The Seventeenth International Symposium on Artificial Life and Robotics 2012 (AROB 17th '12),
B-Con Plaza, Beppu, Oita, Japan, January 19-21, 2012

Identification of continuous-time Hammerstein systems using Gaussian process models trained by particle swarm optimization

Tomohiro Hachino[1] and Shouichi Yamakawa[1]

[1] Kagoshima University, Kagoshima 890-0065, Japan
(Tel & Fax: 81-99-285-8392)
(hachino@eee.kagoshima-u.ac.jp)

Abstract: This paper deals with a nonparametric identification of continuous-time Hammerstein systems using Gaussian process (GP) models. A Hammerstein system consists of a memoryless nonlinear static part followed by a linear dynamic part. The identification model is derived using the GP prior model which is described by the mean function vector and the covariance matrix. This prior model is trained by the separable least-squares approach combining particle swarm optimization with the linear least-squares method to minimize the negative log marginal likelihood of the identification data. Then the nonlinear static part is estimated by the predictive mean function of the GP, and the confidence measure of the estimated nonlinear static part is evaluated by the predictive covariance function of the GP. Simulation results are shown to illustrate the proposed method.

Keywords: continuous-time system, Gaussian process model, Hammerstein system, particle swarm optimization, system identification

1 INTRODUCTION

The Hammerstein system is expressed by a memoryless nonlinear static part followed by a linear dynamic part and has many advantages for control design or stability analysis due to its simple model structure. Numerous identification methods have been proposed for discrete-time Hammerstein systems [1, 2, 3]. Since these approaches are categorized into parametric identification, one needs to use a large number of weighting parameters to describe the nonlinearity and handle a complicated model structure determination. Moreover, no confidence measures for the estimated nonlinear static part are obtained in parametric identification. On the other hand, we have developed a nonparametric identification method based on the Gaussian process (GP) models, which can give the estimated nonlinear function with the confidence measure [4]. However, since most practical systems are usually continuous-time, it is very important to develop an accurate identification method for continuous-time Hammerstein systems.

Therefore, in this paper, we discuss the nonparametric identification of continuous-time Hammerstein systems using the GP model. The identification model is derived using the GP prior model which is described by the mean function vector and the covariance matrix. The prior mean function is represented in linear form of the input and the prior covariance function is expressed by the Gaussian kernel. This prior model is trained by the separable least-squares (LS) approach combining particle swarm optimization (PSO) [5] with the linear LS method to minimize the negative log marginal likelihood of the identification data. PSO is a swarm intelligence optimization technique, which was inspired by the social behavior of a flock of birds or a shoal of fish, and has been empirically shown to be very efficient for optimization. The use of PSO might increase the efficiency of identification due to its simple algorithm. The hyperparameters of the covariance functions and the numerator parameters of the linear dynamic part are represented with the particles and are searched by PSO, while the denominator parameters of the linear dynamic part are estimated by the linear LS method. Then the nonlinear static part is estimated by the predictive mean function of the GP, and the confidence measure of the estimated nonlinear static part is evaluated by the predictive covariance function of the GP. Simulation results show that the accuracy of the proposed method is superior to that of a conventional identification method.

2 STATEMENT OF THE PROBLEM

Consider a single-input, single-output continuous-time nonlinear system described by the Hammerstein model shown in Fig. 1. This system can be mathematically described as

$$\left\{ \begin{array}{l} \displaystyle\sum_{i=0}^{n} a_i p^{n-i} y(t) = b_0 x(t) \qquad (a_0 = 1) \\ x(t) = f(u(t)) + \epsilon(t) \end{array} \right. \tag{1}$$

where $u(t)$ and $y(t)$ are input and output signals, respectively. $x(t)$ is intermediate signal that is not accessible for measurement. $f(\cdot)$ is unknown nonlinear function, which is assumed to be stationary and smooth. $\epsilon(t)$ is assumed to be a zero-mean Gaussian noise with variance σ_n^2. $A(p) = \sum_{i=0}^{n} a_i p^{n-i}$ is the denominator polynomial of the linear dynamic part, where p denotes a differential operator. n is assumed to be known.

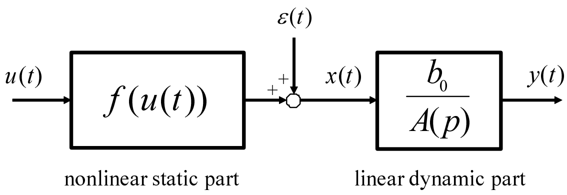

Fig. 1. Continuous-time Hammerstein system

The aim of this paper is to identify the system parameters $\{a_i\}$ and $\{b_j\}$ of the linear dynamic part, and the nonlinear static function $f(\cdot)$ with the confidence measure, from input and output data in the GP framework.

3 IDENTIFICATION MODEL BY THE GP

The following state variable filter $F(p)$ is introduced in order to evaluate higher order derivatives of signals:

$$F(p)=\frac{1}{p^q+\gamma_1 p^{q-1}+\cdots\cdots+\gamma_q} \quad (q>n) \tag{2}$$

Multiplying both sides of Eq. (1) by $F(p)$ yields

$$\begin{cases} \sum_{i=0}^{n} a_i p^{n-i} y^f(t) = b_0 x^f(t) \\ x^f(t) = F(p)f(u(t)) + \epsilon^f(t) \end{cases} \tag{3}$$

where $y^f(t) = F(p)y(t)$, $x^f(t) = F(p)x(t)$ and $\epsilon^f(t) = F(p)\epsilon(t)$. When $F(p)$ has a transport lag characteristic, the filter $F(p)$ and the nonlinear function $f(\cdot)$ are exchangeable and it follows that $F(p)f(u(t)) = f(F(p)u(t)) = f(u^f(t))$. Thus Eq. (3) becomes

$$\begin{cases} \sum_{i=0}^{n} a_i p^{n-i} y^f(t) = b_0 x^f(t) \\ x^f(t) = f(u^f(t)) + \epsilon^f(t) \end{cases} \tag{4}$$

In general the Butterworth filter has approximately a transport lag characteristic for frequencies $\omega \le \omega_c$, where ω_c is the cutoff frequency. Therefore, the Butterworth filter is utilized as the delayed state variable filter $F(p)$ in this paper.

Putting $t = t_1, t_2, \cdots, t_N$ into Eq. (4) yields

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\theta}_a + b_0\boldsymbol{x} \tag{5}$$

where

$$\begin{aligned} \boldsymbol{y} &= [p^n y^f(t_1), p^n y^f(t_2), \cdots, p^n y^f(t_N)]^{\mathrm{T}} \\ \boldsymbol{x} &= [f(u^f(t_1)) + \epsilon^f(t_1), f(u^f(t_2)) + \epsilon^f(t_2), \\ &\qquad \cdots, f(u^f(t_N)) + \epsilon^f(t_N)]^{\mathrm{T}} \\ \boldsymbol{\theta}_a &= [a_1, a_2, \cdots, a_n]^{\mathrm{T}} \\ \boldsymbol{A} &= \begin{bmatrix} -p^{n-1}y^f(t_1) & \cdots & -y^f(t_1) \\ -p^{n-1}y^f(t_2) & \cdots & -y^f(t_2) \\ \vdots & & \vdots \\ -p^{n-1}y^f(t_N) & \cdots & -y^f(t_N) \end{bmatrix} \end{aligned} \tag{6}$$

A GP is a Gaussian random function and is completely described by its mean function and covariance function. We can regard it as a collection of random variables with a joint multivariable Gaussian distribution. Therefore, the function values $\boldsymbol{f}$ can be represented by the GP:

$$\boldsymbol{f} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{u}), \boldsymbol{\Sigma}(\boldsymbol{u}, \boldsymbol{u})) \tag{7}$$

where

$$\begin{aligned} \boldsymbol{f} &= [f(u^f(t_1)), f(u^f(t_2)), \cdots, f(u^f(t_N))]^{\mathrm{T}} \\ \boldsymbol{u} &= [u^f(t_1), u^f(t_2), \cdots, u^f(t_N)]^{\mathrm{T}} \end{aligned} \tag{8}$$

$\boldsymbol{u}$ is the input (variable) of the function $\boldsymbol{f}$, $\boldsymbol{m}(\boldsymbol{u})$ is the mean function vector, and $\boldsymbol{\Sigma}(\boldsymbol{u}, \boldsymbol{u})$ is the covariance matrix. In this paper the mean function is expressed as $m(u^f(t)) = u^f(t)$, i.e., the mean function vector $\boldsymbol{m}(\boldsymbol{u})$ is described as follows:

$$\boldsymbol{m}(\boldsymbol{u}) = \boldsymbol{u} \tag{9}$$

The covariance $\Sigma_{pq} = s(u(t_p), u(t_q))$ is an element of the covariance matrix $\boldsymbol{\Sigma}$, which is a function of $u(t_p)$ and $u(t_q)$. Under the assumption that the nonlinear function $f(\cdot)$ is stationary and smooth, the following Gaussian kernel is utilized in this paper:

$$\begin{aligned} \Sigma_{pq} &= s(u(t_p), u(t_q)) \\ &= \sigma_y^2 \exp\left(-\frac{|u(t_p) - u(t_q)|^2}{2\ell^2}\right) \end{aligned} \tag{10}$$

From Eq. (7), the intermediate signal vector can be written as

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}(\boldsymbol{u}), \boldsymbol{K}(\boldsymbol{u}, \boldsymbol{u})) \tag{11}$$

where

$$\begin{aligned} &\boldsymbol{K}(\boldsymbol{u}, \boldsymbol{u}) = \boldsymbol{\Sigma}(\boldsymbol{u}, \boldsymbol{u}) + \sigma_n^2 \boldsymbol{I}_N \\ &\boldsymbol{I}_N : N \times N \ \text{identity matrix} \end{aligned} \tag{12}$$

and $\boldsymbol{\theta}_c = [\sigma_y, \ell, \sigma_n]^{\mathrm{T}}$ is called the *hyperparameter* vector.

Applying the property of the multivariable Gaussian distribution for the linear transformation to Eqs. (5) and (11), the identification model by the GP is derived as

$$\boldsymbol{y} \sim \mathcal{N}(b_0\boldsymbol{m}(\boldsymbol{u}) + \boldsymbol{A}\boldsymbol{\theta}_a, b_0^2\boldsymbol{K}(\boldsymbol{u}, \boldsymbol{u})) \tag{13}$$

4 IDENTIFICATION

4.1 Training by PSO

First, the GP prior model is trained by optimizing the unknown parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_a^{\mathrm{T}}, b_0, \boldsymbol{\theta}_c^{\mathrm{T}}]^{\mathrm{T}}$. Although this is a nonlinear optimization problem, we can separate the linear optimization part and the nonlinear optimization part. Therefore, in this paper, we propose a separable LS approach combining PSO with the linear LS method. Only $\boldsymbol{\Omega} = [b_0, \boldsymbol{\theta}_c^{\mathrm{T}}, \omega_c]^{\mathrm{T}}$ is represented with the particles and searched by PSO. The proposed training algorithm is as follows:

Step 1: Initialization

Generate an initial population of Q particles with random positions $\boldsymbol{\Omega}^0_{[i]} = [b_{0[i]}, \boldsymbol{\theta}^{\mathrm{T}}_{c[i]}, \omega_{c[i]}]^{\mathrm{T}}$ and velocities $\boldsymbol{V}^0_{[i]}$ $(i = 1, 2, \cdots, Q)$.

Set the iteration counter l to 0.

Step 2: Filtering of the identification data

Construct Q candidates of the state variable filter using $\omega_{c[i]}$. Calculate the filtered input $u^f_{[i]}(t)$, filtered output $y^f_{[i]}(t)$ and their higher-order derivatives, using each candidate of the state variable filter. Then construct Q candidates of $\boldsymbol{y}_{[i]}$, $\boldsymbol{A}_{[i]}$ and $\boldsymbol{u}_{[i]}$ $(i = 1, 2, \cdots, Q)$.

Step 3: Construction of covariance matrix

Construct Q candidates of covariance matrix $\boldsymbol{K}_{[i]}$ using $\boldsymbol{\theta}_{c[i]}$ $(i = 1, 2, \cdots, Q)$.

Step 4: Estimation of $\boldsymbol{\theta}_{a[i]}$

Estimate Q candidates for $\boldsymbol{\theta}_{a[i]}$ corresponding to $\boldsymbol{\Omega}_{[i]}$ $(i = 1, 2, \cdots, Q)$:

$$\boldsymbol{\theta}_{a[i]} = (\boldsymbol{A}^{\mathrm{T}}_{[i]}\boldsymbol{\mathcal{K}}^{-1}_{[i]}\boldsymbol{A}_{[i]})^{-1}\boldsymbol{A}^{\mathrm{T}}_{[i]}\boldsymbol{\mathcal{K}}^{-1}_{[i]}(\boldsymbol{y}_{[i]} - b_{0[i]}\boldsymbol{u}_{[i]}) \quad (14)$$

where $\boldsymbol{\mathcal{K}}_{[i]} = b^2_{0[i]}\boldsymbol{K}_{[i]}$.

Step 5: Evaluation value calculation

Calculate the negative log marginal likelihood of the identification data:

$$\begin{aligned} J(\boldsymbol{\Omega}^l_{[i]}) &= \frac{1}{2}\log|\boldsymbol{\mathcal{K}}_{[i]}| \\ &+ \frac{1}{2}(\boldsymbol{y}_{[i]} - b_{0[i]}\boldsymbol{u}_{[i]} - \boldsymbol{A}_{[i]}\boldsymbol{\theta}_{a[i]})^{\mathrm{T}}\boldsymbol{\mathcal{K}}^{-1}_{[i]} \\ &\times(\boldsymbol{y}_{[i]} - b_{0[i]}\boldsymbol{u}_{[i]} - \boldsymbol{A}_{[i]}\boldsymbol{\theta}_{a[i]}) + \frac{N}{2}\log(2\pi) \end{aligned} \quad (15)$$

Step 6: Update of the best positions $\boldsymbol{pbest}$ and $\boldsymbol{gbest}$

Update $\boldsymbol{pbest}^l_{[i]}$, which is the personal best position, and $\boldsymbol{gbest}^l$, which is the global best position among all particles as follows:

If $l = 0$ then

$$\begin{aligned} &\boldsymbol{pbest}^l_{[i]} = \boldsymbol{\Omega}^l_{[i]} \\ &\boldsymbol{gbest}^l = \boldsymbol{\Omega}^l_{[i_{best}]} \quad i_{best} = arg\min_i J(\boldsymbol{\Omega}^l_{[i]}) \end{aligned} \quad (16)$$

otherwise

$$\boldsymbol{pbest}^l_{[i]} = \begin{cases} \boldsymbol{\Omega}^l_{[i]} & (J(\boldsymbol{\Omega}^l_{[i]}) < J(\boldsymbol{pbest}^{l-1}_{[i]})) \\ \boldsymbol{pbest}^{l-1}_{[i]} & (\text{otherwise}) \end{cases}$$

$$\boldsymbol{gbest}^l = \boldsymbol{pbest}^l_{[i_{best}]} \quad i_{best} = arg\min_i J(\boldsymbol{pbest}^l_{[i]}) \quad (17)$$

Step 7: Update of positions and velocities

Update the particle positions and velocities using Eq. (18):

$$\begin{cases} \boldsymbol{V}^{l+1}_{[i]} = w^l \cdot \boldsymbol{V}^l_{[i]} + c_1 \cdot rand_1() \cdot (\boldsymbol{pbest}^l_{[i]} - \boldsymbol{\Omega}^l_{[i]}) \\ \qquad\qquad + c_2 \cdot rand_2() \cdot (\boldsymbol{gbest}^l - \boldsymbol{\Omega}^l_{[i]}) \\ \boldsymbol{\Omega}^{l+1}_{[i]} = \boldsymbol{\Omega}^l_{[i]} + \boldsymbol{V}^{l+1}_{[i]} \end{cases} \quad (18)$$

where w^l is an inertia factor, c_1 and c_2 are constants representing acceleration coefficients, and $rand_1()$ and $rand_2()$ are uniformly distributed random numbers with amplitude in the range $[0, 1]$.

Step 8: Repetition

Set the iteration counter to $l = l + 1$ and go to Step 2 until the prespecified iteration number l_{max}.

Step 9: Determination of the GP prior model

Determine the vector $\hat{\boldsymbol{\Omega}} = [\hat{b}_0, \hat{\boldsymbol{\theta}}^{\mathrm{T}}_c, \hat{\omega}_c]^{\mathrm{T}}$ and the corresponding denominator parameters of the linear dynamic part $\hat{\boldsymbol{\theta}}_a = [\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_n]^{\mathrm{T}}$ using the best particle position $\boldsymbol{gbest}^{l_{max}}$. Construct the suboptimal prior covariance function:

$$\begin{aligned} &s(u^f(t_p), u^f(t_q)) \\ &\qquad = \hat{\sigma}^2_y \exp\left(-\frac{|u^f(t_p) - u^f(t_q)|^2}{2\hat{\ell}^2}\right) \end{aligned} \quad (19)$$

4.2 Estimation of the nonlinear static function

The estimates of the intermediate signal vector $\hat{\boldsymbol{x}}$ can be evaluated as follows:

$$\hat{\boldsymbol{x}} = \frac{1}{\hat{b}_0}(\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{\theta}}_a) \quad (20)$$

Let the test input vector and the corresponding function value vector and intermediate signal vector be $\boldsymbol{u}_*$, $\boldsymbol{f}_*$ and $\boldsymbol{x}_*$, respectively. Then the posterior distribution for $\boldsymbol{f}_*$ is obtained as

$$\boldsymbol{f}_*|\boldsymbol{u}, \hat{\boldsymbol{x}}, \boldsymbol{u}_* \sim \mathcal{N}(\bar{\boldsymbol{f}}_*, cov(\boldsymbol{f}_*)) \quad (21)$$

where $\bar{\boldsymbol{f}}_*$ is the predictive mean vector and $cov(\boldsymbol{f}_*)$ is the predictive covariance matrix, which are given as follows:

$$\begin{aligned} \bar{\boldsymbol{f}}_* &= \boldsymbol{m}(\boldsymbol{u}_*) + \boldsymbol{\Sigma}(\boldsymbol{u}_*, \boldsymbol{u})\boldsymbol{K}^{-1}(\hat{\boldsymbol{x}} - \boldsymbol{m}(\boldsymbol{u})) \\ cov(\boldsymbol{f}_*) &= \boldsymbol{\Sigma}(\boldsymbol{u}_*, \boldsymbol{u}_*) - \boldsymbol{\Sigma}(\boldsymbol{u}_*, \boldsymbol{u})\boldsymbol{K}^{-1}\boldsymbol{\Sigma}(\boldsymbol{u}, \boldsymbol{u}_*) \end{aligned} \quad (22)$$

Thus, the nonlinear static function of the objective system is estimated as

$$\begin{aligned} \hat{f}(u_*(t)) &= m(u_*(t)) \\ &+ \boldsymbol{\Sigma}(u_*(t), \boldsymbol{u})\boldsymbol{K}^{-1}(\hat{\boldsymbol{x}} - \boldsymbol{m}(\boldsymbol{u})) \end{aligned} \quad (23)$$

and its covariance function $\hat{s}$ is evaluated as

$$\begin{aligned} \hat{s}(u_*(t_p), u_*(t_q)) &= s(u_*(t_p), u_*(t_q)) \\ &- \boldsymbol{\Sigma}(u_*(t_p), \boldsymbol{u})\boldsymbol{K}^{-1}\boldsymbol{\Sigma}(\boldsymbol{u}, u_*(t_q)) \end{aligned} \quad (24)$$

The predictive covariance function $\hat{k}$ of the intermediate signal is obtained as

$$\hat{k}(u_*(t_p), u_*(t_q)) = \hat{s}(u_*(t_p), u_*(t_q)) + \hat{\sigma}^2_n\delta_{pq} \quad (25)$$

where δ_{pq} is the Kronecker delta, which is 1 if $p = q$ and 0 otherwise. Equations (24) and (25) are used as confidence measures of the estimated nonlinear static function and the intermediate signal, respectively.

5 NUMERICAL SIMULATIONS

Consider a continuous-time Hammerstein system described by

$$\begin{cases} \ddot{y}(t) + a_1\dot{y}(t) + a_2 y(t) = b_0 x(t) \\ x(t) = f(u(t)) + \epsilon(t) \\ a_1 = 3.0, \quad a_2 = 1.5, \quad b_0 = 1.0 \\ f(u(t)) = u(t) + 0.5u^3(t) \end{cases} \tag{26}$$

The output signal is generated by a random signal of band-pass 3.0[rad/s]. The sampling period is taken to be $T = 0.05$[s]. $\epsilon(t)$ is zero-mean Gaussian noise with a standard deviation σ_n of 2.1, which means the noise-to-signal ratio (NSR) is 20%. The number of input and output data is $N = 300$. The design parameters for PSO are chosen as follows: particle size: $Q = 50$, inertia factor: $w^l = w_{max} - (w_{max} - w_{min})l/l_{max}$ $(w_{max} = 0.9, w_{min} = 0.4)$, acceleration coefficients $c_1 = 1.0$, $c_2 = 0.9$, maximum iteration number $l_{max} = 200$.

Estimates of of the parameters in the linear dynamic part are $\hat{a}_1 = 3.072$ and $\hat{a}_2 = 1.490$, respectively, where estimate of b_0 is omitted because the final estimated model is normalized by $\hat{b}_0$. Fig. 2 and Fig. 3 show the estimated nonlinear function and the estimated intermediate signal with the double standard deviation confidence intervals, respectively.

Monte-Carlo simulations of 20 experiments are implemented for the proposed method and conventional RBF-based method. Table 1 shows the mean squares errors between the true outputs and the outputs of the estimated models for various values of σ_n on the average of 20 experiments.

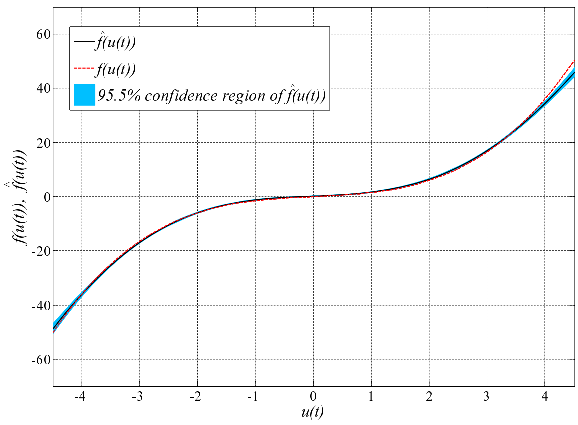

Fig. 2. Estimated nonlinear function

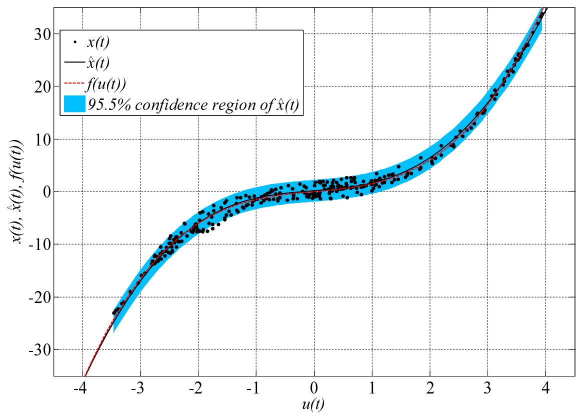

Fig. 3. Estimated intermediate signal

Table 1. Mean squares errors between the true outputs and the outputs of the estimated models

σ_n	NSR [%]	Proposed	RBF-based
1.1	10	2.17e-3	2.26e-3
1.6	15	5.02e-3	6.01e-3
2.1	20	7.41e-3	8.02e-3
2.6	25	1.68e-2	1.99e-2
3.1	30	2.14e-2	2.35e-2

6 CONCLUSIONS

In this paper a novel nonparametric identification method for continuous-time Hammerstein systems has been proposed using the GP model. The GP prior model is trained by the separable LS approach combining the linear LS method with PSO so that the negative log marginal likelihood of the identification data is minimized. The nonlinear static part of the objective Hammerstein system is estimated by the predictive mean function of the GP, and the confidence region of the estimated nonlinear static part is given by the predictive covariance function. Simulation results show that the accuracy of the proposed method is superior to that of a conventional identification method.

REFERENCES

[1] Al-Duwaish H, Karim M. N (1997), A new method for the identification of Hammerstein model, Automatica, 33(10):1871–1875

[2] Adachi S, Murakami H (1995), Generalized predictive control system design based on non-linear identification by using Hammerstein model (in Japanese), Trans. ISCIE, 8(3):115–121

[3] Hatanaka T, Uosaki K, Koga M (2002), Evolutionary computation approach to Hammerstein model identification, Proc. of the 4th Asian Control Conference :1730–1735

[4] Hachino T, Takata H (2010), Identification of discrete-time Hammerstein systems using Gaussian process model, Journal of Signal Processing, 14(2):129–137

[5] Kennedy J, Eberhart R. C (1995), Particle swarm optimization, Proc. of IEEE International Conference on Neural Networks :1942-1948