

Cutting Error Prediction by Multilayer Neural Networks for Machine Tools with Thermal Expansion and Compression

Kenji NAKAYAMA Akihiro HIRANO Shinya KATOH
Tadashi YAMAMOTO† Kenichi NAKANISHI† Manabu SAWADA†

Dept of Information and Systems Eng., Faculty of Eng., Kanazawa Univ.
2-40-20 Kodatsuno, Kanazawa, 920-8667, Japan
e-mail: nakayama@t.kanazawa-u.ac.jp
† Nakamura-Tome Precision Industry Co., Ltd.

Abstract

In training neural networks, it is important to reduce input variables for saving memory, reducing network size, and achieving fast training. This paper proposes two kinds of selecting methods for useful input variables. One of them is to use information of connection weights after training. If a sum of absolute value of the connection weights related to the input node is large, then this input variable is selected. In some case, only positive connection weights are taken into account. The other method is based on correlation coefficients among the input variables. If a time series of the input variable can be obtained by amplifying and shifting that of another input variable, then the former can be absorbed in the latter. These analysis methods are applied to predicting cutting error caused by thermal expansion and compression in machine tools. The input variables are reduced from 32 points to 16 points, while maintaining good prediction within $6\mu\text{m}$, which can be applicable to real machine tools.

1 Introduction

Recently, prediction and diagnosis have been very important in a real world. In many cases, relations between the past data and the prediction, and the symptoms and diseases are complicated nonlinear. Neural networks are useful for these signal processing. Many kinds of approaches have been proposed [1]–[12].

In these applications, observations, which are the past data, the symptoms and so on, are applied to the input nodes of neural networks, and the prediction and the diseases are obtained at the network outputs. In order to train neural networks and to predict the coming

phenomenon and to diagnose the diseases, it should be analyzed what kinds of observations are useful for these purposes. Usually, the observations, which seems to be meaningful by experience, are used. In order to simplify observation processes, to minimize network size, and to make a learning process fast and stable, the input data should be minimized. How to selected the useful input variables have been discussed [11]–[12].

In this paper, selecting methods for useful input variables are proposed. The corresponding connection weights and correlation coefficients among the input data are used. The proposed methods are applied to predicting cutting error caused by thermal expansion and compression in numerical controlled (NC) machine tools. Temperature is measured at many points on the machine tool and in the surroundings. For instance, 32 points are measured, which requires a complicated observation system. It is desirable to reduce the temperature measuring points.

2 Network Structure and Equations

Figure 1 shows a multilayer neural network with a single hidden layer. Relations among the input, hidden layer outputs and the final outputs are shown here.

$$u_j(n) = \sum_{i=1}^N w_{ji}x_i(n) + \theta_j \quad (1)$$

$$y_j(n) = f_h(u_j(n)) \quad (2)$$

$$u_k(n) = \sum_{j=1}^J w_{kj}y_j(n) + \theta_k \quad (3)$$

$$y_k(n) = f_o(u_k(n)) \quad (4)$$

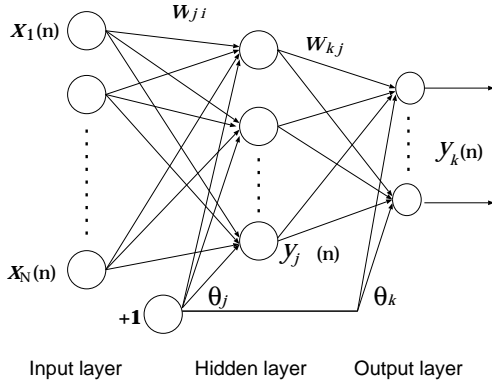


Figure 1: Multilayer neural network with a single hidden layer.

$f_h()$ and $f_o()$ are sigmoid functions. The connection weights are trained through supervised learning algorithms, such as an error back-propagation algorithm.

3 Analysis Methods for Useful Input Data

3.1 Method-Ia: Based on Absolute Value of Connection Weights

The connection weights are updated following the error back-propagation (BP) algorithm. $f_o()$ and $f_h()$ are the 1st order derivative of $f_o()$ and $f_h()$, respectively.

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n) \quad (5)$$

$$\Delta w_{kj}(n) = \alpha \Delta w_{kj}(n-1) + \eta \delta_k y_j(n) \quad (6)$$

$$\delta_k = e_k(n) \dot{f}_o(u_k(n)) \quad (7)$$

$$e_k(n) = d_k(n) - y_k(n), \quad d_k(n) \text{ is a target.} \quad (8)$$

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n) \quad (9)$$

$$\Delta w_{ji}(n) = \alpha \Delta w_{ji}(n-1) + \eta \delta_j x_i(n) \quad (10)$$

$$\delta_j = \dot{f}_h(u_j(n)) \sum_{k=1}^K \delta_k w_{kj}(n) \quad (11)$$

From the above equations, the connection weight $w_{ji}(n)$ is updated by $\eta \delta_j x_i(n)$. Since η is usually a positive small number, then by repeating the updating, $\eta \delta_j x_i(n)$ is accumulated in $w_{ji}(n+1)$. Thus, growth of $w_{ji}(n+1)$ is expressed by $E[\delta_j x_i(n)]$, which is a cross-correlation. On the other hand, as shown in Eq.(11), δ_j expresses the output error caused by the j th hidden unit output. If the input variable $x_i(n)$ is an important factor, then it may be closely related to the output error, and their cross-correlation becomes a large value. For this reason, it can be expected that the connection weights for the important input variables to the hidden units will be grown up in a learning process. Based on this analysis,

the important input variables are selected by using a sum of the corresponding connection weights after the training.

$$S_{i,abs} = \sum_{j=1}^J |w_{ji}| \quad (12)$$

3.2 Method-Ib: Based on Positive Connection Weights

When the input data always take positive numbers, negative connection weights may reduce the input potential $u_j(n)$ in Eq.(1). Furthermore, when a sigmoid function shown in Fig.2 is used for an activation function, negative $u_j(n)$ generates small output, which does not affect the final output. Thus, in this case, the negative connection weights are not useful. Therefore, only positive connection weights are taken into account. The useful temperatures are selected based on $S_{i,posit}$.

$$S_{i,posit} = \sum_{\sigma} w_{\sigma i}, \quad w_{\sigma i} > 0 \quad (13)$$

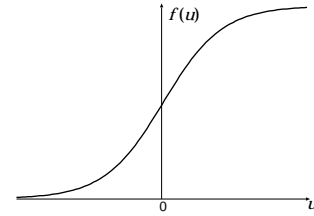


Figure 2: Sigmoid function, whose output is always positive.

3.3 Method II: Based on Cross-correlation among Input variables

Dependency among Input variables

First, we discuss using a neuron model shown in Fig.3. The input potential u is given by

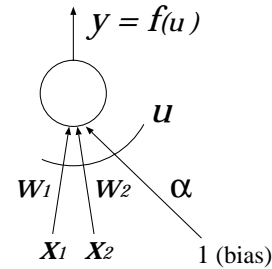


Figure 3: Neuron model.

$$u = w_1 x_1 + w_2 x_2 + \alpha \quad (14)$$

If the following linear dependency is held,

$$x_2 = ax_1 + b \quad a \text{ and } b \text{ are constant} \quad (15)$$

then, u is rewritten as follows:

$$u = w_1x_1 + w_2(ax_1 + b) + \alpha \quad (16)$$

$$= (w_1 + aw_2)x_1 + (bw_2 + \alpha) \quad (17)$$

$$(18)$$

Therefore, by replacing w_1 by $w_1 + aw_2$ and α by $bw_2 + \alpha$, the input variable x_2 can be removed as follows:

$$u = wx_1 + \beta \quad (19)$$

$$w = w_1 + aw_2 \quad (20)$$

$$\beta = bw_2 + \alpha \quad (21)$$

This is an idea behind the proposed analysis method. The linear dependency given by Eq.(15) can be analyzed by using correlation coefficients.

Correlation Coefficients

The i th input variable is defined follows:

$$\mathbf{x}_i = [x_i(0), x_i(1), \dots, x_i(L-1)]^T \quad (22)$$

Correlation coefficient between the i th and the j th variable vectors is given by

$$\rho_{ij} = \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^T (\mathbf{x}_j - \bar{\mathbf{x}}_j)}{\|\mathbf{x}_i - \bar{\mathbf{x}}_i\| \|\mathbf{x}_j - \bar{\mathbf{x}}_j\|} \quad (23)$$

$$\bar{\mathbf{x}}_{i,j} = \frac{1}{L} \sum_{n=0}^{L-1} x_{i,j}(n) \quad (24)$$

If \mathbf{x}_i and \mathbf{x}_j satisfy Eq.(15), then $\rho_{ij} = 1$. In other words, if ρ_{ij} is close to unity, then \mathbf{x}_i and \mathbf{x}_j are linearly dependent. On the other hand, if $\rho_{ij} = 0$, then they are orthogonal to each other.

Combination of Data Sets

The modifications by Eqs.(19)–(21) are common in a MLNN. This means the modifications are the same for all the input data sets. Let the number of the data sets be Q . The input variables are re-defined as follows:

Definition of Input Data for Q Data Sets

$$\mathbf{x}_i^{(q)} = [x_i^{(q)}(0), x_i^{(q)}(1), \dots, x_i^{(q)}(L-1)]^T \quad (25)$$

$$\mathbf{X}^{(q)} = [\mathbf{x}_1^{(q)}, \mathbf{x}_2^{(q)}, \dots, \mathbf{x}_N^{(q)}] \quad (26)$$

$$\mathbf{X}_{total} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \\ \vdots \\ \mathbf{X}^{(Q)} \end{bmatrix} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_N] \quad (27)$$

$$\tilde{\mathbf{x}}_i = \begin{bmatrix} \mathbf{x}_i^{(1)} \\ \mathbf{x}_i^{(2)} \\ \vdots \\ \mathbf{x}_i^{(Q)} \end{bmatrix} \quad (28)$$

$\mathbf{x}_i^{(q)}$ is the input variable vector of the q th input data set. $\mathbf{X}^{(q)}$ is the q th input data set, \mathbf{X}_{total} is a total input data set, which includes all the input data. In $\tilde{\mathbf{x}}_i$, the i th variables at all sampling points, $n = 0, 1, \dots, L-1$ and for all data sets $q = 1, 2, \dots, Q$ are included. Using these notations, the correlation coefficients are defined as follows:

$$\rho_{ij} = \frac{(\tilde{\mathbf{x}}_i - \bar{\tilde{\mathbf{x}}}_i)^T (\tilde{\mathbf{x}}_j - \bar{\tilde{\mathbf{x}}}_j)}{\|\tilde{\mathbf{x}}_i - \bar{\tilde{\mathbf{x}}}_i\| \|\tilde{\mathbf{x}}_j - \bar{\tilde{\mathbf{x}}}_j\|} \quad (29)$$

$$\bar{\tilde{\mathbf{x}}}_i = \frac{1}{LQ} \sum_{q=1}^Q \sum_{n=0}^{L-1} x_i^{(q)}(n) \quad (30)$$

One example of the combined input data is shown in Fig.4, where $Q = 4$.

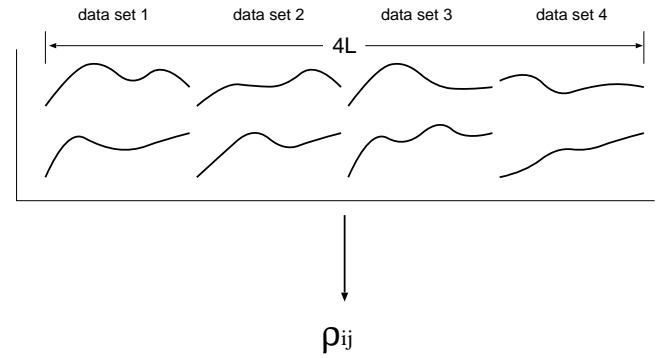


Figure 4: Combined input variable vectors.

Average of Correlation Coefficients

Method-IIa

The correlation coefficients for all combinations of the input variables are calculated by Eq.(29). Furthermore, dependency of the i th variable is evaluated by

$$\bar{\rho}_i^{(1)} = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^N \rho_{ij} \quad (31)$$

$\bar{\rho}_i^{(1)}$ expresses average of the correlation coefficients between the i th variable and all the other variables. Thus, the variables, which have small $\bar{\rho}_i$ are selected for the useful input variables.

Method-IIb

Let \mathbf{x}_σ be the selected variable vectors, and the number of \mathbf{x}_σ be N_1 . The correlation coefficients $\rho_{ij}^{(2)}$ are evaluated once more among the selected σ variable vectors. The variable vectors are further selected based on $\rho_{ij}^{(2)}$. The variables having large $\rho_{ij}^{(2)}$ are removed from the selected set. Instead, the variables, which are not selected in Method-IIa and have small $\bar{\rho}_i^{(1)}$, are selected and added to \mathbf{x}_σ . This process is repeated until all the selected variables have small $\rho_{ij}^{(2)}$.

3.4 Comparison with Other Analysis Methods

There are several methods to extract important components among the input data. One of them is principal component analysis. The other method is vector quantization. In these methods, however, in order to extract these components and vectors, many input data are required. Our purpose is to simplify the observation process for the input data, that is to select useful input variables, which are directly observed. It is difficult to obtain the useful observation data from the principal components and the representative vectors.

4 Prediction of Cutting Error Caused by Thermal Expansion and Compression

Numerical controlled (NC) machine tools are required to guarantee very high cutting precision, for instance tolerance of cutting error is within $10\mu\text{m}$ in diameter. There are many factors, which degrade cutting precision. Among them, thermal expansion and compression of machine tools are very sensitive in cutting precision. In this paper, the multilayer neural network is applied to predicting cutting error caused by thermal effects.

4.1 Structure of NC Machine Tool

Figure 5 shows a blockdiagram of NC machine tool. Distance between the cutting tool and the objective is changed by thermal effects. Temperatures at many

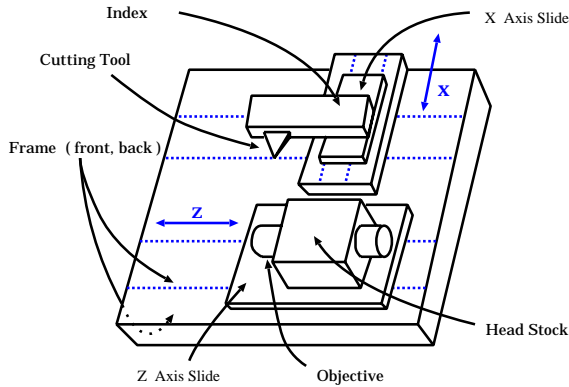


Figure 5: Rough sketch of NC machine tool.

points on the machine tool and in surrounding are measured. The number of measuring points is up to 32 points.

4.2 Multilayer Neural Network

Figure 6 shows the multilayer neural network used predicting cutting error of machine tools. The temperature and deviation are measured as a time series. Thermal expansion and compression of machine tools are also

dependent on hysteresis of temperature change. $x_i(n)$ means the temperature at the i th measuring point and at the n th sampling points on the time axis. Its delayed samples $x_i(n-1), x_i(n-2), \dots$ are generated through the delay elements "T" and are applied to the MLNN. One hidden layer and one output unit are used.

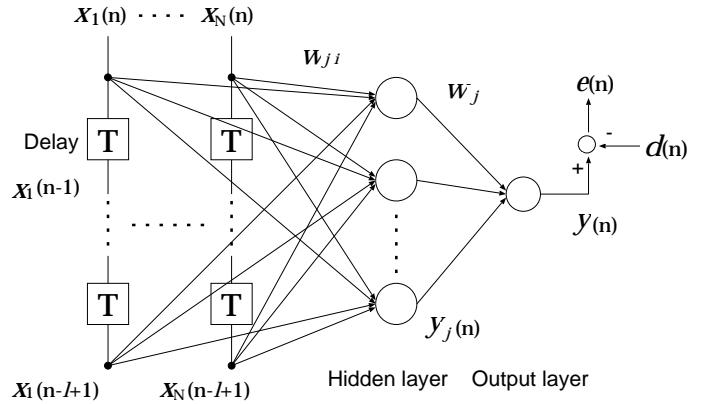


Figure 6: Neural network used for predicting cutting error caused by thermal effects.

4.3 Training and Testing Using All Input variables

Four kinds of data sets are measured by changing cutting conditions. They are denoted D_1, D_2, D_3, D_4 . Since it is not enough to evaluate prediction performance of the neural network, data sets are increased by combining the measured data sets by linear interpolation, denoted $D_{12}, D_{13}, D_{14}, D_{23}, D_{24}, D_{34}$. Some of the measured temperatures in time are shown in Fig.7. Training and testing conditions are shown in Table 1. All measuring points are employed. The data sets except for D_1 are used for training and D_1 is used for testing.

Figure 8 shows a learning curve using all data sets,

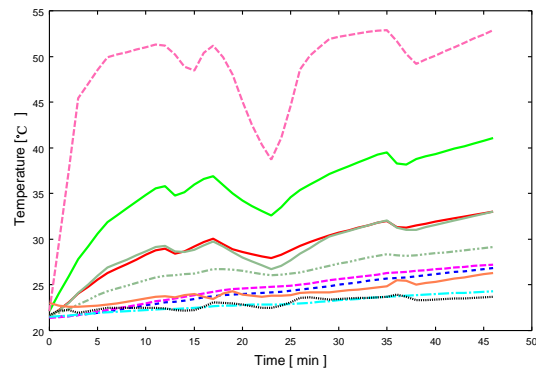


Figure 7: Some of measured temperatures in time.

Table 1: Training and testing conditins

Measuring points	32 points
Training data sets	D_2, D_3, D_4 $D_{12}, D_{13}, D_{14}, D_{23}, D_{24}, D_{34}$
Test data set	D_1
Learning rate η	0.001
Momentum rate α	0.9
Iterations	100,000

except for D_1 , and all measuring points, that is 32 points. The vertical axis means the mean squared error (MSE) of difference between the measured cutting error and the predicted cutting error. It is well reduced.

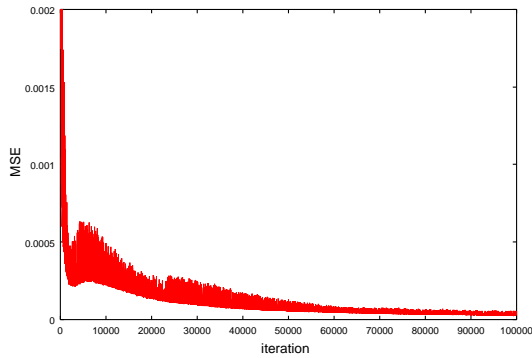


Figure 8: Learning curve of cutting error prediction. All measuring points are used.

Figure 9 shows cutting error prediction under the conditions in Table 1. The prediction error is within $6\mu m$, which satisfies the tolerance $10\mu m$ in diameter.

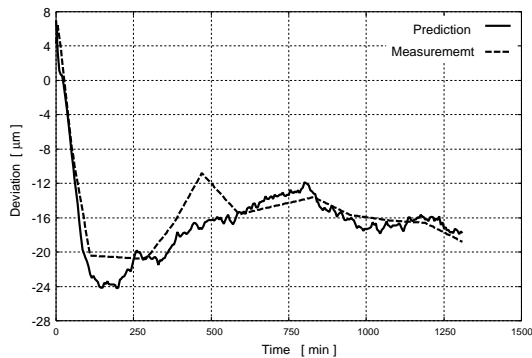


Figure 9: Cutting error prediction. All measuring points are used.

5 Selection of Useful Measuring Points

The useful 16 measuring points are selected from 32 points by the analysis methods proposed in Sec.3.

5.1 Selection Based on Connection Weights

The measuring points are selected based on a sum of absolute value of the connection weights $S_{i,abso}$ (Method-Ia) and on a sum of positive connection weights $S_{i,positi}$ (Method-Ib). Figure 10 shows both sums. The horizontal axis shows the temperature measuring points, that is 32 points. Their prediction are shown in Fig.11. The selection method using $S_{i,positi}$ is superior to the other using $S_{i,abso}$, because the temperature in this experience is always positive. The prediction error is within $6\mu m$.

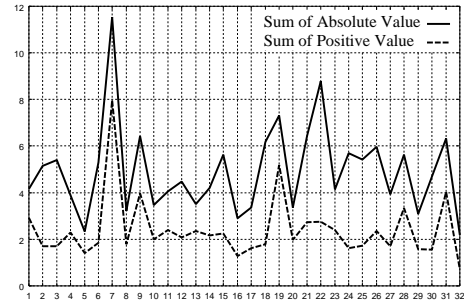


Figure 10: Sum of absolute value of temperature and positive temperature.

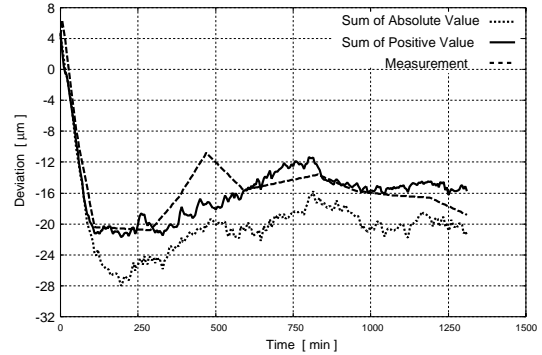


Figure 11: Cutting error prediction with 16 measuring points selected by connection weights.

5.2 Selection Based on Correlation Coefficients

In Method-IIb, after the first selection by Method-IIa, the variables, whose correlation coefficients $\rho_{ij}^{(2)}$ exceed 0.9 are replaced by the variables, which are not selected in the first stage and have small $\bar{\rho}_i^{(1)}$. Simulation results by both methods are shown in Fig.12. The result by Method-IIb "Correlation(2)" is superior to that of Method-IIa "Correlation(1)". Because the former more precisely evaluates the correlation coefficients.

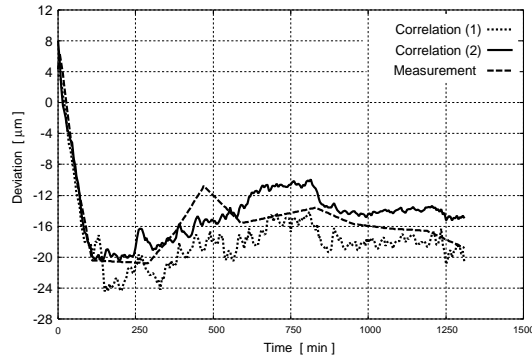


Figure 12: Cutting error prediction with 16 measuring points selected by correlation coefficients.

5.3 Comparison of Selected Measuring Points

Table 2 shows the selected temperature measuring points by four kinds of the methods. Comparing the selected measuring points by Method-Ib and Method-IIb, the following 9 points 4, 11, 12, 14, 15, 21, 22, 28, 31 are common. However, the selected measuring points are not exactly the same. A combination of the measuring points seems to be important.

Table 2: Measuring points selected by four kinds of methods.

Methods	Measuring points
Method-Ia Sum of absolute values	2, 3, 6, 7, 9, 15, 18, 19, 21, 22, 24, 25, 26, 28, 30, 31
Method-Ib Sum of positive weights	1, 4, 7, 9, 11, 12, 13, 14, 15, 19, 21, 22, 23, 26, 28, 31
Method-IIa Correlation(1)	3, 4, 7, 8, 12, 14, 16, 18, 19, 20, 21, 22, 23, 25, 31, 32
Method-IIb Correlation(2)	3, 4, 8, 11, 12, 14, 15, 16, 20, 21, 22, 27, 28, 30, 31, 32

6 Conclusions

Two kinds of methods, selecting the useful input variables, have been proposed. They are based on the connection weights and the correlation coefficients among the input variables. The proposed methods have been applied to predicting cutting error caused by thermal effects in machine tools. Simulation results show precise prediction with the reduced number of the input variables.

References

- [1] S.Haykin, "Neural Networks: A Comprehensive Foundation," Macmillan, New York, 1994.
- [2] S.Haykin and L. Li, "Nonlinear adaptive prediction of nonstationary signals," IEEE Trans. Signal Processing, vol.43, No.2, pp.526-535, February. 1995.
- [3] A. S. Weigend and N. A. Gershenfeld, "Time series prediction: Forecasting the future and understanding the past. in Proc. V. XV, Santa Fe Institute, 1994.
- [4] M. Kinouchi and M. Hagiwara, "Learning temporal sequences by complex neurons with local feedback," Proc. ICNN'95, pp.3165-3169, 1995.
- [5] T.J. Cholewo and J.M. Zurada, "Sequential network construction for time series prediction," in Proc. ICNN'97, pp.2034-2038, 1997.
- [6] A. Atia, N. Talaat, and S. Shaheen, "An efficient stock market forecasting model using neural networks," in Proc. ICNN'97, pp.2112-2115, 1997.
- [7] X.M. Gao, X.Z. Gao, J.M.A. Tanskanen, and S.J. Ovaska, "Power prediction in mobile communication systems using an optimal neural-network structure", IEEE Trans. on Neural Networks, vol.8, No.6, pp.1446-1455, November 1997.
- [8] A.A.M.Khalaf, K.Nakayama, "A cascade form predictor of neural and FIR filters and its minimum size estimation based on nonlinearity analysis of time series", IEICE Trans. Fundamental, vol.E81-A, No.3, pp.364-373, March 1998.
- [9] A.A.M.Khalaf, K.Nakayama, "Time series prediction using a hybrid model of neural network and FIR filter", Proc. of IJCNN'98, Anchorage, Alaska, pp.1975-1980, May 1998.
- [10] A.A.M. Khalaf and K.Nakayama, "A hybrid nonlinear predictor: Analysis of learning process and predictability for noisy time series", IEICE Trans. Fundamentals, Vol.E82-A, No.8, pp.1420-1427, Aug. 1999.
- [11] K.Hara and K.Nakayama, "Training data selection method for generalization by multilayer neural networks", IEICE Trans. Fundamentals, Vol.E81-A, No.3, pp.374-381, March 1998.
- [12] K.Hara and K.Nakayama, "A training data selection in on-line training for multilayer neural networks", IEEE-INNS Proc. of IJCNN'98, Anchorage, pp.227-2252, May 1998.