Automatic Fuzzy Rule Generation Utilizing Genetic Algorithms

Hee Soo Hwang, Kwang Bang Woo

Department of Electrical Engineering, Yonsei University, Seoul 120-749, Korea

ABSTRACT

In this paper, an approach to identify fuzzy rules is proposed. The decision of the optimal number of fuzzy rule is made by means of fuzzy c-means clustering. The identification of the parameters of fuzzy implications is carried out by use of genetic algorithms. For the efficient and fast parameter identification, the reduction technique of search areas of genetic algorithms is proposed. The feasibility of the proposed approach is evaluated through the identification of the fuzzy model to describe an input-output relation of Gas Furnace. Despite the simplicity of the proposed approach the accuracy of the identified fuzzy model of gas furnace is superior as compared with that of other fuzzy models.

Key Words: Identification of fuzzy implications; genetic algorithms; fuzzy clustering; fuzzy modeling

1. INTRODUCTION

In this paper, the identification of fuzzy rules is classified into the optimal number of fuzzy implications and the identification of the parameters defining membership functions in each fuzzy implication. The optimal number of the membership functions of each input variable is determined by soft c-means clustering. The optimal number of fuzzy rules is determined by the multiplication of the cluster numbers of each input variable under the assumption that input variables are mutually independent. The identification of the parameters is carried out utilizing genetic algorithms which display an excellent robustness in complex optimization problems. For an efficient and fast identification of parameters, the reduction technique of search areas of genetic algorithms is proposed. Schematic diagram of the identification procedure is shown in Fig. 1. The feasibility of the proposed approach is evaluated through the identification of the fuzzy model to describe an input-output relation of Gas Furnace by use of data which were previously presented[3].

2. FUZZY IMPLICATION AND REASONING

The format of fuzzy implication and reasoning algorithm is described as follows:

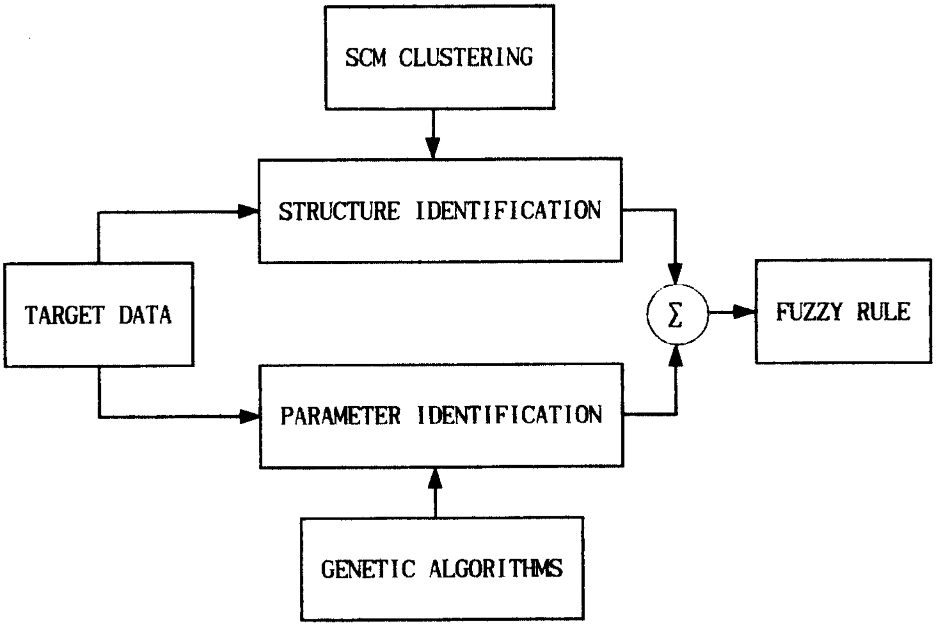


Fig. 1. Schematic diagram of the identification procedure

2.1. Format of implication

We consider an ith fuzzy implication, R^i.

R^i: If x_1 is Small and x_2 is Big, then $y_i = w_i \cdot a_i + b_i$

where Small and Big are fuzzy lables of x_1 and x_2, respectively, w_i the degree of fulfillment of the premise, and a_i and b_i consequent parameters.

2.2. Reasoning algorithm

Suppose implications $R^i(i=1,2)$ of the above format.

R^1: If x_1 is Small and x_2 is Big, then $y_1 = w_1 \cdot a_1 + b_1$

R^2: If x_1 is Big and x_2 is Medium, then $y_2 = w_2 \cdot a_2 + b_2$

Fig. 2 shows the procedure of reasoning, where w_1 and w_2 are caculated by eq. (1). Given input data x_1^0 and x_2^0, the output y^* inferred from above two implications is obtained in terms of the average of y_1 and y_2 with the weights w_1 and w_2.

$$w_1 = \mu_{Small}(x_1^0) \cdot \mu_{Big}(x_2^0) \quad (1)$$

$$w_2 = \mu_{Big}(x_1^0) \cdot \mu_{Medium}(x_2^0) \quad (2)$$

$$y^* = \frac{w_1 \cdot y_1 + w_2 \cdot y_2}{w_1 + w_2} = \frac{w_1 \cdot (w_1 \cdot a_1 + b_1) + w_2 \cdot (w_2 \cdot a_2 + b_2)}{w_1 + w_2}$$

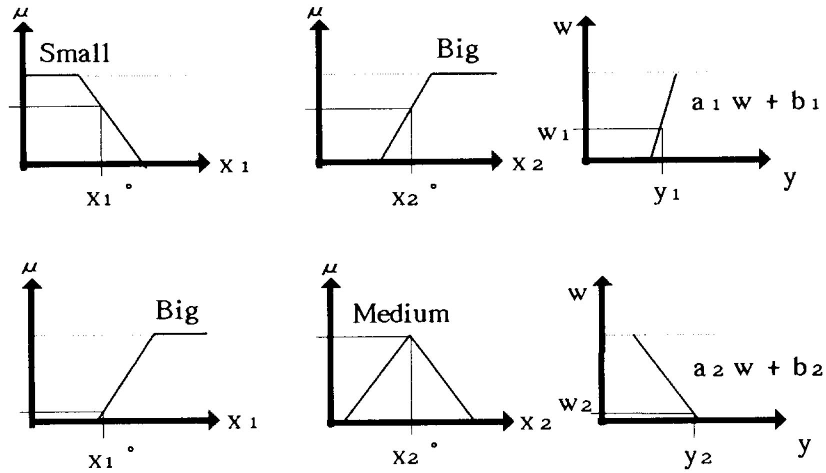


Fig. 2. Procedure of reasoning

3. STRUCTURE IDENTIFICATION

The identification of structure which corresponds to the decision of the optimal number of fuzzy implications to describe an input-output relation is carried out by means of soft c-means clustering which produces a fuzzy c-partition of data set. Our purpose is to find the optimal or sub-optimal numbers of clusters to effectively describe the relation between each input-output value of a system under the assumption that input variables are mutually independent, and then the number of fuzzy implications is determined to consider all possible combinations of the partitioned input spaces of all input variables. For the brevity, we consider a system composed of two inputs x_1 and x_2 and one output y. If the optimal or sub-optimal cluster numbers of x_1-y and x_2-y are c_1 and c_2, repectively, the optimal number of fuzzy implications is $c_1 \cdot c_2$. The validity of partitioning is evaluated by a validity index, S[2] according to the number of clusters. The appropriate values of clusters can be determined at lower value of S.

The soft c-means(SCM) clustering algorithm produces a fuzzy c-partition of the data set $X=\{X_1, X_2, \ldots, X_n\}$. The basic steps of the algorithm used in this paper are given as follows[1]:

Step 1

Fix the number of cluster $c(2 \leq c \leq n)$, where n is number of data items.

Fix $m(1 \leq m \leq \infty)$, set $p=1$, and initialize the fuzzy c partition, $U^{(p-1)}$.

Step 2

Calculate the c cluster centers $\{v_i^{(p)}\}$ with $U^{(p-1)}$ and the formula (3) for the ith cluster center.

$$v_{iL}^{(p)} = \frac{\sum_{k=1}^{n} (\mu_{ik})^m X_{kl}}{\sum_{k=1}^{n} (\mu_{ik})^m}, \quad L=1, \ldots, d \tag{3}$$

where d is dimension of data vector X_k, and $\mu_{ik} = \mu_i(X_k)$ is the membership grade of Xk in fuzzy set μ_i.

Step 3

Update $U^{(p)}$ for k = 1 to n.

① Calculate Ik and Ik'.

$I_k = \{ i \mid 1 \le i \le c, D_{ik} = \|X_k - V_i\| = 0 \}, I_k' = \{1,2,\ldots,c\} - I_k$

② For data item k, compute new membership values.

i) If $I_k = 0$, $\mu_{ik} = D_{ik}^{2/(1-m)}$ if $\mu_{ik} < (\alpha/c)$, $\mu_{ik} = 0$

$$\mu_{ik} = \mu_{ik} / \sum_{i=1}^{c} \mu_{ik}$$

ii) If $I_k' \neq 0$, $\mu_{ik} = 0$ for all $i \in I_k'$, and $\sum_{i \in I_k} \mu_{ik} = 1$.

③ Next k.

Step 4

$$J_m^{(p)} = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^m D_{ik}^2 \tag{4}$$

Compare $J_m^{(p)}$ and $J_m^{(p-1)}$. If $|J_m^{(p)} - J_m^{(p-1)}| < \varepsilon$, stop;

otherwise, set $p = p + 1$, and go to step 2.

Validity measure for fuzzy partitioning used in this paper is as follow[2]:

$$S = \frac{\sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^m \|X_k - V_i\|^2}{n \cdot \min_{i,j} \|V_j - V_i\|^2} \tag{5}$$

For the SCM algorithm with m=2 in eq. (4), the smallest S indeed indicates a valid optimal partition.

4. PARAMETER IDENTIFICATION

The identification of parameters which define the membership functions of the premise and coefficients of the consequent is carried out using genetic algorithms which provide an excellent robustness in complex optimization problems.

4.1 Genetic algorithms

Genetic algorithms are iterative adaptive general purpose search strategies based on the principles of natural population genetics and natural selection. A simple genetic algorithm that yields good results in many practical problems is composed of three operators: reproduction, crossover, and mutation. Reprodution is a process in which individual strings are copied according to their objective function(fitness function) values which we want to maximize. Copying strings according to their fitness values means that strings with a higher value have a higher probability of contributing one more offspring in the next generation. After reproduction, simple crossover may proceed in two steps. First, members of the newly reproduced strings in the mating pool are mated at random. Second, each pair of strings is selected uniformly at random between 1 and string length less one, $L-1$. Two new strings are created by swapping all characters between positions $k+1$ and L inclusively. Mutation is a secondary operator whose use guarantees that the probability of searching a particular sub-region of the solution is never zero. These operators are simplicity itself, involving nothing more complex than random number generation, string copying, and partial string exchanging; yet, despite their simplicity, the resulting search performances is wide-ranging and impressive due to implicit parallellism of genetic algorithm.

4.2 Fitness function

In many problems, the objective is stated as the minimization of some cost function rather than the maximization of some utility or profit. In the parameter identification, our purpose is to minimize the cost function, eq. (6) which is defined as the average of squared errors between target output data and inferred output data.

$$E = \frac{1}{n}\sum_{i=1}^{n}(y_i^0 - y_i^*)^2 \tag{6}$$

where n is the total number of data, y_i^0 target output value, y_i^* output inferred from fuzzy implications,

With genetic algorithms, we use the following cost to fitness transformation:

Fitness function, $f = 1.0/E$ (7)

4.3 Multiparameter coding

One successful method employed in coding multi-parameters of optimization problems is the concatenated, multiparameter, mapped, and fixed point coding. To construct a multiparametr coding, we can simply concatenate many single parameter codings which define the membership functions of the premise and coefficients of the consequent. Each coding has its own sublength L, its own minimum and maximum values, Pmin and Pmax, respectively, and the precision of the decoded value is controlled by $(P_{max}-P_{min})/(2^L-1)$.

4.4 Reduction of search areas

For the efficient and fast convergence of genetic algorithms, the reduction technique of search areas of the parameters is proposed. GAs learn by iteratively generating candidate solutions through genetic operators, such as reproduction, crossover, and mutation and testing the fitness of the solutions as shown in Fig. 3, in which the reduction is carried out through the increment of Pmin and the decrement of Pmax of each parameter, where Pmax's and Pmin's are obtained from the decoded values of m chromosomes whose fitness values are greater than a value α.

```
Generate random population, P(0);
Evaluate fitness of individuals in P(0);
freq = 1, t = 1
While(t is less than max generation) {
     Generate P(t+1) from P(t) as follows:
         select the fittest individuals from P(t);
         recombine them by crossover and mutation;
     Evaluate the fitness of individuals in P(t+1);
     While(fitness value is greater than reduction criterion){
         decode individuals;
         decide the minimum and maximum ranges of parameters;
         if(freq is equal to m) {
            reduce search areas;
            freq=1;
         }
         else freq = freq + 1;
     }
     t = t + 1;
     Retain the fittest individual;
}
```

Fig. 3. Sequence of GAs by reduction of search areas.

5. SIMULATION

In this section, the feasibility of the proposed approach is evaluated through the identification of the fuzzy model to describe an input-output relation of Gas Furnace. Our purpose is to identify the fuzzy model which describe the relation between a gas flow u(t) and the combusted CO_2 concentration y(t) of Gas Furnace using the 299 pairs of data presented by Box and Jenkins[3]. We consider u(t-4) and y(t-1) as input variables of fuzzy implications, and y(t) as

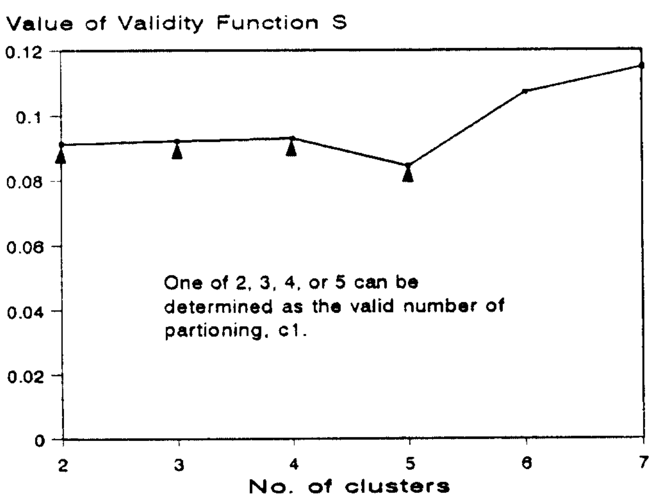


Fig. 4. Graph for the decision of cluster number in u(t-4) vs. y(t)

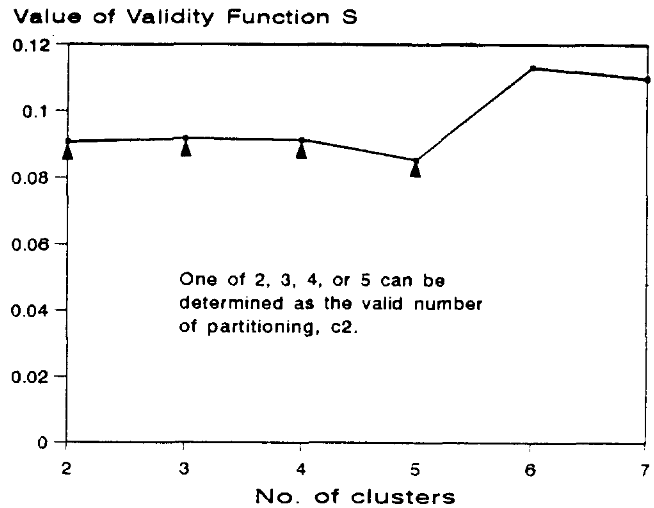


Fig. 5. Graph for the decision of cluster number in y(t-1) vs. y(t)

output variable. A small S indicates a partition in which all the clusters are overall compact, and separate to each other. Thus, our goal is to find the fuzzy c-partition with smaller S and smaller c in order to minimize the number of fuzzy implications. The appropriate numbers of clusters which describe each input-output relation effectively are determined as 3, and 2 for c1 and c2 from the validity measure graphs for fuzzy clustering shown in Fig. 4 and Fig. 5, respectively.

From the clustering results, the membership function of each input variable calculated by step 3 of SCM clustering algorithms from each cluster of x_1 and x_2, are shown in Fig. 6 and

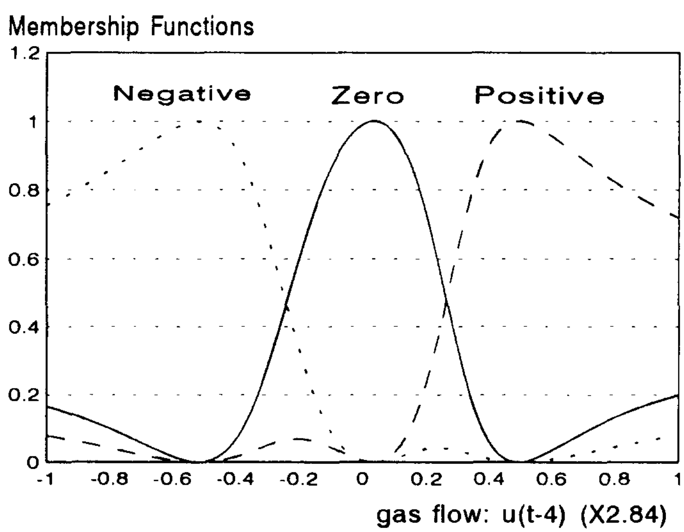


Fig. 6. Membership values of u(t-4) calculated by SCM clustering with c1=3

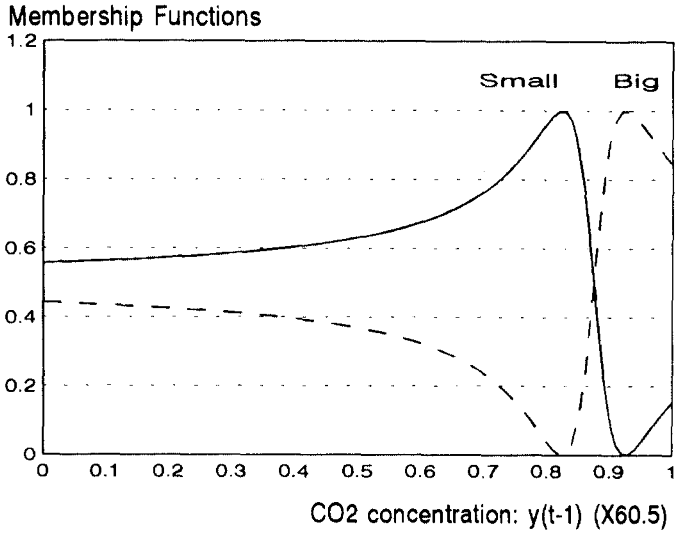
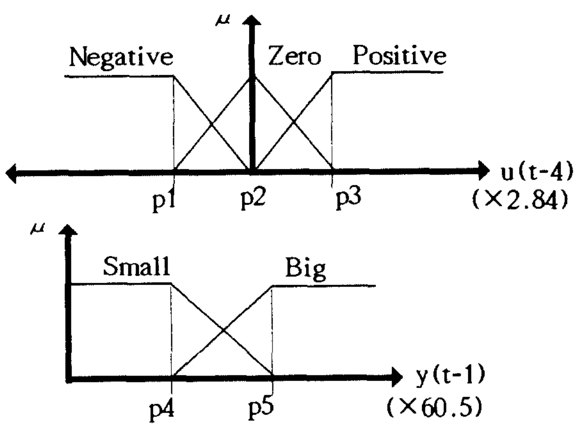


Fig. 7. Membership values of y(t-1) calculated by SCM clustering with c2=2

Fig. 7, in which the membership function of input variables are defined as Positive, Zero, Negative for u(t-4), and Small and Big for y(t-1), respectively.

In order to eliminate the ripples of the membership functions shown in Fig. 6 and Fig. 7, we transform them into triangular and trapezoidal membership functions. Using the transformed membership functions, the fuzzy implications for modeling are composed of as follows:



If u(t-4) is Negative & y(t-1) is Small, then $y = a_1 \cdot w_1 + b_1$
If u(t-4) is Negative & y(t-1) is Big, then $y = a_2 \cdot w_2 + b_2$
If u(t-4) is Zero & y(t-1) is Small, then $y = a_3 \cdot w_3 + b_3$
If u(t-4) is Zero & y(t-1) is Big, then $y = a_4 \cdot w_4 + b_4$
If u(t-4) is Positive & y(t-1) is Small, then $y = a_5 \cdot w_5 + b_5$
If u(t-4) is Positive & y(t-1) is Big, then $y = a_6 \cdot w_6 + b_6$

In the fuzzy implications the parameter $p_1, p_2, \ldots, p_5$ and $a_1, b_1, \ldots, a_6, b_6$ are identified by use of GAs. Initial parameters for GAs are as follows: population size is 50, length of individuals 10, crossover rate 0.6, and mutation rate 0.033. The reduction was carried out at the number of generation, 63 and 82. Fitness values calculated from the best strings in evolving populations are shown in Fig. 8.

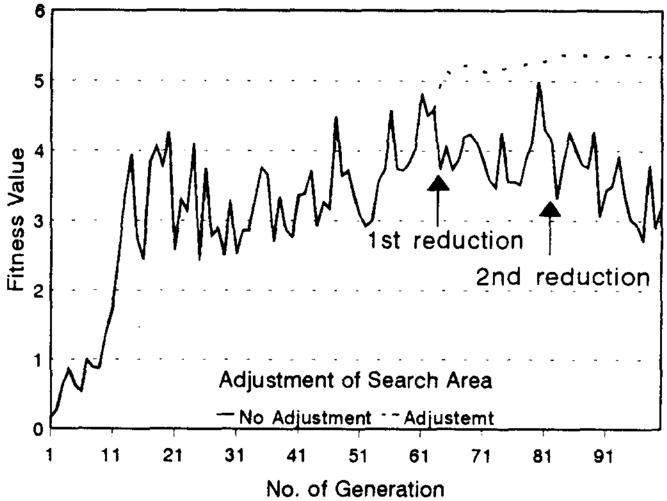


Fig. 8. Graph of fitness values for the reduction technique

Identified parameters are shown in Table 1, and mean square errors for the comparison with other results in Table 2.

Table 1. Identified parameters of fuzzy implications

Premise		Consequent			
p1	-0.695	a1	6.654	b1	44.349
p2	-0.138	a2	1.951	b2	58.025
p3	0.906	a3	2.394	b3	39.94
p4	0.561	a4	0.15	b4	58.478
p5	0.989	a5	-4.86	b5	41.989
		a6	3.412	b6	51.274

Table 2. Comparison with other fuzzy model

Model Name	Mean Square Error	Number of Rules
Tong's model[4]	0.469	19
Pedrycz's model[5]	0.776	20
Xu's model[6]	0.328	25
Sugeno's model[7]	0.355	6
Our model	0.187	6

6. CONCLUSION

In spite of the simplicity of the proposed identification method, the accuracy of the identified fuzzy model to describe input-output relation of gas furnace is superior as compared with that of other fuzzy models, so the proposed method for the automatic generation of fuzzy ruels is also able to be applied to the generation of the fuzzy control rules.

REFERENCE

1. Bezdek, J. C., "Pattern Recognition with Fuzzy Objective Function Algorithms," New York Plenum, 1981.
2. Xuanli, L. X., and Gerado, B., "A Validity Measure for Fuzzy Clustering," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-13, no. 8, pp. 841-847, August, 1991.
3. Box, G. E. P. and et al., "Time Series Analysis, Forcasting and Control," Holden Day, Sanfrancisco, 1970.
4. Tong, R. M., "The evaluation of fuzzy models derived from experimental data," Fuzzy Set and Systems., Vol. 13, 1-12, 1980.
5. Pedrycz, W., "An identification algorithm in fuzzy relational systems," Fuzzy Sets and Systems, Vol. 13, pp. 153-167, 1984.
6. Chen-Wei Xu and Yong-Zailu, "Fuzzy Model Identification and Self-Learning for Dynamic Systems," IEEE Trans. on Syst. Man and Cybern., SMC-17, No. 4, pp. 684-689, 1987.
7. Sugeno, M. and Yasukawa, T., "Linguistic Modeling Based on Numerical Data," IFSA' 91 Brussels, Computer, Management & System Science., pp. 264-267, 1991.
8. Goldberg, D. E., "Genetic Algorithms in Search, Optimization, and Machine Learning," Addison Wesley, 1989.