Revised GMDH-type Neural Networks using Prediction Error Criterions AIC and PSS

Tadashi Kondo
School of Health Sciences, The University of Tokushima,3-18-15
Kuramoto-cho Tokushima 770-8509, Japan

Abstract- **In this study, the revised GMDH (Group Method of Data Handling)-type neural network algorithm using prediction error criterions PSS (Prediction Sum of Squares) and AIC (Akaike's Information Criterion) is described. The revised GMDH-type neural network algorithm can generate optimum multi-layered neural network architectures fitting the complexity of nonlinear system so as to minimize the error criterion defined as PSS or AIC. The revised GMDH-type neural networks have abilities of self-selecting the number of layers, the number of neurons in the hidden layers and the useful input variables. This algorithm is applied to the identification problem of the nonlinear complex system and the results obtained by using prediction error criterions PSS and AIC are compared.**

1. INTRODUCTION

The GMDH-type neural networks and their applications have been proposed in our early works [1]-[3]. The GMDH-type neural networks can automatically organize the neural network architecture by using the heuristic self-organization method [4],[5] and they can also determine such structural parameters as the number of layers, the number of neurons in the hidden layers and the useful input variables. These structural parameters are automatically determined so as to minimize the error criterion defined as PSS [6],[7] or AIC [8].

In the conventional multi-layered neural networks trained by using the back propagation method, AIC can not be used to determine the optimum neural network architecture due to the nonuniqueness of the connection weights [9],[10]. Furthermore the conventional multi-layered neural networks do not have the structural identification ability of optimum neural network architectures and so applying the conventional multi-layered neural networks to practical complex problems and obtaining good prediction accuracy is difficult. In the GMDH-type neural networks, the optimum network architecture is automatically organized so as to minimize the error criterion defined as PSS or AIC by using the heuristic self-organization method.

In this study, the revised GMDH-type neural network algorithm using PSS and AIC is applied to the identification problem of the nonlinear complex system and the results are compared with those obtained by the GMDH algorithm and the conventional multi-layered neural networks. These results show that the revised GMDH-type neural network algorithm using PSS and AIC is useful for the structural identification of the multi-layered network architecture.

II. REVISED GMDH-TYPE NEURAL NETWORKS

The architectures of the revised GMDH-type neural networks are automatically organized by using the heuristic self-organization method. At first, we show the procedures of the heuristic self-organization method because it plays very important roles for organization of the revised GMDH-type neural networks.

A. Heuristic se1f-organization method

The heuristic self-organization method is constructed by the following five procedures:

(1) Separating the original data into the training and test sets.

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

(2) Generating the combinations of the input variables in each layer.

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

(3) Calculating the partial descriptions.

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

(4) Selecting the intermediate variables.

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

(5) Stopping the multi-layered iterative computation.

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

B. Revised GMDH-type neural network algorithm

The revised GMDH-type neural networks have a common feedforward multi-layered architecture and the algorithm can organize the same architectures as the conventional multi-layered neural networks trained by using the back propagation method. Figure 1 shows the architecture of the revised GMDH-type neural networks. This neural networks are also organized by using the heuristic self-organization method. But this algorithm does not need the separation of the original data into the training and test sets because PSS or AIC is used instead of test errors. So all the data are used as the training data.

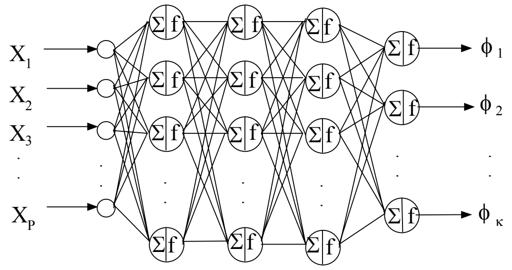

Fig.1 Architecture of revised GMDH-type neural networks.

PSS is described by the following equation:

$$PSS=\sum_{\alpha=1}^{n}(\phi_\alpha - y_\alpha^{*})^2 \qquad (1)$$

where

$$y_\alpha^{*}= b_{0\alpha}+b_{1\alpha}x_{i\alpha}+b_{2\alpha}x_{j\alpha}+b_{3\alpha}x_{i\alpha}x_{j\alpha}+b_{4\alpha}x_{i\alpha}^2+b_{5\alpha}x_{j\alpha}^2, \quad \alpha=1,2,\ldots,n \qquad (2)$$

n is the number of training data, ϕ_α is the α-th observed value for the output variable, $x_{i\alpha}$ is the α-th observed value for the input variable x_i and y_α^{*} is the α-th estimated value obtained by the linear regression analysis of all the data except the α-th datum. In order to compute PSS of Eq. (1), the linear regression analysis must be repeated n times, and the amount of computation increases in the number of data. But PSS of Eq. (1) can be reduced as follows,

$$PSS=\sum_{\alpha=1}^{n}\left((\phi_\alpha - y_\alpha)/(1-\mathbf{x}_\alpha^{T}(X^{T}X)^{-1}\mathbf{x}_\alpha)\right)^2 \qquad (3)$$

where

$$y_\alpha= b_0+b_1x_{i\alpha}+b_2x_{j\alpha}+b_3x_{i\alpha}x_{j\alpha}+b_4x_{i\alpha}^2+b_5x_{j\alpha}^2, \qquad (4)$$

$$\mathbf{x}_\alpha^{T}=[1, x_{i\alpha}, x_{j\alpha}, x_{i\alpha}x_{j\alpha}, x_{i\alpha}^2, x_{j\alpha}^2], \quad \alpha=1,2,\ldots,n \qquad (5)$$

$$X^{T}=[\,\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\,] \qquad (6)$$

y_α is the α-th estimated value obtained by the linear regression analysis of all the data. With this procedure, we need not repeat the regression analysis n times. In the conventional multi-layered neural networks, PSS can not be used to determine the optimum neural network architectures because the back propagation method is used to estimate the connection weights.

In the revised GMDH-type neural network algorithm, we can also use the error criterion defined as AIC to organize the optimum network architecture.

AIC is described by the following equations:

$$AIC = n \log_e S_m^2 + 2(m+1) + C \qquad (7)$$

$$S_m^2=\sum_{\alpha=1}^{n}(\phi_\alpha - y_\alpha)^2/n \qquad (8)$$

$$y_\alpha= b_0+b_1x_{i\alpha}+b_2x_{j\alpha}+b_3x_{i\alpha}x_{j\alpha}+b_4x_{i\alpha}^2+b_5x_{j\alpha}^2, \quad \alpha=1,2,\ldots,n \qquad (9)$$

Here, m is the number of terms in Eq.(9), n is the number of training data and C is a constant.

The procedures for determining the architecture of the revised GMDH-type neural networks conform to the following:

(1) The first layer

$$u_j=x_j \quad (j=1,2,\ldots,p) \qquad (10)$$

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

(2) The second layer

Many combination of r input variables are generated. For each combination, the neuron architectures are automatically selected from the following neurons so as to minimize PSS or AIC value.

i) The first type neuron

Σ: (Nonlinear function)

$$z_k=w_1u_i+w_2u_j+w_3u_iu_j+w_4u_i^2+w_5u_j^2+w_6u_i^3+w_7u_i^2u_j+w_8u_iu_j^2+w_9u_j^3 - w_0\theta_1 \qquad (11)$$

f : (Nonlinear function)

$$y_k= 1/(1+\exp(-z_k)) \qquad (12)$$

Here, $\theta_1=1$ and w_i $(i=0,1,2,\ldots,9)$ are the weights between the first and second layer. The value of r, which is the number of input variables u in each neuron, is set to two for the first type neuron.

ii) The second type neuron

Σ: (Linear function)

$$z_k= w_1u_1+w_2u_2+w_3u_3+\cdots+w_ru_r - w_0\theta_1 \quad (r<p) \qquad (13)$$

f : (Nonlinear function)

$$y_k= 1/(1+\exp(-z_k)) \qquad (14)$$

Here, $\theta_1=1$ and w_i $(i=0,1,2,\ldots,r)$ are the weights between the first and second layer. The 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 $(i=1,2,\ldots,p)$.

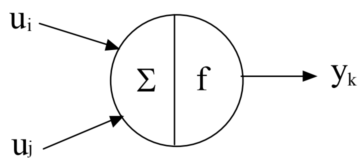

Fig.2 Neuronal architecture with two inputs

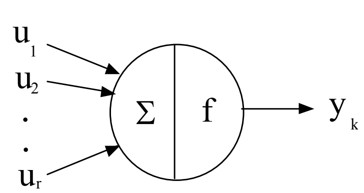

Fig.3 Neuronal architecture with r inputs

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

The weights w_i $(i=0,1,2,\ldots)$ are estimated by the stepwise regression analysis using PSS or AIC.

[The estimation procedure of the weight w_i]

First, the values of z_k are calculated by using the following equation:

$$z_k=log_e(\phi'/(1-\phi')) \quad (15)$$

where ϕ' is the normalized output variable whose values are between zero and one. Then the weights w_i are estimated by using the stepwise regression analysis [11] which selects useful input variables by using PSS or AIC. Only useful variables in Eq.(11) and Eq.(13) are selected by the stepwise regression analysis using PSS or AIC and the optimum neuronal architectures are organized by the selected useful variables.

From these generated neurons, the L neurons which minimize PSS or AIC value are selected. The output variables y_k of L selected neurons are set to the input variables of the neurons in the third layer.

(3) The third and succeeding layers

In the third and succeeding layers, the same computation of the second layer is iterated until PSS or AIC values of L neurons stop decreasing. When iterative computation is terminated, the neural network architectures are produced by the selected neurons in each layer.

By using these procedures, the revised GMDH-type neural networks can be organized. The flowchart for evolving the architecture of the revised GMDH-type neural networks is shown in Fig.4. In this algorithm, if only the second type neurons are used to organize the neural networks, we can construct the same architectures as the conventional multi-layered neural networks trained by using the back propagation method.

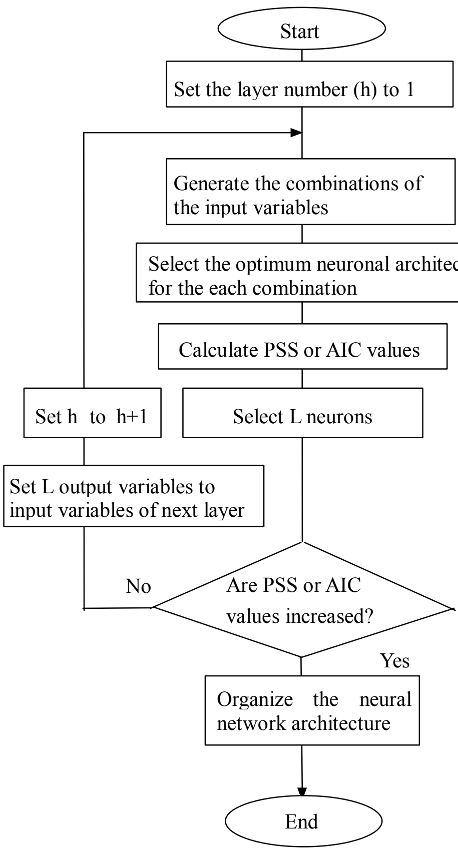

Fig.4 Flowchart for evolving the architecture of the revised GMDH-type neural networks

III. APPLICATION TO THE NONLINEAR SYSTEM IDENTIFICATION

The nonlinear system is assumed to be described by the following equations:

$$\phi=f_1(x_1, x_2, x_3)\,/\,f_2(x_1, x_2, x_3)+\varepsilon \quad (16)$$

$$f_1(x_1, x_2, x_3)=1.0+2.0\,x_1^2x_2+3.0\,x_2^2x_3 \quad (17)$$

$$f_2(x_1, x, x_3)=1.0+2.0exp(x_1)+3.0exp(x_1x_2)+4.0exp(x_3) \quad (18)$$

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

A. The identification results obtained by using the revised GMDH-type neural networks

The neural network architecture identified by the revised GMDH-type neural network algorithm using AIC was the same one identified by the revised GMDH-type neural network algorithm using PSS.

The identification results of the revised GMDH-type neural networks are shown as follows:

(1) Input variables.

Four input variables were used, but the 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) Variation of AIC.

The variation of AIC is shown in Fig.5. AIC values converged at the fourth layer.

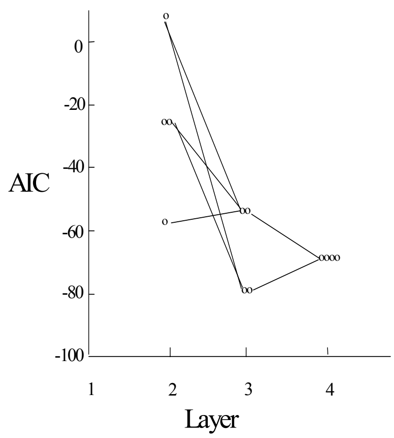

Fig.5 Variation of AIC in the revised GMDH-type neural networks

(4) Variation of PSS.
The variation of PSS is shown in Fig.6. PSS values converged at the fourth layer.

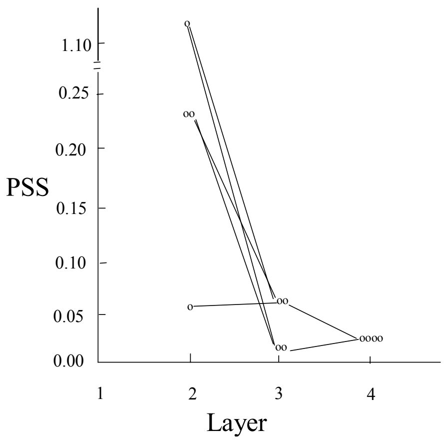

Fig.6 Variation of PSS in the revised GMDH-type neural networks

(5) Architecture of the neural networks.
The calculation of the revised GMDH-type neural networks was terminated at the fourth layer and the neurons of the third layer had minimum PSS and AIC values. Therefore, three layered neural network architecture was organized. The first layer is the input layer and the second layer is the hidden layer and the third layer is the output layer.
(6) Estimation accuracy.
The estimation accuracy was evaluated by using the following equation:

$$J_1 = \sum_{i=1}^{20} \left| \phi_i - \phi_i^* \right| / 20 \qquad (19)$$

where ϕ_i $(i=1,2,\cdots,20)$ were the actual values with the Gaussian white noise ε and ϕ_i^* $(i=1,2,\cdots,20)$ were the estimated values by the revised GMDH-type neural networks. ϕ_i $(i=1,2,\cdots,20)$ ware used to organize the revised GMDH-type neural networks. The value of J_1 is shown in Table1. In this table, GMDH-NN shows the revised GMDH-type neural networks and conventional NN shows the conventional neural networks.

Table 1 Identification results of three methods

Method	J1	J2	Layeres
GMDH-NN (PSS)	0.01462	0.01831	3
GMDH-NN (AIC)	0.01462	0.01831	3
GMDH (PSS)	0.04094	0.03829	5
GMDH (AIC)	0.03991	0.04526	4
Conventional NN	0.02614	0.02813	3

(7) Prediction accuracy.
The prediction accuracy was evaluated by using the following equation:

$$J_2 = \sum_{i=21}^{40} \left| \phi_i - \phi_i^* \right| / 20 \qquad (20)$$

where ϕ_i $(i=21,22,\cdots,40)$ were the actual values with the Gaussian white noise ε and ϕ_i^* $(i=21,22,\cdots,40)$ were the predicted values by the revised GMDH-type neural networks. ϕ_i $(i=21,22,\cdots,40)$ were not used to organize the revised GMDH-type neural networks and were used to check the generalization ability. The value of J_2 is shown in Table1 and is very small. From this prediction result, we can see that the revised GMDH-type neural networks do not overfit the training data and have good generalization ability.
(8) Estimated and predicted values.
The estimated and predicted values of ϕ by the revised GMDH-type neural networks are shown in Fig.7. The estimated values are shown for the data points between the first and the 20-th data entities and the predicted values are shown for the data points between the 21-th and the 40-th data entities. We can see that the estimated and predicted values are very accurate.

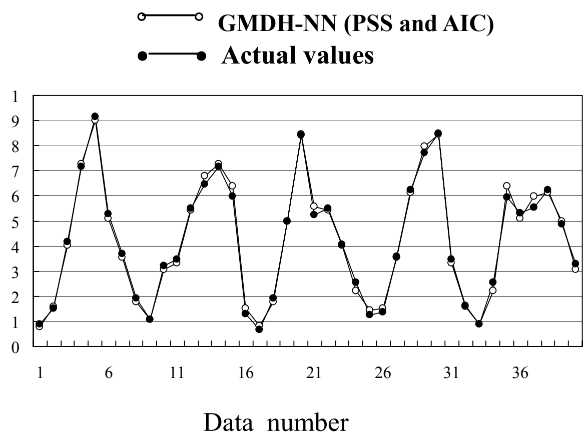

Fig.7 Estimated and predicted values by the GMDH-type neural networks

B. The identification results obtained by the GMDH algorithm.

The polynomial type networks identified by the GMDH algorithm using AIC ware different from those identified by the GMDH algorithm using PSS criterion. The identification results of the GMDH algorithm are shown as follows:

[The identification results by the GMDH algorithm using AIC]
(1) Input variables.
Four input variables were used, but again the 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 AIC
The variation of AIC is shown in Fig.8. AIC values converged at the fourth layer.

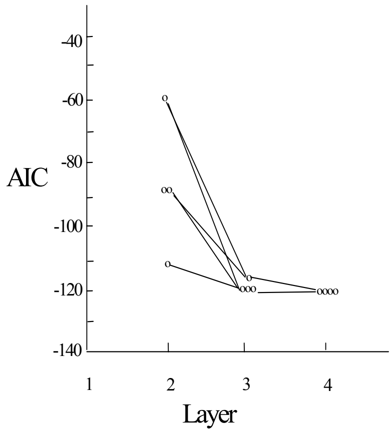

Fig.8 Variation of AIC in the GMDH

(4) Architecture of the networks.

The calculation of the GMDH converged at the fourth layer and the neurons of the fourth layer had the minimum AIC value. Therefore, four layered polynomial type network architecture was organized. The first layer was the input layer and the second and the third layer were the hidden layers and the fourth layer was the output layer.

(5) Estimation accuracy.

The estimation accuracy was evaluated by using Eq.(19) and the value of J_1 is shown in Table1.

(6) Prediction accuracy.

The prediction accuracy was evaluated by using Eq.(20) and the value of J_2 is shown in Table1.

(7) Estimated and predicted values.

The estimated and predicted values of ϕ by the GMDH are shown in Fig.9. The estimation and prediction accuracy were not good compared with the revised GMDH-type neural networks.

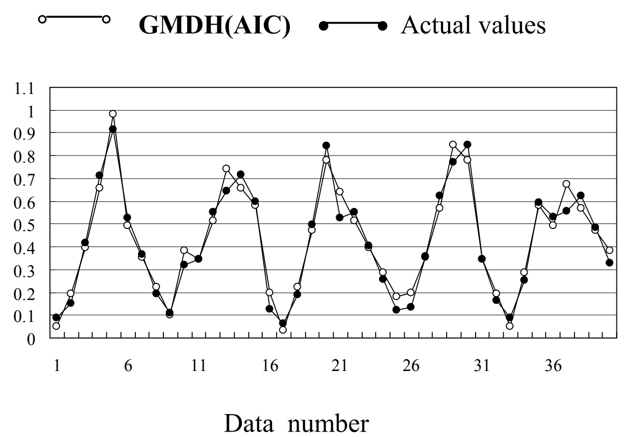

Fig.9 Estimated and predicted values by the GMDH

[The identification results by the GMDH algorithm using PSS criterion]

(1) Input variables.

Four input variables were used, but again the 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

The variation of PSS is shown in Fig.10. PSS values converged at the fifth layer.

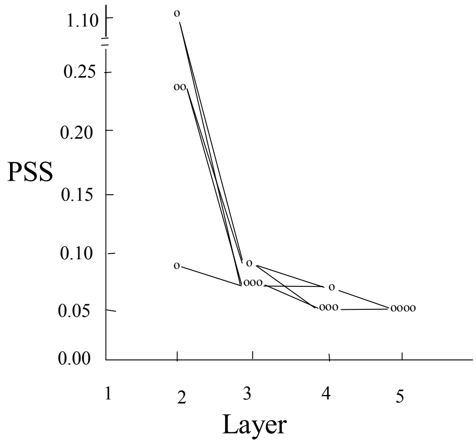

Fig.10 Variation of PSS in the GMDH

(4) Architecture of the networks.

The calculation of the GMDH converged at the fifth layer and the neurons of the fifth layer had the minimum PSS value. Therefore, five layered polynomial type network architecture was organized. The first layer was the input layer and the second, the third and the fourth layer were the hidden layers and the fifth layer was the output layer.

(5) Estimation accuracy.

The estimation accuracy was evaluated by using Eq.(19) and the value of J_1 is shown in Table1.

(6) Prediction accuracy.

The prediction accuracy was evaluated by using Eq.(20) and the value of J_2 is shown in Table1.

(7) Estimated and predicted values.

The estimated and predicted values of ϕ by the GMDH are shown in Fig.11. The estimation and prediction accuracy were not good compared with the revised GMDH-type neural networks.

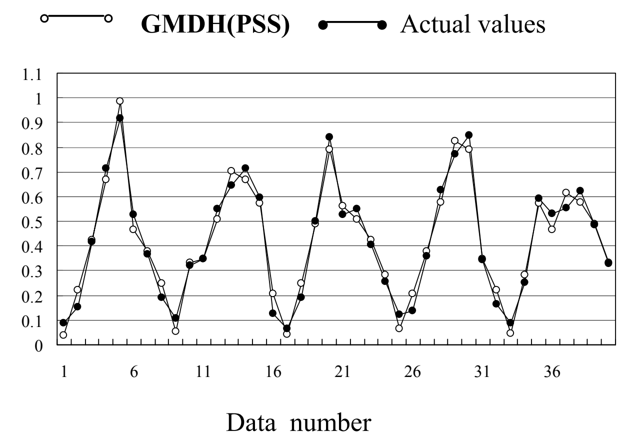

Fig.11 Estimated and predicted values by the GMDH

C. The identification results of the conventional neural networks trained by using the back propagation method.

In the conventional multi-layered neural networks, the neural network was developed as a three layered architecture. Four input variables were used in the input layer and four neurons were used in the hidden layer. The weights of the neural networks were estimated by using the back propagation method. The estimation accuracy was evaluated by using Eq.(19) and the value of J_1 is shown in Table1. The prediction accuracy was evaluated by using Eq.(20) and the value of J_2 is shown in Table1. The estimated and predicted values of ϕ by the conventional neural networks are shown in Fig.12.

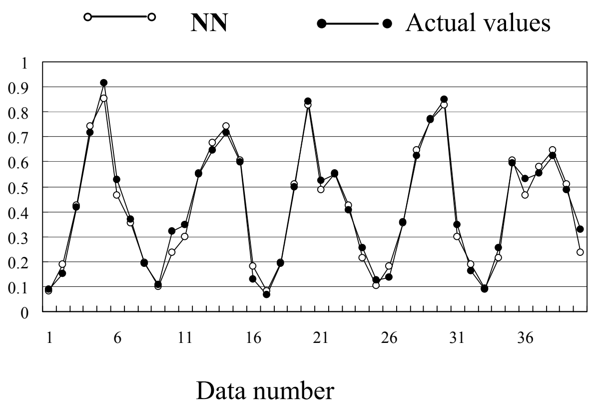

Fig.12 Estimated and predicted values by the conventional neural networks

D. Comparison of the revised GMDH-type neural networks and the other models.

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

VI. CONCLUSION

The revised GMDH-type neural network algorithm using prediction error criterions PSS and AIC can automatically organize a multi-layered neural network architecture fitting the complexity of the nonlinear system by using the heuristic self-organization method. It is very easy to apply this algorithm to the identification problem of the practical complex system because the structural parameters such as the number of layers, the number of neurons in the hidden layers and the useful input variables are automatically determined so as to minimize PSS or AIC. This algorithm was applied to the nonlinear system identification problem and it was shown that this algorithm was a very useful method.

REFERENCES

[1] T. Kondo. 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*, pp.1143-1148,1998.

[2] T. Kondo, A.S. Pandya and J.M. Zurada. GMDH-type neural networks and their application to the medical image recognition of the lungs, *Proc. of the 38th SICE Annual Conference*, pp.1181-1186, 1999.

[3] T. Kondo, A.S. Pandya and J.M. Zurada. Logistic GMDH-type neural networks and their application to the identification of the X-ray film characteristic curve, *Proc. of IEEE International Conference on System, Man, and Cybernetics,* pp.437-442, 1999.

[4] A.G. Ivakhnenko. Heuristic self-organization in problems of engineering cybernetics, *Automatica*, vol.6, No.2, pp.207-219, 1970.

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

[6] D.M. Allen. The relationship between variable selection and data augmentation and a method for prediction, *Technometrics,* vo.16, No.1, pp.125-127, 1974.

[7] H.Tamura and T.Kondo. 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, pp.1095-1111,1980.

[8] H. Akaike. A new look at the statistical model identification, *IEEE Trans. Automatic Control*, vol.AC-19, No.6, pp.716-723, 1974.

[9] K. Hagiwara, N. Toda and S. Usui. On the problem of applying AIC to determine the structure of a layered feed-forward neural network, *Proc. of International Joint Conference on Neural Networks,* vol.III, pp.2263-2266, 1993.

[10] K. Hagiwara et al.. Upperbound of the expected training error of neural network regression for a Gaussian noise sequence, *Neural Networks*, vol.14, pp.1419-1429, 2001.

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