Identification of Optimal Kernel Parameters of RKHS model based on Genetics Algorithm

OKBA TAOUALI, NAJEH TAWFIK

Laboratory of Automatic Signal and Image Processing, National School of Engineers of Monastir, University of Monastir, 5019, Tunisia Rue Ibn El Jazzar 5019 Monastir; Tel: + (216) 73 500511, Fax: + (216) 73 500 514, TUNISIA taoualiok@gmail.com

Abstract—This paper proposes a new approach based on the Genetics Algorithm to determine the optimal kernel parameters of the Reproducing Kernel Hilbert Space (RKHS) model. These parameters are the width of the kernel function and the regularization parameter. The proposed meta-method has been tested to model some benchmarks such a benchmark of DC-Motor [21], a Wiener-Hammerstein benchmark [17], a Feedback's Process Trainer PT326 [20] and a Continuous Stirred Tank Reactor CSTR [22] and the results are satisfactory.

Keywords- RKHS, Kernel method, Regularization Networks, genetics algorithm, optimal parameters, Benchmarks model.

1. Introduction

Recently a new modelling technique of nonlinear systems developed on Reproducing Kernel Hilbert Space (RKHS) is proposed. This technique provided a new model entitled RKHS model [1], [2], [6], [10] and [24]. The solution is obtained by solving a quadratic optimization problem by using the learning algorithms such as support vector machines (SVM) [3], Regularization Networks (RN) [7] and Kernel Principal Component Analysis KPCA [5], [23]. These algorithms known as kernel methods are characterized by the kernel and the regularized parameters which should be determine in order to guaranty good generalization ability. In literature many techniques are used to determine these optimal parameters such as cross validation technique [9], Monte Carlo method [18]. In this paper, we propose a new approach to determine the optimal parameters of RKHS model which is based on the genetic algorithms.

The paper is organized as follows. In section 2, the RKHS model is reminded. The Regularized Networks method is presented in the section 3. Section 4 is devoted to the Genetic Algorithms. The optimization of RKHS Model Parameters with Genetic Algorithms is presented in section 5. Section 6 validates the proposed algorithm on the some benchmarks. Finally section 7 concludes the paper.

2. RKHS Model

Let $E \subset \mathbb{R}^d$ an input space and $k : E \times E \to \mathbb{R}$ is a continuous positive definite kernel. Let F_k is a Hilbert space associated to the kernel k. The kernel k is said to be a Reproducing Kernel of the Hilbert Space F_k if and only if:

$$\begin{cases} \forall x \in E, \quad k(x, .) \in F_k \\ \forall x, t \in E \text{ and } \forall f \in F_k , \\ _{F_k} = f(x) \end{cases}$$
(1)

In literature there exists many reproducing kernels such as the Radial Basis function (RBF):

$$k(x, t) = \exp(-||x-t||^2 / 2\mu^2) ; \forall x, t \in E$$
 (2)

The Exponential Radial Basis function (ERBF) given by:

$$k(x,t) = \exp\left(-\left\|x - t\right\|/2\sigma^2\right) \; ; \; \forall x, t \in E \tag{3}$$

The polynomial kernel written as :

$$k(x,t) = \exp\left(\left(\langle x,t \rangle + 1\right)^p\right) \; ; \; \forall \; x, \; t \; \in \; E \tag{4}$$

The sigmoidal kernel written as :

$$k(x,t) = \tanh((\alpha < x,t > +\beta)) \; ; \; \forall \; x, \; t \; \in E$$
(5)

with $\mu, \sigma, p, \alpha, \beta$ are a kernel parameter chosen in order to obtain a good generalization ability

Let's define the application Φ [1], $\Phi: E \to \mathbb{R}^l$ that transform the input data in the feature space F_k :

$$< \Phi(x), \Phi(t) > = k(x, t) x, t \in E$$
 (6)

Let's consider a set of data $\{x^{(i)}, y^{(i)}\}_{i=1, \dots, M}$ with $x^{(i)} \in \mathbb{R}^n$, $y^{(i)} \in \mathbb{R}$ are respectively the system input and output. The identification problem in the RKHS F_k can be

formulated as a minimization of the regularized empirical risk.

$$f^{*} = \min_{f \in F_{k}} \frac{1}{M} V\left(y^{(i)}, f\left(x^{(i)}\right)\right) + \lambda \|f\|_{F_{k}}^{2}$$

$$= \min_{f \in F_{k}} \frac{1}{M} \sum_{i=1}^{M} \left(y_{i} - f\left(x_{i}\right)\right)^{2} + \lambda \|f\|_{F_{k}}^{2}$$
(7)

where V is a cost function and λ is a regularization parameter chosen in order to guarantee a good prediction error. The solution f^* [7] of the problem (7) is written:

$$f^{*}(x) = \sum_{i=1}^{M} a_{i}^{*} k(x^{(i)}, x)$$
(8)

The solution (8) can be determined using kernel methods such as Regularization Network (RN) [3], Support Vector machine [6] (SVM), Reduce Kernel principal Component Analysis (RKPCA) [1].

3. Regularization Network Method (RN)

The Regularized Network technique is based on the cost function V:

$$V(y^{(i)}, f(x^{(i)})) = (y^{(i)} - f(x^{(i)}))^{2}$$
(9)

The parameters $\{a_i\}$ of the solution f^* are determined as following:

$$a_{i} = \sum_{j=1}^{M} \left(K + \lambda M I_{M} \right)_{i,j}^{-1} y^{(j)}$$
(10)

The relation (10) is written in the matrix form as:

$$\mathbf{A} = \left(K + \lambda M I_{M}\right)^{T} Y \tag{11}$$

where: K is the Gram matrix that satisfy:

$$K_{ij} = k \left(x^{(i)}, x^{(j)} \right), A = \left(a_1 \ a_2 \dots a_M \right)^T$$

and $Y = \left(y^{(1)}, y^{(2)}, \dots, y^{(M)} \right)^T$ (12)

4. The proposed approach based on Genetics Algorithm

In order to choose the optimal kernel and regularized parameters, the classical approaches fix randomly these parameters to obtain a suitable RKHS model. These obtained parameters don't provide an optimal solution.

In literature, many methods of optimization such as a cross-validation technique [19] and Monte Carlo [18] have been proposed. Using Genetic Algorithms (GAs) to obtain an optimal solution has attracted growing interest in many researches works [12], [13]. The novelty of this approach is the assumptions commonly used with conventional methods to ensure convergence of the solution [15], [16]. The application of these tools to determine the RKHS parameter is very interesting. In this section the principle of Genetic Algorithms is presented.

The GAs algorithms are applied to variety problems [13], [15]. The GAs applies an evolutionary approach to inductive learning. GA has been successfully applied to problems that are difficult to solve using conventional techniques such as scheduling problems, network routing problems and financial marketing. After fixing the expression of the objective function to be optimized, probabilistic steps are involved to create an initial population of individuals [14].

5. Optimization of RKHS Model Parameters with Genetic Algorithms

The determination of RKHS model needs to determine the function f^* eq. (8) with an optimal choice of the kernel parameter (for example μ) and the regularization parameter λ . This choice has no unique solution. The methods used in literatures to determine these parameters are all deterministic and the solutions that provided haven't a good prediction error. These methods can generate overfitting of RKHS model.

Solving this problem by Genetic Algorithms provides an optimal solution which guarantee better generalization ability and overcoming the drawbacks of deterministic methods. Using the GAs calculate these parameters requires the to development of an objective function f^* that must take into account, simultaneously, the two parameters μ and λ .

Each couple of parameters used for the elaboration of the final model should minimize the function f^* . In most works, in order to obtain an optimal RKHS model, authors require random values for searching parameters.

In this paper, we determine the function f^* tacking account of the learning phase error that is less than a threshold $\varepsilon = 10^{-3}$. This minimization can expand the number of solutions of the two parameters μ and λ which provides better prediction error. The value of this threshold is chosen empirically and can be adjusted depending on the application processed. After determining the objective function to be minimized, we generate, randomly, a population of individuals $P_i = (\mu_i, \lambda_i)$ abilities of individuals P_i which are evaluated by the function f^* . Individuals with the highest skills are selected to undergo different genetic operators (crossover, mutation and selection). After a set number of generations G_{max} , the genetic algorithm converges to the global optimum.

The optimization steps are as follows.

5.1 Initialization

It is usually random and it is often advantageous to include the maximum knowledge about the problem.

5.2 Evaluation

This step consist to compute the quality of individuals by the allocation a positive value entitled "ability or fitness" to each one. The highest is assigned to the individual that minimizes (or maximizes) the objective function.

The fitness of an individual is calculated as follows:

Fitness(Pos) =
$$2 - P_s + \frac{2(P_s - 1)(Pos - 1)}{Nind - 1}$$
 (13)

The evaluation is characterized by a parameter called selection pressure (P_s) . This method allows P_s values in the range of $\{1,2\}$

5.3 The Selection

This step selects a definite number of individuals of the current population. The selection is probabilistic. It is based on the ability of individuals a way that the best ones have a chance of being selected more than once. In this step is assigned to each individual probability P_i which is proportional to its fitness and defined by:

$$P_i = \frac{F_i}{\sum_{j=1}^M F_j} \tag{14}$$

with F_i the fitness of individual *i* and *M* the size of population.

5.4 The Crossover

The genetic crossover operator creates new individuals. From two randomly selected parents, crossover produces two descendants. This step affects only a limited number of individuals established by the crossover rate (*Pc*) number. Let $X = (x_i)_{1 \le i \le m}$ and $Y = (y_i)_{1 \le i \le m}$ be two individuals. These two parents will produce two offspring $X' = (x'_i)_{1 \le i \le m}$ and $Y' = (y'_i)_{1 \le i \le m}$ according to the equation:

$$\begin{cases} x'_{i} = x_{i} + s_{i} \ r_{i} \ a \frac{y_{i} - x_{i}}{\|Y - X\|} \\ y'_{i} = y_{i} + s_{i} \ r_{i} \ a \frac{x_{i} - y_{i}}{\|Y - X\|} \end{cases}$$
(15)

with : $a = 2^{-ku}$, k : mutation precision $k \in \{4, 5, ..., 20\}$, $u \in [0,1]$, $r_i = r \times domian$, $s_i = \{-1, 1\}$

5.5 The Mutation

Mutation is used mainly to break the stagnation in improvement by introducing new genetic information into the population. It consists in providing a small disruption to a number (Pm) of individuals. The effect of this operator is to counteract the attraction exerted by the best individuals. This allows us to explore other areas of the search space.

Let u_i and l_i be the respective lower and upper bounds for all individuals. Let $X = (x_i)_{1 \le i \le m}$ the individual to mutate that will give the new individual $X' = (x'_i)_{1 \le i \le m}$ according to:

$$x'_{i} = \begin{cases} x_{i} + (l_{i} - x_{i})f(G) & \text{if } r_{1} < 0.5 \\ x_{i} - (x_{i} - u_{i})f(G) & \text{if } r_{1} \ge 0.5 \\ x_{i} & \text{if } x'_{i} \notin [u_{i}, l_{i}] \end{cases}$$
(16)

with :

$$f(G) = \left[r_2 \left(1 - \frac{G}{G_{\text{max}}} \right) \right]^{b_s}$$

 r_1 r_2 : uniform random number between 0 and 1, G: the current generation, G_{max} : the maximum number of generations and b_s : shape parameter

Algorithm:

- 1- Repeat until G_{max} (the maximum number of generations)
- 2- Generate randomly an initial population of *N* pair of kernel parameters (μ, λ)
- 3- for each kernel parameters pair (μ, λ) do
- compute the fitness $f^*(x) = \sum_{i=1}^{M} a_i^* k(x^{(i)}, x)$
- 4- Apply crossover to produce new vectors
- 5- Apply mutation
- 6- Choose better couple (μ, λ)
- 7- Reinsertion of better pair of kernel parameters
- 8- Go to 1 and repeat until optimum value of parameters (μ, λ) .

6. Experimental Results

The proposed genetic approach has been tested to identify a benchmark of DC-Motor [21], a Wiener-Hammerstein benchmark [17], a Feedback's Process Trainer PT326 [20] and a Continuous Stirred Tank Reactor CSTR [22].

The performances of the resulting RKHS model are evaluated using genetic algorithm. This performance is formulated as: The Mean square Error (MSE) which calculates the cumulated error between the process output and the RKHS model output.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} \left(\tilde{y}^{(i)} - y^{(i)} \right)^2$$
(17)

where $y^{(i)}$ and $\tilde{y}^{(i)}$ are respectively the system ant the model outputs.

6.1 Benchmark of DC-Motor

6.1.1 Process description

The process to be identified is sketched by Fig. 1. The data contains two outputs, the first is the angular position (rad) and the second is the angular velocity (rad/s) of the motor shaft. The input u (V) is the voltage applied to the DC-motor. We have collected 300 input/output observations from the process at a sampling time of 0.1 s.



Fig.1. Schematic diagram of a DC-motor

In Figs.2 and 3, the evolution of the input voltage u and the output Angular velocity are respectively presented.



Fig.2. Input signal u (V)



Fig.3. Output signal (rad/s)

6.1.2 Results

The RBF kernel is used to build a RKHS model. The observations used in the identification and validation phases are collected over two separate windows and their numbers are respectively 80 and 220. The input vector $x \in \mathbb{R}^d$ of RKHS model is given by:

$$x^{(i)} = \begin{bmatrix} u^{(i)} & y^{(i-1)} \end{bmatrix}^T$$
(18)

The optimal RKHS parameter are determined using genetic algorithm and the values are respectively $\mu = 21$ and $\lambda = 10^{-4}$.

In Figs. 4 and 5, the RKHS model and the benchmark outputs in the learning and validation phases are presented. We remark that the model output is in concordance with the system output. The Mean Square Error (MSE) is equal to 9.8910^{-4} % in the learning phase and 0.0063% in the validation one. This shows the good performances of the Genetic Algorithms.



Fig.4. Process and RKHS model outputs during the identification phase



Fig.5. Process and RKHS model outputs during the validation phase

6.2 Application to the benchmark: Winner Hammerstein Benchmark

6.2.1 Process description

The process to be identified is sketched by Fig. 6. It consists on an electronic nonlinear system with a Wiener Hammerstein structure that was built by Gerd Vendesteen [17]. This benchmark represents a challenge to identify using kernel methods.



Fig.6. Wiener Hammerstein structure

6.2.2 Results

The input vector x is:

$$x^{(k)} = \begin{cases} u^{(k-1)}, u^{(k-2)}, u^{(k-4)}, \dots, \\ u^{(k-15)}, y^{(k-1)} \end{cases}^T \in \mathbb{R}^{15}$$
(19)

The Exponential Radial Basis Function kernel (ERBF) is used to construct a RKHS model. The

observations uses in the identification and validation phases are respectively 200 and 500 new data. The optimal RKHS parameter are determined using genetic algorithm and the values are respectively $\sigma = 29$ and $\lambda = 1.2 \ 10^{-7}$.

In Figs.7 and 8, the RKHS model and the benchmark outputs in the learning and validation phases are presented. We notice that the model output is in concordance with the system output, indeed the mean Square Error (MSE) is equal to 0.02% in the learning phase and 0.33% in the validation one. This illustrates the advantages of the proposed approach.



Fig.7. Process and RKHS model outputs during the identification phase



Fig.8. Process and RKHS model outputs during the validation phase

6.3 Application to the benchmark: Feedback's Process Trainer PT326

6.3.1 Process description

The Feedback's Process Trainer PT326 (Figs.9) [20] is a benchmark system for identification. The device's function is like a hair dryer where the air is fanned through a tube and heated at the inlet. The air temperature is measured by a thermocouple at the outlet. The input u is the voltage over a mesh of resistor wires to heat the incoming air; the output T is the outlet air temperature.



Fig.9. Process Trainer PT 326

We have collected 600 input/output observations from the process at a sampling time of 0.08 s. In Figs.10 and 11, the evolution of the input voltage u and the output temperature are respectively presented.





Fig.11. Output signal (T°C)

Fig.10 Input signal u (V)

6.3.2 Results

To build the RKHS model we use the RBF kernel. We have used 200 observations in the identification phase and 350 new observations in the validation one. The input vector $x \in \mathbb{R}^d$ of RKHS model is given by:

$$x^{(i)} = \begin{bmatrix} u^{(i)} & y^{(i-1)} \end{bmatrix}^T$$
(20)

The optimal RKHS parameter are determined using genetic algorithm and the values are respectively $\mu = 100$ and $\lambda = 9 \ 10^{-5}$.

In Figs 12 and 13, the RKHS model and the benchmark outputs in the learning and validation phases are presented. We remark that the model output is in concordance with the system output. Therefore, the mean Square Error (MSE) is equal to 0.0027% in the learning phase and 0.0026% in the validation one. This illustrates the advantages of the proposed approach.



Fig.12. Process and RKHS outputs during the identification phase



Fig.13. Process and RKHS outputs during the validation phase

6.4 Chemical reactor modeling

6.4.1 Process description

The system is a Continuous Stirred Tank Reactor CSTR which is a nonlinear system used for the conduct of the chemical reactions [22]. A diagram of the reactor is presented in Figure 14.



 C_h : Concentration product

Fig.14. Chemical reactor Diagram The physical equations describing the system are:

$$\frac{dh(t)}{dt} = w_1(t) + w_2(t) - 0, 2\sqrt{h(t)}$$

$$\frac{dC_b(t)}{dt} = (C_{b1} - C_b(t))\frac{w_1(t)}{h(t)} + (C_{b2} - C_b(t))\frac{w_2(t)}{h(t)} - \frac{k_1 \cdot C_b(t)}{(1 + k_2 \cdot C_b(t))^2}$$
(21)

where: h(t) is the height of the mixture in the reactor of the feed of reactant 1 w_1 (resp, reactant 2, w_2) with concentration Cb_1 (resp. Cb_2). The feed of product of the reaction is w_0 and its concentration is C_b . k_1 and k_2 are consuming reactant rate. The temperature in the reactor is assumed constant and equal to the ambient temperature. We are interested by modelling the subsystem presented in Figure 15.



Fig.15. Considered subsystem

6.4.2 Results

For the purpose of the simulations, the CSTR model of the reactor provided with Simulink of Matlab is used.

The input vector is:

$$x^{(k)} = \left[w_1^{(k)}, cb^{(k-1)}, cb^{(k-2)} \right]^T$$
(22)

The Sigmoid kernel (tahn) is used to construct a RKHS model. The observations uses in the identification and validation phases are respectively 100 and 200 new data. The optimal RKHS parameter are determined using Genetic Algorithm

and the values are respectively $\alpha = 20, \beta = 1$ and $\lambda = 10^{-2}$.

In Figs.16 and 17, the RKHS model and the benchmark outputs in the learning and validation phases are presented. We remark that the model output is in concordance with the system output, indeed the mean Square Error (MSE) is equal to 4.5810^{-8} % in the learning phase and 5.510^{-8} % in the validation one. This show the advantages of the proposed approach.



Fig.16. Process and RKHS outputs during the identification phase



Fig.17.Process and RKHS outputs during the validation phase

7. Conclusion

In this paper, a new approach based on Genetic Algorithms is proposed in order to determine the optimal parameters of the RKHS model. Through several results, we showed the advantages and the effectiveness of the proposed genetic algorithm in term of prediction error. This algorithm has been tested for modelling three benchmarks and the results were satisfactory.

References:

- Taouali. O, Elaissi. I and Messaoud .H, Online Prediction Model based on Reduced Kernel Principal Component Analysis. *Neural Computing and Application journal*, Vol.21, No.1, 2012, pp.161-169.
- [2] Taouali. O, Elaissi. I and Messaoud. H, Design and Comparative Study of online kernel methods identification of nonlinear system in RKHS Space. *Artificial Intelligence Review*, Vol.37, 2012, pp.289-300.
- [3] Veropoulos. K, N. Cristianini and Campbell. C, The Application of Support Vector Machines to Medical Decision Support: A Case Study, *ACAI conference*, Chania, Greek Islands, 1999, pp. 17-21.
- [4] Vapnik. V. N, *Statistical Learning Theory*, Wiley, New York, 1998.
- [5] Scholkopf. B, Smola. A et Muller. K-R. Nonlinear component analysis as Kernel eigenvalue problem, *Neural computation*, Vol.10, 1998, pp. 1299-1319.
- [6] Scholkopf. B, Smola. A, *Learning with Kernel:* Support Vector Machine Regularization, Optimization and Beyond, MIT Press, 2001.
- [7] Wahba. G, An introduction to model building with Reproducing Kernel Hilbert Spaces, Technical report No 1020, Department of Statistics, University of Wisconsin-Madison, 2000.
- [8] Jeffal. I, Taouali. O, Elaissi. I and H. Messaoud, A new online fault detection method based on PCA technique. *IMA Journal of Mathematical Control and Information*, Vol.31, 2014, pp.487-499.
- [9] Elaissi. I, I. jaffal, Taouali. O and Messaoud. H, Online Prediction Model based SVD-KPCA methods, *ISA transactions*, Vol.52, 2013, pp.96-104.

- [10] Taouali. O, Saidi .N and Messaoud. H, Identification of Non linear MISO process using RKHS and Volterra models. Wseas Transactions on System, Vol.8, 2009, pp.723-732.
- [11] Chan . W. C, Chan. C. W, K. C. Cheung, and C.J. Harris, Modelling of nonlinear dynamical systems using support vector neural networks. Engineering Applications of Artificial *Intelligence*, Vol.14, 2001, pp.105-113.
- [12] Goldberg D, Korb B, Deb K, Messy genetic algorithms Motivation, analysis, and first results. *Journal of Complex Systems*, Vol.5, No.3, 1989, pp.493-530.
- [13] Bies. R , Muldoon. F, Pollock. G, Manuck. S, Smith. M, A Genetic Algorithm-Based Hybrid Machine Learning Approach to Model Selection. Journal of Pharmacokinetics and pharmacodynamics, Vol.33, No.2, 2006, pp.196-221.
- [14] Cha. S. H, Tappert. C, A genetic Algorithm for Constructing Compact Binary Decision Trees. Journal of Pattern Recognition Research, Vol.4, No.1, 2009, pp.1-13.
- [15] Akbari. Z, A multilevel evolutionary algorithm for optimizing numerical functions, *International Journal of Industrial Engineering Computations* Vol.2, 2011, pp.419-430.
- [16] Zhang. J, Chung. H, Lo. W, Clustering-Based Adaptive Crossover and Mutation Probabilities for Genetic Algorithms, *IEEE Transactions on Evolutionary Computation*, Vol.11, No.3, 2007, pp.326-335.
- [17] Vandersteen. G, *Identification of linear and nonlinear systems in an errors-in-variables least squares and total least squares framework.* These of doctorat, Université Brussel, Belgique ,1997.
- [18] Jolliffe. I.T, *Principal components analysis*, Springer-Verlag, New York, 1986.
- [19] Taouali. O, Elaissi. I, Villa. N and Messaoud. H, Identification of NonLinear Multivariable Processes Modelled on Reproducing Kernel Hilbert Space: Application to Tennessee Process, ICONS, 2nd IFAC International Conference on Intelligent Control Systems and Signal Processing, Turquie, 2009, pp. 1-6.
- [20] Ljung. L, System Identification: Theory for the User, Englewood Cliffs, NJ: Prentice Hall, 1987.
- [21] Ljung. L, System Identification Toolbox, Create linear and nonlinear dynamic system models from measured input-output data, MathWorks, 2007.

- [22] H. Demuth. M. Beale, M. Hagan, *Neural Network Toolbox 5, User's Guide*, The MathWorks, 2007.
- [23] Nathalie. P and Jean-François. B, Principal Component Analysis for Greenhouse Modelling, Wseas Transactions on Systems, Issue 1, Vol.7, No.1, 2008, pp.24-30.
- [24] Saidi. N, Taouali. O and Messaoud. H, Modelling of a SISO and MIMO nonlinear communication channel using two modelling techniques, *Wseas Transactions on Circuits* and Systems, Vol.8, No.6, 2009, pp.518-528.
- [25] Tian-Zhen. W, Man Xu, Tian-Hao .T, Christophe. C, A Fault Detection Method Based on Dynamic Peak-valley Limit under the Non-Steady Conditions. Wseas Transactions on Circuits and Systems. Vol.12, No.4, 2013, pp.118-128.