ELSEVIER

Computational Statistics & Data Analysis 37 (2001) 209–217

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

Local linear regression for estimating time series data

Quinton J. Nottingham *, Deborah F. Cook

Department of Management Science and Information Technology, Virginia Polytechnic Institute and State University, 1007 Pamplin Hall, Blacksburg, VA 24061-0235, USA

Received 1 December 2000; received in revised form 2 January 2001; accepted 20 February 2001

Abstract

Predicting future performance based on past performance history is a task often undertaken by business process managers. Various statistical and analytical techniques, such as time series and neural network modeling, are available. However, these techniques require the availability of a long time series for the development of a predictive model. Local linear regression (LLR) is an additional nonparametric statistical method that can be used to estimate a time series response variable. The LLR technique does not require a long time series for the development of a predictive model. In fact, the LLR technique can be utilized for prediction once three data points have been collected from the business process. In this work, LLR was evaluated as a tool for predicting future values of process parameters based on historical values. If successful, the LLR technique could be applied in start-up conditions or used as an alternative in some situations to time series modeling. The LLR procedure outperformed traditional time series techniques for the example stationary data sets and had comparable results to the ARIMA model for the example seasonal data set. In addition the LLR technique uses the data that is currently available from a process as its basis for prediction, thus providing a dynamic predictive technique that can continue to function in the presence of process changes.

Keywords: ARIMA; Nonparametric; Predictions

1. Introduction

Business process managers are often interested in predicting the future value of important process parameters based on historical values of those parameters. A mul-

* Corresponding author. Tel.: +1-540-231-7843.
E-mail address: notti@vt.edu (Q. J. Nottingham).

0167-9473/01/$ - see front matter
PII: S 0167-9473(01)00006-8

titude of methods is available for making such predictions, including statistical and analytical techniques. Statistical techniques including time series, regression, and nonparametric regression have been used for developing predictive models (Box and Jenkins, 1976; Härdle, 1990). Analytical neural network techniques have also been used extensively for prediction (Chiu et al., 1995; Cook and Chiu, 1997; Hoskins and Himmelblau, 1988; Saad et al., 1998; Gao et al., 1997).

These methods require the availability of a long time series that is used for model development and model validation before parameter estimates can be obtained and predictions can be made. The ARIMA approach for time series predictive model development is theoretically and statistically appealing. However, the complexity of these models has often hindered their widespread adoption as a forecasting tool in organizations (Makridakis et al., 1983). The development of neural network models requires large training data sets containing examples of multiple process operating conditions. Extensive network design, training and testing are also typically required for model development. Additionally the performance of the time series and neural network models must be regularly monitored to assure that the model continues to represent the process. If a process change occurs, development of a new model is required. When a process change has occurred, a new data set representing the process must be collected and a new model developed and tested, for both time series and neural network techniques.

The ability to develop a predictive model that uses the data that is currently available, does not require a large initial training data set, and responds dynamically to process changes over time is very appealing to many process modelers and managers. Work by Härdle et al. (1997) used nonparametric regression methods for estimating spectral densities, higher order conditional moments or conditional densities for time series data. Tjostheim (1994) used nonparametric regression techniques to test the linearity and independence of time series data. Local linear regression (LLR), the nonparametric procedure presented in this paper, is a technique that meets these specifications. The LLR procedure requires only three data points to obtain an initial prediction and then uses all new data as it becomes available to make future predictions, thus making it a dynamic procedure for predicting time series data. That is, making predictions with the first three measurements from a process, the LLR procedure can dynamically make predictions after each subsequent measurement or observation—rather than waiting until all of the data has been collected.

The proposed LLR technique is described and analyzed as follows. First the theory and implementation of the LLR method is described. Then the simulation of various test data series is described. Finally, results are generated using the LLR technique and the performance of the LLR technique is compared to that of traditional time series methods.

2. Description of the LLR procedure

The nonparametric method considered in this paper for estimating the time series response variable is local linear regression (LLR), a variation of the procedure first