Signal Processing 93 (2013) 3053–3061

Contents lists available at SciVerse ScienceDirect

Signal Processing

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

Kernel autoregressive models using Yule–Walker equations

M. Kallas [a,b], P. Honeine [a,*], C. Francis [b], H. Amoud [c]

[a] Institut Charles Delaunay (CNRS), LM2S, Université de Technologie de Troyes, France
[b] Laboratoire d'analyse des systèmes, Faculté de Génie 1, Université Libanaise, Lebanon
[c] Azm Center for Research in Biotechnology and its Applications, Doctoral School for Sciences and Technology, Lebanese University, Lebanon

ARTICLE INFO

Article history:
Received 2 July 2012
Received in revised form
5 March 2013
Accepted 20 March 2013
Available online 4 April 2013

Keywords:
Kernel machines
Autoregressive model
Time series prediction
Yule–Walker equations
Pre-image problem

ABSTRACT

This paper proposes nonlinear autoregressive (AR) models for time series, within the framework of kernel machines. Two models are investigated. In the first proposed model, the AR model is defined on the mapped samples in the feature space. In order to predict a future sample, this formulation requires to solve a pre-image problem to get back to the input space. We derive an iterative technique to provide a fine-tuned solution to this problem. The second model bypasses the pre-image problem, by defining the AR model with an hybrid model, as a tradeoff considering the computational time and the precision, by comparing it to the iterative, fine-tuned, model. By considering the stationarity assumption, we derive the corresponding Yule–Walker equations for each model, and show the ease of solving these problems. The relevance of the proposed models is studied on several time series, and compared with other well-known models in terms of accuracy and computational complexity.

1. Introduction

The autoregressive (AR), or linear predictive, model is pervasive in science and technology, with an essential role in the analysis of time series in applications ranging from financial forecasting, to meteorological analysis, to speech processing. For instance to maintain a phone conversation, every cell phone estimates a linear model every 20 milliseconds [1]. The AR model defines each sample as a linear combination of previous samples. The problem consists of estimating the coefficients in the linear combination. Essentially, two strategies have been investigated. The model parameters are estimated either by minimizing the mean square error, or by considering the correlation between the samples. The underlying mathematics that govern the AR model are the Yule–Walker equations. The scientific community has made an ever-growing investment to master these equations for the linear prediction [2]. The Yule–Walker equations are the building block of the linear AR model, connecting its parameters to the covariance function of the process. The model parameters are therefore estimated from the covariances of the time series. Forecasting can be considered by applying the resulting predictive model. However, the linearity assumption is often insufficient to explain nonlinear phenomena. A first attempt to derive a nonlinear Yule–Walker like procedure for a specific nonlinear, high-order, model is given in [3]. Nevertheless, up to our knowledge, there is no work that combine the power of the Yule–Walker equations with the proliferating kernel-based methods.

Kernel machines are essentially based on a nonlinear transformation of the data, by using a mapping function from the input space to some feature space, prior to applying a linear procedure in the latter space [4]. Nevertheless, it is not necessary to explicitly define the nonlinear transformation, but implicitly by considering a (positive semi-definite) kernel function. The use of kernel machines has received considerable attention since Vapnik's Support Vector Machines (SVMs) [5]. Many nonlinear

* Corresponding author. Tel.: +33 3 25 71 56 25.
E-mail addresses: maya.kallas@utt.fr (M. Kallas), paul.honeine@utt.fr (P. Honeine), cfrancis@ul.edu.lb (C. Francis), hassan.amoud@gmail.com (H. Amoud).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.03.032

techniques have been derived, such as the kernel principal component analysis, kernel Fisher discriminant analysis, and SVM novelty detection, only to name a few. In the same spirit, some kernel-based methods were considered for the analysis and prediction of time series data [6], including the SVM regression and kernel Kalman filter [7]. A first attempt to use the AR model with the kernel-based machines was introduced in [8], with parameters estimated by minimizing the mean square error. However, the proposed model fails to perform a prediction scheme. While the model parameters are determined in the feature space, the predicted samples need to be evaluated back in the input space, *i.e.* the space of samples. Therefore, a pre-image technique must be used in order to predict future samples, as we have proposed in [9].

In this paper, we derive nonlinear prediction models by taking full advantage of the Yule–Walker equations. This leads to the estimation of the model parameters by using lagged expected kernels. It is worth noting that the concept of expected kernels has shown its efficiency in recent research [10,11]. Two models are under investigation in this paper.

The first model is based on the underlying concept behind kernel machines, namely mapping data from an input space to a feature space. By operating an AR model on the images of the samples, prediction is defined in the feature space. To provide the predicted sample, one needs to get back to the input space, namely to the space of samples. This is the pre-image problem, with a solution (sample in the input space) that has an image as close as possible to the predicted feature (in the feature space). By following recent developments in the resolution of this ill-posed problem [12], we derive an iterative technique to provide a fine-tuned solution to this problem. We propose to bypass the pre-image problem, by deriving another model. In the second model, we propose an hybrid formulation, as a tradeoff considering the computational time and the precision, compared to the iterative, fine-tuned, model.

The rest of the paper is organized as follows: In the next section, we introduce the linear AR model and present the Yule–Walker equations for estimating the model parameters, and give the main idea behind kernel machines in Section 3. The first model is derived in Section 4, by applying the AR model on the images of the samples, and solving the pre-image problem to interpret the prediction in the input space. Section 5 provides pre-image-free techniques, by deriving an AR model on the kernel values. Finally, Section 7 illustrates the efficiency of the proposed models on several time series data, and provides a comparative study with well-known prediction methods.

2. The Yule–Walker equations of the linear autoregressive model

The linear AR model defines each sample as a linear combination of previous samples. Let $x_1, x_2, \ldots, x_n$ be a time series, the p-order AR model is described by

$$x_i = \sum_{j=1}^{p} \alpha_j x_{i-j} + \varepsilon_i, \qquad (1)$$

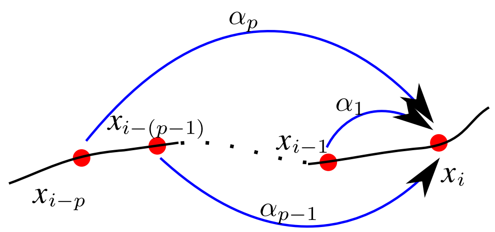

Fig. 1. Illustration of the AR model, where the x_i is defined by a linear combination of the p previous samples x_{i-k}'s, with weight parameters α_1, α_2, …, α_p.

for $i = p+1, \ldots, n$, and where ε_i is the unfitness error, often assumed white Gaussian with zero mean. Fig. 1 illustrates the concept of the AR model. The parameters $\alpha_1, \alpha_2, \ldots, \alpha_p$ are directly connected with the covariance function of the process. One can therefore determine these parameters from the autocorrelation function. This is the essence of the Yule–Walker equations, as illustrated here.

Let the data be centered, thus let μ be the expectation of x_i, namely,

$$\mu = \mathbb{E}[x_i],$$

where $\mathbb{E}[\cdot]$ is the expectation.[1] If we apply the expectation on each side of (1), we get that $(1-\sum_{j=1}^{p}\alpha_j)\mu = \mathbb{E}[\varepsilon_i]$. For any positive lag τ, we can evaluate the autocorrelation function of each time series. Let r be the empirical counterpart of the autocorrelation function of the time series, then $r(\tau) = \sum_{j=1}^{p}\alpha_j\, r(\tau-j)$, for any lag $\tau \geq 1$. Since the autocorrelation function is even, *i.e.*, $r(-\tau) = r(\tau)$, we obtain the matrix form of the Yule–Walker equations

$$\boldsymbol{r} = \boldsymbol{R}\boldsymbol{\alpha},$$

where $\boldsymbol{r} = [r(1)\;\; r(2) \ldots r(p)]^\top$, $\boldsymbol{\alpha} = [\alpha_1\;\; \alpha_2 \ldots \alpha_p]^\top$, and

$$\boldsymbol{R} = \begin{bmatrix} r(0) & r(1) & \ldots & r(p-1) \\ r(1) & r(0) & \ldots & r(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & \ldots & r(0) \end{bmatrix}.$$

Assuming that the $p \times p$ symmetric matrix $\boldsymbol{R}$ is invertible, the coefficients $\boldsymbol{\alpha}$ are estimated by $\boldsymbol{\alpha} = \boldsymbol{R}^{-1}\boldsymbol{r}$. Once the coefficients are estimated, the AR model can be applied to predict future samples, with $x_k = \sum_{j=1}^{p}\alpha_j x_{k-j}$.

While this technique is easy to implement, it is not adapted for nonlinear systems. Next, we derive Yule–Walker-like equations for nonlinear models, within the framework of kernel machines. But before, we prepare the ground by briefly describing the main idea behind the kernel machines.

3. Kernel machines

A kernel is a symmetric and continuous function defined by $\kappa : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$, where $\mathcal{X}$ is an input space. If the kernel verifies $\sum_{i,j}\alpha_i\,\alpha_j\,\kappa(x_i, x_j) \geq 0$ for all $\alpha_i, \alpha_j \in \mathbb{R}$ and all $x_i, x_j \in \mathcal{X}$, then the kernel is positive semi-definite. The Moore–Aronszajn theorem [13] states that each positive semi-definite kernel defines a unique (up to an isometry)

[1] In this paper, all expectations are taken on the index i.