi-step ahead prediction techniques, by keeping small the accumulation error on long forecasting horizons.

2. Local iterated time series prediction

A time series is a sequence of measurements $\varphi(t)$ of an observable φ at equal time intervals. The Takens theorem implies that for a wide class of deterministic systems, there exists a *diffeomorphism* (one-to-one differential mapping) between a finite window of the time series $\{\varphi(t), \varphi(t-1), \ldots, \varphi(t-n+1)\}$ (*lag vector*) and the state of the dynamic system underlying the series. This means that in theory it exists a multi-input single-output mapping (*delay coordinate embedding*) $f : R^n \to R$ so that:

$$\varphi(t+1) = f(\varphi(t), \varphi(t-1), \ldots, \varphi(t-n+1)) \tag{1}$$

where n (*dimension*) is the number of past values taken into consideration. This formulation returns a state space description, where in the n dimensional state space the time series evolution is a trajectory, and each point represents a temporal pattern of length n.

A model of the mapping (1) can be used for two objectives: *one-step* prediction and *iterated* prediction. In the first case, the n previous values of the series are assumed to be available and the problem is equivalent to a problem of function estimation. In the case of iterated prediction, the predicted output is fed back as an input to the following prediction. Hence, the inputs consist of

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 311-316

predicted values as opposed to actual observations of the original time series. A prediction iterated for k times returns a *k-step-ahead* forecasting.

We propose a locally weighted regression method to estimate a one-step-ahead predictor for performing iterated prediction. Typically the data analyst who adopts a local regression approach, has to take a set of decisions related to the model structure (e.g. the number of neighbors, the kernel function, the parametric family, the distance metric). In local learning literature several methods have been proposed to automatically select the adequate configuration [2]. In previous work [4, 3] we studied the PRESS statistic which is a simple, well-founded and economical result of linear statistical analysis to perform *leave-one-out* cross-validation and to assess the performance in generalization of local linear models.

Here, we propose a modelselection criterion based on the local dynamics of the system underlying the observations. This criterion is used to select the best structure of a one-step-ahead estimator with the aim of capturing the long term dynamics underlying the available set of observations. In particular, the criterion will be used to select query-by-query (i.e. at each time step) the best number of neighbors by keeping fixed the shapeof the regression kernel (tricube) and the distance metric (euclidean).

In the following sections we will first introduce some basic results for discrete time dynamic systems and then we will discuss how these results can be used to formulate a model selection criterion.

2.1. Volume contraction in the state space

Consider a generic n-dimensional discrete time dynamic system described by

$$\mathbf{x}(t+1) = F(\mathbf{x}(t)) \tag{2}$$

where the $[n \times 1]$ vector $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ denotes the state at time t.

Let $\Omega(t)$ represent a small rectangular region at time t in the neighborhood of $\mathbf{x}(t)$, whose i^{th}, $i = 1, \ldots, n$, side is the vector $\boldsymbol{\delta}_i = [0 \ldots \delta_i \ldots 0]$. The volume of $\Omega(t)$ is given by $V(t) = \prod_{i=1}^{n} \delta_i$. Let

$$\Omega(t+1) = \{\mathbf{x}(t+1) : \ \mathbf{x}(t) \in \Omega(t) \ \& \ \mathbf{x}(t+1) = F(\mathbf{x}(t))\} \tag{3}$$

be the region defined at time $t+1$ by the set of trajectories of (2) passing at time t through $\Omega(t)$.

It is known in literature [1] that the volume of this n-dimensional region is given by $V(t+1) = V(t) * \det J$ where J denotes the Jacobian matrix of the system (2). Therefore, for a n-dimensional discrete time system (2) the local volume contracts by a factor $\det J$ at each time step.

Let us now consider the linearization $\mathbf{x}(t+1) = A\mathbf{x}(t)$ of the system (2) at time t in the neighborhood of the state $\mathbf{x}(t)$, where A is a $[n \times n]$ matrix of linear coefficients. If we assume that the nonlinear system (2) is well described by the linearization in a neighborhood of $\mathbf{x}(t)$, an approximation of the contraction of the local volume in the neighborhood of $\mathbf{x}(t)$ is given by $\det A$.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 311-316

2.2. A contraction-based criterion of consistency

Our local approach to multi-step-ahead prediction exploits the dynamic properties described in the previous section in order to check the goodness of a local approximation in the neighborhood of the current state.

Given a local predictor, we compute two measures, a *state space* measure and an *estimation* measure, to evaluate the contraction rate of a local volume around the current trajectory .Hence, the consistency between these two measures is checkedto assess the goodness of the local predictor itself. The assumption is that, if the local model is a reliable description of the local dynamics, these twomeasures should be consistent. A large deviation between these measures should be considered as a violation of the hypothesis of locality.

Consider an input/output embedded model (1) of an observed time series φ, whose available realization is described in terms of an input/output dataset D_N. Assume that the vector $\boldsymbol{\varphi}(t) = [\varphi(t), \ldots, \varphi(t-n+1)]^T$ is an accurate reconstruction of the state $\mathbf{x}(t)$.

Let us denote by $\mathcal{L}_k$, $k = k_m, \ldots, k_M$, a local linear description of the dynamics (1) which fits a linear model to the k nearest neighbors of $\boldsymbol{\varphi}(t)$ in D_N. The linearization of the dynamics f on the basis of the historical data is given by

$$\hat{\varphi}_k(t+1) = \hat{a}_0 + \hat{a}_1\varphi(t) + \hat{a}_2\varphi(t-1) + \cdots + \hat{a}_n\varphi(t-n+1)) \tag{4}$$

where $\hat{\varphi}_k(t+1)$ is the one-step-ahead prediction of the local model $\mathcal{L}_k$ and $\hat{a}_i$, $i = 0, \ldots, n$ is the set of parameters estimated by a Locally Weighted Regression procedure.

Our goal is to assess the predictive accuracy of on the basis of the available data. The idea is that for a generic local model $\mathcal{L}_k$ we can derive two independent estimates of the contraction of the state space and that a good criterion to assess the accuracy of $\mathcal{L}_k$ is to check the consistency between these two measures. Let us see the two measures in detail:

State-space measure M^{ss}: The local model (4) can be written as

$$\begin{cases} x_1(t+1) & = \hat{a}_0 + \hat{a}_1 x_1(t) + \hat{a}_2 x_2(t) + \cdots + \hat{a}_n x_n(t) \\ \cdots & \\ x_n(t+1) & = x_{n-1}(t) \end{cases} \tag{5}$$

From Section 2.1. a measure of the contraction rate of the dynamics (4) is the determinant of the Jacobian of (5) which, in this case, equals $\hat{a}_n$. Hence, we define the quantity $\hat{a}_n$ as the *state-space* measure M_k^{ss}.

Estimation measure M^{es}: Let us define with $d(t)$ the distance of the reconstructed state $\boldsymbol{\varphi}(t)$ from the nearest neighbor in the training set D_N. Let $d_k(t+1)$ be the analogous distance of the predicted state

$$\hat{\boldsymbol{\varphi}}_k(t+1) = [\hat{\varphi}_k(t+1), \varphi(t), \ldots, \varphi(t-n+2)]^T \tag{6}$$

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 311-316

from the nearest neighbor in D_N, where the prediction $\hat{\varphi}_k(t+1)$ is given by (4). The quantity

$$M_k^{\mathrm{es}} = \frac{d_k^n(t+1)}{d^n(t)} \tag{7}$$

is the estimation measure of the contraction of the state space

We define with

$$C_k = \frac{1}{(M_k^{\mathrm{es}} - M_k^{\mathrm{ss}})^2} \tag{8}$$

the *consistency* of the two contraction measures for a generic local model $\mathcal{L}_k$ fitted on the k nearest neighbors.

Our approach consists in considering the measure (8) as an assessment measure of the local model $\mathcal{L}_k$. The idea is that the more reliable is the local model, the larger is the consistency between the state space measure and the estimation measure. As a consequence, the measure C_k can be used to select among a class $\mathcal{L}_k$, $k = k_m, \ldots, k_M$, of local linear models, the one which is expected to return the best prediction.

Our local learning procedure can be summarized as follows:

1. The one-step-ahead predictor is a local model of the embedding.
2. The k-step-ahead prediction is performed by iterating a one-step-ahead estimator.
3. The local model is selected in a space of alternative model configurations $\mathcal{L}_k$, each characterized by a different number of neighbors and assessed by a value C_k of the consistency criterion.
4. The local model with the highest consistency is selected.

3. Experiments and final considerations

The local learning iterated approach has been applied both to the prediction of a *real-world* data set (A) and to a computer generated time series (D) from the *Santa Fe Time Series Prediction and Analysis Competition*. In particular, this section evaluates our query-by-query selection of the number of neighbors based on the consistency criterion (8). The number of neighbors is limited to range from 4 to 12. We adopt for the series A an embedding model having the same dimension $m = 16$ proposed in [8] and for the series D an embedding model with $m = 20$ as reported in [10].

Table1 (left) compares the NMSE (Normalized Mean Squared Error) on the A test set of the local predictor based on the consistency criterion (CC) with the local method based on cross-validation (Press) proposed in [4] and with the performance statistics reported by Sauer [8] and Wan [9]. Table1 (right) compares the RMSE (Root Mean Squared Error) on the series D of the

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 311-316

Test A	CC	Press	Sauer	Wan
1-100	0.028	0.029	0.077	0.055
1180-1280	0.051	0.028	0.174	0.065
2870-2970	0.255	0.003	0.183	0.487
3000-3100	0.039	0.030	0.006	0.023

Test D	CC	Press	Zhang
0-24	0.0459	0.0553	0.0665
100-124	0.0524	0.0244	0.0616
200-224	0.2572	0.1073	0.1475
300-324	0.0347	0.0240	0.0541

Table 1: On the left: NMSE of the predictions for time series A. On the right: RMSE of the predictions for time series D.

local predictor based on the consistency criterion (CC) with the local method based on cross-validation (Press) proposed in [4] and with the results of [10].

The experiments show that for long horizons prediction the idea of using dynamic measures, like the consistency criterion, can be competitive with state-of-the-art prediction architectures and with methods based on conventional assessment techniques, like cross-validation.

References

[1] V. I. Arnol'd. *Ordinary differential equations* Springer-Verlag, 1984.

[2] C. G. Atkeson, A. W. Moore, and S. Schaal. Locally weighted learning. *Artificial Intelligence Review*, 11(1–5):11–73, 1997.

[3] G. Bontempi. *Local Learning Techniques for Modeling, Prediction and Control.* PhD thesis, IRIDIA- Université Libre de Bruxelles, 1999.

[4] G. Bontempi, M. Birattari, and H. Bersini. Local learning for iterated time-series prediction. In I. Bratko and S. Dzeroski, editors, *Machine Learning: Proceedings of the Sixteenth International Conference*, pages 32–38, San Francisco, CA, 1999. Morgan Kaufmann Publishers.

[5] U. Dressler and J. D. Farmer. Generalized lyapunov exponents corresponding to higher derivatives. *Physica D*, 59:365–377, 1992.

[6] J. D. Farmer and J. J. Sidorowich. Predicting chaotic time series. *Physical Review Letters*, 8(59):845–848, 1987.

[7] G. Nicolis and I. Prigogine. *Self Organisation in Nonequilibrium Systems.* Wiley, New York, 1977.

[8] T. Sauer. Time series prediction by using delay coordinate embedding. In A. S. Weigend and N. A. Gershenfeld, editors, *Time Series Prediction: forecasting the future and understanding the past*, pages 175–193. Addison Wesley, Harlow, UK, 1994.

[9] E. A. Wan. Time series prediction using a connectionist network with internal delay lines. In A. S. Weigend and N. A. Gershenfeld, editors, *Time Series Prediction: forecasting the future and understanding the past*, pages 195–217. Addison Wesley, Harlow, UK, 1994.

[10] X. Zhang and J. Hutchinson. Simple architectures on fast machines: practical issues in nonlinear time series prediction. In A. S. Weigend and N. A. Gershenfeld, editors, *Time Series Prediction: forecasting the future and understanding the past*, pages 219–241. Addison Wesley, 1994.