Local Polynomial Method for Ensemble Forecast of Time Series

Satish Kumar Regonda[1]
Department of Civil Environmental and Architectural Engineering, University of Colorado, Boulder, CO
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO

Balaji Rajagopalan
Department of Civil Environmental and Architectural Engineering, University of Colorado, Boulder, CO
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO

Upmanu Lall
Department of Earth and Environmental Engineering, Columbia University, New York, NY

Martyn Clark
Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO

Young-Il Moon
Department of Civil Engineering, University of Seoul, Seoul, Korea

Abstract: We present a nonparametric approach based on local polynomial regression for ensemble forecast of time series. The state space is first reconstructed by embedding the univariate time series of the response variable in a space of dimension (D) with a delay time (τ). To obtain a forecast from a given time point t, three steps are involved: (i) the current state of the system is mapped on to the state space, known as the feature vector, (ii) a small number ($K = \alpha * n$, α = fraction (0,1] of the data, n =data length) of neighbors (and their future evolution) to the feature vector are identified in the state space, and (iii) a polynomial of order p is fitted to the identified neighbors, which is then used for prediction. A suite of parameter combinations (D, τ, α, p) is selected based on an objective criterion, called the Generalized Cross Validation (GCV). All of the selected parameter combinations are then used to issue a T-step iterated forecast starting from the current time t, thus generating an ensemble forecast which can be used to obtain the forecast probability density function (PDF). The ensemble approach improves upon the traditional method of providing a single mean forecast by providing the forecast uncertainty. Further, for short noisy data it can provide better forecasts. We demonstrate the utility of this approach on two synthetic (Henon and Lorenz attractors) and two real data sets (Great Salt Lake bi-weekly volume and NINO3 index). This framework can also be used to forecast a vector of response variables based on a vector of predictors.

[1] CADSWES,
UCB 421, University of Colorado
Boulder, CO-80309
e-mail: regonda@colorado.edu