Proceedings of the 2007 IEEE Symposium on Computational
Intelligence in Multicriteria Decision Making (MCDM 2007)

Nonlinear Dynamic System Identification Based on Multiobjectively Selected RBF Networks

Nobuhiko Kondo, Toshiharu Hatanaka
Department of Information and Physical Sciences
Osaka University
Suita, 565–0871, Japan
Email: {nobuhiko,hatanaka}@ist.osaka-u.ac.jp

Katsuji Uosaki
Department of Management and Information Sciences
Fukui University of Technology
Fukui, 910–8505,Japan
Email: uosaki@fukui-ut.ac.jp

***Abstract*— In this paper, nonlinear dynamic system identification by using multiobjectively selected RBF network is considered. RBF networks are widely used as a model structure for nonlinear systems. The determination of its structure that is the number of basis functions is prior important step in system identification, and the tradeoff between model complexity and accuracy exists in this problem. By using multiobjective evolutionary algorithms, the candidates of the RBF network structure are obtained in the sense of Pareto optimality. We discuss an application to system identification by using such RBF networks having Pareto optimal structures. Some numerical simulations for nonlinear dynamic systems are carried out to show the applicability of the proposed approach.**

I. Introduction

Mathematical model of an actual system has important roles in the engineering problems such as control system design, fault detection and diagnosis, signal processing, time series prediction and so on. Though the system model is often constructed based on the physical or chemical laws of the target system, it is hard to build such model for stochastic, large scale, nonlinear or complex systems. System identification i.e. building mathematical model based on the observed input and output data of the system is a promising approach to model such system, so that system identification is a fundamental issue of engineering problem [1].

System identification techniques have been developed for the stochastic linear dynamic systems. However, almost actual existing systems have inherent nonlinear properties such as dead zone and saturation. Hence, the linear system models are not enough to represent such dynamics of nonlinear systems. Nonlinear system identification for the particular objective systems has also studied and a lot of identification algorithms have been developed in the recent two decades [2]. Most of these approaches are ad-hoc because it is not easy to describe wide class of nonlinear properties by the specific model structure.

Artificial neural networks are widely used as model structures to learn a nonlinear mapping based on the training data set in many kinds of application field due to their powerful nonlinear mapping ability. The use of artificial neural networks is also being studied to nonlinear system identification [3], [4]. The primary importance in applying neural networks to the nonlinear system identification is in selecting its structure rather than the connection weights learning algorithms and the network structure is characterized by the number of hidden layer, the number of hidden layer's unit, and the response function. However, a general method of the structure determination has not established, because the optimum structure depends on a class of the objective system and learning algorithm. So the network structure is generally determined by trial and error or a heuristic method.

In addition, there exists tradeoff between model accuracy and complexity in the identification problem. The system model optimized under the specific criterion is not always the optimal model because there are usually several demands to a system model. For example, it required that the model should be easy to handle and well explainable for the modeling data set contaminated by observation noise, but these properties are mutually exclusive. However, almost identification techniques give one best model under the criterion given by using prior knowledge or information. The system identification methodology based on the multi-objective optimization will be useful, however there are few studies from this point of view [5]- [7].

On the other hand, the multi-objective optimization problem is a common problem in real world. Pareto optimality is important concept in such problem that has ambivalent objectives. Multi-objective evolutionary algorithms are much being studied as efficient technique to providing the Pareto optimum solutions with a single run [8], [9]. Nonlinear system identification by using multi-objective evolutionary algorithms have been proposed, these approaches deal with polynomial dynamic system model and give the optimal model set concerning model accuracy and complexity [10], [11].

From this viewpoint, we consider in this study a multi-objective optimization based nonlinear system identification by using multi-obejectively selected RBF (Radial Basis Function) networks, which is a kind of artificial neural networks. RBF network has in their hidden layer a number of basis function which responds locally in input space. The network output is the linear sum of the basis function values. If the parameters of RBF networks, i.e. the number of basis functions and the widths and centers of each basis function, are determined, output layer weights can be calculated with the training data [12]. Therefore we consider the structure determination problem of RBF networks as a multi-objective optimization problem that concerns with the model accuracy, the model complexity and the output layers' weights. Then a method of obtaining the candidates of model as a Pareto optimal set based on evolutionary algorithms is proposed [13]. The designers will be able to select one model from the Pareto optimal model set obtained by the proposed method according to their use or a specific criterion. Alternately by introducing the concept of ensemble learning [14], [15] and neural network ensembles based on Pareto set [16], [17], [18], one RBF network can be obtained by constructing the obtained Pareto optimal RBF networks. Kondo et. al. applied Pareto optimal RBF networks to pattern classification problem [19]. Then, for the nonlinear regression, multi-objective identification of static nonlinear input output relation by using RBF network was proposed by Hatanaka et.al. [20].

From this viewpoint, we consider nonlinear system identification based on the multi-objective optimization using the evolutionary algorithms. In this paper, we deal with the dynamic nonlinear system modeling by using RBF network, and then we propose an identification technique using the Pareto optimum RBF network ensemble. Numerical simulation studies are carried out to show the applicability of the proposed technique to nonlinear system identification.

II. Multi-objective Genetic Algorithms and Pareto RBF Networks

A. Genetic Algorithm

GA (Genetic algorithm) is one of the stochastic search or optimization methods, originally proposed by Holland [21], which has been invented based on the natural genetics and evolution. The outline of simple GA procedure is a following way. Initially, the initial population of individuals having a string as the "chromosome" is generated randomly. Each element of the chromosome is called "gene". The "fitness", which is a measure of adaptation to environment, is evaluated for each individual. Then, "selection" operation leaving individuals to next generation is performed based on the fitness value, and "crossover" and "mutation" are operated on the selected individuals to generate a new population by transforming chromosomes into offspring's ones. This procedure is continued until the terminate condition is satisfied. This algorithm is conforming to the mechanism of evolution, in which the genetic information changes for every generation and the individuals which adapt to environment better survive preferentially. Since GA is a stochastic parallel search and GA requires only fitness value based on the objective functions, GAs are attracting attentions as a solver of multi-objective optimization problems due to parallel search.

B. Multi-objective Optimization

In the actual optimization problems, there generally exists tradeoff among the objective functions. And so two concept, "domination" and the "Pareto optimum," are considered in multi-objective optimization.
Let's consider the multi-objective optimization problem such as

$$\begin{array}{ll} \max & f_1(\boldsymbol{x}), \cdots, f_n(\boldsymbol{x}) \\ \text{subject to} & g_j(\boldsymbol{x}) \geq 0, j = 1, 2, \cdots, k \end{array}$$

where, $\boldsymbol{x}$ represents m dimensional decision variable $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)^T$ and $f_i(\boldsymbol{x}), \quad i = 1, 2, \cdots, n$ denote n objective functions. $g_i(\boldsymbol{x}), \quad i = 1, 2, \cdots, k$ are the constraint conditions.
First, $\boldsymbol{x}_1$ is said to "dominate" $\boldsymbol{x}_2$, if and only if

$$\forall i = 1, 2, \ldots, n \quad f_i(\boldsymbol{x}_1) \leq f_i(\boldsymbol{x}_2)$$

and

$$\exists j = 1, 2, \ldots, n \quad f_j(\boldsymbol{x}_1) < f_j(\boldsymbol{x}_2).$$

Then, $\boldsymbol{x}_0$ which is not dominated by any other $\boldsymbol{x}$ is called "Pareto optimal solution". Pareto optimal solution is considered to be the best solution comprehensively and many Pareto optimal solutions exist simultaneously, in general. Considering tradeoff among the objective functions, on multi-objective optimization problems it is important to obtain a Pareto optimal solution set.

C. Multi-objective GA based on $\boldsymbol{rank}$

A parameter $rank$ is introduced in order to apply the concepts of domination and Pareto optimum to GA. Though there are some ranking methods such as Fonseca's ranking method [8]. According to Fonseca's ranking, a $rank$ of an individual $\boldsymbol{x}_i$ on a generation t is:

$$rank(\boldsymbol{x}_i, t) = 1 + p_i^{(t)}$$

where p_i is the total number of individuals which dominate $\boldsymbol{x}_i$. By evaluating $rank$ for each individual and selecting based on

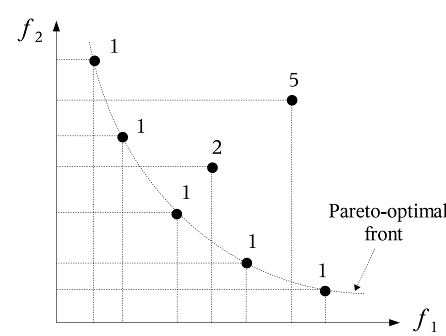


Fig. 1. Fonseca's ranking method

it, a population can evolve toward a Pareto optimal solution set. Since GA is a multi-point search algorithm, GA is expected to find a Pareto optimal set in a single simulation run.

D. RBF Network

RBF (Radial Basis Function) network is constructed of three layers as shown in Fig.2 and has basis functions which respond locally in the input space. Basis function $\phi_j(\boldsymbol{x})$ in this study is defined by Gaussian function,

$$\phi_j(\boldsymbol{x}) = \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{c}_j)^T(\boldsymbol{x}-\boldsymbol{c}_j)}{2\sigma_j^2}\right) \tag{1}$$

Here, $\boldsymbol{x}$ is input variable, $\boldsymbol{c}_j$ is center vector, and σ_j^2 is a parameter which decides function width. Using this $\phi_j(\boldsymbol{x})$, RBF network is constructed as:

$$u(\boldsymbol{x}) = w_0 + \sum_{j=1}^{m} w_j \phi_j(\boldsymbol{x}) \tag{2}$$

Here, m is the number of hidden units, i.e., the basis functions, and w_j are the output layer weights. RBF network will be determined if the parameters $m, \boldsymbol{c}_j, \sigma_j$, and w_j are estimated based on the data observed from the system. In this study, these parameters are estimated by two GAs. The parameters σ_j are assumed to be constant value for simplicity.

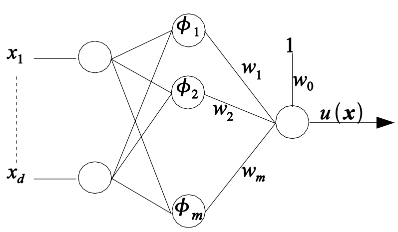


Fig. 2. RBF network

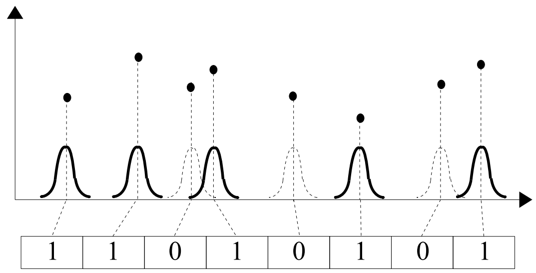


Fig. 3. Chromosome representation

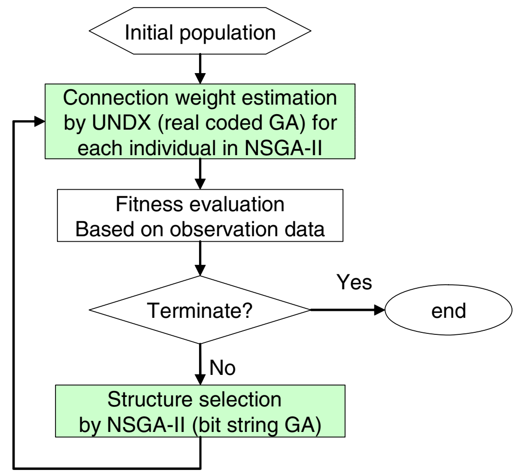


Fig. 4. Overall flowchart of the proposed approach

E. Genetic Representation

Multi-objective GA is applied to structure selection of RBF network. The problem of network structure selection is to determine the number of basis functions and their center positions. Here we assume that the center of basis function is the position of the training data points. The chromosomes of GA population indicate that each data point is employed as a center of basis functions or not. That is "1" represents that a basis function is located at the corresponding training data point, as shown in Fig.refchromo. By this setting, the length of the chromosome becomes equal to the number of training data, the number of "1" gene in the chromosome indicates the number of basis functions and the locus of the "1" shows the center position of the basis functions. Then, the connection weights are estimated by real-coded GA, in which each individual represents straightforward the weight vectors consisted by $w_j, (j = 0, 1, \ldots, m)$.
Thus, two stage GAs are used to give Pareto optimal RBF network set in this method, the first one is real coded GA estimating the connection weights for the candidate of RBF network, the second GA is multi-objective binary coded GA to examine the structure candidates of the RBF networks. The overall flow diagram is shown in Fig.4.

After estimating all the parameters of the network, $rank$ is assigned for each individual by the concept of multi-objective optimization problem, in which the objective functions are to be optimized. Then Pareto optimal individuals will be obtained.

III. Nonlinear System Identification by Pareto Ensemble Network

A. Problem Statement

Here we consider the following single output single input nonlinear ARX model (Auto Regressive model with eXogenous input),

$$y_t = g(\phi_t) + e_t \tag{3}$$

$$\phi_t = (y_{t-1}, \ldots, y_{t-n_y}, u_{t-1}, \ldots, u_{t-n_u})^T \tag{4}$$

where, $g(\cdot)$ represents an unknown nonlinear function, $y_t \in \Re$ and $u_t \in \Re$ are system output and input, respectively, and $e_t \in \Re$ indicates noise term with zero mean and infinite variance. ϕ_t is called as a regression vector composed of delayed inputs and outputs, n_y is unknown maximum delay of output and n_u is unknown maximum delay of input. The overall flow of system identification is shown in Fig.6 [1]. The proposed method performs two steps of "model structure determination" and

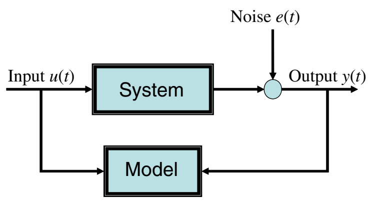


Fig. 5. System identification

"parameter estimation" simultaneously under the multi criteria. The criteria are selected by considering model accuracy and complexity.

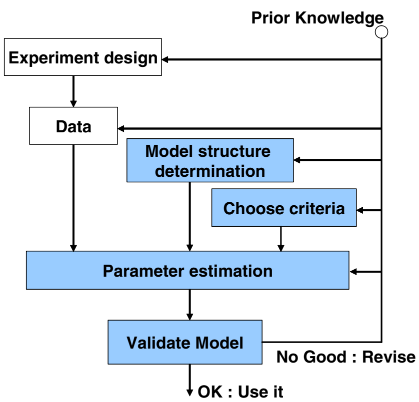


Fig. 6. Overall flow of system identification

B. Fitness Evaluation

It is generally demanded that the mathematical models not only can explain the relationship between input and output enough but also is simple in order to have the generalization ability. Then in this study three evaluation criteria are set for the evaluation of NSGA-II which determines the network architecture.

The first fitness is the number of basis function. This fitness indicates the complexity of the model. The second fitness is log MSE which indicates the extent of a fit of the model to the training data. MSE(Mean Squared Error) is defined as :

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \hat{y}_i\}^2 \tag{5}$$

Here, y_i is the observed output, $\hat{y}_i$ is the model output. The third fitness is the sum of the absolute value of weights and represents the smoothness of the network. These three evaluation criteria are to be minimized. Then, the uniform crossover and the bit reversal mutation are used as genetic operations in NSGA-II.

In real-coded GA, about genetic operation, UNDX (Unimodal Normal Distribution Crossover) [22] is applied in the proposed method. UNDX generates two offsprings by normal random numbers which is determined by three parents, as shown in Fig. 7. Basically offsprings are generated by normal distribution around segment connecting two parents. The third parent is used to determine the standard deviation of normal distribution. MSE is used for evaluation in UNDX to estimates the connection weights.

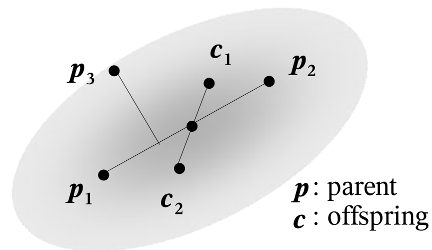


Fig. 7. UNDX

MGG[23] is adopted as the generation alternation model of real-coded GA used in the proposed method. MGG is said to have an ability to preserve the diversity of population. MGG procedure is as follows.

1) Plurality of real number vector is generated at random as the initial population.
2) Two parents are selected at random from population.
3) $2n_c$ offsprings are generated by applying UNDX to two parents n_c times. Here the third parent which determines the standard deviation of normal distribution is selected from population.
4) Fitness values of each offspring are calculated, then two individuals are selected from the set which is composed of two parents and all offsprings, then two parents are replaced by the selected two individual. The individuals selected here are elite and the individual selected by roulette selection in which the elite was pruned.
5) Continue 2–4 until the end condition is met.

C. Pareto RBF Network Ensemble

Various models based on the above mentioned three criterion can be obtained by the proposed method, so the designers will be able to select one model flexibly. On the other hand, there are the demand to obtain one model with good generalization ability. For instance, model selection by information criteria has been studied.

Recently the ensemble learning is receiving much attentions in the field of machine learning. In the ensemble learning, a monolithic model is constructed by combining several models. While some learning methods to make models constructing ensemble have been proposed, in this study an model ensemble is constructed by Pareto optimal models obtained by the proposed method.

Suppose that the number of Pareto models is L and the output of j-th network is $y_j(\boldsymbol{x})$, then the output of ensemble network $y^{EN}(\boldsymbol{x})$ is :

$$y^{EN}(\boldsymbol{x}) = \sum_{j=1}^{L}\alpha_j y_j(\boldsymbol{x}) \tag{6}$$

Here, $alpha_j$ is the weight on the output of j-th network. In the following numerical simulations, $alpha_j$ is assumed to be $1/L$ for each j, for simplicity.

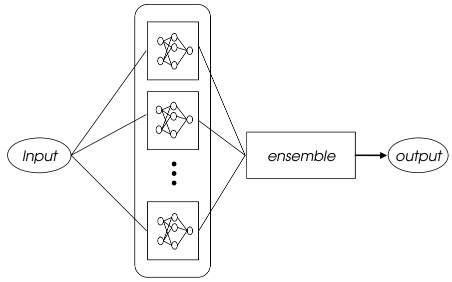


Fig. 8. RBF network ensemble

IV. Numerical Examples

To show the effect of the proposed approach we show the following two examples.

In the first simulation, we assume that the true system is described by

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}, u_{t-1}, u_{t-2}) + e_t. \tag{7}$$

Where,

$$f(x_1, x_2, x_3, x_4, x_5) = \frac{x_1 x_2 x_3 x_5 (x_3 - 1) + x_4}{1 + x_2^2 + x_3^2}, \tag{8}$$

and e_y indicates observation noise with mean 0 and variance σ_ε^2. Then, $u_t, (t = 1, 2, \cdots)$ are sampled by uniform distribution over $[-1, 1]$ we obtained 50 set of input and output observation data. Here, assume that the maximum delay n_y and n_u are known for simplicity.

The control parameters for genetic algorithms are listed in the TableI, we carried out simulation runs in the case of $\sigma_\varepsilon^2 = 0.1$ for three level widths of RBF function.

TABLE I
Configurations for two GAs

	multi-objective GA	UNDX
crossover rate	0.7	–
mutation rate	0.1	–
population size	50	30

The MSE values are shown in the Table. II and the trajectories of the observed output and the predicted output by using an ensemble model (6) are indicated in Fig.9, 10 and 11.

TABLE II
Numerical simulation results (1)

	Number of generations	
RBF width (σ^2)	5	50
1.0	0.201851	0.202434
2.0	0.198471	0.201319
5.0	0.210963	0.214082

Then, we assume that the true system is described by

$$\begin{aligned} y_t \quad = \quad & -0.5 y_{t-2} + 0.7 u_{t-1} y_{t-1} + 0.6 u_{t-2}^2 \\ & + 0.2 y_{t-1}^3 - 0.7 u_{t-2}^2 y_{t-2}. \end{aligned} \tag{9}$$

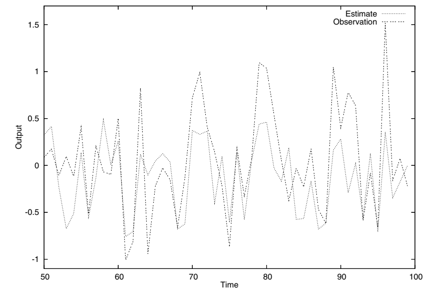


Fig. 9. Trajectories of observed output and one step prediction by the provided model ($\sigma^2 = 1.0$)

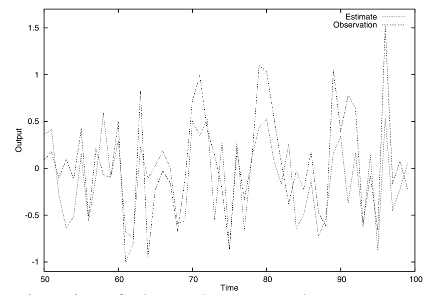


Fig. 10. Trajectories of observed output and one step prediction by the provided model $\sigma^2 = 2.0$

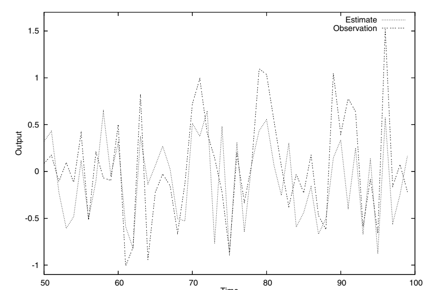


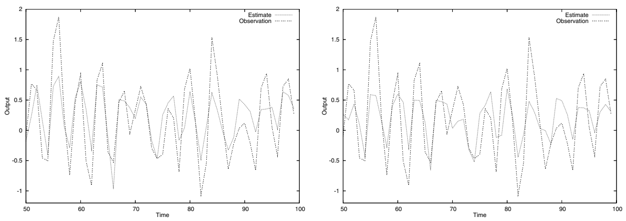
Fig. 11. Trajectories of observed output and one step prediction by the provided model $\sigma^2 = 5.0$

We carried out simulation runs with same condition as Simulation 1, but the maximum delay n_y and n_u are assumed 2 and 3, respectively. The MSE values are shown in the Table. III and the trajectories of the observed output and predicted output by using an ensemble model (6) are indicated in Fig.12, 13 and 14.

As shown above two results, the proposed method has applicability to nonlinear dynamic system identification. Though the user's preference for tradeoff or specific setting of the ensemble weights are not discussed, it will be selected by means of the purpose of identification.

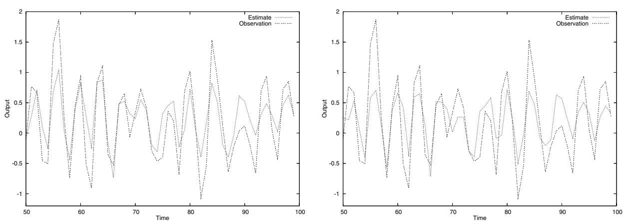
TABLE III
Numerical simulation results (2)

		Number of generations	
RBF width (σ^2)	delay	5	50
1.0	2	0.179057	0.189671
	3	0.251702	0.250289
2.0	2	0.169839	0.169682
	3	0.213453	0.212877
5.0	2	0.197117	0.192981
	3	0.223191	0.218099



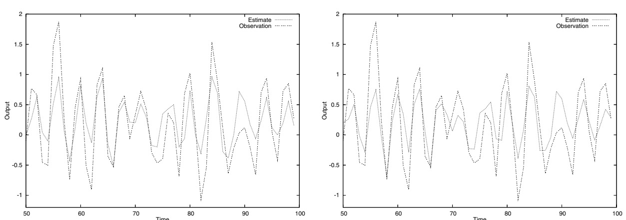
(a) $n_y = n_u = 2$ (b) $n_y = n_u = 3$

Fig. 12. trajectories of observed output and one step prediction by the provided model $\sigma^2 = 1.0$



(a) $n_y = n_u = 2$ (b) $n_y = n_u = 3$

Fig. 13. trajectories of observed output and one step prediction by the provided model $\sigma^2 = 2.0$



(a) $n_y = n_u = 2$ (b) $n_y = n_u = 3$

Fig. 14. trajectories of observed output and one step prediction by the provided model $\sigma^2 = 5.0$

V. CONCLUSIONS

In this study, we have shown a method of obtaining a Pareto optimal RBF network set based on multi-objective evolutionary algorithms. Then we have constructed an ensemble network by the Pareto optimal RBF networks applying them to nonlinear system identification, and a performance of the ensemble network as a nonlinear system model has been also considered. Numerical simulation results indicate that the ensemble network has an ability to identify nonlinear systems without specific prior knowledge or information about the objective systems.

The proposed method has applicability to nonlinear dynamic system identification. To find the user's preference for tradeoff and appropriate settings of the ensemble weights are the further issues. These may depend on the application field or the purpose of identification. Reduction of the computational costs, improvement of the ensemble technique and comparison to the conventional approaches are also the future works and they are under investigation.

REFERENCES

[1] L. Ljung : *System Identification : Theory for the User*, Prentice-Hall, NJ, 1999.

[2] O. Nelles : *Nonlinear System Identification*, Springer, Berlin, 2000.

[3] S. Chen and S. A. Billings : Neural networks for nonlinear dynamic system modelling and identification, *International Journal of Control*, Vol.56, No.2, pp. 319-346, 1992.

[4] J. Sjoberg, Q. Zhang et.al. : Nonlinear black-box modeling in system identification : a unified overview, *Automatica*, Vol.31, No.12, pp.1691-1724, 1995.

[5] T. A. Johansen : Multi-objective identification of FIR models, *Proceedings of IFAC Symposium on System Identification SYSID2000*, Vol.3, pp.917-922, 2000.

[6] T. A. Johansen and R. Babuška : Multi-Objective Identification of Takagi-Sugeno Fuzzy Models, *IEEE Transactions on Fuzzy Systems*, Vol.11, No.6, pp.847-860, 2003.

[7] K. Rodriguez-Vazquez, C. M. Fonseca, and P. J. Fleming : Identifying the Structure of NonLinear Dynamic Systems Using Multiobjective Genetic Programming, *IEEE Transactions on systems, man and cybernetics Part A*, Vol. 34, No. 4, 2004.

[8] C. M. Fonseca and P. J. Fleming, P. J : Genetic algorithms for multiobjective optimization : Formulation, discussion and generalization, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp.416-423, 1993.

[9] K. Deb : *Multi-objective Optimization using Evolutionary Algorithms*, John Wiley & Sons, New York, 2001.

[10] R.-V. Katya C. M. Fonseca and P. J. Flemming : Multiobjective Genetic Programming : A Non-Linear System Identification Application, *Genetic Programming 97 Conference*, pp.207-212, 1997.

[11] T. Hatanaka, K. Uosaki and Y. Hossaka : Application of Evolutionary Multi-Objective Optimization to Nonlinear System Identification, *Proceedings of the SICE Kansai Branch Symposium 2002*, pp.54-56, 2002. (in Japanese)

[12] T. Hatanaka, K. Uosaki and T. Kuroda : Structure Selection of RBF Neural Network Using Information Criteria, *Proceedings of Fifth International Conference on Knowledge-Based Intelligent Information Engineering Systems and Allied Technologies*, pp.166-170, 2001.

[13] T. Hatanaka, N. Kondo and K. Uosaki : "Multi-objective structure selection for radial basis function networks based on genetic algorithm," *Proceedings of 2003 Congress on Evolutionary Computation*, pp. 1095–1100, 2003.

[14] Y. Jin, T. Okabe and B. Sendhoff : Neural network regularization and ensembling using multi-objective evolutionary algorithms, *Proceedings of the 2004 IEEE Congress on Evolutionary Computation*, pp.1-8, 2004.

[15] M.M. Islam, X. Yao and K. Murase : A constructive algorithm for training cooperative neural network ensembles, *IEEE Transactions on Neural Networks*, Vol.14, No.4, pp.820-834, 2003.

[16] H. A. Abbass : Pareto Neuro-Ensemble, *Proceedings of the 16th Australian Joint Conference on Artificial Intelligence*, pp.554-566, 2003.

[17] H. A. Abbass : Pareto Neuro-Evolution : Constructing Ensemble of Neural Networks Using Multi-objective Optimization, *Proceedings of the 2003 IEEE Congress on Evolutionary Computation*, Vol.3, pp.2074-2080, 2004.

[18] J. E. Fieldsend and S. Singh : Pareto evolutionary neural networks, *IEEE Transactions on Neural Networks*, Vol.16, No.2, pp.338-354, 2005.

[19] N. Kondo, T. Hatanaka and K. Uosaki : Pattern Classification by Evolutionary RBF Networks Ensemble Based on Multi-objective Optimization, *Proceedings of 2006 IEEE World Congress on Computational Intelligence*, pp.5226–5232, 2006.

[20] T. Hatanaka, N. Kondo and K. Uosaki : "Multi Objective Structure Selection for RBF Networks and Its Application to Nonlinear System Identification," Yaochu Jin eds, *Multi–Objective Machine Learning*, Springer, pp.491–506, 2006.

[21] J. Holland : *Adaption in Natural and Artificial Systems*, The University of Michigan Press, MI, 1975.

[22] I. Ono and S. Kobayashi : A Real-coded Genetic Algorithm for Function Optimization Using Unimodal Normal Distribution Crossover, *Proceedings of 7th International Conference.on Genetic Algorithms*, pp.246-253, 1997.

[23] H. Sato, I. Ono and S. Kobayashi : A New Generation Alternation Model of Genetic Algorithms and Its Assessment, *Journal of Japanese Society for Artificial Intelligence*, Vol.12, No.5, pp.734-744, 1997.