A COMPARISON BETWEEN KRIGING AND RADIAL BASIS FUNCTION NETWORKS FOR NONLINEAR PREDICTION

J.-P. Costa, L. Pronzato and E. Thierry

Laboratoire I3S, CNRS-UNSA,
Les Algorithmes/Bât. Euclide,
2000 route des Lucioles, Sophia-Antipolis,
06410 Biot, FRANCE
{*costa, pronzato, et*}@*i3s.unice.fr*

ABSTRACT

Predictions by Kriging and radial basis function (RBF) networks with gaussian Kernels are compared. Kriging is a semi–parametric approach that does not rely on any specific model structure, which makes it much more flexible than approaches based on parametric behavioural models. On the other hand, accurate predictions are obtained for short training sequences, which is not the case for nonparametric prediction methods based on neural networks. Examples are presented to illustrate the effectiveness of the method.

1. INTRODUCTION

We consider the situation where the relationship beetween the input and output sequences $\{x_k\}$ and $\{y_k\}$ of a SISO system S is dominated by nonlinear characteristics. When using a parametric nonlinear model, first one has to choose a suitable model structure [1]. Second, once a structure has been chosen, one has to estimate its parameters. Traditional parametric representations for nonlinear unknown structures are the Volterra, Wiener or NARMAX (Nonlinear AutoRegressive Moving Average model with eXogenous inputs) models. They generally involve a very large number of unknown parameters [2, 3], so that one has to collect a large amount of data (training sequence) to be able to estimate these parameters.

Linear prediction by Kriging can be considered as a general statistical tool for modeling spatial observations, with or without observation errors [4, 5, 6]. Kriging is based on a semi–parametric model which allows much more flexibility than parametric models, since no specific model structure is used : the model contains a linear regression part (parametric) and a non–parametric part considered as the realization of a random process. Assuming that the process is Gaussian, the parameters of its covariance matrix can be estimated by maximum likelihood. It happens that the choice of the linear regression has little influence on the predictive properties of the model see, e.g., [7]. The memory length m_x of the input thus corresponds to the only important prior choice concerning the unknown structure, and a prior over–estimation of m_x only results in heavier computations.

An alternative approach to parametric, or semi–parametric, models is to use a nonparametric model. Nonparametric approaches based on neural networs, for instance Radial Basis Function (RBF) networks, are becoming more and more popular. A reason is that, in principle, neural networks can approximate any continuous behaviour with arbitrary precision [8]. However, besides the problem of having to choose the structure of the network, remains the major issue of choosing a suitable (and long enough) training sequence [9].

If the linear part of the Kriging model is reduced to a constant term, Kriging corresponds to a RBF network with gaussian kernels, with the centers of the kernels corresponding to the inputs in the training data set, see section 3.2.

2. PREDICTION BY KRIGING

Let $\{\mathbf{x}_k\}$ and $\{y_k\}$ be the input and output sequences of a system S, which are observed for $k = 1, \ldots, n$. Prediction by Kriging consists in interpolating these data by constructing the best linear unbiased predictor at new unsampled values of $\mathbf{x}$, that is $\mathbf{x}_{n+1}, \mathbf{x}_{n+2}, \ldots$

2.1. Prediction

When there are no observation errors, the observations y_k are modelled by

$$y_k = \mathbf{f}^T(\mathbf{x}_k)\beta + z(\mathbf{x}_k)\,,$$

where $\mathbf{x}_k$ denotes the vector formed by lagged scalar inputs

$$\mathbf{x}_k = (x_k, x_{k-1}, \ldots, x_{k-m_x+1})^T\,,$$