# Constructive Methods of Fuzzy Rules for Function Approximation

Michiharu Maeda[1] and Hiromi Miyajima[2]

[1]Department of Control and Information Systems Engineering
Kurume National College of Technology, Kurume, Japan
maedami@kurume-nct.ac.jp
[2]Department of Electrical and Electronics Engineering, Faculty of Engineering
Kagoshima University, Kagoshima, Japan
miya@eee.kagoshima-u.ac.jp

**Abstract**: This paper describes novel methods to construct fuzzy inference rules with gradient descent. The present methods have a constructive mechanism of the rule unit that is applicable in two parameters: the central value and the width of the membership function in the antecedent part. The first approach is to create the rule unit at the nearest position from the input space, for the central value of the membership function in the antecedent part. The second is to create the rule unit which has the minimum width, for the width of the membership function in the antecedent part. Experimental results are presented in order to show that the proposed methods are effective in difference on the inference error and the number of learning iterations.

## 1. Introduction

A number of approaches have been studied on fuzzy modeling, with the object of automatically constructing fuzzy inference rules utilizing the learning process [1], [2]. They are aimed at drastically reducing the processing labor by applying the learning function to the tuning of fuzzy inference rules [3]–[9]. With respect to the constitution of fuzzy inference rules, the reduction methods [10] have been introduced in order to minimize the inference error and to shorten the learning process. In the techniques, the rule which seems to be little influence on the inference error is removed, and the methods proceed with learning by gradient descent. Thus, the extraction of appropriate rules become possible by reducing the number of redundant rules. However, when the reduction techniques were adopted, the different results were shown according to the application of various functions.

In this paper, we present novel methods of rule construction in fuzzy modeling by using gradient descent. the present approaches are described with constructive methods applicable to two parameters: the central value and the width of the membership function in the antecedent part. The first approach is to create the rule unit at the nearest position from the input space, for the central value of the membership function in the antecedent part. The second is to create the rule unit which has the minimum width, for the width of the membership function in the antecedent part. Experimental results are presented in order to show that the validity of the present methods is confirmed.

## 2. Fuzzy Inference Rules and Reasoning Methods

The simplified fuzzy reasoning which treats the consequent part as the reasoning method at the real number is used. When the input is $(x_1, x_2, \cdots, x_n)$ and the output is $y$, the procedure is expressed as follows:

$R^i$ : If $x_1$ is $M_{i1}$ and $\cdots$ and $x_n$ is $M_{in}$ then $y$ is $w_i$,

where $M_{ij}$ is a membership function in the antecedent part, and $w_i$ is a real number in the consequent part. The membership function $M_{ij}$ in the antecedent part is set separately for every rule, and has the index $i$ for the rule number. The membership function $M_{ij}$ in the antecedent part is an isosceles triangle, and can be expressed using the central value $a_{ij}$ and the width $b_{ij}$ in the following equation.

$$M_{ij}(x_j) = \begin{cases} 1 - \frac{2|x_j - a_{ij}|}{b_{ij}} & (|x_j - a_{ij}| \le b_{ij}/2) \\ 0 & (|x_j - a_{ij}| > b_{ij}/2). \end{cases} \quad (1)$$

The membership value $\mu_i$ of the $i$-th rule is obtained by the following equation.

$$\mu_i = \prod_{j=1}^{n} M_{ij}(x_j). \quad (2)$$

Therefore, the reasoning result $y$ consists of

$$y = \frac{\sum_{i=1}^{\gamma} \mu_i w_i}{\sum_{i=1}^{\gamma} \mu_i}. \quad (3)$$

The function which shows the shape of the membership function is adjusted by the delta rule with the central value $a_{ij}$, the width $b_{ij}$, and the actual value $w_i$ in the consequent part. It is possible to consider the delta rule as a minimization problem of the objective function $E$ which shows the error between the output value (i.e., $y$ of fuzzy reasoning) of fuzzy systems and the desired output value $y_r$, as shown in the following equation.

$$E = \frac{1}{2}(y - y_r)^2. \quad (4)$$

In order to decrease the value of the objective function $E$, when input-output data $(x_1, \cdots, x_n, y_r)$ of the $n+1$ dimension are given, the gradients ($\partial E/\partial a_{ij}$, $\partial E/\partial b_{ij}$, $\partial E/\partial w_i$) of objective function $E$ are calculated on $a_{ij}$, $b_{ij}$, and $w_i$. Subsequently, the values of $a_{ij}$, $b_{ij}$, and $w_i$ are updated according to

$$\Delta a_{ij} = -\eta_a \frac{\partial E}{\partial a_{ij}} \tag{5}$$

$$\Delta b_{ij} = -\eta_b \frac{\partial E}{\partial b_{ij}} \tag{6}$$

$$\Delta w_i = -\eta_w \frac{\partial E}{\partial w_i}, \tag{7}$$

where $\eta_a$, $\eta_b$, and $\eta_w$ are learning constants. $\partial E/\partial a_{ij}$, $\partial E/\partial b_{ij}$, and $\partial E/\partial w_i$ are calculated as

$$\frac{\partial E}{\partial a_{ij}} = \frac{\mu_i}{\sum_{i=1}^{\gamma} \mu_i}(y - y_r)(w_i - y) \cdot \mathrm{sgn}(x_j - a_{ij}) \frac{2}{b_{ij} M_{ij}(x_j)} \tag{8}$$

$$\frac{\partial E}{\partial b_{ij}} = \frac{\mu_i}{\sum_{i=1}^{\gamma} \mu_i}(y - y_r)(w_i - y) \cdot \frac{1 - M_{ij}(x_j)}{b_{ij} M_{ij}(x_j)} \tag{9}$$

$$\frac{\partial E}{\partial w_i} = \frac{\mu_i}{\sum_{i=1}^{\gamma} \mu_i}(y - y_r), \tag{10}$$

where

$$\mathrm{sgn}(\theta) = \begin{cases} -1 & (\theta < 0) \\ 0 & (\theta = 0) \\ 1 & (\theta > 0). \end{cases}$$

By giving input-output data one after another and repeating the learning process, the shape of the membership function in which the value of objective function $E$ becomes minimal is determined. The adjustment of the shape of the membership function is carried out until the inference error $D(t)$ shown in the following equation is less than the desired value $\delta$, for the given input-output data $(x_1^p, x_2^p, \cdots, x_n^p, y_r^p)$, $p = 1, 2, \cdots, P$.

$$D(t) = \frac{1}{P} \sum_{p=1}^{P} (y^p - y_r^p)^2, \tag{11}$$

where $y^p$ is an output of fuzzy reasoning.

## 3. Self-Tuning and Constructive Methods

In this section, the self-tuning and the constructive methods are described. The membership value in the antecedent part consists of two parameters: The central value and the width. With respect to these, the new rule constructions of fuzzy systems seem to be regard as the constructive approaches presented as follows.

*I* For the central value of the membership function in the antecedent part, the rule unit nearest from the central input space is created for all central values.

*II* For the width of the membership function in the antecedent part, the rule unit which has the minimum width is created for all weights.

By altering the constructive standards like these, the fuzzy reasoning systems which differ in the inference error and the number of learning iterations are constructed. Essentially, these parameters are important components which renew the value by gradient descent, when the fuzzy system is constituted. The new techniques will be effective if it is proven that appropriate results can be obtained using those techniques. The constructive method is presented as follows. To begin with, a few rules are given and learning is carried out for the input-output data prepared in advance. Then, rules are created sequentially to reach a prespecified number, and self-tuning is performed until the termination condition is satisfied. The constructive algorithm is presented as follows.

[*Constructive algorithm*]

*Step A1* Initialization:
Give central value $a_{ij}$ and width $b_{ij}$ in the antecedent part, real number $w_i$ in the consequent part, creation threshold $\delta_c$, termination threshold $\delta_T$, maximum number of learning iterations $T_{max}$, initial rule number $R_I$, and final number of rules $R_T$. Set $t \leftarrow 0$ and $\gamma \leftarrow R_I$.

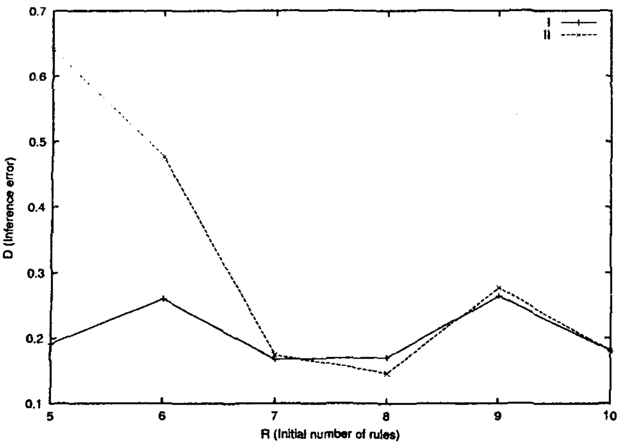
*Step A2* Self-tuning:
**(A2.1)** Let $p = 1$.
**(A2.2)** Let $s_p$ be an index selected at random among $\{1, 2, \cdots, P\}$, for all $s_i \neq s_j$.
**(A2.3)** Allot the input-output data $(x_1^{s_p}, x_2^{s_p}, \cdots, x_n^{s_p}, y_r^{s_p})$.
**(A2.4)** Derive the output of fuzzy inference $y^p$ performed by the simplified fuzzy reasoning.
**(A2.5)** Adapt $w_i$ according to Eq. (7) and repeat the fuzzy reasoning at A2.4.
**(A2.6)** Adapt $a_{ij}$ and $b_{ij}$ of the membership functions in the antecedent part, according to Eqs. (5) and (6), respectively.
**(A2.7)** If $p < P$, then set $p \leftarrow p+1$ and go to A2.2, otherwise set $t \leftarrow t+1$ and go to Step A3.

*Step A3* Rule construction:
**(A3.1)** Calculate $D(t)$ and $\Delta D(t)$ according to Eqs. (11) and (12), respectively.
**(A3.2)** If $\gamma < R_T$ and $\Delta D(t) \leq \delta_c$, then go to A3.4.
**(A3.3)** If $\gamma = R_T$, then go to Step A4, otherwise go to Step A2.
**(A3.4)** Create the $k$-th rule according to the constructive method. Set $\Re \leftarrow \Re + \{k\}$ and $\gamma \leftarrow \gamma + 1$. Go to Step A2.

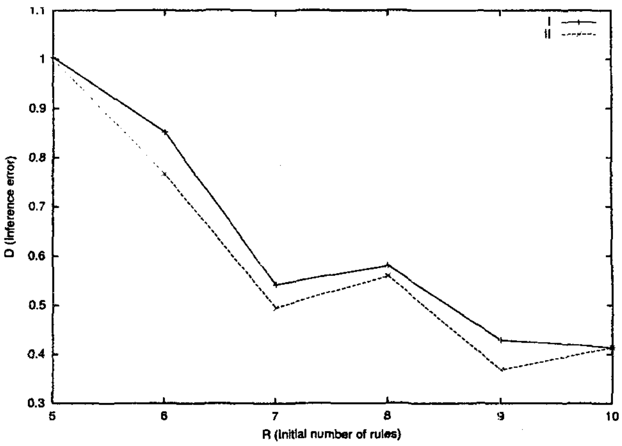
*Step A4 Termination condition:*
If $t = T_{max}$ or $\Delta D(t) \leq \delta_T$, then terminate, otherwise go to Step A2.

At steps A3.1 and A4, the value $\Delta D(t)$ is calculated as follows.

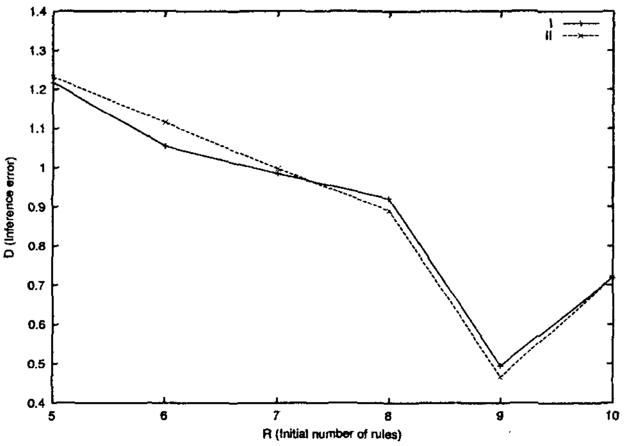
$$\Delta D(t) = \frac{|D(t) - D(t-1)|}{D(t-1)} \tag{12}$$



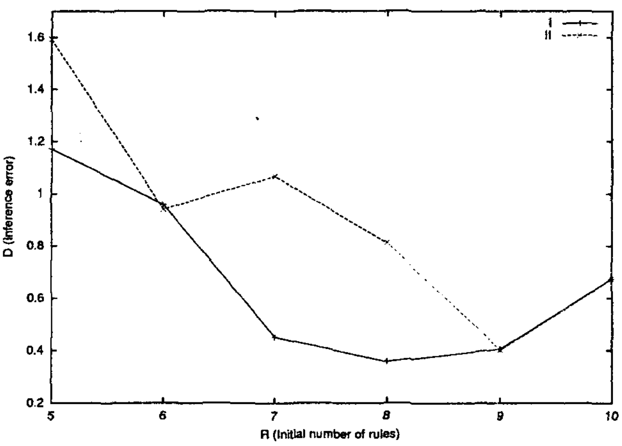

(a) System i




(b) System ii




(c) System iii




(d) System iv

Figure 1. Relation between inference error ($\times 10^{-4}$) and initial number of rules for each constructive model. The results are averages of 1000 trials. When $R = 10$, all the constructive models are equivalent to the conventional model because there are no existing rules to create.

## 4. Numerical Experiments

We perform the function approximation by using the above models. The systems are identified by the data as fuzzy inference rules, with the utilization of input-output data from the known function as follows:

$$\text{(i)} \quad y = \frac{\cos \pi x + 1}{2}$$

$$\text{(ii)} \quad y = \begin{cases} -x & (x \leq 0) \\ \sin \pi x & (x > 0) \end{cases}$$

$$\text{(iii)} \quad y = |\sin \pi x|$$

$$\text{(iv)} \quad y = |\cos \pi x|$$

The domain of each variable $x$ and output $y_r$ normalize within $[-1, 1]$ and $[0, 1]$, respectively. The parameters are chosen as follows: $\eta_a = 0.1$, $\eta_b = 0.1$, $\eta_w = 0.2$, $P = 100$, $\delta_c = 10^{-2}$, $\delta_T = 10^{-4}$, $T_{max} = 100000$, and $R_T = 10$.

Figure 1 shows the influence of the inference error on the initial number of rules for each model. For system i, models I and II are better in the inference error when the initial number of rules is 7 or 8. For system ii, model II is superior when the initial number of rules is 9. For system iii and iv, both of models I and II are best when the initial number of rules is large (e.g. $R = 9$). The effectiveness differs among the present techniques according to the inference error. Figure 2 shows the relation between number of learning iterations and initial number of rules for each constructive model. The effectiveness also differs among the present techniques according to the number of learning iterations

## 5. Conclusions

In this paper, we have presented novel constructive methods of fuzzy inference rules with gradient descent and have examined their validity through numerical experiments. The approaches were the construction mechanisms of rule unit that was applicable in two parameters: the central value and the width of the membership function in the antecedent part. The result of numerical experiments was that the present models led to different effects according to the application of various functions.

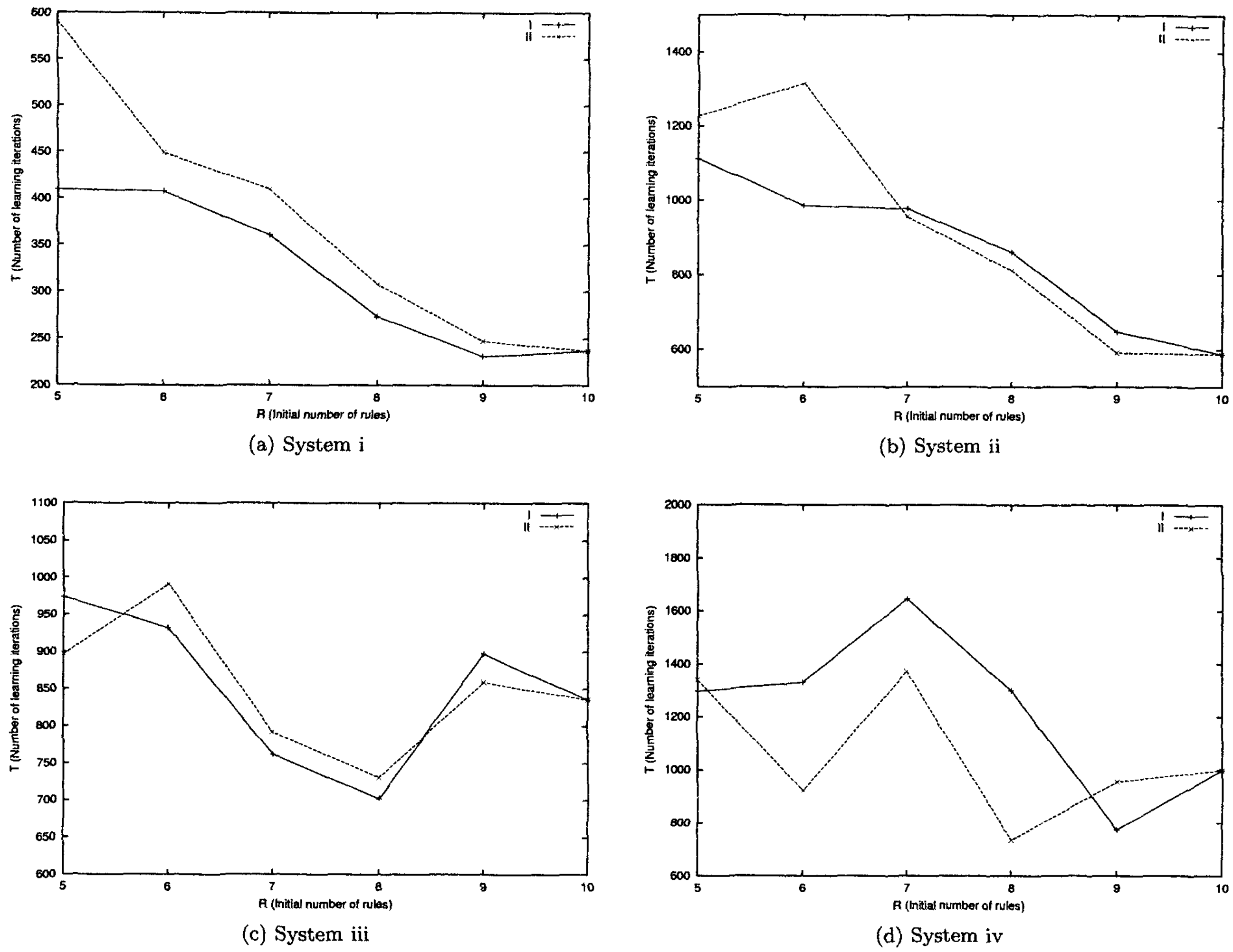



Figure 2. Relation between number of learning iterations and initial number of rules for each constructive model. The results are averages of 1000 trials. When $R = 10$, all the constructive models are equivalent to the conventional model because there are no existing rules to create.

For the future works, we will study more effective approaches.

## References

[1] R.R. Yager and L.A. Zadeh, Fuzzy sets, neural networks, and soft computing, Van Nostrand Reinhold, 1994.

[2] N.K. Kasabov, Foundations of neural networks, fuzzy systems, and knowledge engineering, The MIT Press, 1996.

[3] H. Ichihashi and T. Watanabe, "Learning control by fuzzy models using a simplified fuzzy reasoning," Journal of Japan Society for Fuzzy Theory and Systems, vol.2, pp.429–437, 1990.

[4] H. Takagi, T. Kouda, and Y. Kojima, "Neural-networks designed on approximate reasoning architecture," Journal of Japan Society for Fuzzy Theory and Systems, vol.3, pp.133–141, 1991.

[5] N. Nomura, I. Hayashi, and N. Wakami, "A self-turning method of fuzzy reasoning by delta rule and its application to a moving obstacle avoidance," Journal of Japan Society for Fuzzy Theory and Systems, vol.4, pp.379–388, 1992.

[6] H. Miyajima, S. Fukumoto, and Y. Ishizuka, "Fuzzy modeling based on deleting rules," IEICE Trans., vol.J77-A, pp.1555–1562, 1994.

[7] H. Miyajima, K. Kishida, and S. Fukumoto, "Constructive, destructive and simplified learning methods of fuzzy inference," IEICE Trans. Fundamentals, vol.E78-A, pp.1331–1338, 1995.

[8] K. Kishida, H. Miyajima, and M. Maeda, "Destructive fuzzy modeling using neural gas network," IEICE Trans. Fundamentals, vol.E80-A, pp.1578–1584, 1997.

[9] S. Fukumoto, H. Miyajima, K. Kishida, and Y. Nagasawa, "An investigation of fuzzy model using AIC," IEICE Trans. Fundamentals, vol.E80-A, pp.1698–1703, 1997.

[10] M. Maeda and H. Miyajima, "Fuzzy modeling in some reduction methods of inference rules," IEICE Trans. Fundamentals, vol.E84-A, pp.820–828, 2001.