Available online at www.sciencedirect.com
ScienceDirect
Neurocomputing 70 (2007) 2331–2341
NEUROCOMPUTING
www.elsevier.com/locate/neucom

CATS benchmark time series prediction by Kalman smoother with cross-validated noise density

Simo Särkkä*, Aki Vehtari, Jouko Lampinen

Helsinki University of Technology, P.O. Box 9203, FIN-02015 HUT, Finland

Available online 22 February 2007

Abstract

This article presents the winning solution to the CATS time series prediction competition. The solution is based on classical optimal linear estimation theory. The proposed method models the long and short term dynamics of the time series as stochastic linear models. The computation is based on a Kalman smoother, in which the noise densities are estimated by cross-validation. In time series prediction the Kalman smoother is applied three times in different stages of the method.

Keywords: CATS benchmark; Bayesian filtering; Optimal filtering; Kalman filter; Kalman smoother

1. Introduction

This article[1] presents the winning solution to the time series prediction competition, the CATS benchmark [17], which was organized as a special session of the IJCNN 2004 conference. The solution is based on the classical Kalman smoother with cross-validated process noise variances. In addition to presenting the winning solution, this article also discusses the connection of optimal filtering to Bayesian inference, and to the Gaussian process regression models used in Bayesian neural network literature.

1.1. CATS benchmark

The goal of the CATS competition [17] was to provide a new benchmark for the problem of time series prediction and to compare different methods and models that can be used for the prediction. The proposed time series is the *CATS benchmark* (Competition on Artificial Time Series).

This artificial time series with 5000 data was given. Within those 100 values were missing. These missing values were divided into 5 blocks

- elements 981–1000;
- elements 1981–2000;
- elements 2981–3000;
- elements 3981–4000;
- elements 4981–5000.

The purpose was to predict the 100 missing values based on the other data. The performance criterion was the mean square error, which was computed on the 100 missing values. The time series is shown in Fig. 1.

1.2. Optimal linear filtering

The success of *optimal linear filtering* is mostly due to the seminal article of Kalman [13], which describes a recursive solution to the optimal discrete linear filtering problem. Although the original derivation of the *Kalman filter* was based on the least squares approach, the same equations can be derived from pure probabilistic Bayesian analysis. The Bayesian analysis of Kalman filtering is well covered in the classic book by Jazwinski [12] and more recently in the book by Bar-Shalom et al. [3].

*Corresponding author. Tel.: +358 40 757 0730; fax: +358 9 451 4830.
E-mail address: simo.sarkka@hut.fi (S. Särkkä).
[1]This article is an extended version of the article [25].

0925-2312/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2005.12.132

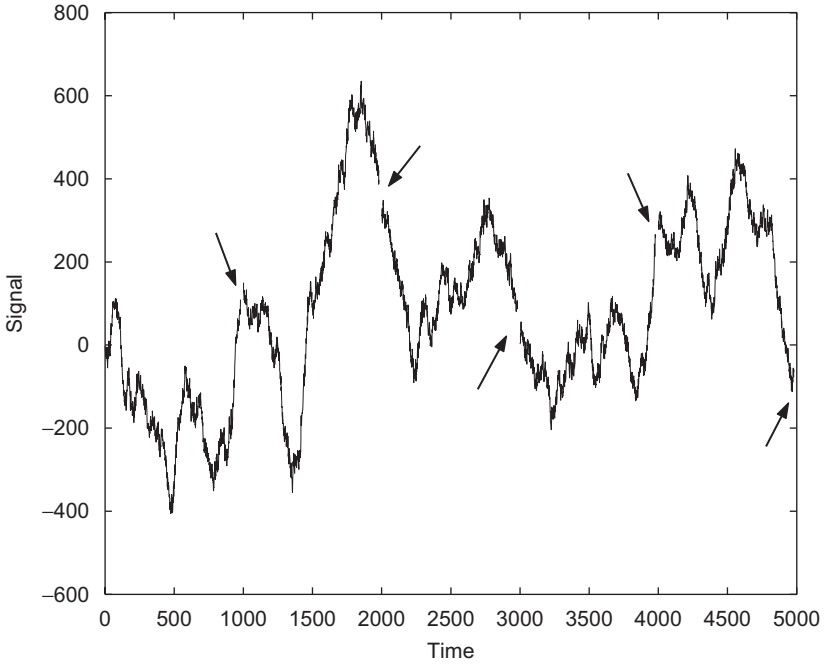

Fig. 1. The CATS benchmark time series. The purpose of the competition was to predict the missing data (marked with arrows) such that the mean squared error is minimized.

Kalman filtering, mostly because of its least squares interpretation, has been widely used in stochastic optimal control. A practical reason to this is that the inventor of Kalman filter, Rudolph E. Kalman, has also made several contributions [14] to the theory of *linear quadratic Gaussian* (LQG) regulators, which are fundamental tools of stochastic optimal control [19,27].

As discussed in the book by West and Harrison [32], in the sixties, Kalman filter like recursive estimators was also used in the Bayesian community and it is not clear whether the theory of Kalman filtering or the theory of *dynamic linear models* (DLM) was the first. Although these theories were originally derived from slightly different starting points, they are equivalent. Because of its useful connection to the theory and history of stochastic optimal control, this article approaches the Bayesian filtering problem from the Kalman filtering point of view.

In the early stages of its history, the Kalman filter was soon discovered to belong to the class of Bayesian estimators [11], with the resulting generalized theory called *non-linear filtering theory* [12]. An interesting historical detail is that while Kalman and Bucy were formulating the linear theory in the United States, Stratonovich was doing the pioneering work on the probabilistic (Bayesian) approach in Russia [12,29].

An optimal discrete filter, such as the Kalman filter, solves the *discrete-time* filtering problem, which means that the underlying physical phenomenon is modeled as a discrete-time process. However, because Nature is continuous, a physically more realistic approach is *continuous-discrete filtering* [12], where state dynamics are modeled as continuous-time stochastic processes, that is, *stochastic differential equations* [16,23] and measurements are assumed to be obtained at discrete time steps. The dynamic model in this paper is also first designed as a continuous-time process and then discretized to allow for consistent prediction over intervals of varying length, that is, for non-uniform sampling of measurements.

2. Optimal estimation

In this section we review the formulation of optimal filtering and smoothing as recursive Bayesian estimation and introduce the notation used in this article. The equivalent formulation is used in classical estimation theory and optimal filtering literature (e.g., [12]).

Optimal non-linear discrete-time estimation considers generic state space models of the form

$$\begin{aligned} \mathbf{x}_k &\sim p(\mathbf{x}_k|\mathbf{x}_{k-1}), \\ \mathbf{y}_k &\sim p(\mathbf{y}_k|\mathbf{x}_k), \end{aligned} \tag{1}$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the unknown hidden state and $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement at time step k. The dynamic model $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ defines the Markov model for state transitions. The measurement model $p(\mathbf{y}_k|\mathbf{x}_k)$ defines the distribution of measurements for given state configurations. At the initial time step $k = 0$, the state is assumed to have the prior distribution $p(\mathbf{x}_0)$.

2.1. The optimal filtering equations

The goal of *filtering* is to compute the *posterior distribution* of the state $\mathbf{x}_k$ at time step k given the history