International Conference on Control, Engineering & Information Technology (CEIT'14)
Proceedings - Copyright IPCO-2014
ISSN 2356-5608

Comparative study of Kernel Canonical Correlation Analysis and Kernel Partial Lealst Square with application

Souilem_n@yahoo.com, elaissi_ilyes@yahoo.fr, taoualiokba@yahoo.fr, hassani.messaoud@enim.rnu.tn

Abstract— This paper treats the identification of nonlinear systems using RKHS models, with Kernel Canonical Correlation Analysis (KCCA) technique. KCCA finds common semantic features between mapped input-output data in high dimension nonlinear spaces, and then use the common features to represent the data. We use KCCA and Kernel Partial Least Square (KPLS) technique in nonlinear system identification benchmark, a Wiener-Hammerstein, and we compare results.

Keywords— RKHS, Kernel method, KPLS, KCCA, Wiener-Hammerstein

I. Introduction

During last years, a lot of researchs were performed in the field of nonlinear system identification using Kernel methods [5, 12], many techniques have been developed and widely used in different applications, such as Support Vector Machines (SVM) [3], Kernel Patial Least Square (KPLS) [10], Kernel Principal Component Analysis (KPCA) [11], Kernel Independent Component Analysis (KICA) [2], ...

In this paper we are interested in using Kernel Canonical Correlation Analysis (KCCA) in regression problems. KCCA is the kernel extension of Canonical Correlation Analysis (CCA) with positive definite kernels [1] and is mainly used in classification problems [5, 6, 7]. CCA was proposed by Hotelling in 1936 [8], to solve the problem of finding vector bases for two sets of variables such that the correlations between the projections of the variables onto these bases are mutually maximized. The solution then results in solving an eigenvector problem. Kernel methods principle maps the data into a higher dimensional feature space, KCCA is
then applied on the mapped input-output data and extract their common information allowing the construction of the RKHS model.
The partial least squares (PLS) [15, 17] creates orthogonal score vectors (component, latent vectors)by using the existing correlations between different sets of variables (blocks of data) while also keeping most of the variance of both sets.
Kernel PLS (KPLS) is the kernel extension of PLS and deals with the data mapped into a higher dimensional feature [10, 16].
In this paper we use KCCA and KPLS techniques to identify RKHS models for nonlinear systems, with applications on a Wiener-Hammerstein benchmark [13].

The paper is organized as follows. In sections 2, 3 and 4 we present the Kernel Canonical Correlation Analysis, the eigenvector problem is firstly solved with a complete Cholesky decomposition, and then with a Partial Gram Shmidt Orthogonalisation [3]. In section five, we perform a numerical simulation on benchmark application, and a comparison study on KCCA and KPLS results is presented.

II. Kernel Canonical Correlation Analysis (KCCA)

Due to its linearity, CCA may not extract useful descriptors of the data in nonlinear cases. KCCA offers a solution by mapping the data into a higher dimensional feature space.
Consider the set of data $D=\{x_i,\ y_i\}_{i=1,\ \dots,\ N}$, N the number of observations. We define $k_x : E_x^{\ 2} \to IR$ and $k_y : E_y^{\ 2} \to IR$ to be continuous positive definite kernels[9, 14, 1]. It exists sequences of orthonormal eigenfunctions $\left(\varphi_{x_1},\ \varphi_{x_2},\ \dots,\ \varphi_{x_l}\right)$ in $L^2\left(E_y\right)$ and $\left(\varphi_{y_1},\ \varphi_{y_2},\ \dots,\ \varphi_{y_{l'}}\right)$ in $L^2\left(E_x\right)$ (where l and l' can be infinite) so that:

$$k_x\left(x,\ x'\right)=\sum_{j=1}^{l}\sigma_j\ \varphi_{x_j}\left(x\right)\ \varphi_{x_j}\left(x'\right)\quad ;\quad x,\ x'\in E_x$$
(1)

$$k_y\left(y,\ y'\right)=\sum_{j=1}^{l'}\sigma_j\ \varphi_{y_j}\left(y\right)\ \varphi_{y_j}\left(y'\right)\quad ;\quad y,\ y'\in E_y$$
(2)

Let $F_x \subset L^2\left(E_x\right)$ (resp. $F_y \subset L^2\left(E_y\right)$) be a Hilbert space associated to the kernel k_x (resp. k_y) and defined by:

$$F_x=\left\{g_x\in\ L^2\left(E_x\right)\ /\ g_x=\sum_{i=1}^{l}w_i\,\varphi_{x_i}\ \ and\ \ \sum_{j=1}^{l}\frac{w_j^{\ 2}}{\sigma_j}<+\infty\right\}$$
(3)

$$F_y=\left\{g_y\in\ L^2\left(E_y\right)\ /\ g_y=\sum_{i=1}^{l}w_i\,\varphi_{y_i}\ \ and\ \ \sum_{j=1}^{l}\frac{w_j^{\ 2}}{\sigma_j}<+\infty\right\}$$
(4)

Let the applications Φ_x and Φ_y:

$$\Phi_x : E_x \to \mathbb{R}^l$$

$$x \mapsto \Phi_x(x) = \begin{pmatrix} \varphi_{x\,1}(x) \\ . \\ . \\ . \\ \varphi_{x\,l}(x) \end{pmatrix}$$

(5)

$$\Phi_y : E_y \to \mathbb{R}^{l'}$$

$$\mathrm{y} \mapsto \Phi_y(y) = \begin{pmatrix} \varphi_{y\,1}(y) \\ . \\ . \\ . \\ \varphi_{y\,l}{'}(y) \end{pmatrix}$$

(6)

We have then:

$$k_x(x, x') = \langle \Phi_x(x), \Phi_x(x') \rangle, \quad x,\ x' \in E_x$$
$$k_y(y, y') = \langle \Phi_y(y), \Phi_y(y') \rangle, \quad y,\ y' \in E_y$$

Let the Gram matrices $K_x \in IR^{N\times N}$ and $K_y \in IR^{N\times N}$:

$(K_x)_{ij} = k_x(x_i, x_j)$ and $(K_y)_{ij} = k_y(y_i, y_j)$

We consider u_x, v_y defined by :

$$u_x = \langle \Phi_x(x), f_x \rangle = f_x^{\ T}\, \Phi_x(X)$$

(7)

$$v_y = \langle \Phi_y(y), f_y \rangle = f_y^{\ T}\, \Phi_y(Y)$$

(8)

With $\Phi_x(X)$ and $\Phi_y(Y)$ are the mappings of X, and Y with the applications Φ_x and Φ_y respectively. We suppose that $\Phi_x(X)$ and $\Phi_y(Y)$ are centered, and with $f_x \in F_x$, $f_y \in F_y$ are defined as:

$$f_x = \sum_{i=1}^{N} \alpha_i\, \Phi_x(x_i)$$

(9)

$$f_y = \sum_{i=1}^{N} \beta_i\, \Phi_y(y_i)$$

(10)

$$\alpha_i,\ \beta_i \in IR,\ i = 1, ..., N$$

The correlation between u_x and v_y is defined by

$$\rho = corr(\langle \Phi_x(x),\ f_x \rangle,\ \langle \Phi_y(y),\ f_y \rangle)$$

(11)

$$= \frac{\mathrm{cov}(\langle \Phi_x(x),\ f_x \rangle,\ \langle \Phi_y(y),\ f_y \rangle)}{\mathrm{var}(\langle \Phi_x(x),\ f_x \rangle)\,\mathrm{var}(\langle \Phi_y(y),\ f_y \rangle)} = \frac{\mathrm{cov}(u_x, v_y)}{\mathrm{var}(u_x)\ \mathrm{var}(v_y)}$$

with:

$$\mathrm{cov}(u_x, v_y) = \frac{1}{N}\sum_{k=1}^{N} \left\langle \Phi_x(x_i),\ \sum_{i=1}^{N} \alpha_i \Phi_x(x_i) \right\rangle \ \left\langle \Phi_y(y_i),\ \sum_{i=1}^{N} \beta_i \Phi_y(y_i) \right\rangle$$

$$\mathrm{var}(u_x) = \frac{1}{N}\sum_{k=1}^{N} \left\langle \Phi_x(x_i),\ \sum_{i=1}^{N} \alpha_i \Phi_x(x_i) \right\rangle \ \left\langle \Phi_x(x_i),\ \sum_{i=1}^{N} \alpha_i \Phi_x(x_i) \right\rangle$$

Let

$$\alpha = (\alpha_1,\ \alpha_2,\ ...,\ \alpha_N)^T$$

(12)

$$\beta = (\beta_1,\ \beta_2,\ ...,\ \beta_N)^T$$

(13)

Using the kernel trick, and performing some easy calculus, we have:

$$\mathrm{cov}(u_x, v_y) = \frac{1}{N}\, \alpha^T\, K_x\, K_y\, \beta$$

(14)

$$\mathrm{var}(u_x) = \frac{1}{N}\, \alpha^T\, K_x\, K_x\, \alpha$$

(15)

$$\mathrm{var}(v_y) = \frac{1}{N}\, \beta^T\, K_y\, K_y\, \beta$$

(16)

KCCA then solves the following problem:

$$\max_{\alpha,\ \beta \in\ IR^N} \frac{\alpha^T\, K_x\, K_y\, \beta}{\sqrt{\alpha^T\, K_x^{\ 2} \alpha}\sqrt{\beta^T\, K_y^{\ 2} \beta}}$$

(17)

subject to :

$$\alpha^T\, K_x^{\ 2} \alpha = 1 \ \text{ and } \ \beta^T\, K_y^{\ 2} \beta = 1$$

The Lagrangian is

$L(\lambda_x, \lambda_y, \alpha, \beta) =$

$$\alpha^T K_x K_y \beta - \frac{\lambda_\alpha}{2}(\alpha^T K_x^{\ 2} \alpha - 1) - \frac{\lambda_\beta}{2}(\beta^T K_y^{\ 2} \beta - 1)$$

The derivative of the Lagrangian :

$$\frac{\partial L}{\partial \alpha} = K_x K_y \beta - \lambda_\alpha K_x{}^2 \alpha = 0$$
(18)

$$\frac{\partial L}{\partial \beta} = K_y K_x \alpha - \lambda_\beta K_y{}^2 \beta = 0$$
(19)

multiplying relation (18) by α^T and (19) by β^T, and subtracting them, we find

$$\alpha^T K_x K_y \beta - \alpha^T \lambda_\alpha K_x{}^2 \alpha - \beta^T K_y K_x \alpha + \beta^T \lambda_\beta K_y{}^2 \beta = 0$$

Or: $\lambda_\beta \beta^T K_y{}^2 \beta - \lambda_\alpha \alpha^T K_x{}^2 \alpha = 0$

then $\lambda_\beta = \lambda_\alpha$, let $\lambda_\beta = \lambda_\alpha = \lambda$

We consider K_x and K_y are non singular, we have :

$$\beta = \frac{K_y^{-1} K_x \alpha}{\lambda}$$
(20)

From relation (20) and (18) we find

$$K_x K_y K_y^{-1} K_x \alpha - \lambda^2 K_x K_x \alpha = 0$$

Hence $K_x K_x \alpha - \lambda^2 K_x K_x \alpha = 0$

Then

$$I \alpha = \lambda^2 \alpha$$
(21)

From equation (21), we can deduce that $\lambda = 1$ for every vector α, hence if K_x and K_y are invertible, then perfect correlation can be formed, suggesting learning is trivial and applying KCCA in this fashion will not then provide useful results.

Instead of solving (17), we solve a regularized form, we introduce a control on the flexibility of the projection mapping by penalizing the norms of the associated weights, and we convexly combine the CCA term with the regularization term, we obtain:

$$\max_{\alpha,\ \beta \in IR^N} \frac{\alpha^T K_x K_y \beta}{\sqrt{\alpha^T K_x{}^2 \alpha + \tau \|f_x\|^2}\ \sqrt{\beta^T K_y{}^2 \beta + \tau \|f_y\|^2}} \qquad \tau > 0$$
(22)

According to the definitions of f_x and f_y, in (9) and (10):

$$\|f_x\|^2 = \langle f_x, f_x \rangle = \left\langle \sum_{i=1}^{N} \alpha_i \Phi_x(x_i),\ \sum_{j=1}^{N} \alpha_j \Phi_x(x_j) \right\rangle$$
$$= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j k_x(x_i, x_j) = \alpha^T K_x \alpha$$

similarly $\|f_y\|^2 = \langle f_y, f_y \rangle = \beta^T K_y \beta$

Then the KCCA regularized problem becomes

$$\max_{\alpha,\ \beta \in IR^N} \frac{\alpha^T K_x K_y \beta}{\sqrt{\alpha^T K_x{}^2 \alpha + \tau \alpha^T K_x \alpha}\ \sqrt{\beta^T K_y{}^2 \beta + \tau \beta^T K_x \beta}}$$
(23)

Subject to

$\alpha^T K_x{}^2 \alpha + \tau \alpha^T K_x \alpha = 1$ and $\beta^T K_y{}^2 \beta + \tau \beta^T K_x \beta = 1$

The Langrangien is

$$L(\lambda_\alpha, \lambda_\beta, \alpha, \beta) = \alpha^T K_x K_y \beta - \frac{\lambda_\alpha}{2}(\alpha^T K_x{}^2 \alpha + \tau \alpha^T K_x \alpha - 1)$$
$$- \frac{\lambda_\beta}{2}(\beta^T K_y{}^2 \beta + \tau \beta^T K_y \beta - 1)$$

We calculate $\frac{\partial L}{\partial \alpha}$ and $\frac{\partial L}{\partial \beta}$ we obtain

$$\frac{\partial L}{\partial \alpha} = K_x K_y \beta - \lambda_\alpha (K_x{}^2 \alpha + \tau K_x \alpha) = 0$$
(24)

$$\frac{\partial L}{\partial \beta} = K_y K_x \alpha - \lambda_\beta (K_y{}^2 \beta + \tau K_y \beta) = 0$$
(25)

We multiply (24) by α^T and (25) by β^T and subtracting both equations, we obtain:

$$\alpha^T K_x K_y \beta - \lambda_\alpha \alpha^T (K_x{}^2 \alpha + \tau K_x \alpha) -$$
$$\beta^T K_y K_x \alpha + \lambda_\beta \beta^T (K_y{}^2 \beta + \tau K_y \beta) = 0$$

Then

$$\lambda_\beta (\beta^T K_y{}^2 \beta + \tau \beta^T K_y \beta) - \lambda_\alpha (\alpha^T K_x{}^2 \alpha + \tau \alpha^T K_x \alpha) = 0$$

Then $\lambda_\beta = \lambda_\alpha$, let $\lambda_\beta = \lambda_\alpha = \lambda$

We consider K_x and K_y non singular, then equation (25) gives:

$$\beta = \frac{(K_y + \tau I)^{-1} K_x \alpha}{\lambda}$$
(26)

The equations (24) and (26) give
$K_x\, K_y (K_y + \tau I)^{-1}\, K_x\, \alpha = \lambda^2 K_x (K_x + \tau I)\alpha$

Then

$$K_y (K_y + \tau I)^{-1}\, K_x\, \alpha = \lambda^2 (K_x + \tau I)\alpha$$

we finally obtain the eigenvector problem

$$(K_x + \tau I)^{-1}\, K_y (K_y + \tau I)^{-1}\, K_x\, \alpha = \lambda^2 \alpha$$
(27)

we obtain an eigenvector problem of the form $A_3\ x = \lambda_3\ x$, with

$$A_3 = (K_x + \tau I)^{-1}\, K_y (K_y + \tau I)^{-1}\, K_x$$
(28)

Unfortunately, the matrix A_3 in (28) is not symmetric and this may yields imaginary eigenvalues. Besides, matrix inversion in the case of large $K_y = R_y R_y^T$ training sets may lead to computational problems.

III. First level heading KCCA regularisation with partial Gram Chmidt orthogonalisation

To overcome the problem we apply Partial Gram Shmidt Orthogonalisation (PGSO) [3] or equivalently incomplete Cholesky decomposition to reduce the dimensionality of the kernel matrices.
We will decompose the kernel matrices K_x and K_y via the complete Cholesky decomposition as:

$$K_x = R_x R_x^T$$
$$K_y = R_y R_y^T$$

where R_x and R_y is a lower triangular matrices, gives

the relation (24) et (25) can be written as

$$R_x R_x^T\ \ R_y R_y^T\ \beta - \lambda (R_x R_x^T R_x R_x^T + \tau R_x R_x^T)\ \alpha = 0$$
(29)

$$R_y R_y^T\ \ R_x R_x^T\ \alpha - \lambda (R_y R_y^T R_y R_y^T + \tau R_y R_y^T)\ \beta = 0$$
(30)

Multiplying (29) with R_x^T and (30) with R_y^T, we obtain

$$R_x^T R_x R_x^T\ \ R_y R_y^T\ \beta - \lambda\, R_x^T\, (R_x R_x^T R_x R_x^T + \tau R_x R_x^T)\ \alpha = 0$$
(31)

$$R_y^T R_y R_y^T\ \ R_x R_x^T\ \alpha - \lambda\, R_y^T (R_y R_y^T R_y R_y^T + \tau R_y R_y^T)\, \beta = 0$$
(32)

we define $Z_{xx} = R_x^T R_x$, $Z_{yy} = R_y^T R_y$, $Z_{xy} = R_x^T R_y$, $Z_{yx} = R_y^T R_x$, $\tilde{\alpha} = R_x^T \alpha$, $\tilde{\beta} = R_y^T \beta$

the relation (31) and (32) can be written as

$$Z_{xx} Z_{xy}\ \ \tilde{\beta} - \lambda\ Z_{xx} (Z_{xx} + \tau I)\ \tilde{\alpha} = 0$$
(33)

$$Z_{yy} Z_{yx}\ \ \tilde{\alpha} - \lambda\ Z_{yy} (Z_{yy} + \tau I)\ \tilde{\beta} = 0$$
(34)

From equation (34) we write

$$\tilde{\beta} = \frac{(Z_{yy} + \tau I)^{-1} Z_{yy}^{-1}\, Z_{yy}\, Z_{yx}\ \tilde{\alpha}}{\lambda}$$

Then

$$\tilde{\beta} = \frac{(Z_{yy} + \tau I)^{-1}\, Z_{yx}\ \tilde{\alpha}}{\lambda}$$
(35)

Using the relation (33) and (35), we obtain

$$Z_{xx} Z_{xy}\ (Z_{yy} + \tau I)^{-1}\, Z_{yx}\, \tilde{\alpha} = \lambda^2\ Z_{xx} (Z_{xx} + \tau I)\ \tilde{\alpha}$$
(36)

Then

$$Z_{xy}\ (Z_{yy} + \tau I)^{-1}\, Z_{yx}\, \tilde{\alpha} = \lambda^2\ (Z_{xx} + \tau I)\ \tilde{\alpha}$$
(37)

Let B be the lower triangular matrix of the complete Cholesky decomposition of $Z_{xx} + \tau I$ such that $Z_{xx} + \tau I = B\, B^T$ and let $\hat{\alpha} = B^T\, \tilde{\alpha}$, the equation (27) then becomes

$$B^{-1}\, Z_{xy}\ (Z_{yy} + \tau I)^{-1}\, Z_{yx}\, (B^{-1})^T\, \hat{\alpha} = \lambda^2\ \hat{\alpha}$$
(38)

It is a symmetric eigenvector problem of the form $A_4\ x = \lambda_4\, x$ with $A_4 = B^{-1}\, Z_{xy}\ (Z_{yy} + \tau I)^{-1}\, Z_{yx}\, (B^{-1})^T$ .The eigenvector problem (38) is equivalent to the eigenvector problem (27), with $\hat{\alpha} = B^T R_x^T\, \alpha$.

From the equations (7), (8) we have $u_x = f_x^T \Phi_x(X)$ and $v_y = f_y^T \Phi_y(Y)$, the problem is to find matrix A, such that

$$\min_A \left\| f_y^T \Phi_y(Y) - A f_x^T \Phi_x(X) \right\|^2$$
(39)

Using relations (12) and (13) we can write

$$f_x^T \Phi_x(X) = \alpha^T K_x$$
(40)

$$f_y^T \Phi_y(Y) = \beta^T K_y$$
(41)

The problem can now be written as follows:

$$\min_A \left\| \beta^T K_y - A \alpha^T K_x \right\|^2$$
(42)

The matrix $K_y = Y Y^T$, we obtain

$$Y Y^T = \left(\beta \beta^T\right)^{-1} \beta A \alpha^T K_x$$
(43)

The estimate output can be found from the equation (43).

IV. Simulations

We proceed to the identification of a Wiener Hammerstein using KCCA and KPLS methods, and we compare the performances of the RKHS models obtained with the both methods. The vector x of RKHS model have the structure:

$$x_i = \left(u(i-1), u(i-2), y(i-1)\right)^T$$
(44)

The system to be modelled is sketched by Figure 1. It consists on an electronic nonlinear system with a Wiener Hammerstein structure that was built by Gerd Vendesteen [13]. This process was adopted as a nonlinear system benchmark in SYSID 2009.

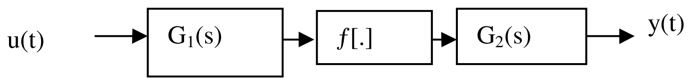

Fig. 1. Wiener Hammerstein benchmark.

To build the RKHS model we use the ERBF kernel (Extended Radial Basis Function) defined as:

$$k(x, x') = e^{-\gamma \|x - x'\|} \qquad \text{With } \gamma = 50$$
(45)

In simulation of both methods KCCA and KPLS, we have used 300 observations in the identification phase and 1600 other observations in the validation one.

In figure 2 we plot the RKHS model outputs using KCCA methods and the output process in the validation phase. As previously we notice the concordance between the outputs. The Means Square Error (MSE) % is equal to 0,1 %.

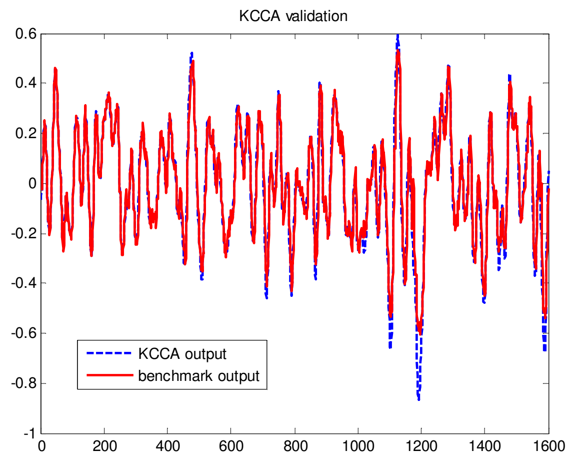

Fig. 2. Validation phase using KCCA.

In figure 3 we plot the RKHS model outputs using KPLS methods and the output process in the validation phase. As previously we notice the concordance between the outputs. The Means Square Error (MSE) % is equal to 0, 01 %.

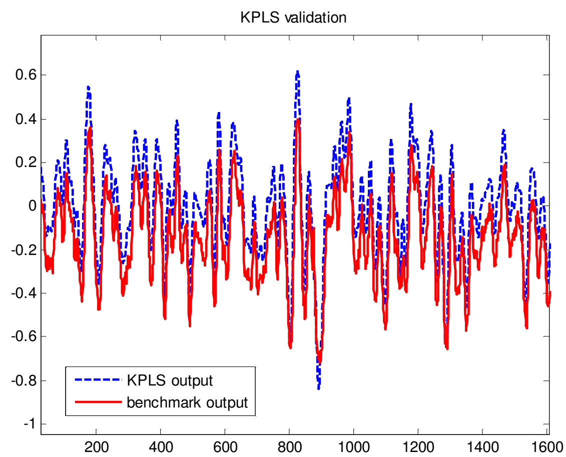

Fig. 3. Validation phase using KPLS.

In Table I we present the performances of both kernel methods. The first performance is the generalization ability evaluated by the Means Square Error (MSE) % in the identification phase and the validation one and the second concerns the compute time.

Table 1. Performances of kernel methods KCCA and KPLS

	KCCA	KPLS
MSE learning %	0,003	0,0034
MSE validation %	0,1	0,0161
Compute time (seconds)	1,8218	0,8281

VI. Conclusion

In this paper we have presented the KCCA method, it is a kernel method used to extract common features between highly non linear mapped data. We have used KCCA in the regression case to identify benchmark system. A comparative study with KPLS method has been achieved and shows the effectiveness of the method.

References

[1] N. Aronszajn, Theory of reproducing kernels, Transaction American Mathematical Society. 68, (3), 337–404, 1950.

[2] F. R. Bach And M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research* 3, 1-48, 2002.

[3] N. Cristianini, J. Shawe-Taylor, and H. Lodhi. Latent semantic kernels. Journal of intelligent Informations Systems, 18(2/3): 127-152. Special Issue on Automated Text Categorization, 2002.

[4] I. Elaissi, " Modélisation, Identification et Commande Prédictivedes systèmes non linéaires par utilisations des espaces RKHS", Thèse de doctorat, ENIT Tunis, Tunisia, 2009.

[5] K. Fukumizu et al. , " Statistical consistency of Kernel Canonical Correlation Analysis, " *Journal of Machine Learning Research*, vol. 8, pp.361-383, 2007.

[6] D. R. Hardoon et al., « Canonical correlations analysis : An overview with application to learning methods,"*Neural Computation*, vol. 16, no. 12, pp. 2639-2664,2004.

[7] S. Y. Huang, M. H. Lee, C. K. Hsiao, Nonlinear measures of association with kernel canonical correlation analysis and applications, Journal of Statistical Planning and Inference 139 (7) 2162- 2174, 2009.

[8] H. Hotelling, Relations between two sets of variates, Biometrika 28 (3/4) 321-377, 1936.

[9] J. Mercer, Functions of positive and negative type and their Connection with the theory of integral equations, Philosophical Transactions of the Royal Society. **209**, 415–446, 1909.

[10] R. Rosipal and L.J. Trecho "Kernel Partial Least Squares in Reproducing Kernel Hilbert Spaces" Journal of machine learning research 2; pp. 97-123, 2001.

[11] B. Scholkopf, Smola. A et Muller. K-R. Nonlinear component analysis as Kernel eigenvalue problem, Neural computation. v10, pp. 1299-1319, 1998.

[12] O. Taouali, I. Elaissi, T. Garna and H. Messaoud. Supervised Learning with Kernel methods. NOLASC '10, The 9th WSEAS International Conference on Non Linear Analysis, Nonlinear System and Chaos, Sousse, Tunisia, 3-6 may 2010

[13] G. Vandersteen Identification of linear and nonlinear systems in an errors-in-variables least squares and total least squares framework, PhD thesis, Vrije Universiteit Brusse, 1997.

[14] G. Wahba, An introduction to model building with Reproducing Kernel Hilbert Spaces. Technical report No 1020 (Department of Statistics, University of Wisconsin-Madison,2000).

[15] H. Wold. Soft Modelling by Latent Variables, the Nonlinear Iterative Partial Least Squares Approach. In J. Gani, editor, Perspectives in Probability and Statistics, pages 520-540. Academic Press, London, 1975.

[16] H. Wold: "Estimation of principal components and related models by iterative least squares" In *Multivariate Analysis*, Academic Press, NY pp. 391–420. 1966.

[17] S. Wold, H. Rube, H. Wold, and W.J. Dunn III. The collinearity problem in linear regression. The PLS approach to generalized inverse. SIAM journal of scientific and statistical computations, 5:735-743, 1984.