ISA Transactions 52 (2013) 96–104

Contents lists available at SciVerse ScienceDirect

ISA Transactions

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

Online prediction model based on the SVD–KPCA method

Ilyes Elaissi, Ines Jaffel, Okba Taouali*, Hassani Messaoud

Unité de Recherche d'Automatique, Traitement de Signal et Image (ATSI), Monastir 5000, Tunisia

ARTICLE INFO

Article history:
Received 24 April 2012
Received in revised form
24 August 2012
Accepted 29 September 2012
Available online 24 October 2012
This paper was recommended for publication by Dr. A.B. Rad

Keywords:
SLT
RKHS
RKPCA
Online SVD–KPCA

ABSTRACT

This paper proposes a new method for online identification of a nonlinear system modelled on Reproducing Kernel Hilbert Space (RKHS). The proposed SVD–KPCA method uses the Singular Value Decomposition (SVD) technique to update the principal components. Then we use the Reduced Kernel Principal Component Analysis (RKPCA) to approach the principal components which represent the observations selected by the KPCA method.

1. Introduction

Many kernel methods such as Support Vector Machine (SVM)[14], Regularization Network (RN) [3], Kernel Principal Component analysis (KPCA) [2] and Kernel Partial Least Squares (KPLS) [11] have been proposed in recent years for applications in classification [32], diagnostic [31] and nonlinear regression problems [16,30]. In their original forms, most of these methods cannot operate in online way because of the difficulties accompanying them such as the time and memory complexities (because of the growing kernel matrix) and the need to avoid overfitting. Recently a theoretical foundation for online learning algorithm with kernel method in reproducing kernel Hilbert spaces was proposed [4,12,13,17,18,20,21]. The online kernel algorithm is more useful when the system to be identified is time-varying, because these algorithms can automatically track changes of system model with time-varying and time lagging characteristic.

In this paper we propose a new method for online identification of a nonlinear system parameters modelled on Reproducing Kernel Hilbert Space (RKHS). This method uses the SVD technique to update the principal components and then the Reduced Kernel Principal Component Analysis (RKPCA) to select the observations data to approach the Principal Components Analysis updated by the SVD method. The selected observations are used to build an RKHS model with a reduced parameter number. The proposed online identification method updates the list of the retained principal components, and then the RKHS model by evaluating the error between the output model and the process one. The proposed technique may be very helpful to design an adaptive control strategy of nonlinear systems.

The paper is organized as follows: in Section 2, we remind the Reproducing Kernel Hilbert Space (RKHS). Section 3 is devoted to the modelling in RKHS. The Reduced Kernel Principal Component Analysis RKPCA method is presented in Section 4. In Section 5, we propose the new online SVD–KPCA method. The proposed algorithm has been tested to identify a chemical reactor [7] and a Tennessee Eastman process [25].

2. Reproducing kernel Hilbert space

Let $E \subset \Re^d$ an input space and $L^2(E)$ the Hilbert space of square integrable functions defined on E. Let $k : E \times E \to \Re$ be a continuous positive definite kernel. It is proved [5,8] that it exists a sequence of an orthonormal eigen functions $(\psi_1,\psi_2,\ldots,\psi_l)$ in $L^2(E)$ and a sequence of corresponding real positive eigenvalues $(\sigma_1,\sigma_2,\ldots,\sigma_l)$ (where l can be infinite) so that

$$k(x,t) = \sum_{j=1}^{l} \sigma_j \psi_j(x) \psi_j(t); x,t \in E \tag{1}$$

Let $F_k \subset L^2(E)$ be a Hilbert space associated to the kernel k and defined by

$$F_k = \left\{ f \in L^2(E) / f = \sum_{i=1}^{l} w_i \varphi_i \text{ and } \sum_{j=1}^{l} \frac{w_j^2}{\sigma_j} < \infty \right\} \tag{2}$$

* Corresponding author.
E-mail addresses: ilyes.elaissi@yahoo.fr (I. Elaissi), inesjaffel@yahoo.fr (I. Jaffel), taoualiokba@yahoo.fr (O. Taouali), hassani.messaoud@enim.rnu.tn (H. Messaoud).

0019-0578/$ - see front matter © 2012 ISA. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.isatra.2012.09.007

where $\varphi_i = \sqrt{\sigma_i}\,\psi_i$ $i=1,\ldots,l$. The scalar product in the space F_k is given by

$$\langle f,g\rangle_{F_k} = \langle \sum_{i=1}^{l} w_i\varphi_i, \quad \sum_{j=1}^{l} z_j\varphi_j\rangle_{F_k} = \sum_{i=1}^{l} w_i z_i \tag{3}$$

The kernel k is said to be a reproducing kernel of the Hilbert space F_k if and only if the following conditions are satisfied.

$$\begin{cases} \forall x\in E, & k(x,.)\in F_k \\ \forall x\in E \quad \text{and} \quad \forall f\in F_k, & \langle f(.),k(x,.)\rangle_{F_k} = f(x) \end{cases} \tag{4}$$

Where $k(x,.)$ means $k(x,x')$ $\forall x'\in E$

F_k is called reproducing kernel Hilbert space (RKHS) with kernel k and dimension l. Moreover, for any RKHS, there exists only one positive definite kernel and vice versa [9].

Among the possible reproducing kernels, we mention the Radial Basis function (RBF) defined as

$$k(x,t) = \exp\left(-\|x-t\|^2/2\sigma^2\right); \forall x,t\in E \tag{5}$$

with σ a fixed parameter.

3. RKHS models

Consider a set of observations $\{x^{(i)},y^{(i)}\}_{i=1,\ldots,M}$ with $x^{(i)}\in\Re^n$, $y^{(i)}\in\Re$ are respectively the system input and output. According to the statistical learning theory (SLT) [14,15] the identification problem in the RKHS F_k can be formulated as a minimization of the regularized empirical risk. Thus it consists in finding the function $f^*\in F_k$ such that

$$f^* = \sum_{j=1}^{l} w_j^*\varphi_j = \min_{f\in F_k}\frac{1}{M}\sum_{i=1}^{M}\left(y^{(i)}-f\left(x^{(i)}\right)\right)^2 + \lambda\|f\|_{F_k}^2 \tag{6}$$

where M is the measurement number and λ is a regularization parameter chosen in order to ensure a generalization ability to the solution f^*. According to the representer theorem [8], the solution f^* of the optimization problem (6) is a linear combination of the kernel k applied to the M measurements $x^{(i)}$, $i=1,\ldots,M$, as

$$f^*(x) = \sum_{i=1}^{M} a_i^* k\left(x^{(i)},x\right) \tag{7}$$

To solve the optimization problem (6) we can use some kernel methods such that Support Vector Machine (SVM) [10], Least Square Support Vector Machine (LSSVM) [6], Regularization Network (RN) [3], Kernel Partial Least Square (KPLS) [11], etc. In [2], the Kernel Principal Component Analysis KPCA was proposed. This method reconsiders the regularization idea by finding the solution to the identification problem in some subspaceF_{kpca} spanned by the so called principal component analysis.

In the next section we present the Reduced KPCA in which we approximate the retained principal component given by the KPCA method by a set of vectors of observations. These vectors point to the directions of the largest variance with the retained principal component.

4. RKPCA method

Let a nonlinear system with an input $u\in\Re$ and an output $y\in\Re$ from which we extract a set of observations $\{u^{(i)},y^{(i)}\}_{i=1,\ldots,M}$. Let F_k an RKHS space with kernel k. To build the input vector $x^{(i)}$ of the RKHS model we use the NARX (Nonlinear auto regressive with eXogeneous input) structure as

$$x^{(i)} = \left\{u^{(i)},u^{(i-1)},\ldots,u^{(i-m_u)},y^{(i-1)},\ldots,y^{(i-m_y)}\right\}^T;\ m_u,\ m_y\in\text{naturals}; \tag{8}$$

The set of observations becomes $D=\{x^{(i)},y^{(i)}\}_{i=1,\ldots,M}$ where $x^{(i)}\in\Re^{m_u+m_y+1}$ and $y^{(i)}\in\Re$. and the RKHS model of this system based on (7) can be written as

$$\tilde{y}^{(j)} = \sum_{i=1}^{M} a_i k\left(x^{(i)},x^{(j)}\right) \tag{9}$$

Let the application Φ

$$\Phi: E\to\Re^l$$

$$x\mapsto\Phi\left(x\right) = \begin{pmatrix} \varphi_1(x) \\ . \\ . \\ . \\ \varphi_l(x) \end{pmatrix} \tag{10}$$

where φ_i are given in (2).

The Gram matrix K associated to the kernel k is an M-dimensional square matrix, so that

$$K_{i,j} = k\left(x^{(i)},x^{(j)}\right) \quad \text{for } i,j=1,\ldots,M \tag{11}$$

The kernel trick [9] is so that

$$\langle\Phi(x),\Phi(x')\rangle = k(x,x')\forall x,x'\in E \tag{12}$$

We assume that the transformed data $\{\Phi(x^{(i)})\}_{i=1,\ldots,M}\in\Re^l$ are centered [2]. The empirical covariance matrix of the transformed data is symmetrical and l-dimensional. It is written as the following:

$$C_\phi = \frac{1}{M}\sum_{i=1}^{M}\Phi\left(x^{(i)}\right)\Phi\left(x^{(i)}\right)^T, C_\phi\in\Re^{l\times l} \tag{13}$$

Let l' the number of the eigenvectors $\{V_j\}_{j=1,\ldots,l'}$ of the matrix C_ϕ that corresponding to the non zeros positive eigenvalues $\{\lambda_j\}_{j=1,\ldots,l'}$. It is proved in [2] that the number l' is less or equal to the measurement number M.

Due to the large size l of C_ϕ, the computing of $\{V_j\}_{j=1,\ldots,l'}$ can be difficult. The KPCA method shows that these $\{V_j\}_{j=1,\ldots,l'}$ are related to the eigenvectors $\left\{\beta_j\right\}_{j=1,\ldots,l'}$ of the gram matrix K according to [1]

$$V_j = \sum_{i=1}^{M}\beta_{j,i}\Phi\left(x^{(i)}\right), j=1,\ldots,l' \tag{14}$$

Where $(\beta_{j,i})_{j=1,\ldots,p}$ are the components of $\left\{\beta_j\right\}_{j=1,\ldots,l'}$ associated to their nonzero eigenvalues $\mu_1>\ldots>\mu_{l'}$

The principle of the KPCA method consists in organizing the eigenvectors $\left\{\beta_j\right\}_{j=1,\ldots,l'}$ in the decreasing order of their corresponding eigenvalues $\left\{\mu_j\right\}_{j=1,\ldots,l'}$. The principal components are the p first vectors $\{V_j\}_{j=1,\ldots,p}$ associated to the highest eigenvalues and are often sufficient to describe the structure of the data [1,2]. The number P satisfies the Inertia Percentage criterion IPC given by

$$P^* = \arg(IPC\geq 99) \tag{15}$$