A Hybrid Clustering and Gradient Descent Approach for Fuzzy Modeling

Ching-Chang Wong and Chia-Chong Chen

***Abstract*— In this paper, a hybrid clustering and gradient descent approach is proposed for automatically constructing a multi-input fuzzy model where only the input–output data of the identified system are available. The proposed approach is composed of two steps: structure identification and parameter identification. In the process of structure identification, a clustering method is proposed to provide a systematic procedure to determine the number of fuzzy rules and construct an initial fuzzy model from the given input–output data. In the process of parameter identification, the gradient descent method is used to tune the parameters of the constructed fuzzy model to obtain a more precise fuzzy model from the given input–output data. Finally, two examples of nonlinear system are given to illustrate the effectiveness of the proposed approach.**

***Index Terms*— Fuzzy modeling, gradient descent method, unsupervised clustering algorithm.**

I. Introduction

FUZZY modeling has been studied to deal with complex, ill-defined and uncertain systems, in which the conventional mathematical model fails to give satisfactory results. The main goal of fuzzy modeling is to describe the input–output behavior of a given system by a set of fuzzy inference rules [1]–[14]. This approach can be regarded as a process of system identification. In general, the identification of a fuzzy system model consists of two major phases: structure identification and parameter identification [4], [7], [12], [14]. The first phase is the identification of the structure of the fuzzy model, and the second is the tuning of the parameter values of the fuzzy model. Generally speaking, structure identification mainly involves the determination of the structure of the fuzzy system, the number of fuzzy rules, and the membership functions of the premise and consequent fuzzy sets in each rule. Therefore, the purpose of structure identification is to construct an initial fuzzy model to describe the inherent structure of the given input–output data. In the process of the parameter identification, a parameter learning procedure is applied to obtain a more precise fuzzy model.

When we have a lot of input–output data from the observation of the identified system, and we have no other information about the system, how to determine the structure of the fuzzy model becomes an important issue. In general, if only the input–output data are available for the system, then clustering is one of the most promising techniques to construct the structure of the fuzzy system. Several clustering techniques [15] had been used to extract rules to construct the initial rule-base, such as fuzzy c-means (FCM) algorithm [4], [5], [10], and the mountain method [14], [16], [17], [18]. The basic idea is to group the input–output pairs into clusters and use one rule for one cluster; that is, the number of rules equals the number of clusters. However, the disadvantage of the FCM algorithm is that the number of clusters must be predetermined. If the number of clusters is given, the clustering results of the FCM algorithm are also influenced by the choice of initial cluster centers and the distance measure. On the other hand, the mountain method is proposed for approximately estimating cluster centers [4], [14], [18]. The procedures of the method are the drawing of grid lines on the data space, the construction of a mountain function from the data, and then the destruction of the mountain to obtain the cluster centers. The mountain method does not require predetermination of the number of clusters, but it needs to let intersecting points (grid points) of grid lines as the candidates for cluster centers. Therefore, the closeness of approximation of the actual centers is very sensitive to the denseness of the grid lines. The more the grid lines the closer the approximation, but more calculations are needed. The computation work increases exponentially with the dimension of the problem because the method must evaluate the mountain function over all grid points [4]. In addition, the mountain method cannot classify the data set such that it cannot analyze the shapes or the sizes of clusters. Thus, the width of the membership functions of the fuzzy model is set to be the same [14]. Consequently, the obtained fuzzy model cannot clearly describe the behavior of the given input–output data. In order to avoid these drawbacks, we propose a clustering algorithm by which the clusters are automatically generated and the data points are appropriately classified. It is worth emphasizing that the proposed clustering method does not need any prior assumption about the structure of the data. The inherent structure of the input–output data can be described by the proposed clustering algorithm. Therefore, in the process of structure identification, the number of fuzzy rules and a rough estimate of the membership functions of each rule can be obtained by the proposed clustering algorithm.

When an initial fuzzy model is constructed in the process of structure identification, a parameter learning procedure is successively applied to obtain a more precise fuzzy model in the process of parameter identification. Generally speaking, the process of parameter identification is a process for tuning the

Manuscript received December 22, 1997; revised September 16, 1998 and June 2, 1999. This work was supported in part by the National Science Council of the Republic of China under Contract NSC-88-2213-E-032-024. This paper was recommended by Associate Editor A. Kandel.

The authors are with the Department of Electrical Engineering, Tamkang University, Taipei 25137, Taiwan, R.O.C.

Publisher Item Identifier S 1083-4419(99)08049-8.

TABLE I
NUMERICAL VALUES OF THE ORIGINAL DATA POINTS AND MOVABLE VECTORS AT THE END OF EACH ITERATION

i	x_i (Original Data)	v_i (1st Iteration)	v_i (2nd Iteration)	v_i (3rd Iteration)	v_i (6th Iteration)
1	(0.6000,0.7000)	(0.6000,0.6728)	(0.6000,0.6346)	(0.6000,0.6044)	(0.6000,0.6000)
2	(0.7000,0.6000)	(0.6728,0.6000)	(0.6346,0.6000)	(0.6044,0.6000)	(0.6000,0.6000)
3	(0.6000,0.6000)	(0.6000,0.6000)	(0.6000,0.6000)	(0.6000,0.6000)	(0.6000,0.6000)
4	(0.6000,0.5000)	(0.6000,0.5272)	(0.6000,0.5654)	(0.6000,0.5956)	(0.6000,0.6000)
5	(0.5000,0.60000	(0.5272,0.6000)	(0.5654,0.6000)	(0.5956,0.6000)	(0.6000,0.6000)
6	(0.2000,0.5000)	(0.2000,0.4728)	(0.2000,0.4346)	(0.2000,0.4044)	(0.2000,0.4000)
7	(0.3000,0.4000)	(0.2728,0.4000)	(0.2346,0.4000)	(0.2044,0.4000)	(0.2000,0.4000)
8	(0.2000,0.4000)	(0.2000,0.4000)	(0.2000,0.4000)	(0.2000,0.4000)	(0.2000,0.4000)
9	(0.2000,0.3000)	(0.2000,0.3272)	(0.2000,0.3654)	(0.2000,0.3956)	(0.2000,0.4000)
10	(0.1000,0.4000)	(0.1272,0.4000)	(0.1654,0.4000)	(0.1956,0.4000)	(0.2000,0.4000)
11	(0.8000,0.3000)	(0.8000,0.2728)	(0.8000,0.2346)	(0.8000,0.2044)	(0.8000,0.2000)
12	(0.9000,0.2000)	(0.8728,0.2000)	(0.8346,0.2000)	(0.8044,0.2000)	(0.8000,0.2000)
13	(0.8000,0.2000)	(0.8000,0.2000)	(0.8000,0.2000)	(0.8000,0.2000)	(0.8000,0.2000)
14	(0.8000,0.1000)	(0.8000,0.1272)	(0.8000,0.1654)	(0.8000,0.1956)	(0.8000,0.2000)
15	(0.7000,0.2000)	(0.7272,0.2000)	(0.7654,0.2000)	(0.7956,0.2000)	(0.8000,0.2000)

parameters of the fuzzy model to minimize the squared error function. In this paper, the gradient descent method is applied to tune the parameters of the constructed fuzzy model. The remainder of this paper is organized as follows. In Section II, we propose a new clustering method for automatically generating an initial fuzzy model from the given input–output data. Multidimensional Gaussian functions and real values are respectively used to describe the fuzzy sets in the premise and consequent part of the fuzzy rules. In Section III, the gradient descent method is applied to learn the behavior of the input–output data so that a precise fuzzy model is identified. In Section IV, two examples of nonlinear system are used to illustrate the effectiveness of the proposed approach. Finally, Section V concludes the paper.

II. STRUCTURE IDENTIFICATION BY THE PROPOSED CLUSTERING METHOD

In general, the process of structure identification of a fuzzy model is to obtain the information that includes the structure of the fuzzy system, the number of inference rules of the fuzzy model and rough estimates of the parameters describing the fuzzy sets in the premise and consequent parts. We first propose an alternative approach to extract the fuzzy rules based upon clustering analysis of the input–output data. The clustering method is described as follows.

Let $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a set of n vectors in a p-dimensional feature space, where $\mathbf{x}_i = (x_i(1), x_i(2), \ldots, x_i(p))$ is a vector. Our objective is to cluster the given data set into several groups such that similar points are grouped into the same cluster while dissimilar points are in different clusters. We find that vectors with high relational grades will have the same characteristics, thus we can group those vectors as a cluster. Therefore, the idea is that we choose a point as a reference vector and find those vectors which have high relational grades with the reference vector. Then we can replace the reference vector by the average of those vectors with high relational grades with the reference vector. In this way, the replaced vector tends toward the center of cluster. The unsupervised algorithm can be proposed as follows.

Step 1) Define n movable vectors $\mathbf{v}_i$ $(i = 1, 2, \ldots, n)$ and let $\mathbf{v}_i = \mathbf{x}_i$, that is, $\mathbf{x}_i$ is the initial value of $\mathbf{v}_i$.

Step 2) Calculate the relational grades between the reference vector $\mathbf{v}_i$ and the comparative vector $\mathbf{v}_j$ by

$$r_{ij} = \exp\left(-\frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{2\sigma^2}\right), \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n \tag{1}$$

where $\|\mathbf{v}_i - \mathbf{v}_j\|$ represents the Euclidean distance between $\mathbf{v}_i$ and $\mathbf{v}_j$; and σ is the width of the Gaussian function.

Step 3) Modify the relational grades between the reference vector $\mathbf{v}_i$ and the comparative vector $\mathbf{v}_j$ by the

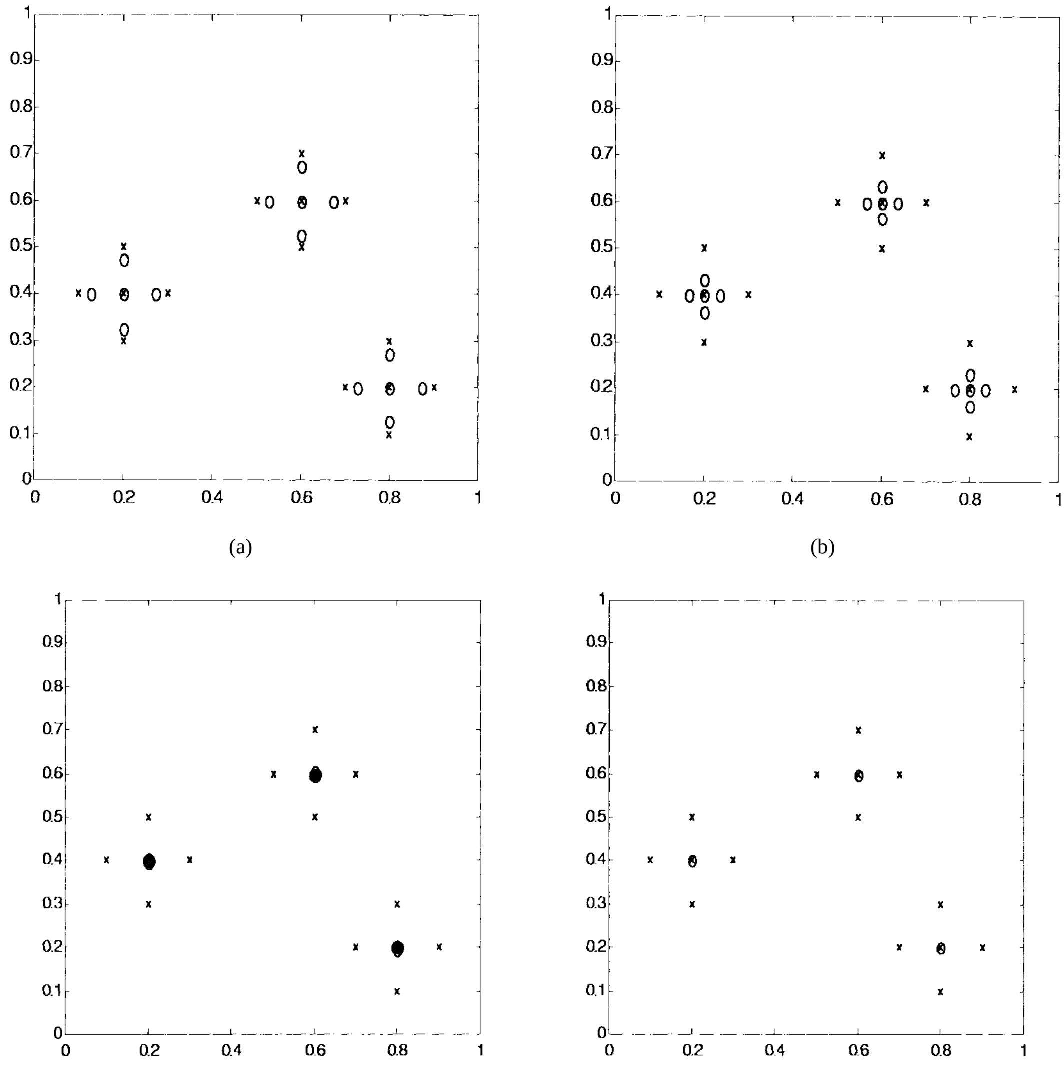

Fig. 1. Intermediate results for Table I at the end of each iteration. (a) First iteration; (b) second iteration; (c) third iteration; and (d) sixth iteration (convergence).

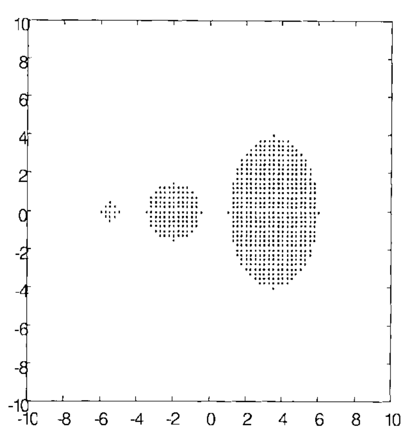

Fig. 2. Synthetic data set.

following formula

$$r_{ij} = \begin{cases} 0, & \text{if } r_{ij} < \zeta \\ r_{ij}, & \text{otherwise.} \end{cases} \quad (2)$$

where ζ is a small constant.

Step 4) Calculate $\mathbf{v}_i' = (v_i'(1), v_i'(2), \ldots, v_i'(p))$ by the following equations

$$\mathbf{v}_i' = \frac{\sum_{j=1}^n r_{ij}\mathbf{v}_j}{\sum_{j=1}^n r_{ij}}, \quad i = 1, 2, \ldots, n. \quad (3)$$

Step 5) If all the vectors $\mathbf{v}_i'$ are the same as $\mathbf{v}_i$, $i = 1, 2, \ldots, n$, then go to Step 6; otherwise let $\mathbf{v}_i = \mathbf{v}_i'$, $i = 1, 2, \ldots n$, and go to Step 2.

Step 6) Based on the final results $\mathbf{v}_i$, $i = 1, 2, \ldots n$, we can determine that the number of clusters is equal to the number of convergent vectors, the original data with the same convergent vector are grouped into the same cluster, and the convergent vector is the cluster center.

It is evident from (1) that the closer the comparative vector $\mathbf{v}_j$ to the reference vector $\mathbf{v}_i$ is, the higher the relational grade r_{ij} is. That is, the value of r_{ij} is maximum for $\mathbf{v}_j$ that coincides with $\mathbf{v}_i$ and the value of r_{ij} decreases exponentially with the distance of $\mathbf{v}_j$ away from $\mathbf{v}_i$. The idea of this procedure is to group the data points with high relational grades as a

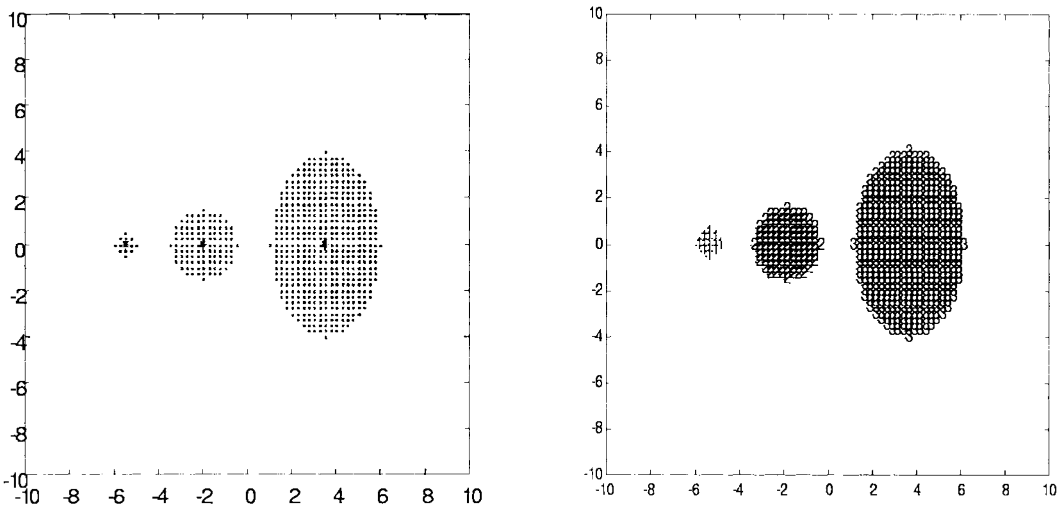

Fig. 3. Results obtained by the proposed clustering algorithm for the synthetic data set.

cluster. Thus, we modify the relational grade in Step 3 such that the vectors with low relational grades cannot give any effect on the movable vector in Step 4. In this paper, we choose $\xi = 0.01$. According to the algorithm, these movable vectors will gradually converge to some vectors. Therefore, the number of convergent vectors is the number of clusters, and the convergent vector is viewed as the center of the corresponding cluster. It is worth emphasizing that the proposed clustering method does not require predetermination of the number of clusters and the appropriate initial cluster centers. Furthermore, the data points can be classified according to the convergent vector of each data point. That is, the data points are classified as the same group when their convergent vectors are the same.

In order to illustrate the process of the proposed clustering algorithm, we consider a set of data points in R^2 space. The numerical values of the original data points and movable vectors at the end of each iteration using the proposed algorithm are listed in Table I. The simulation results are shown in Fig. 1, where $\mathbf{x}_i$ and $\mathbf{v}_i$ $(i = 1, 2, \ldots, 15)$ are represented as "x" and "o", respectively. We find that the moveable vectors respectively converge to three points in the 6th iteration. This indicates that there are three clusters, and the three cluster centers are at $(0.6, 0.6)$, $(0.2, 0.4)$ and $(0.8, 0.2)$, which are consistent with the actual cluster centers of the given data. Furthermore, from the results in Table I, we can easily find that the data set is classified into three groups $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$, $\{\mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{10}\}$ and $\{\mathbf{x}_{11}, \mathbf{x}_{12}, \mathbf{x}_{13}, \mathbf{x}_{14}, \mathbf{x}_{15}\}$ according to the convergent vector of each data point. In order to illustrate the effectiveness of the clustering algorithm, we also consider a data set containing three clusters of various shapes and sizes shown in Fig. 2. After the proposed clustering algorithm and the FCM algorithm are respectively applied to the synthetic data set, the obtained results are respectively shown in Figs. 3 and 4, where "*" signs indicate the locations of the resulting cluster centers. In the estimation of the cluster centers, the cluster centers estimated by the proposed method is more reasonable than that by the FCM method. In the classification of the data, the data set is classified correctly by the proposed approach, but many data points are misclassified by the FCM method. On the other hand, the mountain method with the more grid lines can obtain the reasonable cluster centers as the proposed method. However, the mountain method cannot classify the data set, hence it cannot analyze the shapes or sizes of clusters. Consequently, in the sense of fuzzy modeling, the proposed clustering algorithm is better than the FCM algorithm and the mountain method. In the following, we will describe how to use the obtained cluster centers to construct an initial fuzzy model to approximate the behavior of the input–output data.

Let us assume that n input–output data $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ are obtained from the observation of the identified system and each data point is represented by $\mathbf{x}_i = (x_i(1), x_i(2), \ldots, x_i(p))$, where $(x_i(1), x_i(2), \ldots, x_i(p-1))$ is the input vector of the ith input–output pair and $x_i(p)$ is the corresponding output. When c clusters are obtained by the proposed clustering algorithm for the given data set, and the corresponding cluster centers are $\mathbf{c}_m = (c_m(1), c_m(2), \ldots, c_m(p))$, $m = 1, 2, \ldots, c$, the rule-base of an initial fuzzy model can be constructed as follows:

$$\text{Rule } m\text{: If } \mathbf{x} \text{ is } A_m \text{ then } y \text{ is } y_m. \quad m = 1, 2, \ldots, c \tag{4}$$

where c is the number of fuzzy rules, $\mathbf{x} = (x_i(1), x_i(2), \ldots, x_i(p-1))$ is the input variable, y is the output variable, A_m is a multidimensional fuzzy set in the premise part, and y_m is a real number in the consequent part. In each rule, the real number y_m in the consequent part is described by

$$y_m = c_m(p) \tag{5}$$

and the fuzzy set A_m in the premise part is described by the multidimensional Gaussian membership function

$$A_m(\mathbf{x}) = \exp\left(-\frac{\sum_{k=1}^{p-1}(x(k) - c_m(k))^2}{2\delta_m^2}\right), \quad m = 1, 2, \ldots, c \tag{6}$$

where δ_m is the width of the Gaussian function. It is clear that the rule number and the premise and consequent pa-

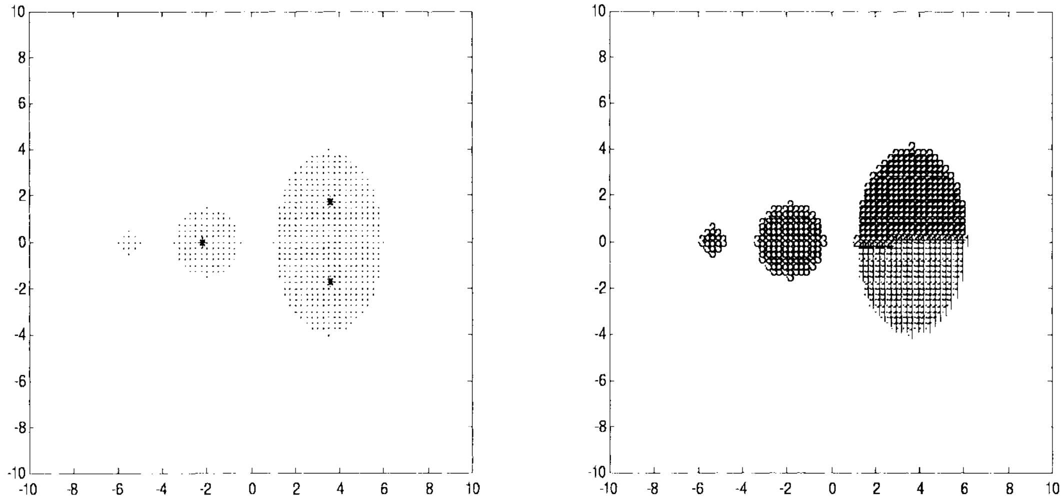

Fig. 4. Results obtained by the FCM algorithm for the synthetic data set.

rameters of the initial fuzzy model are determined by the cluster number c and the corresponding cluster center $\mathbf{c}_m = (c_m(1), c_m(2), \ldots, c_m(p-1), c_m(p))$, $m = 1, 2, \ldots, c$, which are obtained by the proposed clustering algorithm. In order to determine the appropriate value of δ_m, we let

$$A_m(\mathbf{x}_m^*) = u \tag{7}$$

where $\mathbf{x}_m^*$ is the farthest data point from the cluster center $\mathbf{c}_m$ in the mth cluster, and $u \in [0, 1]$ is a constant value. That is, we let the membership value $A_m(\mathbf{x}_m^*)$ be equal to the value of u. Therefore, from (6) and (7), an appropriate value of δ_m can be directly derived by

$$\delta_m = \sqrt{\frac{-\sum_{k=1}^{p-1}(x_m^*(k) - c_m(k))^2}{2\ln(u)}}, \quad m = 1, 2, \ldots, c. \tag{8}$$

Based on experience, we suggest $u \in [0.1, 0.3]$ and we adopt $u = 0.2$ in this paper. From the above discussion, we know that the proposed clustering method can provide a systematic approach to construct an initial fuzzy model from the given input–output data.

III. Parameter Identification by the Gradient Descent Method

In Section II, a rough fuzzy model is automatically generated from the given input–output data by using the proposed clustering algorithm. By using the weight averaged method, the output of the constructed fuzzy model can be described by

$$y = \frac{\sum_{m=1}^{c} A_m(\mathbf{x}) \cdot y_m}{\sum_{m=1}^{c} A_m(\mathbf{x})}. \tag{9}$$

From (5), (6), and (9), we know that the output of the constructed fuzzy model consists of the following real parameters: $c_m(1), c_m(2), \ldots, c_m(p-1), c_m(p)$ and δ_m $(m = 1, 2, \ldots, c)$. In which, $c_m(1), c_m(2), \ldots, c_m(p-1)$, and δ_m are the premise parameters describing the fuzzy set A_m, and $c_m(p)$ is the consequent parameter describing the fuzzy singleton y_m. It is clear that the performance of the constructed fuzzy model that approximates the behavior of the given input–output data will be influenced by these premise and consequent parameters. Consequently, in the process of parameter identification, a parameter learning procedure is used such that the procedure can further use the given input–output data to finely tune the parameters of the constructed fuzzy model. In this paper, the gradient descent method is applied to tune these parameters to minimize the following squared error function

$$E_i = \frac{1}{2}(y - y_i)^2 \tag{10}$$

where $y_i = x_i(p)$ is the desired output data for the input data $x = (x_i(1), x_i(2), \ldots, x_i(p-1))$ and y is the output of the fuzzy model corresponding to the same input data. We denote the instantaneous error e as

$$\begin{aligned} e = y - y_i = y - x_i(p) &= \frac{\sum_{m=1}^{c} A_m(\mathbf{x}) \cdot y_m}{\sum_{m=1}^{c} A_m(\mathbf{x})} - x_i(p) \\ &= \sum_{m=1}^{c} g_m \cdot y_m - x_i(p) \end{aligned} \tag{11}$$

where g_m is the normalized membership value of the mth rule and is described by

$$g_m = \frac{A_m(\mathbf{x})}{\sum_{j=1}^{c} A_j(\mathbf{x})}, \quad m = 1, 2, \ldots, c. \tag{12}$$

According to the gradient descent method [4], [6], [12], [14], the rules for learning the premise and consequent parameters are

$$\begin{aligned} c_m(i+1, k) = c_m(i, k) &- \alpha e g_m (y_m - y)\left(\frac{x_i(k) - c_m(i,k)}{\delta_m^2}\right), \\ & k = 1, 2, \ldots, p-1, \end{aligned} \tag{13}$$

$$\delta_m(i+1) = \delta_m(i) - \alpha e g_m (y_m - y) \times \left(\frac{\sum_{k=1}^{p-1}(x_i(k) - c_m(i,k))^2}{\delta_m^3}\right), \tag{14}$$

$$c_m(i+1, p) = c_m(i, p) - \alpha g_m e \tag{15}$$

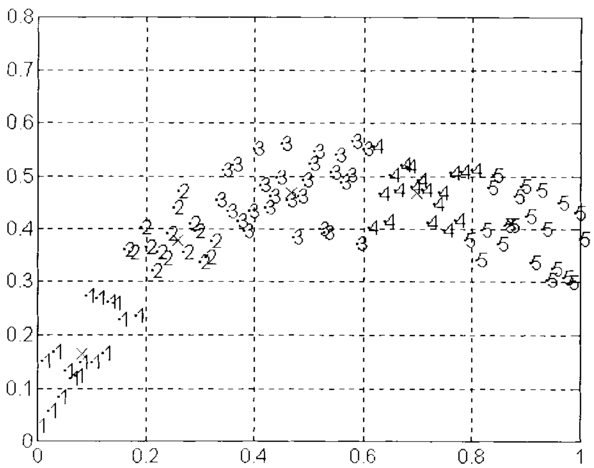

Fig. 5. Classification results of the proposed clustering algorithm.

where α denotes the learning rate. In the learning procedure, the parameter values $c_m(1), c_m(2), \ldots, c_m(p-1), \delta_m$ and $c_m(p)$ obtained by the clustering method are considered as initial values $c_m(0,1), c_m(0,2), \ldots, c_m(0,p-1)$, $\delta_m(0)$ and $c_m(0,p)$. The above recursive expressions provide an algorithm for tuning the parameters $c_m(1), c_m(2), \ldots, c_m(p-1)$, δ_m, and $c_m(p)$ until the same parameters as the previous iteration are obtained.

In summary, the procedure of the hybrid clustering and gradient descent approach is that first the proposed clustering algorithm is used to construct a rough fuzzy model in the structure identification step and then the gradient descent method is used to adjust the parameters of the constructed fuzzy model in the parameter identification step.

IV. Numerical Examples

The effectiveness of the proposed hybrid clustering and gradient descent method is illustrated by the following examples:

Example 1:

We consider a set of input–output data generated by the nonlinear system [14]

$$y = \frac{0.9x}{1.2x^3 + x + 0.3} + \eta \tag{16}$$

with x in the interval $[0, 1]$ and a noise component η with amplitude 0.2. In the structure identification step, the classification results of the proposed clustering approach are shown in Fig. 5, where "×" signs indicate the locations of the resulting cluster centers. This figure shows that the data points are classified into five groups and there are five cluster centers. Consequently, we can construct a rule-base of the initial fuzzy model associated with the obtained five clusters as follows:

$$\text{Rule } m\text{: If } x \text{ is } A_m \text{ then } y \text{ is } y_m. \quad m = 1, 2, \ldots, 5 \tag{17}$$

where x and y are the input and output variables, respectively. The parameter values of the premise and consequent fuzzy sets are listed in Table II. Fig. 6 shows the output of the fuzzy model that is obtained by the proposed clustering approach. It can be seen that the output closely approximates the original nonlinear equation. For comparison, the mountain method is

TABLE II
Parameter Values of the Initial Fuzzy Model Determined by the Clustering Algorithm for Example 1

m	Premise part $(c_m(1), \delta_m)$	Consequent Part $c_m(2)$
1	(0.0803,0.0882)	0.1652
2	(0.2577,0.0555)	0.3799
3	(0.4653,0.0885)	0.4693
4	(0.6987,0.0658)	0.4683
5	(0.8713,0.0873)	0.4114

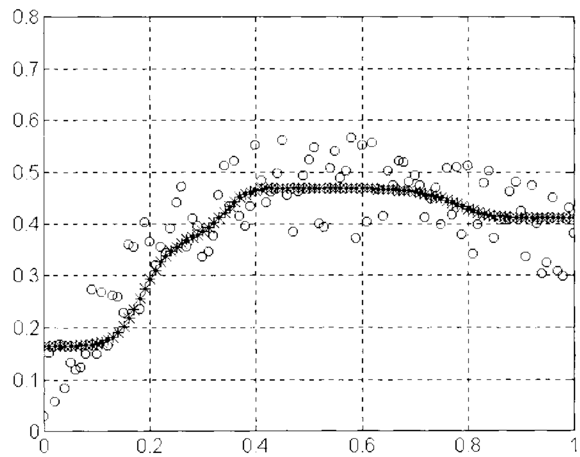

Fig. 6. Output of the initial fuzzy model determined by the proposed clustering algorithm for Example 1.

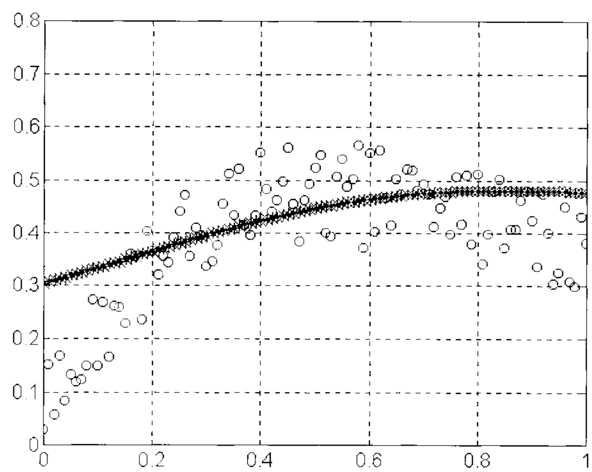

Fig. 7. Output of the initial fuzzy model determined by the mountain method for Example 1.

applied to the same data set, and the output of the obtained fuzzy model is shown in Fig. 7. We find that the output of the initial fuzzy model determined by the mountain method [14] is a very crude estimate of (16).

In the parameter identification step, using the values of the cluster centers obtained by the clustering algorithm as initial values of $c_m(0,1)$, $\delta_m(0)$, and $c_m(0,2)$, and a learning rate $\alpha = 0.03$, we obtain the final parameter values of the fuzzy

TABLE III
FINAL PARAMETER VALUES OF THE FUZZY MODEL DETERMINED BY THE HYBRID CLUSTERING AND GRADIENT DESCENT APPROACH FOR EXAMPLE 1

	Premise part	Consequent Part
m	$(c_m(1), \delta_m)$	$c_m(2)$
1	(0.0517,0.0613)	0.0818
2	(0.2040,0.0837)	0.3398
3	(0.4678,0.1558)	0.4735
4	(0.7227,0.1315)	0.5157
5	(0.9104,0.1451)	0.3489

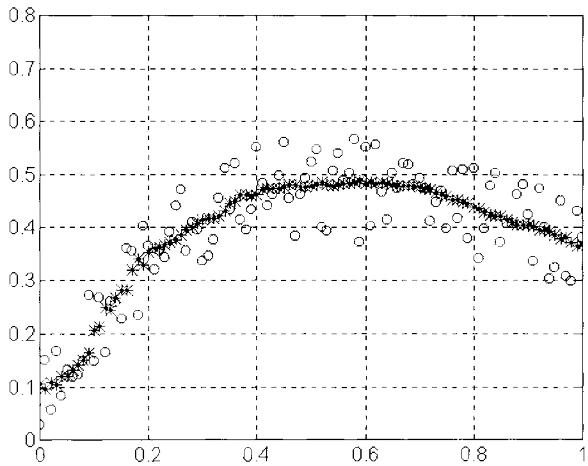

Fig. 8. Output of the final fuzzy model determined by the proposed hybrid clustering and gradient descent approach for Example 1.

TABLE IV
PARAMETER VALUES OF THE INITIAL FUZZY MODEL DETERMINED BY THE CLUSTERING ALGORITHM FOR EXAMPLE 2

	Premise Part	Consequent Part
m	$(c_m(1), c_m(2), \delta_m)$	$c_m(3)$
1	(-0.8477,0.1697,0.1793)	-0.1152
2	(-0.8478,0.5000,0.1118)	-0.1154
3	(-0.8477,0.8303,0.1793)	-0.1152
4	(-0.5005,0.5000,0.2472)	-0.7330
5	(0.0000,0.1446,0.3286)	0.0000
6	(0.0000,0.8552,0.3286)	0.0000
7	(-0.2536,0.5000,0.0926)	-0.3724
8	(0.0000,0.6478,0.0824)	0.0000
9	(0.2536,0.5000,0.0926)	0.3724
10	(0.5005,0.5000,0.2472)	0.7330
11	(0.8477,0.1697,0.1793)	0.1152
12	(0.8477,0.8303,0.1793)	0.1152
13	(0.8478,0.5000,0.1118)	0.1154

model, which are listed in Table III. In Fig. 8, we show the output of the obtained fuzzy model after parameter tuning by the gradient descent method. We can see that a rather close approximation of the original nonlinear system is obtained by the combination of the proposed clustering and the gradient descent method.

TABLE V
FINAL PARAMETER VALUES OF THE FUZZY MODEL DETERMINED BY THE HYBRID CLUSTERING AND GRADIENT DESCENT APPROACH FOR EXAMPLE 2

	Premise Part	Consequent Part
m	$(c_m(1), c_m(2), \delta_m)$	$c_m(3)$
1	(-1.1942,-0.0713,0.4194)	0.0256
2	(-0.8568,0.5560,0.0331)	-0.1903
3	(-0.6927,1.3602,0.1940)	0.0353
4	(-0.4700,0.4101,0.2006)	-1.2667
5	(0.1269,-0.2405,0.5255)	-0.0231
6	(0.2926,1.3169,0.2361)	-0.0137
7	(-0.2295,0.5107,0.0046)	-0.3800
8	(0.0037,0.6533,0.0244)	0.0093
9	(0.2313,0.5590,0.0381)	0.3991
10	(0.4656,0.4245,0.2033)	1.2331
11	(1.1275,0.3253,0.1108)	-0.0557
12	(0.9244,1.0432,0.0521)	-0.0037
13	(1.1901,0.6156,0.1443)	-0.0431

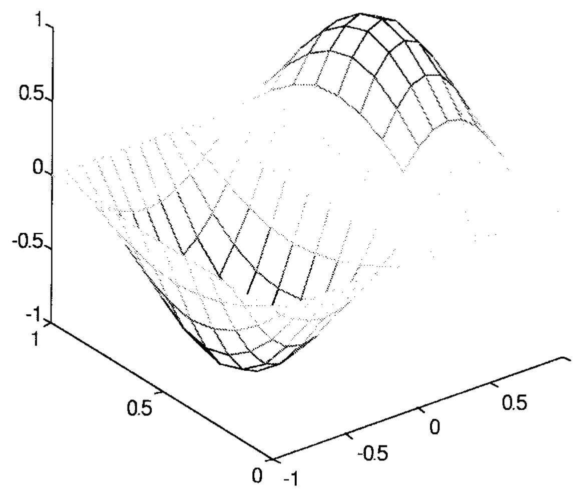

Fig. 9. Original input–output data set for Example 2.

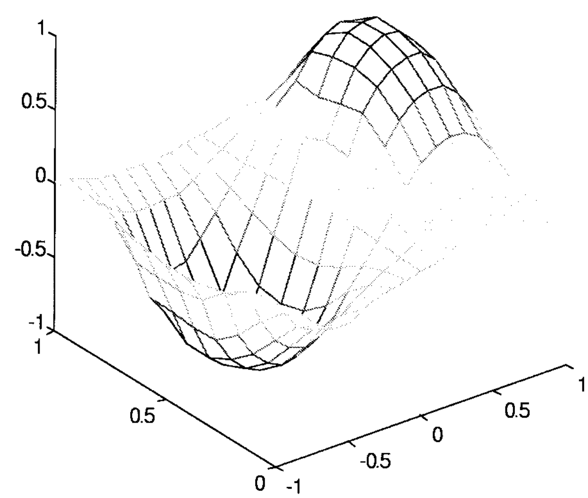

Fig. 10. Output of the final fuzzy model determined by the proposed hybrid clustering and gradient descent approach for Example 2.

Example 2: Consider a set of three-dimensional data generated by the nonlinear system [1]

$$x_3 = \sin(\pi x_1) \cdot \sin(\pi x_2) \tag{18}$$

with x_1 in the interval $[-1, 1]$ and x_2 in the interval $[0, 1]$. From the evenly distributed grid points of the input range $[-1, 1] \times [0, 1]$ to the nonlinear system, $21 \times 11 = 231$ input–output data points are obtained. In the structure identification step, 13 cluster centers are determined by the proposed clustering algorithm. Consequently, we can construct a rule-base of the initial fuzzy model associated with the 13 clusters as follows:

$$\text{Rule } m\text{: If } \mathbf{x} \text{ is } A_m \text{ then } y \text{ is } y_m. \quad m = 1, 2, \ldots, 13 \tag{19}$$

where $\mathbf{x} = (x_1, x_2)$ and $y = x_3$ are the input and output variables, respectively. The parameter values of the premise and consequent fuzzy sets are listed in Table IV. In the parameter identification step, we choose the values of cluster centers obtained by the clustering algorithm as initial values of $c_m(0,1), c_m(0,2)$, $\delta_m(0)$, and $c_m(0,3)$ and a learning rate $\alpha = 0.03$ for the parameter tuning procedure. The obtained final parameter values of the fuzzy model are listed in Table V. The outputs of the original nonlinear system and the fuzzy model described by Table V are plotted in Figs. 9 and 10, respectively. It is clear that the fuzzy model determined by the proposed approach has a very good approximation.

From the above results of two examples, we can illustrate that the proposed approach is effective in fuzzy modeling.

V. Conclusion

In this paper, we proposed a hybrid clustering and gradient descent approach to build a multi-input fuzzy model where only the input–output data of the identified system are available. The proposed clustering algorithm can automatically generate an initial fuzzy model, in which multidimensional membership functions and real values are respectively used to describe the premise and consequent fuzzy sets. Then, the gradient descent method is used to tune some adjustable parameters such that the obtained fuzzy model has a higher accuracy. The key idea is to generate an initial fuzzy model from the information of the number of clusters and the corresponding cluster obtained by a proposed clustering algorithm. The proposed clustering algorithm does not need any prior assumption about the data. It needs not to determine the appropriate cluster centers in the initialization step. The locations of the clusters can be efficiently determined and the data can be appropriately classified by the proposed clustering method. Therefore, it can be efficiently applied in fuzzy modeling to describe the behavior of the identified system. Obviously, the width (σ) of the Gaussian function in (1) plays an important role in the proposed clustering algorithm, because the specified value of σ will affect the number of clusters in the data analysis and the rule number in the fuzzy modeling. Generally speaking, the optimal value for the coefficient is application dependent and requires some experimentation by trial-and-error. But, in the topic of fuzzy modeling, the appropriate value of σ can be easily determined by the required performance index (e.g., mean squared error). From the simulation results and the comparison, it is clear that the proposed clustering method is better than the FCM clustering algorithm and the proposed hybrid method is useful for fuzzy modeling.

References

[1] Y. J. Chen and C. C. Teng, "Rule combination in a fuzzy neural network," *Fuzzy Sets Syst.*, vol. 82, pp. 161–166, 1996.

[2] M. Delgado, A. F. Gomez-Skarmeta, and F. Martin, "A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling," *IEEE Trans. Fuzzy Syst.*, vol. 5, pp. 223–232, Apr. 1997.

[3] H. Ishibuchi, R. Fujioka, and H. Tanaka, "Neural networks that learn from fuzzy if-then rules," *IEEE Trans. Fuzzy Syst.*, vol. 5, May 1997.

[4] J. S. Jang, C. T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*. Englewood Cliffs, NJ: Prentice Hall, 1997.

[5] Y. H. Joo, H. S. Hwang, K. B. Kim, and K. B. Woo, "Fuzzy system modeling by fuzzy partition and GA hybrid schemes," *Fuzzy Sets Syst.*, vol. 86, pp. 279–288, 1997.

[6] E. Kim, M. Park, S. Ji, and M. Park, "A new approach to fuzzy modeling," *IEEE Trans. Fuzzy Syst.*, vol. 5, pp. 328–337, June 1997.

[7] B. Kosko, *Fuzzy Engineer*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[8] A. Kroll, "Identification of functional fuzzy models using multidimensional reference fuzzy sets," *Fuzzy Sets Syst.*, vol. 80, pp. 279–288, 1996.

[9] M. Sugeno and G. T. Kang, "Structure identification of fuzzy model," *Fuzzy Sets Syst.*, vol. 28, pp. 15–33, 1988.

[10] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7–31, 1993.

[11] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 6, pp. 1414–1427, 1992.

[12] L. X. Wang, *A Course in Fuzzy Systems and Control*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[13] C. C. Wong and N. S. Lin, "Rule extraction for fuzzy modeling," *Fuzzy Sets Syst.*, vol. 88, pp. 23–30, 1997.

[14] R. R. Yager and D. P. Filev, *Essentials of Fuzzy Modeling and Control*. New York: Wiley, 1994.

[15] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithm*. New York: Plenum, 1981.

[16] J. M. Barone, D. P. Filev, and R. R. Yager, "Mountain method based fuzzy clustering: Methodological considerations," *Int. J. Gen. Syst.*, vol. 23, no. 4, pp. 281–305, 1995.

[17] R. P. Velthuizen, L. O. Hall, L. P. Clake, and M. L. Silbiger, "An investigation of mountain method clustering for large data sets," *Pattern Recognit.*, vol. 30, no. 7, pp. 1121–1136, 1997.

[18] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, pp. 1279–1284, Aug. 1994.

Ching-Chang Wong received the B.S. degree in electronic engineering from Tamkang University, Taipei, Taiwan, R.O.C., in 1984. He received the M.S. and Ph.D. degrees in electrical engineering from the Tatung Institute of Technology, Taipei, in 1986 and 1989, respectively.

He is currently a Professor with the Department of Electrical Engineering, Tamkang University. His research interests include fuzzy systems, grey systems, genetic algorithms, and intelligent control.

Chia-Chong Chen received the B.S. and M.S. degrees in electrical engineering from Tamkang university, Taipei, Taiwan, R.O.C., in 1995 and 1997, respectively. He is currently pursuing the Ph.D. degree in electrical engineering at Tamkang University.

His current research interests include fuzzy systems, clustering algorithm and genetic algorithms.