Multi-Layered GMDH-Type Neural Network Self-Selecting Optimum Neural Network Architecture and Its Application to Nonlinear System Identification

Tadashi Kondo[1], Junji Ueno[2]

[1,2] School of Health Sciences, University of Tokushima, 3-18-15Kuramoto-cho Tokushima, 770-8509, Japan

[1]kondo@medsci.tokushima-u.ac.jp

Abstract. *In this study, a new multi-layered Group Method of Data Handling (GMDH)-type neural network self-selecting optimum neural network architecture is proposed. We call this algorithm as revised GMDH-type neural network algorithm self-selecting optimum neural network architecture. Revised GMDH-type neural network algorithm has an ability of self-selecting optimum neural network architecture from three neural network architectures such as sigmoid function neural network, radial basis function (RBF) neural network and polynomial neural network. Revised GMDH-type neural network also has abilities of self-selecting the number of layers, the number of neurons in hidden layers and useful input variables. This algorithm is applied to the nonlinear system identification problem and it is shown that this algorithm is useful for the nonlinear system identification because optimum neural network architecture is automatically organized.*

Keywords

GMDH, Neural network, System identification

1 Introduction

Group Method of Data Handling (GMDH)-type neural networks and their applications have been proposed in our early works [1]-[3]. GMDH-type neural networks can automatically organize the optimum neural network architecture by heuristic self-organization method [4],[5] and they can also determine such structural parameters as the number of layers, the number of neurons in hidden layers and useful input variables. But, these conventional GMDH-type neural network algorithms do not have an ability of selecting the type of neural network architectures such as sigmoid function neural network, radial basis function (RBF) neural network and polynomial neural network.

In this study, revised GMDH-type neural network algorithm self-selecting optimum neural network architecture is proposed. In this algorithm, the optimum neural network architecture is automatically selected from three neural network architectures such as sigmoid function neural network, RBF neural network and polynomial neural network. Furthermore, structural parameters such as the number of layers, the number of neurons in hidden layers and useful input variables are automatically selected to minimize prediction error criterion defined as Prediction Sum of Squares (PSS) [6]. Revised GMDH-type neural network algorithm proposed in this paper is applied to the nonlinear system identification and results show that revised GMDH-type neural network algorithm is useful for the nonlinear system identification.

2 Multi-Layered GMDH-Type Neural Network

Architectures of GMDH-type neural network are automatically organized by heuristic self-organization method. First, we show procedures of heuristic self-organization method because it plays very important roles for organization of GMDH-type neural network.

2.1 Heuristic Se1f-Organization

Heuristic self-organization method is constructed by the following six procedures:

(1) Separating original data into training and test sets

Original data is separated into training and test sets. Training data is used for estimating parameters of partial descriptions which describe partial relationships of the nonlinear system. Test data is used for organizing complete description which describes complete relationships between input and output variables of the nonlinear system.

(2) Generating combinations of input variables in each layer

All combinations of two input variables (x_i, x_j) are generated in each layer. The number of combinations is $\frac{p!}{(p-2)!2!}$. Here, p is the number of input variables.

(3) Calculating partial descriptions

For each combination, partial descriptions of the nonlinear system can be calculated by applying regression analysis to training data. Output variables of partial descriptions are called as intermediate variables.

(4) Selecting intermediate variables

L intermediate variables which give L smallest test errors calculated using test data are selected from generated intermediate variables.

(5) Iterating calculations from 2 to 5

Select L intermediate variables are set to input variables of the next layer and calculations from procedure 2 to 5 are iterated. The multilayered architecture is organized.

(6) Stopping multilayered iterative calculation

When errors of test data in each layer stop decreasing, iterative calculation is terminated. Finally, complete description of the nonlinear system is constructed by partial descriptions generated in each layer.

Heuristic self-organization method is a kind of the evolutional computations. Architectures of the GMDH-type neural network are automatically organized by this method.

2.2 Multi-layered GMDH-Type Neural Network Algorithm

Multi-layered GMDH-type neural network has a common feedforward multilayered architecture. Figure 1 shows architecture of revised GMDH-type neural network. This neural network is organized by heuristic self-organization method.

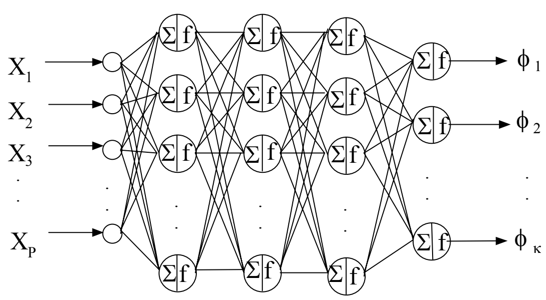

Fig. 1. Architecture of revised GMDH-type neural network

Procedures for determining architecture of revised GMDH-type neural network conform to the following:

2.2.1 The first layer

$u_j=x_j \ (j=1,2,\ldots,p)$ (1)

where $x_j \ (j=1,2,\ldots,p)$ are input variables of the nonlinear system, and p is the number of input variables. In the first layer, input variables are set to output variables.

2.2.2 The second layer

All combinations of r input variables are generated. For each combination, optimum neuron architectures are automatically selected from the following two neurons.

Architectures of the first and second type neurons are shown in Fig.2 and Fig.3 respectively. Optimum neuron architecture for each combination is selected from the first and second type neuron architectures.

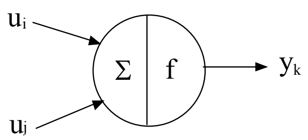

Fig.2. Neuron architecture with two inputs

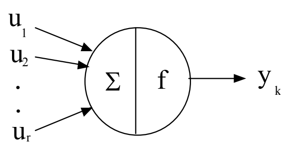

Fig.3. Neuron architecture with r inputs

Revised GMDH-type neural network algorithm proposed in this paper can select optimum neural network architecture from three neural network architectures such as sigmoid function neural network, RBF neural network and polynomial neural network.

Neuron architectures of the first and second type neurons in each neural network architecture are shown as follows.

a) Sigmoid function neural network

i) The first type neuron

Σ: (Nonlinear function)

$$z_k=w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j+w_8u_iu_j^2+w_9u_j^3-w_0\theta_1 \quad (2)$$

f : (Nonlinear function)

$$y_k=\frac{1}{1+e^{(-z_k)}} \quad (3)$$

ii) The second type neuron

Σ: (Linear function)

$$z_k=w_1u_1+w_2u_2+w_3u_3+\cdots+w_ru_r-w_0\theta_1 \quad (r<p) \quad (4)$$

f : (Nonlinear function)

$$y_k=\frac{1}{1+e^{(-z_k)}} \quad (5)$$

b) Radial basis function neural network

i) The first type neuron

Σ: (Nonlinear function)

$$z_k=w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j+w_8u_iu_j^2+w_9u_j^3-w_0\theta_1 \quad (6)$$

f : (Nonlinear function)

$$y_k=e^{(-z_k^2)} \quad (7)$$

ii) The second type neuron

Σ: (Linear function)

$$z_k=w_1u_1+w_2u_2+w_3u_3+\cdots+w_ru_r-w_0\theta_1 \quad (r<p) \quad (8)$$

f : (Nonlinear function)

$$y_k=e^{(-z_k^2)} \quad (9)$$

c) Polynomial neural network

i) The first type neuron

Σ: (Nonlinear function)

$$z_k=w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j+w_8u_iu_j^2+w_9u_j^3 - w_0\theta_1 \quad (10)$$

f : (Linear function)

$$y_k= z_k \quad (11)$$

ii) The second type neuron

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+ \cdots +w_ru_r - w_0\theta_I \;\; (r<p) \quad (12)$$

f : (Linear function)

$$y_k= z_k \quad (13)$$

In the first type neuron, $\theta_I=1$ and $w_i\ (i=0,1,2,\ldots,9)$ are weights between the first and second layer. Value of r, which is the number of input variables u in each neuron, is set to two for the first type neuron.

In the second type neuron, $\theta_I=1$ and $w_i\ (i=0,1,2,\ldots,r)$ are weights between the first and second layer. Value of r, which is the number of input variables u in each neuron, is set to be greater than two and smaller than p for the second type neuron. Here p is the number of input variables x_i $(i=1,2,\ldots,p)$.

Weights $w_i\ (i=0,1,2,\ldots)$ in each neural network architecture are estimated by stepwise regression analysis [7] using PSS.

Estimation procedure of weight w_i :

First, values of z_k are calculated for each neural network architecture as follows.

i) Sigmoid function neural network

$$z_k = \log_e\left(\frac{\phi'}{1-\phi'}\right) \quad (14)$$

ii) RBF neural network

$$z_k = \sqrt{-\log_e \phi'} \quad (15)$$

iii) Polynomial neural network

$$z_k=\phi \quad (16)$$

where ϕ' is normalized output variable whose values are between zero and one and ϕ is output variable.

Then weights w_i are estimated by stepwise regression analysis [7] which selects useful input variables using PSS. Only useful variables in Eq.(2), Eq.(4), Eq.(6), Eq.(8), Eq.(10) and Eq.(12) are selected by stepwise regression analysis using PSS and optimum neuron architectures are organized by selected useful variables.

PSS is described by the following equation:

$$PSS = \sum_{\alpha=1}^{n}(\phi_\alpha - z_\alpha^*)^2 \quad (17)$$

where

$$z_\alpha^*=w_{1\alpha}u_{i\alpha}+w_{2\alpha}u_{j\alpha}+w_{3\alpha}u_{i\alpha}u_{j\alpha}+w_{4\alpha}u_{i\alpha}^2+w_{5\alpha}u_{j\alpha}^2+w_{6\alpha}u_{i\alpha}^3+w_{7\alpha}u_{i\alpha}^2u_{j\alpha}+w_{8\alpha}u_{i\alpha}u_{j\alpha}^2+w_{9\alpha}u_{j\alpha}^3 - w_{0\alpha}\theta_I \quad \alpha=1,2,\ldots,n \quad (18)$$

n is the number of training data, ϕ_α is the α-th observed value for the output variable, $u_{i\alpha}$ is the α-th observed value for the input variable u_i and z_α^* is the α-th estimated value obtained by the multiple regression analysis of all the data except the α-th datum. In order to calculate PSS in (17), the multiple regression analysis must be repeated n times, and the amount of calculations increases in the number of data. But PSS in (17) can be reduced as follows,

$$PSS = \sum_{\alpha=1}^{n}\frac{\phi_\alpha - z_\alpha}{(1-\underline{u}_\alpha^T(U^TU)^{-1}\underline{u}_\alpha)^2} \quad (19)$$

where

$$z_\alpha=w_1u_{i\alpha}+w_2u_{j\alpha}+w_3u_{i\alpha}u_{j\alpha}+w_4u_{i\alpha}^2+w_5u_{j\alpha}^2+w_6u_{i\alpha}^3+w_7u_{i\alpha}^2u_{j\alpha}+w_8u_{i\alpha}u_{j\alpha}^2+w_9u_{j\alpha}^3 - w_0\theta_I, \quad \alpha=1,2,\ldots,n \quad (20)$$

$$\underline{u}_\alpha^T=[1, u_{i\alpha}, u_{j\alpha}, u_{i\alpha}u_{j\alpha}, u_{i\alpha}^2, u_{j\alpha}^2], \qquad \alpha=1,2,\ldots,n \quad (21)$$

$$U^T=[\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_n] \quad (22)$$

z_α is the α-th estimated value obtained by the multiple regression analysis of all the data. With this procedure, we need not repeat the regression analysis n times. PSS criterion do not contain the statistical assumption in the regression model.

For each combination, three neuron architectures which are sigmoid function neuron, RBF neuron and polynomial neuron, are generated and L neurons which minimize test error calculated using test data are selected for each neuron architecture. From these L selected neurons for each neuron architecture, mean test errors of L neurons are calculated. Then, neural network architecture which has minimum mean test error is selected as revised GMDH-type neural network architecture from three neural network architectures such as sigmoid function neural network, RBF neural network and polynomial neural network.

After the type of revised GMDH-type neural network architecture is selected, output variables y_k of L selected neurons are set to input variables of neurons in the third layer.

2.2.3 The third and successive layers

In the second layer, optimum neural network architecture is selected from three neural network architectures. In the third and successive layers, only one neuron architecture, which is sigmoid function neuron or RBF neuron or polynomial neuron, is used for calculation and the same calculation of the second layer is iterated until PSS values of L neurons with selected neuron architecture stop decreasing. When iterative calculation is terminated, neural network architecture is produced by L selected neurons in each layer.

By using these procedures, revised GMDH-type neural network self-selecting optimum neural network architecture is organized. Revised GMDH-type neural network proposed in this paper has an ability of self-selecting optimum neural network architecture. Therefore, neural network architecture is automatically selected from three neural network architectures. Furthermore, structural parameters such as the number of layers, the number of neurons in hidden layers and useful input variables are automatically selected to minimize prediction error criterion defined as PSS.

3 Application to Nonlinear System Identification

Nonlinear system is assumed to be described by the following equations:

$$\phi = f_1(x_1, x_2, x_3)\,/\,f_2(x_1, x_2, x_3)+\varepsilon \tag{23}$$

$$f_1(x_1, x_2, x_3)=1.0+2.0\,x_1^2x_2+3.0\,x_2^2x_3 \tag{24}$$

$$f_2(x_1, x, x_3)=1.0+2.0\exp(x_1)+3.0\exp(x_1x_2)+4.0\exp(x_3) \tag{25}$$

Here, ϕ is output variable and $x_1 \sim x_3$ are input variables and ε is Gaussian white noise which is N(0, 0.005^2). An additional input, x_4, is added as input variable of neural network to check that revised GMDH-type neural network can detect and eliminate useless input variables. Neural network is organized by twenty training data. Twenty other data are used to check prediction and generalization ability. Identification results of revised GMDH-type neural network are compared with those of GMDH algorithm, conventional neural network trained by back propagation algorithm and conventional RBF neural network.

3.1 Identification Results of Revised GMDH-Type Neural Network

Identification results of revised GMDH-type neural network are shown as follows:

(1) Input variables

Four input variables were used, but useless input variable x_4 was automatically eliminated.

(2) Number of selected neurons in each layer

Four neurons were selected in each hidden layer.

(3) Selection of neural network architecture

Test errors of three neuron architectures in the second layer are shown in Fig.4. RBF neurons had the smallest test errors in three neuron architectures and RBF neural network architecture was selected as revised GMDH-type neural network architecture. Calculations in the third and successive layers were carried out using only RBF neuron architecture.

(4) Variation of PSS

Variation of PSS is shown in Fig.5. PSS values converged at the fourth layer.

(5) Architecture of neural network

Calculation of revised GMDH-type neural network was terminated at the fourth layer and neurons of the third layer had minimum PSS value. Three layered neural network architecture was organized. The first layer is input layer and the second layer is hidden layer and the third layer is output layer.

(6) Estimation accuracy

Estimation accuracy was evaluated by the following equation:

$$J_1 = \frac{1}{20}\sum_{i=1}^{20}\left|\phi_i - \phi_i^*\right| \tag{26}$$

where ϕ_i *(i =1,2,⋯,20)* are actual values with Gaussian white noise ε and ϕ_i^* *(i=1,2,⋯,20)* are estimated values by revised GMDH-type neural network. ϕ_i *(i =1,2,⋯,20)* were used to organize revised GMDH-type neural network. Value of J_1 is shown in Table1. In this table, GMDH-type NN(RBF) shows revised GMDH-type neural network proposed in this paper and GMDH(Polynomial) shows GMDH algorithm and conventional NN (Sigmoid) shows conventional neural network trained by back propagation algorithm and conventional NN (RBF) shows conventional RBF neural network.

(7) Prediction accuracy

Prediction accuracy was evaluated by using the following equation:

$$J_2 = \frac{1}{20}\sum_{i=21}^{40}\left|\phi_i - \phi_i^*\right| \tag{27}$$

where ϕ_i *(i =21,22,⋯,40)* are actual values with Gaussian white noise ε and ϕ_i^* *(i=21,22,⋯,40)* are predicted values by revised GMDH-type neural network. ϕ_i *(i =21,22,⋯,40)* were not used to organize revised GMDH-type neural network and were used to check generalization ability. Value of J_2 is shown in Table1 and is very small. From this prediction result, we can see that revised GMDH-type neural network do not overfit training data and have good generalization ability.

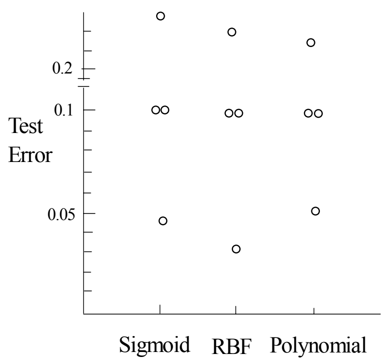

Fig.4. Variation of test error

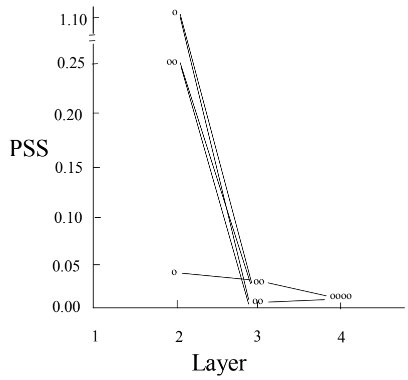

Fig.5. Variation of PSS in revised GMDH-type neural network

Tab. 1. Identification results of four models

Method	J1	J2	Layer
GMDH-type NN (RBF)	0.00928	0.01075	3
GMDH (Polynomial)	0.04094	0.03829	5
Conventional NN (Sigmoid)	0.02614	0.02813	3
Conventional NN (RBF)	0.03282	0.03071	3

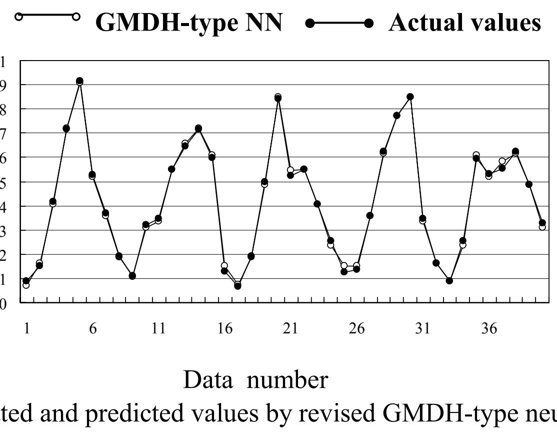

Fig.6. Estimated and predicted values by revised GMDH-type neural network

(8) Estimated and predicted values

Estimated and predicted values of ϕ by revised GMDH-type neural network are shown in Fig.6. Estimated values are shown for the data points between the first and 20-th data entities and predicted values are shown for the data points between the 21-th and 40-th data entities. We can see that estimated and predicted values are very accurate.

3.2 Identification Results of GMDH

Identification results obtained by GMDH are quoted from [2].

(1) Input variables

Four input variables were used, but again useless input variable x_4 was automatically eliminated .

(2) Number of selected intermediate variables

Four intermediate variables were selected in each selection layer.

(3) Variation of PSS

Variation of PSS is shown in Fig.7. PSS converged at the fifth layer.

(4) Architecture of GMDH network

Calculation of GMDH converged at the fifth layer and neurons of the fifth layer had the minimum PSS value. Five layered polynomial network architecture was organized. The first layer is input layer and the second, third and fourth layer are hidden layers and the fifth layer is output layer.

(5) Estimation accuracy

Estimation accuracy was evaluated by Eq.(26) and value of J_1 is shown in Table1.

(6) Prediction accuracy

Prediction accuracy was evaluated by Eq.(27) and value of J_2 is shown in Table1.

(7) Estimated and predicted values

Estimated and predicted values of ϕ by GMDH are shown in Fig.8. Estimation and prediction accuracy were not good compared with revised GMDH-type neural network.

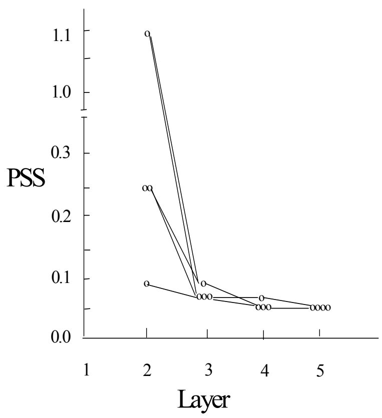

Fig.7. Variation of PSS in GMDH

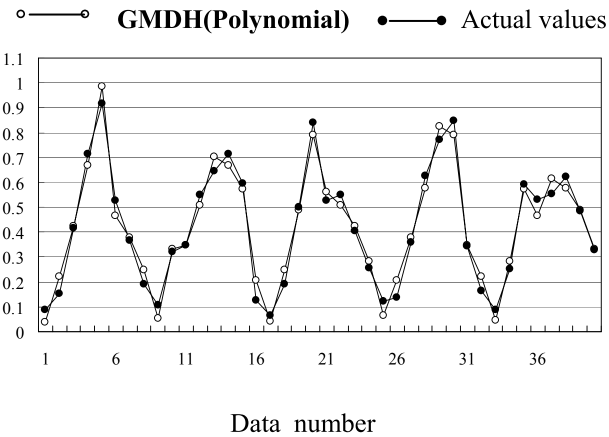

Fig.8. Estimated and predicted values by GMDH

3.3 Identification Results of Conventional Neural Network Trained by Back Propagation Algorithm

Identification results obtained by the conventional neural network trained using back propagation algorithm are quoted from [2]. In conventional multilayered neural network, the neural network was developed as a three layered architecture. Four input variables were used in input layer and four neurons were used in hidden layer. Weights of neural network were estimated by back propagation algorithm. Initial values of the weights were set to random values. The learning calculations of the weights were iterated at 10,000 times. Estimation accuracy was evaluated by Eq.(26) and value of J_1 is shown in Table1. Prediction accuracy was evaluated by Eq.(27) and value of J_2 is shown in Table1. Estimated and predicted values of ϕ by conventional neural network are shown in Fig.9.

3.4 Identification Results of Conventional RBF Neural Network

In conventional RBF neural network, the neural network was developed as a three layered architecture. Four input variables were used in input layer and twenty neurons were used in hidden layer. Weights of neural network were estimated by regression analysis. Estimation accuracy was evaluated by Eq.(26) and value of J_1 is shown in Table1. Prediction accuracy was evaluated by Eq.(27) and value of J_2 is shown in Table1. Estimated and predicted values of ϕ by conventional RBF neural network are shown in Fig.10.

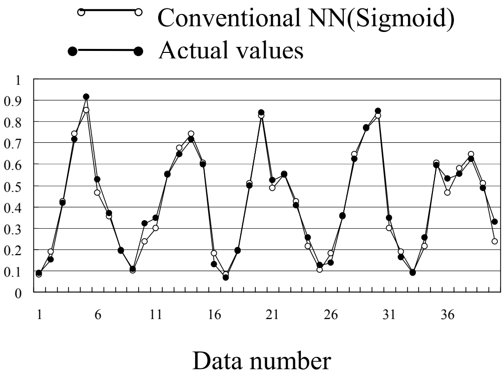

Fig.9. Estimated and predicted values by conventional neural network

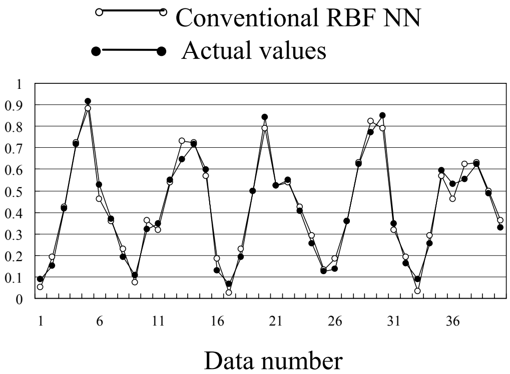

Fig.10. Estimated and predicted values by conventional RBF neural network

3.5 Comparison of Revised GMDH-Type Neural Network and Other Models

Identification results of revised GMDH-type neural network were compared with those of GMDH algorithm, conventional multilayered neural network trained by back propagation algorithm and conventional RBF neural network. From these identification results, both estimation and prediction errors (J_1 and J_2) of revised GMDH-type neural network were the smallest of four identified models and estimated and predicted values of ϕ by revised GMDH-type neural network are very accurate. From these results, we can see that revised GMDH-type neural network algorithm is a very accurate identification method for the nonlinear system.

4 Conclusion

In this paper, a revised GMDH-type neural network algorithm self-selecting optimum neural network architecture was proposed. In this algorithm, optimum neural network architecture is automatically selected from three neural network architectures such as sigmoid function neural network, RBF neural network and polynomial neural network. This algorithm was applied to the nonlinear system identification and it was shown that revised GMDH-type neural network algorithm was a useful method for the nonlinear system identification.

References

[1] Kondo T.: GMDH neural network algorithm using the heuristic self-organization method and its application to the pattern identification problem, *Proc. of the 37th SICE Annual Conference*, p.1143-1148,1998.

[2] Kondo T.: Revised GMDH-type neural networks using prediction error criterions AIC and PSS, *Proc. of SCIS & ISIS 2004*, *WP-6-4,* p.1-6, 2004.

[3] Kondo T., Pandya A.S: Identification of the multi-layered neural networks by revised GMDH-type neural network algorithm with PSS criterion, *Knowledge based intelligent information and engineering systems*, p.1051-1059, 2004.

[4] Farlow S.J.ed.: *Self-organizing methods in modeling, GMDH-type algorithm*, Marcel Dekker, Inc., New York, 1984.

[5] Ivakhnenko A.G.: Heuristic self-organization in problems of engineering cybernetics, *Automatica*, Vo.6, No.2, p.207-219, 1970.

[6] Tamura H., Kondo T.: Heuristics free group method of data handling algorithm of generating optimum partial polynomials with application to air pollution prediction, Int. J. System Sci., Vol.11, No.9, p.1095-1111, 1980.

[7] Draper N.R., Smith H.: *Applied Regression Analysis*, John Wiley and Sons, New York, 1981.