Chaos, Solitons and Fractals 94 (2017) 8–15

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Regularized forecasting of chaotic dynamical systems

Erik M. Bollt*

Department of Mathematics, Clarkson University, Potsdam, New York 13699, USA

ARTICLE INFO

Article history:
Received 27 January 2016
Revised 9 October 2016
Accepted 11 October 2016

Keywords:
Chaos
Forecasting
Inverse problems
Tikhonov regularization
Time delay embedding

ABSTRACT

While local models of dynamical systems have been highly successful in terms of using extensive data sets observing even a chaotic dynamical system to produce useful forecasts, there is a typical problem as follows. Specifically, with k-near neighbors, kNN method, local observations occur due to recurrences in a chaotic system, and this allows for local models to be built by regression to low dimensional polynomial approximations of the underlying system estimating a Taylor series. This has been a popular approach, particularly in context of scalar data observations which have been represented by time-delay embedding methods. However such local models can generally allow for spatial discontinuities of forecasts when considered globally, meaning jumps in predictions because the collected near neighbors vary from point to point. The source of these discontinuities is generally that the set of near neighbors varies discontinuously with respect to the position of the sample point, and so therefore does the model built from the near neighbors. It is possible to utilize local information inferred from near neighbors as usual but at the same time to impose a degree of regularity on a global scale. We present here a new global perspective extending the general local modeling concept. In so doing, then we proceed to show how this perspective allows us to impose prior presumed regularity into the model, by involving the Tikhonov regularity theory, since this classic perspective of optimization in ill-posed problems naturally balances fitting an objective with some prior assumed form of the result, such as continuity or derivative regularity for example. This all reduces to matrix manipulations which we demonstrate on a simple data set, with the implication that it may find much broader context.

1. Introduction – local models, local forecasting

Forecasting chaotic dynamical systems, from measured data, is a topic that has seen a great deal of activity, at least for the last thirty years, facilitated by the time-delay embedding methods. In the time-delay embedding literature, forecasting from observed states, embedding the states and then fitting local models based on regression to the behavior of k-near neighbors (kNN) was put forward and somewhat matured by the mid1990s, [4–10]. So, in [13] we discuss the role of local models in model selection as it relates to spatial scale, and some of that is reviewed here. In particular we have been interested in how local modeling [11], and see also [12], relates to local polynomial models.

Specifically note that local models are built for the transformation based on observing the orbits of near neighbors, and hoping that there are a lot of sampled orbit segments due to recurrence and a long orbit sample, then in any small neighborhood there would be many samples. The “k” in kNN means to collect those k points from the data set that are closest to the forecast point. If there is noise, or otherwise, model error, then a degree of smoothing is implied by a least squares estimation of the local transformation. While a higher ordered model will tend to well fit more terms in a Taylor polynomial estimation of the local model, there are inherently many more parameters to be fitted when using a high degree polynomial, and so a much larger k would be required, and hence correspondingly the neighborhood would be larger. So as discussed in [13], there is a balancing between fine scale, data density, and smoothing when performing local modeling by least squares regression alone; here we add to this discussion that regularity can be emphasized directly by using Tikhonov regularization concepts derived from convex optimization theory, [18,19] and also found in advanced matrix analysis, [16], to find locally useful forecasts, which also have good global regularity properties, and with less data than perhaps a kNN method on its own.

When a global model is not be available to forecast evolution of a given point, a useful forecast may still be possible by observing the evolution of k nearby points using regression to appropriately “average” between them. In terms of a polynomial basis, this approach develops a least squares regression of the first few terms of a Taylor's series for the unknown global model, in the neighborhood of the point to be forecast, using the k neighbors as data.

* Corresponding author.
E-mail address: bolltem@clarkson.edu

http://dx.doi.org/10.1016/j.chaos.2016.10.007

Enough near neighbors must be chosen to allow for the minimal fitting of the polynomial model, and furthermore, somewhat more than the minimal number of points should be chosen to allow for some degree of smoothing. This is akin to the familiar statistical issues faced when fitting a line to noisy data; more than 2 points should be used to confidently specify the line. Scale of the model is a major issue: there are competing demands between local truncation error that push toward small neighborhoods, but smoothing and confidence push toward larger k, leading to larger neighborhoods when using finite data sets. This trade-off was the topic of [13]. In this paper, we furthermore address an important issue overlooked in all previous studies on kNN local modeling, which is that if a model is developed for each neighborhood based on near neighbors, then since two nearby points may have a different set of near neighbors, this leads to lack of smoothness (regularity) of the forecasts. We address this problem here, by expanding our previous work to include regularization by utilizing concepts of Tikhonov regularization theory.

Consider a dynamical system,

$$\begin{aligned} T &: M \to M, \\ y_{n+1} &= T(y_n). \end{aligned} \tag{1.1}$$

Let $y \in M \subset \mathbb{R}^d$. Assume that from the nonlinear dynamical system, we have a large collection of observed iterations as an orbit of Eq. (1.1), $\{y_i\}_{i=0}^N$, such that $y_{i+1} = T(y_i)$. Here, T will stand for a discrete time mapping as the transformation, throughout this paper, and note that if we have a continuous time process, then the discrete time mapping may come either by Poincare' section, or by time delay embedding, of a flow. For the modeling discussion below, these actually need not be a single orbit, but for the regularity discussion to follow, it is best if we include that assumption now. Furthermore assume that there is uniformly enough regularity of T so that there exists a Taylor's expansion through order-K, which we will exploit in the next Section 2. The standard discussion of local modeling is to put forward that these local polynomials can be estimated by nearby sampled points and their images, generally by a regression method. However, we also generally expect that these local models will vary continuously, or vary continuously with respect to higher ordered derivatives of T, with respect to variations in the sample point. We will show here that this desirable and physically expected property can be emphasized with Tikhonov regularity theory.

Now the idea is that for any point w as an initial condition that we may wish to forecast but that may not be amongst the observed orbit values, $w \notin \{y_i\}_{i=0}^N$, we proceed with local models built from first collecting near neighbors to w, amongst the data. A standard way to forecast a dynamical system, when presented with many previous states, is to collect ''k-near'' neighbors (kNN) in the phase space, and in some manner, average, regress, or otherwise associate the current forecast to those previous forecasts. The simplest version of these associations would be the method of analogues [9] from classical weather forecasting, namely forecasts are identified with the most same measured state. From [11–13], we review local forecasting in terms of local polynomial models.

Note that perhaps we may either estimate a discrete time map $T(y)$ from many observations as just stated, with the hope that there is low dimensionality, or a popularly common scenario is that we will only observe a single scalar time series, measured from a vector valued, y and the time delay embedding representation will be used. That is, a time-series from a "chaotic" dynamical system allows a data-only analysis by embedding attractor reconstruction, [1,2,4,6,7,10]. Recall that if an autonomous dynamical system,

$$\dot{x} = F(x), x(t) \in \mathbb{R}^n, \text{ and } x(t_0) = x_0, \tag{1.2}$$

has an invariant attractor A then an experimentalist who does not know the underlying global model Eq. (1.2) may not even know which are the correct variables to measure. Generally, a single-data channel can be considered to be a scalar measurement function $h[x(t)] : \mathbb{R}^n \to \mathbb{R}$. Given a set of measurements $\{h[x(t_i)]\}_{i=0}^N$, with uniformly spaced time samples t_i, the time-delay embedding is a vector,

$$\boldsymbol{y}(t) =< h[x(t)], h[x(t-\tau)], h[x(t-2\tau)], \ldots, h[x(t-d\tau)] >, \tag{1.3}$$

and one generally chooses τ to be some multiple of the sampling rate $\Delta t = t_{i+1} - t_i$. Takens proved [3] that, for topologically generic measurement function h, if the attractor A is a smooth m-dimensional manifold, then if one chooses the delay dimension to be $d \geq 2m+1$, then Eq. (1.3) is an embedding, meaning there exists a one-to-one function $G : A \to \mathbb{R}^d$, and G is a diffeomorphism. Sauer, *et. al* [8] proved an extension to allow for nonsmooth A, and even fractal A. To reconstruct the attractor, both of these results assume that the data is clean, and the data set is arbitrarily long. Neither assumption is physically realizable, but nonetheless, time-delay reconstruction has found many applications to nonlinear modeling and to prediction. See [1,2,4,6,9,10].

Local linear regression of the observed evolution of k-nearest neighbors $\{\boldsymbol{y}_j(t)\}_{j=1}^k$, to their images $\{\boldsymbol{y}_j(t+\tau)\}_{j=1}^k$, has emerged as the most popular method to predict "the next $y(t)$." The idea is that a Taylor's series of the (unknown) function f_τ, which evolves (flows) initial conditions $y(t)$, according to the differential equation, Eq. (1.2), is well approximated by the linear truncation, *if the near neighbors are "near enough."* Error analysis, such as that found in [10], is based on this local-truncation error, and therefore considers the Luyapunov exponents. There is naturally a conflict of demands since on the one hand, a) small local truncation error demands that neighborhoods be small, and therefore k must not be chosen too large, using a fixed (linear) model, but on the other hand, b) statistical fluctuations demand that k be chosen large enough to infer a degree of smoothing. The problem we study here is that it is well know that those points which are the near neighbors to any given sample point may not vary continuously with position in space. So the predictions likewise may vary discontinuously. Therefore we have developed a perspective here to emphasize that regularity is a desirable property. In many ways, this work should be considered as analogous to the standard local forecasting, but simply an enhanced alternative version. The emphasized regularity of forecasts therefore improves plausibility of forecasts in that there will be fewer jumps between forecasts of nearby initial conditions due simply to the artificial reason that the near neighbors set may differ.

2. Basis for local polynomial regression

Assuming that the transformation T has enough regularity to justify a Taylor polynomial at each point w, to the degree sought. For example, a local affine model of T at w, $T|_w$ is,

$$y = T_0 + DT \cdot h, \tag{2.1}$$

regressed over k-nearest neighbors of w, $\{y_{k_j}\}_{j=1}^k \subset \{y_i\}_{i=0}^N$, where DT is related to the Jacobian derivative in a neighborhood of w and $h = w - y$. This may be thought of as a local truncation of a Taylor's series. We index the k-nearest points to w by k_j, ordered $k_1 < k_2 < .. < k_k$ monotonically with respect to distance from w, assuming an underlying metric space. For a "good fit," just as realized by any Taylor polynomial, fit is better if h is small. So we would demand that $\{y_i\}_{i=0}^N$ "fills" the space adequately so that for any w we are likely to select that the k-nearest data points will be "close-enough" for a good estimate. A sufficient condition for a long orbit