ISA Transactions 57 (2015) 205–210

Contents lists available at ScienceDirect

ISA Transactions

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

Dimensionality reduction of RKHS model parameters

Okba Taouali*, Ilyes Elaissi, Hassani Messaoud

Laboratory of Automatic Signal and Image Processing, Tunisia

ARTICLE INFO

Article history:
Received 16 October 2014
Received in revised form
30 January 2015
Accepted 4 February 2015
Available online 10 March 2015
This paper was recommended for publication by Prof. A.B. Rad.

Keywords:
RKHS
System identification
Kernel method
KPLS
RKPLS
Process Trainer

ABSTRACT

This paper proposes a new method to reduce the parameter number of models developed in the Reproducing Kernel Hilbert Space (RKHS). In fact, this number is equal to the number of observations used in the learning phase which is assumed to be high. The proposed method entitled Reduced Kernel Partial Least Square (RKPLS) consists on approximating the retained latent components determined using the Kernel Partial Least Square (KPLS) method by their closest observation vectors. The paper proposes the design and the comparative study of the proposed RKPLS method and the Support Vector Machines on Regression (SVR) technique. The proposed method is applied to identify a nonlinear Process Trainer PT326 which is a physical process available in our laboratory. Moreover as a thermal process with large time response may help record easily effective observations which contribute to model identification. Compared to the SVR technique, the results from the proposed RKPLS method are satisfactory.

1. Introduction

Recently many kernel methods developed in the RKHS space in offline phase [4–7] or in online scenario [8,9,19] such as, online RKPCA, Support Vector Machine (SVM), Regularization Network (RN), Kernel Principal Component analysis (KPCA), Kernel Partial Least Squares (KPLS) have been proposed for applications in classification [1], diagnosis [12,22] and nonlinear regression problems [3–5]. Despite its linearity with respect to its parameters, the RKHS model suffers from huge complexity as the number of its parameters is related to the number of observations used in learning phase and which can be large depending on the process features and the hardness of its nonlinearity. The resulting model may be useless essentially in real time control. The problem of parameter reduction number of RKHS model is under study and only few research works have underlined them except some works such as [6,20]. In this paper, a new kernel method entitled Reduced Kernel Partial Least Square (RKPLS), to reduce the RKHS model parameters number is proposed in which only the set of observations that approximate the retained latent components is considered. The selected observations are used to build the reduced RKHS model which can be very helpful to design real time control strategies. The proposed RKPLS method has been applied to identify a nonlinear systems, the highly nonlinear Trainer Process PT326 [21].

The paper is organized as follows. In Section 2, the Reproducing Kernel Hilbert Space (RKHS) is presented. Section 3 is devoted to the modeling in RKHS. The Partial Least Square method (PLS) and its nonlinear version (KPLS) are presented in Sections 4 and 5. In Section 6, the new proposed Reduced KPLS method is detailed. Section 7 validates the proposed algorithm on the above benchmark. Finally Section 8 concludes the paper.

2. Reproducing Kernel Hilbert Space (RKHS)

Let $E \subset \mathbb{R}^n$ an input space, $L^2(E)$ the Hilbert space of square integrable functions defined on E and $k : E \times E \to \mathbb{R}$ is a continuous positive definite kernel. It is proved [13,14] that there exists a sequence of an orthonormal eigen functions $(\psi_1,\ \psi_2, ..., \psi_l)$ in $L^2(E)$ and a sequence of corresponding real positive eigenvalues $(\sigma_1, \sigma_2, ..., \sigma_l)$ (where l can be infinite) so that:

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

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

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

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

http://dx.doi.org/10.1016/j.isatra.2015.02.003
0019-0578/

where $\alpha_i \in \mathbb{R}$, $\varphi_i = \sqrt{\sigma_i}\psi_i$ $i = 1, ..., l$. The inner product of two functions f and g in the space F_k is given by:

$$<f, g>_{F_k} = <\sum_{i=1}^{l} \alpha_i \varphi_i, \sum_{j=1}^{l} z_j \varphi_j>_{F_k} = \sum_{i=1}^{l} \alpha_i z_i \tag{3}$$

The kernel k given in (1) is said to be a reproducing kernel of the Hilbert space F_k if and only if the following conditions are satisfied.

$$\begin{cases} \forall x \in E, \quad k(x, .) \in F_k \\ \forall x, t \in E \text{ and } \forall f \in F_k, < f(t), k(x,t) >_{F_k} = f(x) \end{cases} \tag{4}$$

In such a case, 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 [15].

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

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

with σ is a fixed parameter.

3. RKHS models

Let's consider a set of observations $\{x^{(i)}, y^{(i)}\}_{i=1,...,M}$ with $x^{(i)} \in \mathbb{R}^n$, $y^{(i)} \in \mathbb{R}$ are respectively the system input and output. According to the statistical learning theory (SLT) [15] the identification problem in the RKHS F_k can be formulated as a minimization of the regularized empirical risk [14]. 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 number of measurement used in learning phase and λ is a regularization parameter chosen in order to ensure a generalization ability to the solutionf^*. According to the representer theorem [14], the solution f^* of the optimization problem (6) is a linear combination of the kernel k applied to the M measurements $x^{(i)}$, $i = 1, ..., M$, as:

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

The parameter number of the model is equal to the observation number. Despite its linearity with respect to parameters, the model (7) suffers from the large parameter number which may disable its use. In literature some methods for reduction parameter number have been proposed such as Support Vector Machine on Regression (SVR) [2], Reduced Kernel Principal Component Analysis [20].

To reduce significantly the parameter number of the RKHS model, this paper propose a new method entitled RKPLS which is a reduced version of the Kernel Partial Least Square (KPLS) [7] technique based on the famous Partial Least Square (PLS) method [7]. Before detailing the proposed method the PLS and the KPLS methods are reminded in both next sections.

Let's define the application Φ:

$$\Phi : E \to \mathbb{R}^l$$

$$x \mapsto \Phi(x) = \begin{pmatrix} \varphi_1(x) \\ . \\ . \\ . \\ \varphi_l(x) \end{pmatrix} \tag{8}$$

where φ_i are given in (2).

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

$$K_{i,j} = k\left(x^{(i)}, x^{(j)}\right) \quad \text{for} \quad i,j = 1, ..., M \tag{9}$$

The inner product of the transformed observation [14] is:

$$< \Phi(x), \Phi(t) > = k(x,t)x, \ t \in E \tag{10}$$

4. Partial Least Square method

4.1. Principle

Let consider the observations data set $\{x^{(i)}, y^{(i)}\}_{i=1}^{M}$, where $x^{(i)} \in E \subset \mathbb{R}^n$, $y^{(i)} \in \mathbb{R}$ are respectively the input and output of a linear model. It is assumed that these observations satisfy:

$$\frac{1}{M}\sum_{i=1}^{M} x^{(i)} = 0, \frac{1}{M}\sum_{i=1}^{M} y^{(i)} = 0 \quad \text{and} \quad \sum_{i=1}^{M} \left(y^{(i)}\right)^2 = 1 \tag{11}$$

The PLS method consists on determining a set of vectors called latent components that represent as most as possible the variations of the input and the output observations [11]. The identification of PLS model parameters results from the resolution of the following optimization problem:

$$\min_{w \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^{M} \left(y^{(i)} - w^T x^{(i)}\right)^2 \tag{12}$$

where w is the coefficient vector to be identified.

The model used to describe the system is given by:

$$y = \sum_{j=1}^{n} w_j x_j \tag{13}$$

where y and x_j are respectively the model output and the jth component of the ith input $x^{(i)}$

When the dimension n of the input space is larger than the observations number M, the solution of the problem (12) can leads to an overfitting.

The PLS technique formulates the problem (12) into a constrained minimization problem so that the coefficient vector is normed.

$$\begin{cases} \min\limits_{w \in \mathbb{R}^n} \quad \frac{1}{2}\sum_{i=1}^{M} \|x^{(i)} - y^{(i)} w\|^2 \\ s.t \\ w^T w = 1 \end{cases} \tag{14}$$

The solution of the problem (14) is called first latent component or first latent vector that approximates the input and the output observations [23].

The problem (14) can be limited as:

$$\begin{cases} \min\limits_{w \in \mathbb{R}^n} \quad -\sum_{i=1}^{M} y^{(i)} w^T x^{(i)} \\ s.t \\ w^T w = 1 \end{cases} \tag{15}$$

The matrix form of the problem (15) is written as following:

$$\begin{cases} \min\limits_{w \in \mathbb{R}^n} -Y^T(Xw) \\ s.t \\ w^T w = 1 \end{cases} \tag{16}$$