Multi–Objective Structure Selection for Radial Basis Function Networks Based on Genetic Algorithm

Toshiharu Hatanaka, Nobuhiko Kondo and Katsuji Uosaki
Department of Information and Physical Science
Graduate School of Information Science and Technology, Osaka University
2–1 YamadaOka, Suita, 565–0871, Japan
hatanaka@ist.osaka-u.ac.jp

Abstract- Radial basis function (RBF) network is well known as a good performance approach to nonlinear system modeling. Though structure selection of RBF network is an important issue, the framework of this problem has not been established. In this paper, we propose multi–objective structure selection method for RBF networks based on MOGA (multi–objective genetic algorithm). The structure of RBF networks is encoded to the chromosomes in GA, then evolved toward to Pareto–optimum for multi–objective functions concerned with model accuracy and complexity. Some numerical simulation results indicate the applicability of the proposed approach.

1 Introduction

The black box modeling for unknown systems based on the observed input and output data set is called system identification [ljung99]. System identification has an important role in a lot of engineering problems including control system design, model based fault detection and diagnosis, signal processing, time series prediction and so on. Hence many identification algorithms have been developed in order to give a good model of actual systems. A great deal of research in system identification concerns an algorithm that gives the best model under a priori provided criterion. However, since there are occasionally several demands to a system model, a model optimized under the specific criterion may not necessarily be the optimal model. For example, the models should be simple and well explainable for the observation data, but these properties are mutually exclusive [hatanaka02] [ukai01].

On the other hand, several approaches to nonlinear system modeling using artificial neural networks have been proposed in the last two decades [oliver01] [yang98]. But a general method to determine the structure of neural network has not been established. The problem of static nonlinear system modeling using RBF (Radial Basis Function) network, which is a kind of neural network, was considered and structure selection of the RBF network for nonlinear system modeling has been studied from the point of view of selection of suitable information criteria [hatanaka01]. With RBF network, if number of the basis functions is determined, the parameters of RBF networks, i.e. widths and the center of each basis functions and each weights of network, can be calculated with the training data. Then the estimate of system outputs can be calculated as a linear sum of basis functions. This parameter settings make great influence on the accuracy of function approximation. Though a modeling method which has capability to select a model according to the specific purpose is desirable, in existing methods, it is not taken into consideration that they provide the flexibility of such model selection.

From this point of view, we consider structure selection problem of RBF network as a multi–objective optimization problem of a complexity of structure and a description capability and propose a method of obtaining the candidates of model as a Pareto–optimal set based on MOGA (multi–objective genetic algorithm). Then we demonstrate its applicability by numerical simulation examples.

In the section 2, outlines of the mechanism of GA and MOGA, and genetic operations using in this study are described. Then, proposed basis function selection algorithm for RBF network is introduced in the section 3. Some numerical study results including 2 dimensional case are shown in the section 4 and concluding remarks are given at the last section.

2 Multi–Objective GA

Genetic algorithm (GA) is search or optimization algorithm which is invented based on genetics and evolution. Initially, the initial population of individuals which have a binary digit string as the “chromosome” is generated at random. Each bit of chromosome is called “gene” [back96]. The “fitness”, which is a measure of adaptation to environment, is

calculated for each individual. Then, "selection" operation leaving individuals to next generation is performed based on fitness value, and then "crossover" and "mutation" are performed on the selected individuals to generate new population by transforming parent's chromosomes into offspring's ones. This procedure is continued until the end 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.

2.1 Multi–Objective GA

In the multi–objective optimization problems, two concept, "domination" and the "Pareto–optimum", are considered. First, $\boldsymbol{x}_1$ is said to "dominate" $\boldsymbol{x}_2$ if

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

and

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

And $\boldsymbol{x}_0$ which is not dominated by any other $\boldsymbol{x}$ is called the "Pareto–optimal solution" [deb01] [fonseca93]. Pareto–optimal solution is considered to be the best solution comprehensively. And many Pareto–optimal solutions exist generally. Considering trade–off among the objective functions, on multi–objective optimization problems it is important to obtain a Pareto–optimal solution set. A parameter *rank* is introduced in order to apply the concepts of domination and Pareto–optimum to GA. Though there are some ranking methods, this study adopts Fonseca's ranking method [fonseca93]. According to Fonseca's ranking method, 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 calculating this *rank* for each individual and selecting based on it, a population can evolve toward a Pareto–optimal solution set.

2.2 Genetic operations

We explain genetic operations used in this study.

- selection
  The tournament selection is used for the selection operation. The tournament selection is an operation that extracts some individuals from parent population at random and leaves an individual of the highest rank to the next generation. This is continued until the population of next generation is filled.
- crossover
  The uniform crossover is used as the crossover operation. Let the crossover probability p_c. First, a random binary string(mask) which has 1 at the probability of p_c is created. Then for each locus,if the mask is 1,the gene of parent 1 becomes the gene of offspring 1, and the gene of parent 2 becomes the gene of offspring 2. It becomes reverse if the mask is 0.
- mutation
  The bit reversal mutation is used for the mutation operation. It reverses the bit of each locus at the mutation probability p_m.

parent 1		offspring 1
0011011010	mask	0010000011
	1100100110	
parent 2	$\longrightarrow$	offspring 2
1010100111		1011111110

Figure 1: An example of the uniform crossover

3 Structure Selection of RBF Network

3.1 RBF Network

RBF (Radial Basis Function) network has basis functions as typified by Gauss function. Basis function $\phi_j(\boldsymbol{x})$ in this study is defined by Gaussian,

$$\phi_j(\boldsymbol{x}) = \exp\left(-\frac{\sum_{i=1}^{d}(x_i - c_{ij})^2}{2\sigma_j^2}\right). \tag{1}$$

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 is the weight on each basis function. If m, $\boldsymbol{c}_j$ and σ_j^2 are set, then all of w_j are set by least-squares method.

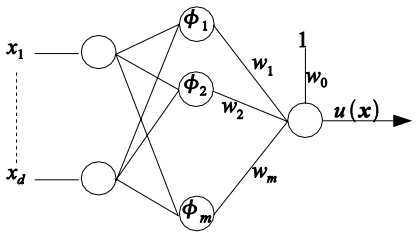

Figure 2: RBF network:The input data are distributed to hidden layer. In the hidden layers the basis functions are calculated. Then the approximation function which is linear sum of basis functions is obtained as output.

3.2 Selection of basis functions by GA

In this study, GA is used for setting of the number of basis functions and their centers . Each basis function's center is put on input variable space at regular intervals. And they are assigned gene of GA respectively. Then if a gene is 1 the basis function which corresponds to it is used, or if a gene is 0 the basis function which corresponds to it is not used. By this, the number of basis functions and its centers are determined for each individual in the GA population. On all basis function the width parameters are fixed to the same value , for simplicity. Then the weights on each basis function are calculated by least-squares method. Thereafter, approximating function is obtained. After obtained approximating function, following MSE(Mean Squared Error) is calculated for each individual:

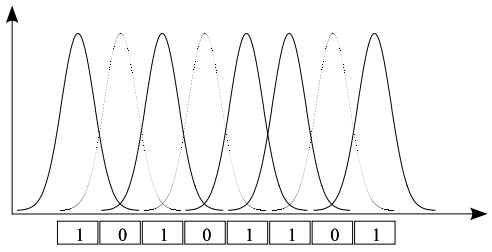

Figure 3: Chromosome:Each basis function's center is put on input variable space at regular intervals. And they are assigned gene of GA respectively. Then if a gene is 1 the basis function which corresponds to it is used, or if a gene is 0 the basis function which corresponds to it is not used.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\{y_i - u(\boldsymbol{x}_i)\}^2 \tag{3}$$

MSE is expressing the extent of a fit of an approximation function to the training data. By the concept of multi–objective optimization problem , in which two fitnesses are to be minimized, *rank* is assigned for each individual. The first fitness is MSE , and the second fitness is the number of basis function. Then Pareto–optimal population will be obtained by calculating based on *rank* in accordance with MOGA algorithm.

4 Numerical Simulations

In the simulation, $\boldsymbol{x}_i$ is the input variable and y_i is the output variable, and training data set is obtained by

$$y_i = v(\boldsymbol{x}_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0,\sigma^2), \quad i = 1,2,\ldots,n \tag{4}$$

where $v(\boldsymbol{x}_i)$ is the true function of the system. $\boldsymbol{x}_i$ is produced by uniform random number of a certain interval. The error term ε_i is produced under $\sigma^2 = 0.04$. The parameters of GA are shown in Table.1

Table 1: parameter settings for GA in the simulation 1

population size	50
tournament size	2
crossover rate	0.7
mutation rate	0.01

4.1 Simulation 1

Let the true function be:

$$v(x_i) = \sin(5\pi x_i), \tag{5}$$

where $x_i, i = 1,2,\cdots$, are sampled from uniform distribution over $[0,1]$. At first, the number of training data is set to 100, the width parameter σ_j^2 is set to 0.01, and the size of chromosome is set to 50. The simulations calculating until 300 generation are performed 20 times. The Pareto–optimal individuals which are obtained on each simulation are shown in Table 2. For illustration, Pareto–optimal individuals obtained in the 7th simulation run are showed in Table 3. Examples of obtained approximation functions by RBF networks are shown in Figure 5,6 and 7, respectively. For each individual in the Pareto–front given by MOGA,

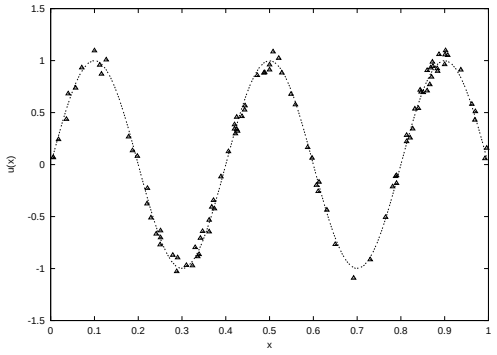

Figure 4: True function of simulation 1

Table 2: Number of individual in the Pareto–front

simulation run	number of Pareto	simulation run	number Pareto
1	9	11	11
2	13	12	10
3	14	13	13
4	11	14	10
5	13	15	13
6	12	16	12
7	14	17	11
8	10	18	14
9	13	19	10
10	10	20	11

MSE value for test data is evaluated by 20 times simulation runs and its results are shown in Figure 8. Next, the number of generation is changed to 1000 and the similar simulations are performed 10 times. Consequently, 13.5 Pareto–optimal individuals are obtained on the average.

4.2 Simulation 2

In this case we assume that the true function is following two dimensional function:

$$v(\boldsymbol{x}) = \sin(\pi x_1) + \cos(2\pi x_2) \tag{6}$$

where x is produced by uniform random number of $[0,2]$. The length of chromosome is set to 100. About the basis functions, a plain of $\boldsymbol{x}$ is divided into 100, then the centers of each basis function are distributed into their centers. If each gene is assigned a number as 1,2, ... from the left, each gene is assigned the basis function ,as shown in Figure 1. At first, the number of training data is set to 400, the width

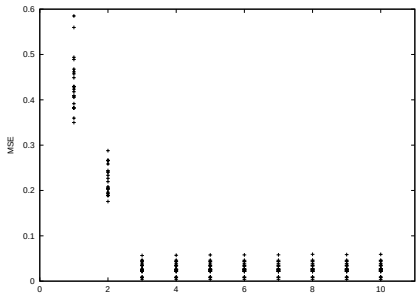

Figure 5: Model validation for selected individuals: Horizontal axis indicates index of individuals and *MSE*s are plotted for 20 runs of validation

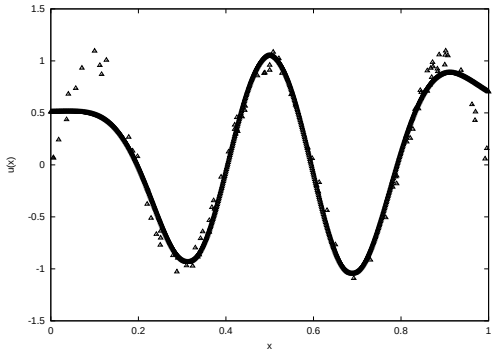

Figure 6: Individual 3 : The structure of this approximation function is very simple, but it doesn't fit to observed data much well.

parameter σ_j^2 is set to 0.04. The simulations calculating until 300 generation are performed. For illustration, Pareto–optimal individuals which are obtained in this implements are showed below:

4.3 Discussion

If the diversity of Pareto individuals is high, much flexibility on selection of the approximating function is made. We discuss the results from that point of view. When the number of generation is great, the number of Pareto individuals is also great. Therefore, it is expectable that the more various Pareto individual are obtained, so that the number of generation increases. However, since the computational burden of GA is large, it is required that various individuals in the Preto front are obtained at smaller generation.

Table 3: An example of chromsomes in the Pareto–front

individual	chromosome	number of basis	*MSE*
1	0100000000000000000010000000000000000000000010000	3	0.281465
2	0000000000000000000000000010000000000000010100100000	4	0.201757
3	0000000000000001100001000000001100000000001000000	6	0.038217
4	0000001100010000000000000011000000000000001100000000	7	0.005819
5	0000000110000000000100000010100000000000001100010100	9	0.004182
6	0000000110000000000100000010100000001000110001010 0	10	0.004124
7	0000000101000001000001000111000110000000101000000	11	0.003785
8	0000000110000100001100100000100000000010110001010 0	12	0.003715
9	0000000101000101000110000001000100000011110000010 0	13	0.003704
10	000000111100010000011000000100010000001111000000001	14	0.003697
11	0000000100000001101000001011000010001011010110010 0	15	0.003536
12	0000011101000001001111100011000100000010010010010 0	17	0.003473
13	0000001111000001001101000111001110000011111001001 00	20	0.003417
14	0001001110000100001110000011100110000011110010110 1	21	0.003350

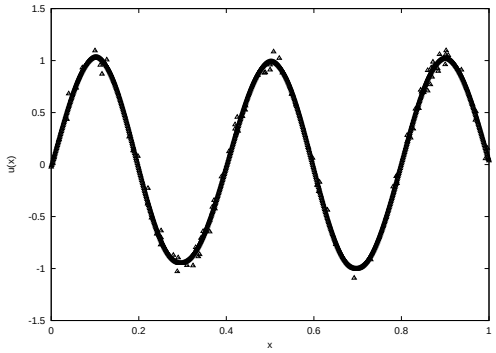

Figure 7: Individual 6 : This approximation function fits to observed data well, but its structure is complex.

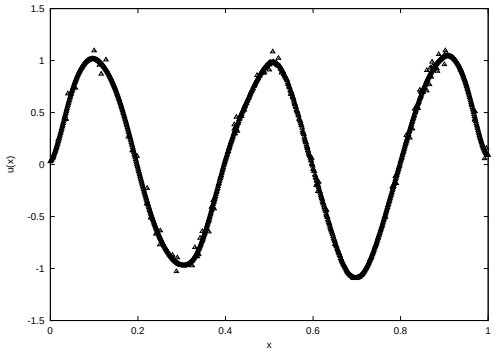

Figure 8: Individual 14 : This approximation function fits to observed data very well, but its structure becomes more complex than individual 6.

5 Conclusions

We have proposed a novel approach to the structure selection problem of RBF network from a viewpoint of multi–objective optimization. By applying the proposed algorithm, wide variety of RBF network structures, in the sense of the pareto optimum solution, are obtained and the availability of the proposed approach is shown by numerical simulation results. Though this method has to be improved in some respects such as uniformly property of the Pareto set, convergence speed of GA, and applications to the actual system modeling, and these are now under investigation.

Bibliography

[ljung99] L. Ljung : *System Identification: Theory for the User*, (2nd Edition), Prentice–Hall, 1999.

[hatanaka02] T. Hatanaka, K.Uosaki and Y.Hossaka: Application of Evolutionary Multi–Objective Optimization Nonlinear System Identification, *Proceedings of the SICE Kansai Branch Symposium 2002*, pp.54–56, 2002. (in Japanese)

[ukai01] N. Ukai, T. Hatanaka, K.Uosaki and H. Kurakake : Process Modeling Using Multi–Objective Genetic

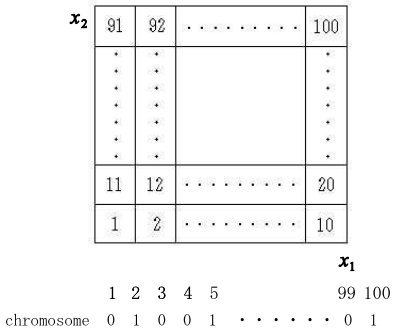

Figure 9: Correspondence between the center of basis functions and the locus in the chromosome

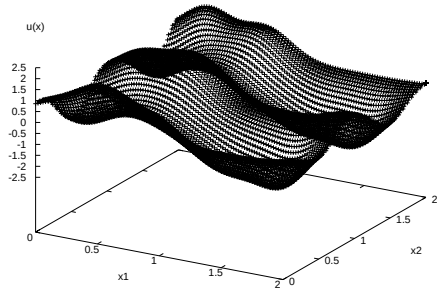

Figure 11: An example of obtained RBF network that has 6 basis functions and its *MSE* is 0.371367. The surface curve is given by the estimated RBF network by one of the individual in the Pareto front.

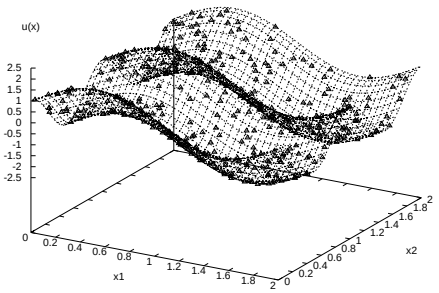

Figure 10: True function of simulation 2. The surface curve is given by the true function and the dotted points indicate the observed data points.

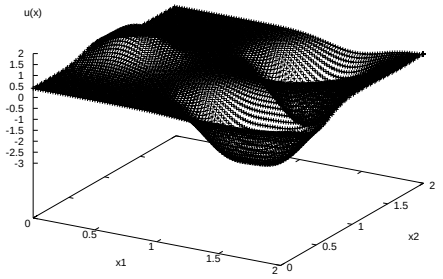

Figure 12: An example of obtained RBF network that has 27 basis functions and its *MSE* is 0.018647. The surface curve is given by the estimated RBF network.

Programming, *SICE 1st Annual Conference on Control Systems*, pp.685–688, 2001. (in Japanese)

[oliver01] Oliver Nelles : *Nonlinear System Identification*, Springer, 2001.

[yang98] Zi–Jiang Yang: Identification of Nonlinear Systems, *Journal of The Society of Instrument and Control Engineers*, Vol.37, No.4, pp.249–255, 1998. (in Japanese)

[hatanaka01] 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*, 2001.

[back96] T. Bäck : *Evolutionary Computation*, Oxford University Press, 1996.

[deb01] Kalyanmoy Deb : *Multi–objective Optimization using Evolutionary Algorithms*ĄCJohn Wiley & Sons, 2001.

[fonseca93] C.M. Fonseca and P.J. Fleming : Genetic algorithms for multi–objective optimization : Formulation,discussion and generalization, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pp.416–423, 1993.