Identification of Dynamical System under Unknown Density of Noise by Neural Network

Yasuhide KOBAYASHI and Tsuyoshi OKITA
Faculty of Information Sciences, Hiroshima City-University
Ozuka Numata-cho Asa-minami-ku, Hiroshima-shi. 731-31 Japan

Abstract
The paper considers a problem of the identification for a dynamical system on the assumption that the density of an observation noise is not available. The system parameters are estimated by a least squares method. The system structure is estimated by using a neural network.

1 Introduction
Identification methods of system structure have been developed in many fields\(^1\). In those methods, the density of noise is supposed to be a Gaussian or well known density. In fact, the probability density of an observation noise might be unknown or non-Gaussian.

In this paper, we consider the identification for a dynamical system, on the assumption that the density of an observation noise is not available. In this case, the estimation methods of a system structure based on those statistics are unable to be applied. Therefore, we estimate the system structure by using a neural network.

2 Statement of the problem
We consider a linear system expressed with the following equation.

\[
\begin{align*}
  x(k+1) &= \begin{pmatrix} a_0^T & b_0^T \end{pmatrix} \begin{bmatrix} x(k) \\ u(k) \end{bmatrix} \\
  x(0) &= x_0
\end{align*}
\]

where, \( x(k) \) is a state vector, \( u(k) \) is an input vector, and the dimensions of state and input are \( n, r \), respectively, and unknown.

\[
\begin{align*}
  x(k) &= (x(k), x(k-1), \ldots, x(k-n+1))^T \\
  u(k) &= (u(k), u(k-1), \ldots, u(k-r+1))^T
\end{align*}
\]

\( \Theta = (a_0^T, b_0^T)^T \)

and also, \( \Theta \) is unknown system parameter of \( l_0 (=n+r) \) dimension.

On the other hand, an observation equation is given in the following manner.

\[
y(k) = h^T x(k) + v(k), \quad k \geq 1
\]

where, \( h = (1, 0, 0, \ldots, 0)^T \).

Furthermore, observation noise \( v(k) \) is assumed to be white, and the density is not available. An observation series at time \( k \) is given in the following equation.
\[ Y_k = \{ u(0), y(1), u(1), \ldots, u(k-1), y(k) \} \] (3)

In the following, we develop to estimate the structure and parameters of system, on the assumption that the density of an observation noise is not available.

3 System identification

3-1 Estimation of system parameters

When a priori information about the system structure is poor, models \( m = 1, 2, \ldots, M \) are set up from lower order. In the case that these values are available, it is feasible to estimate the system parameters by using a performance index based on the probability density and statistics of noise. However, the probability density and statistical values are not available here. The system parameters are estimated by a least square method. We use the output error that is superior to the equation error in low frequency characteristics.

\[ I_n = \frac{1}{k} \sum_k (y(k') - y_m(k'))^2 \] (4)

3-2 Estimation of system structure

We consider the maximum likelihood function for each model to select the optimal model structure.

\[ P(Y_k | m) = \prod_k P(y(k) | \theta_m, Y_{k-1}) = \prod_k \{ P(y(k) | \theta_m, Y_{k-1}) P(\theta_m | m, Y_{k-1}) d\theta_m \} \] (5)

If the observation noise is Gaussian, then the maximum likelihood function is given as follows.

\[ P(Y_k | m) = \prod_k \frac{1}{\sqrt{2\pi (\sigma_v^2 + c_m^T A_m c_m)}} \exp\left[ -\frac{(y(k) - y_m(k | k-1))^2}{2(\sigma_v^2 + c_m^T A_m c_m)} \right] \] (6)

where \( c_m = (\partial y(k) / \partial \theta_n)^T \) and \( A_m = E[(\theta_n - \theta_{on})(\theta_n - \theta_{on})^T | Y_{k-1}] \)

\[ = E\{ \partial \ln P(Y_{k-1} | \theta_n) \} \left( \frac{\partial \ln P(Y_{k-1} | \theta_n)}{\partial \theta_n} \right) \]

\[ = \sigma_v^2 \{ \Sigma (\frac{\partial y(k)}{\partial \theta_n})^T (\frac{\partial y(k)}{\partial \theta_n}) \}^{-1} \] (7)

and \( \sigma_v^2 \) is the variance of the observation noise.

The dominant part of equation (6) is exponential function, and the prediction error variance of output is in inverse proportion to data number \( k \), and approximated with the following equation (7).

\[ \{ (y(k) - y_m(k | Y_{k-1}))^2 \} \approx \sigma_v^2 + c_m^T A_m c_m \approx \sigma_v^2 \left( 1 + \ell / (k-1) \right) \] (8)

where \( \sigma_v^2 \) is the approximated value of the variance of observation noise. \( \ell \) is the dimension of model parameters.

\[ \sigma_v^2 \approx I_n \left( 1 - \ell / (k-1) \right)^{-1} \] (9)

Thus, the exponential part of equation (6) is approximated as follows. We estimate the structure based on the following index \( J_n \) that is normalized by the prediction error variance.
The minimum value of the models can be obtained from equation (10).

\[ J_0 = \min_{n} J_n \] (11)

3-3 Constitution of neural network

The equation (10) is derived under the assumption of Gaussian noise. However, the probability density of observation noise might be non-Gaussian. The model structure given by the approximated equation (10) has a tendency to select a higher order model more than the true system structure, when the data is short. So, we have to compensate the approximation error which gives the influences to the estimation of system structure.

Neural nets (N.N.) offer the ability of learning adaptation to the nonlinear problems. We use neural network of 3 layers, to compensate these approximation errors. The input signal \( X_m \) \((m=1,2,\ldots, M)\) of the neural network is supposed to be maximum value for the model that give the minimum value of equation (10).

\[ X_m = \frac{(1-\alpha) J_0}{J_n - \alpha J_0}, \quad 0<\alpha<1 \] (12)

The input signal \( X_m \) gives 1 for the model that the index \( J_n \) is minimum, otherwise the values are positive but less than 1.

The output signal \( Z_n \) \((n=1,2,\ldots, N)\) of middle layer is given by the following equation.

\[ Z_n = f(\sum_{i=m}^{n} W_{in} X_m) \] (13)

And the output signal \( O_i \) \((i=1,2,\ldots, M)\) of N.N. is given by the following equation.

\[ O_i = f(\sum_{n=1}^{N} W_{in} Z_n) \] (14)

where, \( f \) denotes the next logistic function.

\[ f(*) = 1/(1+\exp(-*)) \] (15)

We adopt the model as the best structure that the output signal of N.N. is maximum.

\[ L (i) = \max_i \] (16)

3-4 Learning of system structure

When certain input pattern \( X^p=(X_1^p, X_2^p, \ldots, X_M^p) \) was applied to N.N., we suppose that an output vector \( O^p=(O_1^p, O_2^p, \ldots, O_M^p) \) was obtained. Corresponding to this output, evaluation function \( E \) is shown with a teacher signal \( T^p \) as given in the following equation.

\[ E = \frac{1}{2} \sum_{p} (T^p-O^p)^2 \] (17)

where \( p \) is the number of data used for learning.

The partial derivative to minimize the equation (17) is given by \( \delta \) formula of back propagation (B.P.) method\(^3\). The adjustment of each connection coefficient of N.N. is given by the following equations.

\[ \tau (q) = \tau (q-1) + \Delta \tau (q) \]

\[ \Delta \tau (q) = -\eta \left( \frac{\partial E}{\partial \tau} \right) (q-1) + \beta \Delta \tau (q-1) \] (18)

where, \( \eta, \beta \) are positive constants of appropriate values, \( \tau \) is a parameter vector.
consists of each connection coefficient $w_{:\text{in}}$, $w_{:\text{in}}$. $\Delta r$ is an adjustment value of a network, and $q$ is the number of iterations.

We compose the N.N. and learn in advance with various noise sequences.

4. Numerical examples
4.1 Outline of the simulation

We considered 5 system structures, and the following 5 models were used for learning of N.N..

1. model 1  n=1 r=1 $\theta_0=(0.91, 0.1) \, T$
2. model 2  n=2 r=1 $\theta_0=(1.2, -0.6, 0.8) \, T$
3. model 3  n=2 r=2 $\theta_0=(1.2, -0.6, 0.8, 1.0) \, T$
4. model 4  n=3 r=2 $\theta_0=(1.2, 0.5, -0.8, 1.0, -0.8) \, T$
5. model 5  n=3 r=3 $\theta_0=(1.8, -1.3, 0.4, 1.0, 0.5, 0.8) \, T$

The input for the model is generated from Gaussian distribution $N(0, 1)$. As the density of an observation noise is not available, we considered 3 typical types of distribution given as follows.

a. Gaussian distribution $N(0, 1)$
b. Uniform distribution $U(-0.5, 0.5)$
c. Beta distribution $B(0.5, 0.5)$

The variance of the observation noise from each distribution are normalized $\sigma_\epsilon^2=1.0$. The learning data for N.N. were collected in a range $11<k<300$ for 6 noise sequences and 3
distribution types. Thus, the number of data that used learning is about 5,000 for each system. The B.P. method was used to minimize error evaluation function.

The systems to identify are set up as follows.

1. \( m=1 \) \( n=1 \) \( r=1 \) \( \theta_0=(0.8,3.0) \)
2. \( m=2 \) \( n=2 \) \( r=1 \) \( \theta_0=(1.5,-0.8,1.5) \)
3. \( m=3 \) \( n=2 \) \( r=2 \) \( \theta_0=(1.5,-0.8,1.0,1.2) \)
4. \( m=4 \) \( n=3 \) \( r=2 \) \( \theta_0=(0.9,-0.58,-0.25,1.0,-1.0) \)
5. \( m=5 \) \( n=3 \) \( r=3 \) \( \theta_0=(2.2,-1.8,0.6,1.4,-0.3,0.5) \)

Table 1 Numbers of maximum signal for each model in a sequence (30<k<1000).

<table>
<thead>
<tr>
<th>Number of models m</th>
<th>1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian input ( X_m )</td>
<td>2 448 520 0 0</td>
</tr>
<tr>
<td>True model ( m=2 ) (fig.3) output ( O_m )</td>
<td>2 968 0 0 0</td>
</tr>
<tr>
<td>Uniform input ( X_m )</td>
<td>0 0 727 77 166</td>
</tr>
<tr>
<td>True model ( m=3 ) (fig.4) output ( O_m )</td>
<td>0 23 929 17 1</td>
</tr>
</tbody>
</table>

Fig. 3 Input and output for N.N. \( m=2 \) with Gaussian distribution \( \sigma_y^2=9 \).

Fig. 4 Input and output for N.N. \( m=3 \) with uniform distribution \( \sigma_y^2=9 \).

Fig. 5 Selected model for system \( m=2 \) with \( N(0,9) \) (fig.3) by AIC.
The variances of observation noise are set up as follows, and the another noise sequence are given for identification.

a. $\sigma_v^2=1.0$ (case a)
b. $\sigma_v^2=9.0$ (case b)

4.2 Simulation Results

Case a Fig.1 shows the input and output signal of N.N. when the system is $m=2$ with beta distribution noise. The input signal $X_2$ corresponded to the true model structure is almost maximum. The output signal discriminates the true system structure clearly. Fig.2 also shows the input and output signal of N.N. when the system is $m=4$ with beta distribution noise. Similar results are given in the other systems and distributions.

Case b In cases of a large noise, the input signal of a true model is not always maximum, when the data is short. Fig.3 shows the input and output signals of N.N. when the system is $m=2$ with Gaussian noise. The input signal $X_3$ is rather larger than $X_2$, then a higher order model more than the true system structure has the maximum input signal in some periods.

However, the output signal $O_2$ of true structure is maximum. Table 1 shows the numbers of input and output signals that give the maximum value for each model in a sequence. The output signal compensates the approximation errors, and discriminates the true system structure clearly. Fig.4 shows the similar case for $m=3$ with uniform distribution noise. Table 1 also shows the results in the case of Fig.4. The output signal discriminates the true system structure clearly even when observation noise is large. In cases of other distribution and models, similar results were given.

We compared with the model order selection method based on Akaike's information criterion (AIC). For the case showed in fig.3, the model selected by AIC is given in fig.5. AIC selected the redundant model in some cases.

5 Conclusion

We developed the system identification by neural network, on the assumption that the density of an observation noise is not available. The output signal of N.N. discriminates the true system structure clearly even when observation noise is large. The method presented here does not need much information on the noise statistics as some existing methods. Then the robust estimation of system structure is given by the proposed method.

References

3) N. Negishi et. al: The Neural Network's Learning Algorithm Based on the Manner of Minimizing the Maximum Error by a Non-Differentiable Optimization Technique, Trans. of SICE, vol.29-10 (1993)