Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

A New Cluster Validity Criterion for Fuzzy C-Regression Models Clustering and Its Application to Fuzzy Model Identification

Chung-Chun Kung*. Jui-Yiao Su**

** Department of Electrical Engineering, Tatung University, Taipei, Taiwan, R.O.C. (Tel: +886-2-25925252-3470-100; e-mail: cckung@ttu.edu.tw).*
*** Department of Electrical Engineering, Tatung University, Taipei, Taiwan, R.O.C. (e-mail: camussu@gmail.com).*

Abstract: In this paper, a new cluster validity criterion for fuzzy c-regression models (FCRM) clustering algorithm with affine linear functional cluster representatives is proposed. The proposed cluster validity criterion calculates the overall compactness and separateness of the FCRM partition and then determines the appropriate number of clusters. Besides, its application to fuzzy model identification is discussed. A T-S fuzzy model identification algorithm is proposed to extract compact number of IF-THEN rules from data. Two simulation examples are provided to demonstrate the potential of the proposed cluster validity criterion and the accuracy of the constructed T-S fuzzy model.

1. INTRODUCTION

Fuzzy clustering algorithms, which are able to find out clusters from mixed data, provide systematic procedures to partition data space and therefore extract rules. In 1993, Hathaway and Bezdek proposed the fuzzy c-regression models (FCRM) clustering algorithm (see Hathaway and Bezdek, 1993) to fit switching regression models for certain types of mixed data. Instead of assuming that a single model accounts for all data pairs, the FCRM assumes that the given data are drawn from c different regression models or hyper-plane-shaped clusters. The measure of goodness is based on the fitness of the input-output data to these regression models. Minimization of the objective function in the FCRM clustering algorithm yields simultaneous estimates for the parameters of regression models together with a fuzzy c-partition of the data.

Recently, Kim *et al.* successfully applied the FCRM clustering algorithm to extract T-S fuzzy models (e.g. Takagi and Sugeno, 1985) from given data (see Kim *et al.,* 1997). Each regression model is essentially a prototype that describes a local characteristic behaviour of the unknown system and the number of clusters is just the number of fuzzy rules. However, for an unknown system, the appropriate number of clusters (rules) is supposed to be unknown by users (see Chuang *et al.,* 2001). The number of fuzzy rules is an important factor that affects the performance of a fuzzy model. While too many redundant rules result in a complex fuzzy model and increase implement difficulties, too few rules produce a less powerful one that may be insufficient to achieve the objective. In Kim's approach, the number of clusters (rules), c, is increased and the fine-tuning procedures are repeated until the model performance is checked and acceptable.

A related important issue to the fuzzy clustering algorithms is the cluster validity criterion, which deals with the significance of the structure imposed by a fuzzy clustering algorithm. There are many cluster validity criteria available, including Bezdek's partition coefficient, partition entropy (see e.g. Bezdek, 1974, 1981; Pal and Bezdek, 1995), and Xie-Beni index (Xie and Beni, 1991) etc. But all of them are not designed for the FCRM clustering with hyper-plane-shaped cluster representatives.

In this paper we adopt the compactness-to-separation ratio concept in Xie-Beni index and design a new cluster validity criterion for the FCRM clustering algorithm with affine linear functional cluster representatives. The numerator of the new cluster validity criterion combines the average flatness index (Babuska, 1998) with the objective function in FCRM to reflect *the compactness validity function* of the entire partition. The denominator of it defines *the separation validity function* as the "shift" from origin in y-axis and the absolute value of standard inner-product of unit normal vectors representing different hyper-planes. Therefore, we can judge the difference between regression models or hyper-plane-shaped clusters.

While applying the new cluster validity criterion to determine the appropriate number of needed clusters for FCRM, we improve Kim's fuzzy modelling approach (Kim *et al.,* 1997) to construct a T-S fuzzy model with compact number of rules.

The framework of this paper is organized as follows. In section 2, we briefly review the FCRM clustering algorithm with affine linear functional cluster representatives and propose a new validity criterion for it. In section 3, a new T-S fuzzy model identification algorithm is presented. In section 4, a numerical example is given to illustrate the potential of the proposed validity criterion and another example is given to illustrate the accuracy and effectiveness of the proposed fuzzy model identification algorithm. Conclusions are stated in Section 5.

2. REVIEW OF CLUSTER ANALYSIS AND DESIGN OF NEW CLUSTER VALIDITY CRITERION

2.1 Review of Cluster Analysis

Let $S=\{(\mathbf{x}_1,y_1),\cdots,(\mathbf{x}_N,y_N)\}=\{(\mathbf{x}_h,y_h)\,|\,h=1,\cdots,N\}$ be a set of N input-output data to be clustered, each independent input vector $\mathbf{x}_h=[x_{1h},\cdots,x_{nh}]^T\in\mathbf{X}_1\times\mathbf{X}_2\times\cdots\times\mathbf{X}_n\subset\Re^n$ has a corresponding dependent output $y_h\in\mathbf{Y}\subset\Re$, where $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n$ are the domains of the input variables and $\mathbf{Y}$ denotes the domain of the output. The FCRM clustering algorithm assumes that the given input-output data are drawn from c different affine linear regression models:

$$\begin{aligned} y_h &= f^i(\mathbf{x}_h,\boldsymbol{\theta}_i) \\ &= \theta_{i1}x_{h1}+\theta_{i2}x_{h2}+\cdots+\theta_{in}x_{hn}+\theta_{i0} \\ &= [\mathbf{x}_h{}^T\ 1]\boldsymbol{\theta}_i, \qquad i=1,\cdots,c, \end{aligned} \tag{1}$$

where $\boldsymbol{\theta}_i=[\theta_{i1},\theta_{i2},\cdots\theta_{in},\theta_{i0}]^T\in\Re^{n+1}$, θ_{i0} is a constant that represents the bias or offset term. The parameter vectors $\boldsymbol{\theta}_i$ are needed to be determined. Label vectors assigned to each data pair can be arrayed as a $(c\times N)$ fuzzy c-partition matrix, $U=[\mu_{ih}]$, in which μ_{ih} is regarded as the membership of each input-output data pair $(\mathbf{x}_h,y_h)$ belonging to the ith fuzzy cluster. All μ_{ih} are constrained labels (Ruspini, 1970):

$$\begin{aligned} &0\le\mu_{ih}\le 1, \quad \forall\, i,h \\ &\sum_{i=1}^{c}\mu_{ih}=1, \quad h=1,\cdots,N \\ &0<\sum_{h=1}^{N}\mu_{ih}<N, \quad i=1,\cdots,c. \end{aligned} \tag{2}$$

The distance (measure of fitness) from every sampled $(\mathbf{x}_h,y_h)$ to the ith affine linear regression model with parameter $\boldsymbol{\theta}_i$ are defined as follows

$$d_{ih}(\boldsymbol{\theta}_i)=\left|f^i(\mathbf{x}_h,\boldsymbol{\theta}_i)-y_h\right|=\left|[\mathbf{x}_h^T\ 1]\boldsymbol{\theta}_i-y_h\right|. \tag{3}$$

The objective function in FCRM clustering algorithm is then defined as follows:

$$J_m(U,\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_c)=\sum_{h=1}^{N}\sum_{i=1}^{c}(\mu_{ih})^m d_{ih}^2(\boldsymbol{\theta}_i) \tag{4}$$

where $m\in(1,\infty)$ is the weighting exponent. Minimization of the objective function (4) yields a fuzzy c-partition of the data, together with estimates for the c parameter vectors simultaneously.

The FCRM clustering algorithm is executed in the following steps (see Hathaway and Bezdek, 1993):

Step 1 Assign the number of the clusters c. Set the weighting exponent $m>1$. Pick a termination threshold $\varepsilon>0$ and an initial partition $U^{(0)}=[\mu_{ih}^{(0)}]$. Set iteration index $r=0$.

Step 2 At each rth iteration, calculate c parameter vectors $\boldsymbol{\theta}_i$ that minimize the objective function (4) by using the weighted least square (WLS) algorithm (e.g. Ljung and Soderstrom, 1983) to $(\mathbf{x}_h,y_h)$, $h=1,\cdots,N$.

Step 3 Update $U^{(r)}$ to $U^{(r+1)}$ as follows:

$$\mu_{ih}^{(r+1)}=\begin{cases} 1\Big/\displaystyle\sum_{j=1}^{c}\left(\frac{d_{ih}(\boldsymbol{\theta}_i)}{d_{jh}(\boldsymbol{\theta}_j)}\right)^{\frac{2}{m-1}} & \text{for } I_h=\varnothing \\ \dfrac{1}{n_h} & \text{for } I_h\ne\varnothing,\ i\in I_h \\ 0 & \text{for } I_h\ne\varnothing,\ i\notin I_h \end{cases} \tag{5}$$

where $I_h=\{i\,|\,1\le i\le c,\ \ d_{ih}(\boldsymbol{\theta}_i)=0\}$ and n_h is the number of elements in I_h.

Step 4 Check for termination in convenient induced matrix norms:

If $\left\|U^{(r)}-U^{(r+1)}\right\|\le\varepsilon$, stop;

otherwise, set $r=r+1$ and return to Step 2.

2.2 Design of New Cluster Validity Criterion

For fuzzy c-means (FCM) clustering algorithm (e.g. Hoppner *et al.*, 1999), one commonly used cluster validity criterion called the Xie-Beni index (see Xie and Beni, 1991) is designed on the concept of compactness-to-separation ratio. The numerator of Xie-Beni index is a compactness validity function that fits the objective function of the FCM and reflects the compactness of clusters. The denominator is a separation validity function that measures the separation status of clusters. The smaller the separation validity function value is, the more probability there will be redundant cluster representative in the existed representatives. We adopt the compactness-to-separation ratio concept and propose a new cluster validity criterion for FCRM with affine linear functional cluster representatives.

1) The Compactness Validity Function: Define the fuzzy covariance matrix of the ith cluster as follows:

$$\mathbf{F}_i=\sum_{h=1}^{N}(\mu_{ih})^m(\mathbf{z}_h-\mathbf{v}_i)(\mathbf{z}_h-\mathbf{v}_i)^T\Big/\sum_{h=1}^{N}(\mu_{ih})^m; \quad 1\le i\le c, 1\le h\le N, \tag{6}$$

where $\mathbf{z}_h=[\mathbf{x}_h^T,y_h]^T=[x_{1h},\cdots,x_{nh},y_h]^T\in\Re^{n+1}$ is the observation consisting of the hth sampled input-output data. $\mathbf{v}_i$ is the *centers* of the ith cluster calculated by

$$\mathbf{v}_i=\sum_{h=1}^{N}(\mu_{ih})^m\mathbf{z}_h\Big/\sum_{h=1}^{N}(\mu_{ih})^m; \quad 1\le i\le c, 1\le h\le N \tag{7}$$

Defined the *flatness index* as the ratio between the smallest and the largest eigenvalue of $\mathbf{F}_i$ (see Babuska and Verbruggen, 1995):

$$t_i=\lambda_{i-\min}/\lambda_{i-\max} \tag{8}$$

where $\lambda_{i-\min}$ is the smallest eigenvalue of $\mathbf{F}_i$ and $\lambda_{i-\max}$ is the largest one, respectively. The flatness index has low values for clusters which are large and flat. For the entire partition, the *average flatness index* is measured by (see Babuska, 1998):

$$t_A = \frac{1}{c}\sum_{i=1}^{c}(\lambda_{i-\min}/\lambda_{i-\max}) \quad (9)$$

When data describe a functional relationship, the clusters are usually flat (Babuska, 1998). We thus combine the average flatness index (9) with the objective function in (4) to obtain the *compactness validity function*, f_{com} :

$$\begin{aligned} f_{com} &\equiv t_A \cdot J_m(U, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_c) \\ &= \frac{1}{c}\sum_{i=1}^{c}\frac{\lambda_{i-\min}}{\lambda_{i-\max}} \cdot \frac{1}{N}\sum_{h=1}^{N}\sum_{i=1}^{c}(\mu_{ih})^m([\mathbf{x}_h^T\ 1]\boldsymbol{\theta}_i - y_h)^2 \quad (10) \\ &= \frac{1}{cN}\sum_{h=1}^{N}\sum_{i=1}^{c}(\mu_{ih})^m t_i([\mathbf{x}_h^T\ 1]\boldsymbol{\theta}_i - y_h)^2 \end{aligned}$$

which prefers a few flat clusters to a large number of small ones; if both settings lead to approximately the same objective function value.

2) The Separation Validity Function: The affine linear regression model defined in (1) can be regarded as a shift of linear hyper-plane $y_h = \theta_{i1}x_{h1} + \theta_{i2}x_{h2} + \cdots + \theta_{in}x_{hn}$ by scale θ_{i0} in $y-axis$. By removing θ_{i0}, we have new linear hyper-planes $y_h = \theta_{i1}x_{h1} + \theta_{i2}x_{h2} + \cdots + \theta_{in}x_{hn}$ all pass through the origin $\mathbf{0}$. We rewrite these linear regression models as follows:

$$\boldsymbol{\zeta}^T\mathbf{n}_i = 0 \quad (10)$$

where $\boldsymbol{\zeta} = [x_{h1}, \cdots, x_{hn}, y_h]^T \in \Re^{n+1}$ is a varying vector on the ith linear hyper-plane and $\mathbf{n}_i = [\boldsymbol{\theta}_i^T, -1]^T \in \Re^{n+1}$ represents the normal vector of it. The corresponding *unit normal vector* of each hyper-plane in (10) can then be defined as

$$\mathbf{u}_i = \mathbf{n}_i / \|\mathbf{n}_i\| \quad (11)$$

where $\|\cdot\|$ denotes the Euclidean norm.

Denote the Euclidean inner product of $\mathbf{u}_i$ and $\mathbf{u}_j$ as $<\mathbf{u}_i, \mathbf{u}_j>$, then $|<\mathbf{u}_i, \mathbf{u}_j>|$ means the projection length (see Friedberg *et al.,* 1989) of $\mathbf{u}_i$ on $\mathbf{u}_j$. Since the only factor that influences the projection length is the angle between $\mathbf{u}_i$ and $\mathbf{u}_j$, we thus use $|<\mathbf{u}_i, \mathbf{u}_j>|$ to measure the angle between two regression models. $|<\mathbf{u}_i, \mathbf{u}_j>| = 0$ implies their linear hyper-planes are orthogonal, while $|<\mathbf{u}_i, \mathbf{u}_j>| = 1$ implies the coincidence of them.

The "shift term" between two affine linear regression models is judged by γ_{ij} as follows:

$$\gamma_{ij} = |\theta_{i0} - \theta_{j0}| / \Delta\gamma_{\max}, \text{ for } i, j = 1, \cdots, c;\ i \neq j, \quad (12)$$

where $\Delta\gamma_{\max} = \max_{\substack{i=1,\cdots,c \\ j=1,\cdots,c}} |\theta_{i0} - \theta_{j0}|$. It is noticed that γ_{ij} has been normalized, i.e., $\gamma_{ij} \in [0,\ 1]$.

Accordingly, the *separation validity function* for affine linear regression models is then designed as follows:

$$f_{sep} \equiv \min_{\substack{i \neq j \\ i=1,\cdots,c \\ j=1,\cdots,c}} \frac{\gamma_{ij} + k_2}{|<\mathbf{u}_i, \mathbf{u}_j>| + k_1} \quad (13)$$

where k_1 k_2 are rather small real positive constants that prevents the function from being divided by zero or being zero. Obviously, f_{sep} in (13) also fits the concept of separation measure criterion: the smaller the separation validity function value is, the more probability there will be a redundant cluster in that one cluster is quite similar to another one.

3) The New Cluster Validity Criterion: The proposed new cluster validity criterion is defined by the compactness-to-separation ratio as follows:

$$F_{NEW} \equiv \frac{f_{com}}{f_{sep}} = \frac{\sum_{h=1}^{N}\sum_{i=1}^{c}(\mu_{ih})^m t_i([\mathbf{x}_h^T\ 1]\boldsymbol{\theta}_i - y_h)^2}{cN\left(\min_{i \neq j}\frac{\gamma_{ij} + k_2}{|<\mathbf{u}_i, \mathbf{u}_j>| + k_1}\right)} \quad (14)$$

The optimal number c is chosen when F_{NEW} reaches its minimum. In practice, the appropriate number c is chosen at which the first local minimum of F_{NEW} has occurred; moreover, when the cluster validity index decreases monotonically, we can choose c at which a significant change in its curvature has occurred (see Babuska, 1998; Xie and Beni, 1991).

3. THE T-S FUZZY MODEL IDENTIFICATION ALGORITHM

The T-S fuzzy model discussed in this paper is of the following form (e.g. Takagi and Sugeno, 1985; Wang, 1997):

$$\begin{aligned} R^i: &\quad \text{IF } x_1 \text{ is } A_1^i \text{ and} \cdots \text{and } x_n \text{ is } A_n^i \\ &\quad \text{THEN } y^i = \theta_{i1}x_1 + \theta_{i2}x_2 + \cdots + \theta_{in}x_n + \theta_{i0} \end{aligned} \quad (15)$$

where $i = 1, 2, \cdots, c$, R^i denotes the ith IF-THEN rule and c is the numbers of rules in the rule base. x_q, $q = 1, \cdots, n$, are individual input variables, and A_q^i are bell-typed fuzzy sets with mean and standard deviation, i.e.,

$$A_q^i(z) = \exp\left\{-\frac{1}{2}(z - \alpha_q^i)^2 / (\beta_q^i)^2\right\}. \quad (16)$$

$y^i \in \Re$ is the output of each rule. θ_{ik}, $k = 1, \cdots, n$, are parameters of the linear function and θ_{i0} denotes a scalar offset. Given $\mathbf{x} = [x_1, \cdots, x_n]^T$, if the method of singleton fuzzifier, product fuzzy inference, and center average defuzzifier (see e.g. Wang, 1997) is employed, the output of the T-S fuzzy model $\hat{y}$ is inferred as follows:

$$\hat{y} = \sum_{i=1}^{c}\phi^i y^i \quad (17)$$

where

$$\phi^i = w^i(\mathbf{x}) \Big/ \sum_{i=1}^{c} w^i(\mathbf{x}), \tag{18}$$

$$w^i(\mathbf{x}) = A_1^i(x_1) \times \cdots \times A_n^i(x_n) = \prod_{q=1}^{n} A_q^i(x_q). \tag{19}$$

The procedure of our T-S fuzzy model identification algorithm is outlined in the following steps:

T-S fuzzy model identification algorithm

Step 1 Get experimental input-output data $(\mathbf{x}_h, y_h)$, $h = 1, \cdots, N$, from the unknown system. Choose the initial number of clusters $c = c_{MIN}$.

Step 2 Apply the FCRM clustering algorithm to partition the product space of the given input-output data into c linear functional clusters.

Step 3 Set $c = c + 1$ and repeat Step 2 to Step 3 until $c = c_{MAX}$, the termination number of clusters.

Step 4 Cluster validation: Use the proposed new cluster validity criterion F_{NEW} in (14) to determine the appropriate number of needed clusters

Step 5 Construct the prototypes of fuzzy rules: the parameter estimations of α_q^i and β_q^i can be roughly obtained from the fuzzy partitions matrix U by the *axis-orthogonal projection method* (see e.g. Babuska, 1998):

$$\alpha_q^i = \sum_{h=1}^{N} \mu_{ih} x_{qh} \Big/ \sum_{h=1}^{N} \mu_{ih} \tag{20}$$

$$\beta_q^i = \sqrt{\sum_{h=1}^{N} \mu_{ih}(x_{qh} - \alpha_q^i)^2 \Big/ \sum_{h=1}^{N} \mu_{ih}}\,. \tag{21}$$

The parameters $\boldsymbol{\theta}_i = [\theta_{i1}, \theta_{i2}, \cdots \theta_{in}, \theta_{i0}]^T$ can inherit from the affine linear functional cluster representatives in FCRM.

Step 6 Fine-tuning of the parameters: define a cost function $J = \frac{1}{2}(y(k) - \hat{y}(k))^2$. By the gradient descent method (e.g. Wang, 1994, 1997), the antecedent and consequent parameters in the T-S fuzzy model can be finely tuned to minimize J by the following equations:

$$\alpha_q^i(k+1) = \alpha_q^i(k) + \Delta\alpha_q^i(k) \tag{22}$$

$$\beta_q^i(k+1) = \beta_q^i(k) + \Delta\beta_q^i(k) \tag{23}$$

$$\theta_q^i(k+1) = \theta_q^i(h) + \Delta\theta_q^i(k) \tag{24}$$

$$\theta_0^i(k+1) = \theta_0^i(k) + \Delta\theta_0^i(k) \tag{25}$$

where $\Delta\alpha_q^i(k)$, $\Delta\beta_q^i(k)$, and $\Delta\theta_q^i(k)$ denote the adjustments at each learning step k as follows (we drop the argument k for brevity):

$$\Delta\alpha_q^i = 2\eta_1(y - \hat{y})(y^i - \hat{y})\phi^i \left(\frac{x_q - \alpha_q^i}{{\beta_q^i}^2} \right) \tag{26}$$

$$\Delta\beta_q^i = 2\eta_2(y - \hat{y})(y^i - \hat{y})\phi^i \left(\frac{x_q - \alpha_q^i}{{\beta_q^i}^2} \right)^2 \tag{27}$$

$$\Delta\theta_q^i = \eta_3(y - \hat{y})\phi^i x_q \tag{28}$$

$$\Delta\theta_0^i = \eta_3(y - \hat{y})\phi^i\,. \tag{29}$$

Where η_1, η_2, and η_3 are positive real-valued constants denoting the step-size.

The deduction of the above fine-tuning laws (22)-(29) are introduced in Wang (1994, 1997). Pick two termination thresholds $\varepsilon_{\alpha\beta} > 0$ and $\varepsilon_a > 0$, we can apply the fine-tuning laws (22)-(23) recursively until the termination conditions, $\max_{\substack{i=1,\cdots c\\ q=1,\cdots n}} \left\{ \left|\Delta\alpha_q^i\right|, \left|\Delta\beta_q^i\right| \right\} < \varepsilon_{\alpha\beta}$ and $\max_{\substack{i=1,\cdots c\\ q=0,\cdots n}} \left\{ \left|\Delta\theta_q^i\right| \right\} < \varepsilon_a$, are satisfied.

4. SIMULATIONS

4.1 Example 1: Mixed data classification

To validate the new cluster validity criterion F_{NEW}, we consider the following example and compare the result with Bezdek's partition coefficient v_{PC}:

$$v_{PC} = \sum_{h=1}^{N} \sum_{i=1}^{c} (\mu_{ih})^2 \Big/ N \tag{30}$$

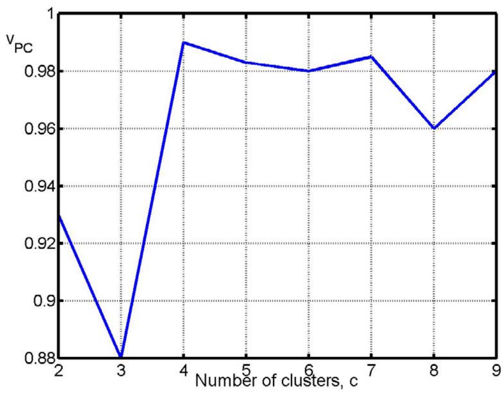
The appropriate number c is chosen when largest v_{PC} appears.

Given a mix of four linear equations with exogenous white Gaussian random noise ε_i $(i = 1, 2, 3, 4)$ having zero mean and variance 0.25:

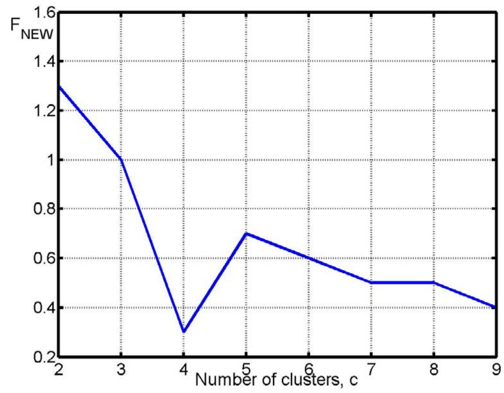
$$\begin{aligned} y &= [\mathbf{x}^T 1]\boldsymbol{\theta}_1^+ + \varepsilon_1 = 2x_1 - 3x_2 + 4x_3 - 4 + \varepsilon_1, \\ y &= [\mathbf{x}^T 1]\boldsymbol{\theta}_2^+ + \varepsilon_2 = 2x_1 - 3x_2 + 4x_3 + 14 + \varepsilon_2, \\ y &= [\mathbf{x}^T 1]\boldsymbol{\theta}_3^+ + \varepsilon_3 = -x_1 + 1x_2 + 2x_3 - 3 + \varepsilon_3, \\ y &= [\mathbf{x}^T 1]\boldsymbol{\theta}_4^+ + \varepsilon_4 = -3x_1 + 5x_2 + x_3 + 10 + \varepsilon_4 \end{aligned} \tag{31}$$

we randomly generate 800 training input vector $\mathbf{x}$ with each element uniformly distributed in the range [-5,5] and then apply each 200 of them to the four linear equations correspondingly. By the FCRM algorithm, we obtain cluster representatives and partition matrix U for different c. we set $k_1 = 0.001$ and $k_2 = 0.001$ in (14) and consider the following two cases:

Case 1: $m = 2$. The plot of cluster index vs. cluster number is depicted in Fig. 1. We see that both the Bezdek's partition coefficient and the new cluster index indicate the correct answer $c = 4$. The affine linear regression models obtained by the FCRM algorithm are listed below:



(a)



(b)

Fig. 1. The plot of cluster index vs. cluster number in *Example 1 Case 1*.

$$\begin{aligned} y = [\mathbf{x}^T 1]\boldsymbol{\theta}_1 &= \ \ 2.0005x_1 - 2.9972x_2 + 3.9981x_3 - 4.0039, \\ y = [\mathbf{x}^T 1]\boldsymbol{\theta}_2 &= \ \ 2.0013x_1 - 3.0012x_2 + 3.9976x_3 + 13.9941, \\ y = [\mathbf{x}^T 1]\boldsymbol{\theta}_3 &= -0.9963x_1 + 0.9979x_2 + 2.0022x_3 - 3.0111, \\ y = [\mathbf{x}^T 1]\boldsymbol{\theta}_4 &= -3.0019x_1 + 5.0016x_2 + 1.0026x_3 + 9.9990, \end{aligned}$$

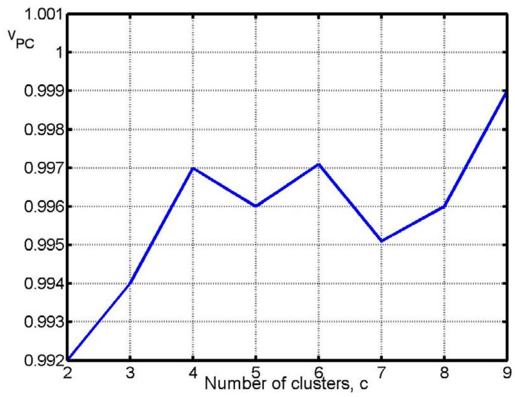
which are quite close to the nominal linear equations.

Case 2: $m = 1.05$. The partition result is close to the *hard c-partition.* (Bezdek, 1981). The plot of cluster index vs. cluster number is depicted in Fig. 2. We see that F_{NEW} can find the correct number of clusters, 4, but v_{PC} can't recognize 4 or 7 as the correct number of cluster. This numerical example illustrates that the proposed F_{NEW} can be applied for a wider range of m and is reliable to validate the FCRM partition.

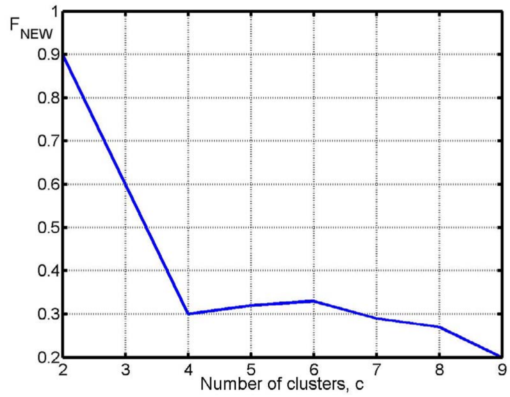
4.2 Example 2: Fuzzy modeling for nonlinear plant

Consider a nonlinear system described by the following second-order difference equation (Wang and Yen, 1999; Setnes and Roubos, 2000):

$$\begin{aligned} y(k+1) &= f(y(k), y(k-1)) + u(k) \\ &= \frac{y(k-1)y(k-2)(y(k-1)-0.5)}{1 + y^2(k-1) + y^2(k-2)} + u(k) \end{aligned} \tag{32}$$



(a)



(b)

Fig. 2. The plot of cluster index vs. cluster number in *Example 1 Case 2*.

Our objective is to build a T-S fuzzy model that can serve an approximation of $f(\bullet)$ in (32) with high accuracy and use as few IF-THEN rules as possible.

Choose $y(k)$ and $y(k-1)$ as the antecedent variables, then the fuzzy model is described as follows:

$$\begin{aligned} R^i: \quad & \text{IF } y(k) \text{ is } A_1^i \text{ and } y(k-1) \text{ is } A_2^i \\ & \text{THEN } y^i(k+1) = \theta_1^i y(k) + \theta_2^i y(k-1) + \theta_0^i \end{aligned} \tag{33}$$

where A_q^i are bell-shaped fuzzy sets with mean α_q^i and standard deviation β_q^i for $q = 1, 2$; $\boldsymbol{\theta}_i = [\theta_1^i, \theta_2^i, \theta_0^i]^T$ are the consequent parameter vectors. We choose a hybrid input signal with partly uniformly distributed white random signal and partly sinusoidal one as the training input signal (see e.g. Wang and Yen, 1999; Setnes and Roubos, 2000), i.e., $u(k)$ is a uniformly distributed random signal in the range $[-1, 1]$ for $1 \le k \le 200$, and $u(k) = \sin(2\pi k/25)$ for $200 \le k \le 400$. The number of training data $N = 400$. The termination threshold in the FCRM is chosen as $\varepsilon = 0.0001$, and the learning step-size for η_1, η_2, and η_3 is set to be 0.005, 0.005, and 0.5, respectively. The plot of F_{NEW} vs. cluster number c is shown in Fig. 3 with $k_1 = k_2 = 0.001$. We find $c = 3$ provides a good choice for the number of clusters. The parameters of the antecedent and the consequent parts are listed in Table 1 and

Table 2, respectively. Define the mean square error as $\text{MSE} = \sum_{k=1}^{N} (y(k) - \hat{y}(k))^2 \Big/ N$. The comparative results in the literatures are listed in Table III. We see that the proposed identification method is capable of obtaining competent results using fewer rules than other approaches reported in the literatures.

5. CONCLUSIONS

In this paper, a new cluster validity criterion F_{NEW} designed for the FCRM clustering algorithm with affine linear functional cluster representatives is proposed. A modification of Kim's fuzzy modelling approach (Kim *et al.*, 1997) is proposed as well to construct a T-S fuzzy model with compact number of rules.A The simulation results illustrate that F_{NEW} is applied for a wider range of m and the T-S fuzzy model obtained by the proposed fuzzy model identification algorithm is able to well approximate the discrete-time nonlinear plant with satisfactory results.

ACKNOWLEDGMENTS

This work was supported by the National Science Council of Taiwan, Republic of China, under grant 96-2221-E-036-033-MY2.

REFERENCES

Babuska, R. and H. Verbruggen (1995). New approach to constructing fuzzy relational models from data. In: *Proceedings of 3rd European congress on Intelligent Techniques and soft Computing EUFIT'95,* Aachen, Germany, 583-587.

Babuska, R. (1998). *Fuzzy Modeling for Control*. Boston: Kluwer Academic Publishers.

Bezdek, J. C. (1974). Cluster validity with fuzzy set. *Journal of Cybernetic*, **3**, 58-72.

Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum.

Chuang, C. C., S. F. Su and S. S. Chen (2001). Robust TSK fuzzy modeling for function approximation with outliers. *IEEE Transactions on Fuzzy Systems*, **9**, 810-821.

Friedberg, S. H., A. J. Insel and L. E. Spence (1989). *Linear Algebra*. NJ: Prentice-Hall.

Hathaway, R. J. and J. C. Bezdek (1993). Switching regression models and fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, **1**, 195-204.

Hoppner, F., F. Klawonn, R. Kruse, and T. Runkler (1999). *Fuzzy cluster analysis, methods for classification, data analysis and image recognition.* John Wiley & Sons.

Kim, E., M. Park, S. Ji, and M. Park (1997) A new approach to fuzzy modeling. *IEEE Transactions on Fuzzy Systems*, **5**, 328-337.

Ljung, L. and T. Soderstrom (1983). *Theory and practice of recursive identification*. MIT Press.

Pal, N.R. and J.C. Bezdek (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy Systems*, **3**, 370-379.

Ruspini, E. (1970). Numerical method for fuzzy clustering. *Information Science,* **2**, 319-350.

Setnes, M. and H. Roubos (2000). GA-fuzzy modeling and classification: complexity and performance. *IEEE Transactions on Fuzzy Systems,* **8,** 509-522.

Takagi, T. and M. Sugeno (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man and Cybernetics*, **15**, 116-132.

Wang, L. X. (1994). *Adaptive Fuzzy Systems and Control: Design and Stability Analysis*. Prentice-Hall.

Wang, L. X. (1997). *A Course in Fuzzy Systems and Control*. New York: Prentice-Hall.

Wang, L. and J. Yen (1999) Extracting fuzzy rules for system modeling using a hybrid of genetic algorithms and Kalman filter. *Fuzzy Sets and Systems* , **101**, 353-362.

Xie, X. L. and G.A. Beni (1991). Validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **3**, 841-846.

Table 1 List of antecedent parameters in *Example 2*

	α_1^i	β_1^i	α_2^i	β_2^i
$i=1$	-0.9175	0.4732	-0.8998	0.3259
$i=2$	1.4970	0.7024	0.2014	0.5541
$i=3$	0.3879	0.3337	0.2867	0.7132

Table 2 List of consequent parameters in *Example 2*

	θ_1^i	θ_2^i	θ_0^i
$i=1$	0.4603	0.1938	0.1926
$i=2$	-0.0953	0.2673	0.0900
$i=3$	-0.0087	-0.0638	-0.0082

Table 3 Comparative results of Example 2

Ref.	No. of rules	No. of sets	Consequent	MSE
Setnes & Roubos	7	14 Triangular	Singleton	3.0e-3
	5	8 Triangular	Affine Linear	7.5e-4
	4	4 Triangular	Affine Linear	1.2e-3
Wang & Yen	28	40 Gauss	Singleton	3.3e-4
The proposed method	3	6 Gauss	Affine Linear	5.2e-4

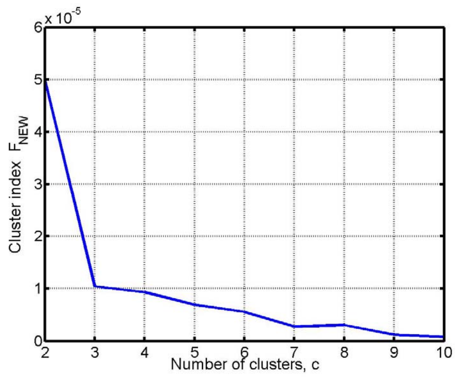


Fig. 3. The plot of F_{NEW} vs. cluster number c in *Example 2.*