Gaussian Process Model Identification Using Artificial Bee Colony Algorithm and Its Application to Modeling of Power Systems

Tomohiro Hachino, Hitoshi Takata, Shigeru Nakayama, Ichiro Iimura, Seiji Fukushima, Yasutaka Igarashi

Abstract—This paper presents a nonparametric identification of continuous-time nonlinear systems by using a Gaussian process (GP) model. The GP prior model is trained by artificial bee colony algorithm. The nonlinear function of the objective system is estimated as the predictive mean function of the GP, and the confidence measure of the estimated nonlinear function is given by the predictive covariance of the GP. The proposed identification method is applied to modeling of a simplified electric power system. Simulation results are shown to demonstrate the effectiveness of the proposed method.

Keywords—Artificial bee colony algorithm, Gaussian process model, identification, nonlinear system, electric power system.

I. INTRODUCTION

Practical systems such as electric power systems are essentially continuous-time nonlinear systems. Development of accurate identification algorithm for such systems is indispensable for precise analysis or control design. Identification based on the continuous-time model has received a little attention owing to difficulty of handling the higher-order derivatives of input and output data. For this approach, identification methods based on neural network model [1], orthogonal least-squares (LS) estimator [2], radial basis function model [3], [4], and automatic choosing function model [5] have been reported. Since these methods are categorized into the parametric identification, one needs many weighting parameters of any basis functions to describe the nonlinearity. Moreover, any confidence measures for the estimated nonlinear functions are not given in such identification methods.

In recent years, the Gaussian process (GP) model has been introduced for the modeling of the nonlinear dynamic systems [6], [7] and the prediction in time series analysis [8], [9]. The GP model was originally utilized for the regression problem by O’Hagan [10] and has recently much attention for regression or classification problem [11]–[13]. Some applications using the GP model have been reported for human motion modeling [14], or predictive control of gas-liquid separation plant [15]. The GP model is a non-parametric model and fits naturally into Bayesian framework. Since it has fewer parameters than parametric models such as the neural network model, we can describe the nonlinearity of the objective system in a few parameters. Moreover, the GP gives us not only the mean function but also the covariance function. Therefore, in this paper, we propose a nonparametric identification of continuous-time nonlinear systems using the GP model.

The hyperparameters included in the GP prior model have to be appropriately determined based on the identification data. Generally this training becomes nonlinear optimization problem. In this paper, the separable LS approach combining the linear LS method with artificial bee colony (ABC) algorithm is presented for this training. ABC algorithm is an optimization algorithm inspired by an intelligent behavior of honeybee swarms and has high potential for both global and local optimizations [16]. This algorithm consists of search by the three types of bees; the employed bees, the onlooker bees, and the scout bees. ABC algorithm consists of only the basic arithmetic operations and does not require complicated coding and genetic operations such as crossovers and mutations for the genetic algorithm. Moreover, the performance of ABC algorithm is better than or similar to those of other population-based algorithms in spite of a few setting parameters [16], [17]. These advantages suggest that the use of ABC algorithm increases efficiency when the GP prior model for identification is trained.

This paper is organized as follows. In Section II the problem is formulated. In Section III the GP prior model for the identification is derived. In Section IV the training method of the GP prior model is presented using ABC algorithm, and the nonlinear function with the confidence measure is estimated from the GP posterior distribution. In Section V numerical simulation for a simplified electric power system is carried out to illustrate the effectiveness of the proposed identification method. Finally some conclusions are given in Section VI.
Consider a single-input, single-output, continuous-time nonlinear system described by

\[ p^n y(t) = f(w(t)) - \sum_{i=1}^{n} a_i p^{n-i} y(t) + \sum_{j=1}^{m} b_j p^{m-j} u(t) + \varepsilon(t) \]  

(1)

where \( u(t) \) and \( x(t) \) are the true input and output signals, respectively. \( y(t) \) is the noisy output that is corrupted by the measurement noise \( \varepsilon(t) \). \( f(\cdot) \) is an unknown nonlinear function, which is assumed to be stationary and smooth. \( p \) denotes the differential operator. \( n, n_i (i = 1, 2, \ldots, \alpha), m \) and \( m_j (j = 1, 2, \ldots, \beta) \) are assumed to be known. The purpose of this paper is to identify the parameters \( \{a_i\} \) and \( \{b_j\} \) of the linear terms and the nonlinear function \( f(\cdot) \) with the confidence measure, from the true input and noisy output data in the GP framework.

### II. STATEMENT OF THE PROBLEM

Equation (1) can be rewritten as

\[ p^n y(t) = f(w(t)) - \sum_{i=1}^{n} a_i p^{n-i} y(t) + \sum_{j=1}^{m} b_j p^{m-j} u(t) + \varepsilon(t) \]

(2)

where \( \varepsilon(t) \) is an error caused by the measurement noise \( \varepsilon(t) \).

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

\[ F(p) = \frac{1}{p^n + \gamma_1 p^{n-1} + \ldots + \gamma_q} \quad (q > n) \]

(3)

Multiplying both sides of (2) by \( F(p) \) yields

\[ p^n y(t) = F(p)f(w(t)) - \sum_{i=1}^{n} a_i p^{n-i} y(t) + \sum_{j=1}^{m} b_j p^{m-j} u(t) + \varepsilon(t) \]

(4)

where \( w(t) = [p^{n-1} y(t), p^{n-2} y(t), \ldots, p^{-m} y(t), p^{m-1} u(t), p^{m-2} u(t), \ldots, p^{m-q} u(t)]^T \)

\[ y(t) = x(t) + \varepsilon(t) \]

where \( \varepsilon(t) \) is assumed to be zero mean Gaussian noise with variance \( \sigma^2_n \).

Putting \( t = t_1, t_2, \ldots, t_N \) into (5) yields

\[ y(t) = v + G\theta_t \]

(6)

where \( y = [p^n y(t_1), p^n y(t_2), \ldots, p^n y(t_N)]^T \)

\[ v = [f(w(t_1)), f(w(t_2)), \ldots, f(w(t_N)), \varepsilon(t_1), \varepsilon(t_2), \ldots, \varepsilon(t_N)]^T \]

\[ \theta_t = [a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m]^T \]

\[ G = [g(t_1), g(t_2), \ldots, g(t_N)]^T \]

\[ g(t) = [-p^{n-1} y(t), \ldots, -p^{n-q} y(t), p^{m-1} u(t), \ldots, p^{m-q} u(t)]^T \]

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 \( f \) can be represented by the GP:

\[ f \sim N(m(w), \Sigma(w, w)) \]

(8)

where

\[ m(w(t)) = (w(t))^T \theta_m \]

(9)

\[ \theta_m = [\theta_{n_1}, \theta_{n_2}, \ldots, \theta_{n_\alpha}, \theta_{m_1}, \theta_{m_2}, \ldots, \theta_{m_\beta}]^T \]

(10)

where \( \theta_m \) is the unknown parameter vector for the mean function. Thus, the mean function vector \( m(w) \) is described as follows:

\[ m(w) = w^T \theta_m \]

(11)

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

\[ \Sigma_{pq} = \exp \left( -\frac{||w(t_p) - w(t_q)||^2}{2\ell^2} \right) \]

(12)
where $\| \cdot \|$ denotes the Euclidean norm. Equation (12) means that the covariance of the outputs of the nonlinear function depends only on the distance between the inputs $w^f(t_p)$ and $w^f(t_q)$. A high correlation between the outputs of the nonlinear function occurs for the inputs that are close to each other. The overall variance of the random function can be controlled by $\sigma_y$, and the characteristic length scale of the process can be changed by $l$.

From (8), the vector $v$ of the noisy function values in (6) can be written as

$$v \sim \mathcal{N}(m(w), K(w, w))$$

where

$$K(w, w) = \Sigma(w, w) + \sigma_n^2 I_N$$

$$I_N : N \times N \text{ identity matrix}$$

and $\theta_v = [\sigma_y, l, \sigma_n]^T$ is called the hyperparameter vector. From (6) and (13), the GP model for the identification is derived as

$$y \sim \mathcal{N}(m(w) + G\theta_1, K(w, w))$$

In the following, $K(w, w)$ is written as $K$ for simplicity.

IV. IDENTIFICATION

A. Training of GP Prior Model by ABC Algorithm

At the first stage of the identification, the GP prior model is trained by optimizing the unknown parameter vector $\theta = [\theta_{ml}, \theta_{V1}, \theta_{V2}]^T$. This training is carried out by minimizing the negative log marginal likelihood of the identification data:

$$J = -\log p(y|w, G, \theta)$$

$$= \frac{1}{2} \log |K| + \frac{1}{2} (y - Z\theta_{ml})^T K^{-1} (y - Z\theta_{ml})$$

$$+ \frac{N}{2} \log (2\pi)$$

where

$$Z = \begin{bmatrix} w^T & : & G \end{bmatrix}$$

$$\theta_{ml} = [\theta_{ml}^T, \theta_{V1}, \theta_{V2}]^T$$

Although this problem is a nonlinear optimization one, we can separate the linear optimization part and the nonlinear optimization part. The partial derivative of (16) with respect to the parameter vector $\theta_{ml}$ is as follows:

$$\frac{\partial J}{\partial \theta_{ml}} = -Z^TK^{-1}y + Z^TK^{-1}Z\theta_{ml}$$

Note that if the candidates of the hyperparameter vector $\theta_v$ of the covariance function are given, the candidates of the covariance matrix $K$ can be constructed. Therefore, the parameter vector $\theta_{ml}$ can be estimated by the linear LS method from (18):

$$\theta_{ml} = (Z^TK^{-1}Z)^{-1}Z^TK^{-1}y$$

However even if the parameter vector $\theta_{ml}$ is known, the optimization with respect to $\theta_v$ is a complicated nonlinear problem and might suffer from the local minima problem. Therefore, in this paper, we propose a method that combines the linear LS method with ABC algorithm. Only $\Omega = \theta_v = [\sigma_y, l, \sigma_n]^T$ is represented with the positions of the food sources and searched by ABC algorithm. The detailed training algorithm is as follows:

Step 1: Initialization

(i-1) Generate an initial population of $N_s$ bees with random positions of the food sources $\Omega_{[i]} (i = 1, 2, \cdots, N_s)$ from (20):

$$\Omega_{ij} = \Omega_{min,j} + \text{rand}[0, 1] \cdot (\Omega_{max,j} - \Omega_{min,j})$$

where $N_s$ denotes the size of the employed bees or onlooker bees and $\Omega_{ij}$ is the $j$th element of the vector $\Omega_{[i]}$. $\Omega_{min,j}$ and $\Omega_{max,j}$ are the minimum and maximum values for $\Omega_{ij}$, respectively. rand[0, 1] is uniformly distributed random number with amplitude in the range [0, 1].

(i-2) Set the iteration counter $l$ to 1.

(i-3) Set the counter for abandonment $\text{trial}_l$ to 0. The counter $\text{trial}_l$ shows the number of times that the solution $\Omega_{[i]}$ is not improved by the employed and onlooker bees.

Step 2: Construction of the covariance matrix

Construct $N_s$ candidates of the covariance matrix $K_{[i]}$ using $\Omega_{[i]}$ ($i = 1, 2, \cdots, N_s$).

Step 3: Estimation of $\theta_{ml}$

Estimate $N_s$ candidates for $\theta_{ml[i]} (i = 1, 2, \cdots, N_s)$ from (19):

$$\theta_{ml[i]} = (Z^TK^{-1}_{[i]}Z)^{-1}Z^TK^{-1}_{[i]}y$$

Step 4: Fitness value calculation

Calculate the negative log marginal likelihood of the identification data:

$$J_i(\Omega_{[i]}) = \frac{1}{2} \log |K_{[i]}| + \frac{1}{2} (y - Z\theta_{ml[i]})^T K_{[i]}^{-1} \times (y - Z\theta_{ml[i]}) + \frac{N}{2} \log (2\pi)$$

and the fitness value $F_i(\Omega_{[i]}) = \exp(-J_i(\Omega_{[i]}))$.

Step 5: Search by the employed bees

(5-1) Determine the new positions of the food sources $V_{[i]} = \theta_{ml[i]}$ around $\Omega_{[i]}$ for the employed bees from (23):

$$V_{ij} = \Omega_{ij} + \text{rand}[-1, 1] \cdot (\Omega_{ij} - \Omega_{kj})$$

where $V_{ij}$ is the $j$th element of the vector $V_{[i]}$ and $k$ is a random integer selected from $\{1, 2, \cdots, N_s\}$, where $k \neq i$.

(5-2) Construct $N_s$ candidates of the covariance matrix $K_{[i]}$ using $V_{[i]} (i = 1, 2, \cdots, N_s)$.

(5-3) Estimate $N_s$ candidates for $\theta_{ml[i]} (i = 1, 2, \cdots, N_s)$ from (19).

(5-4) Calculate the objective function value:

$$J_i(V_{[i]}) = \frac{1}{2} \log |K_{[i]}| + \frac{1}{2} (y - Z\theta_{ml[i]})^T K_{[i]}^{-1} \times (y - Z\theta_{ml[i]}) + \frac{N}{2} \log (2\pi)$$

and the fitness value $F_i(V_{[i]}) = \exp(-J_i(V_{[i]}))$.

(5-5) If $F_i(\Omega_{[i]}) < F_i(V_{[i]})$, update $\Omega_{[i]}$, $\theta_{ml[i]}$, and $F_i(\Omega_{[i]})$ by $V_{[i]}$, $\theta_{ml[i]}$, and $F_i(V_{[i]})$, respectively, and set $\text{trial}_l = 0$. If $\text{trial}_l > \text{max}\_\text{trial}$, generate a new initial population, and if not, set $l = l + 1$.
Otherwise set trial = trial + 1. This procedure is called
“greedy selection”.

**Step 6: Search by the onlooker bees**

(6-1) Choose one position of the food source for each
onlooker bee from \( \Omega_{[i]} \) \( (i = 1, 2, \cdots, N_s) \) through
“roulette-wheel” slots weighted in proportion to the fitness
value of the employed bee. Namely each onlooker bee
selects one position of the food source from the
employed bees \( F_i(\Omega_{[i]})/\sum_{p=1}^{N_s} F_p(\Omega_{[p]}) \).

(6-2) Calculate the new positions of the food sources \( V_{[i]} \)
corresponding to the selected positions \( \Omega_{[i]} \) from (23).

(6-3) Construct the covariance matrix \( K_{[i]} \) using
\( V_{[i]} \) \((i = 1, 2, \cdots, N_s)\).

(6-4) Estimate \( N_s \) candidates for \( \theta_{ml[i]} \) \((i = 1, 2, \cdots, N_s)\) from (19).

(6-5) Calculate the fitness value \( F_i(V_{[i]}) = \exp(-J_i(V_{[i]})) \)
from (24).

(6-6) Carry out the greedy selection with the same way of step
5 (5-5).

**Step 7: Search by the scout bees**

If the counter for abandonment \( trial \) is greater or equal
to the prespecified number \textit{limit}, carry out the following
procedure.

(7-1) Differentiate the corresponding employed bee into the
scout bee and generate the new position of the food source
\( \Omega_{[i]} \) for the scout bee randomly from (20).

(7-2) Construct the covariance matrix \( K_{[i]} \) using the
corresponding \( \Omega_{[i]} \).

(7-3) Estimate \( \theta_{ml[i]} \) from (19).

(7-4) Calculate the fitness value \( F_i(\Omega_{[i]}) = \exp(-J_i(\Omega_{[i]})) \)
from (22).

This step means that if the solution is not improved \textit{limit}
times through search by the employed and onlooker bees, the
corresponding employed bee gives up to search around his
food source and transforms himself to the scout bee to search
around randomly selected food source. Since the number \textit{limit}
is usually set to be the product of the employed bee size and
the dimension of the search space \([16]\), this number is taken
to be \textit{limit} \( = N_s \times 3 \) in this paper.

**Step 8: Repetition**

Set the iteration counter to \( l = l + 1 \) and go to step 5 until
the prespecified iteration number \textit{l}_{max}.

**Step 9: Determination of the GP prior model**

Determine the vector \( \Omega = \theta_c = [\hat{\sigma}_y, \hat{\ell}, \hat{\sigma}_n]^T \) and the
corresponding parameter vector \( \theta_{ml} = [\hat{\theta}_{ml}^y, \hat{\theta}_{ml}^{\ell}, \hat{\theta}_{ml}^n]^T \) using the
best position of the food source. Construct the suboptimal prior
mean function and covariance function:

\[
\begin{align*}
 m(w^{f}(t)) &= (w^{f}(t))^T \theta_{ml} \\
 s(w^{f}(t_p), w^{f}(t_q)) &= \hat{\sigma}_y^2 \exp\left(-\frac{||w^{f}(t_p) - w^{f}(t_q)||^2}{2\hat{\ell}^2}\right) \\
 k(w^{f}(t_p), w^{f}(t_q)) &= s(w^{f}(t_p), w^{f}(t_q)) + \hat{\sigma}_n^2 \delta_{pq},
\end{align*}
\]

where \( s(w^{f}(t_p), w^{f}(t_q)) \) is an element of covariance matrix
\( \Sigma \), \( k(w^{f}(t_p), w^{f}(t_q)) \) is an element of covariance matrix \( K \),
and \( \delta_{pq} \) is the Kronecker delta, which is 1 if \( p = q \) and 0
otherwise.

**B. Estimation of the Nonlinear Function**

For a new input \( w^{f}(t) \) and the corresponding function \( f(w^{f}(t)) \), we have the following joint Gaussian distribution:

\[
\begin{align*}
 y(t) \sim & \mathcal{N}(m(w) + G \hat{\theta}_l, \Sigma_{w, w}^{f}(t)) \end{align*}
\]

From the formula for conditioning a joint Gaussian distribution \([18]\), the posterior distribution for specific test data is

\[
f(w^{f}(t)) \sim \mathcal{N}(\hat{f}(w^{f}(t)), \hat{\sigma}^2(t))
\]

which is used for the confidence measure of the estimated
nonlinear function.

**V. ILLUSTRATIVE EXAMPLE**

Consider an electric power system \([19]\) described by

\[
\begin{align*}
 \ddot{x}(t) + a_1 \dot{x}(t) &= f(z(t)) \\
 f(z(t)) &= -\frac{P_w}{M} + \frac{P_n}{M} \\
 &= -\frac{P_{em}}{M} (1 + u(t)) \sin x(t) + \frac{P_n}{M} \\
 y(t) &= x(t) + \epsilon(t)
\end{align*}
\]

where \( x(t) = \delta(t) \): phase angle, \( u(t) = \Delta E_{fd}(t) \): increment
of excitation voltage, \( M \): inertia coefficient, \( D \): damping
coefficient, \( P_c \): generator output power, \( P_n \): turbine output
power. In numerical example, \( M = D = 0.06, P_{em} = 1.0, P_n = 0.8 \) and \( a_1 = \frac{D}{M} \) = 1.0 are set. The measurement
noise \( \epsilon(t) \) is white Gaussian noise, where noise-to-signal
ratio is about 1.5%. The number of input and output data
for identification is taken to be \( N = 800 \). The third-order
Butterworth filter with the cutoff frequency \( \omega_c = 10 \) [rad/s]
is utilized as a delayed state variable filter. The setting parameters
of ABC algorithm are chosen as follows:

(i) employed bee size \( N_c = 50 \)
(ii) maximum iteration number \( l_{max} = 100 \)

The hyperparameter vector of the covariance function has been
determined by ABC algorithm as \( \theta_c = [\hat{\sigma}_y, \hat{\ell}, \hat{\sigma}_n]^T = [36.190, 0.342, 0.180]^T \).
Estimate of the parameter in the linear
term is \( \hat{a}_1 = 0.962 \), which is very close to the true value
\( a_1 = 1.0 \). The true nonlinear function \( f(z(t)) \), the estimated
nonlinear function \( \hat{f}(z(t)) \), the absolute error between
\( f(z(t)) \) and \( \hat{f}(z(t)) \), and the double standard deviation
confidence interval (95.5% confidence region) around the
estimated nonlinear function are shown in Figs. 1 ~ 4, respectively, where the thick curves depict the trajectories of the identification data. Clearly the estimated nonlinear function $\hat{f}(z(t))$ is shown to be very close to the true nonlinear function $f(z(t))$ on the data region. The confidence region of the estimated nonlinear function grows as $z(t)$ goes away from the data region. On the other hand, the confidence region of the estimated nonlinear function is very small on the data region. Fig. 5 shows the true output $x(t)$ and the output $\hat{x}(t)$ by the estimated model, where the outputs were generated by the inputs for validation. This figure indicates that $\hat{x}(t)$ matches $x(t)$ considerably. Consequently, we can confirm that the proposed method gives an accurate model of the objective electric power system.

VI. CONCLUSIONS

In this paper, we have proposed an identification method of continuous-time nonlinear systems using the GP model. The GP prior model is trained by the aid of ABC algorithm so that the negative log marginal likelihood of the identification data is minimized. The proposed identification method is categorized into the nonparametric identification and does not need the determination of the model structure. Since ABC algorithm has a few setting parameters, the proposed training algorithm is efficient for system identification. Simulation results show that the proposed method can be successfully applied to modeling of the electric power system.

REFERENCES


