EXTRACTING FUZZY SPARSE RULE BASE BY CARTESIAN REPRESENTATION AND CLUSTERING

Yeung Yam[1], Vladik Kreinovich[2], and Hung T. Nguyen[3]

[1]Department of Mechanical & Automation Engineering
The Chinese University of Hong Kong
Shatin, NT
Hong Kong, China
email yyam@mae.cuhk.edu.hk

[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
email vladik@cs.utep.edu

[3]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
email hunguyen@nmsu.edu

Abstract

Sparse rule base and interpolation have been proposed as possible solution to alleviate the geometric complexity problem of large fuzzy set. So far, however, there's no formal method available to extract sparse rule base. This paper combines the recently introduced Cartesian representation of membership functions and a mountain method-based clustering technique for extraction. A case study is included to demonstrate the effectiveness of the approach.

Keywords: Fuzzy sparse rule base, Cartesian representation, fuzzy clustering, mountain method

1. Introduction

Sparse rule base and interpolation have been proposed as possible solution to alleviate the geometric complexity problem associated with large rule set [1] [2]. By definition, a sparse rule base is one which has its antecedent supports covering only a subset of the input universe. In the case when an observation has no overlapping with any membership functions, interpolation technique is employed to extract a conclusion. The goal of the sparse rule base is hence to capture the most essential information embedded in the full set that it seeks to replicate using interpolation. So far, a number of results on interpolation methods have been published [2] [3] [4] [5]. More recently, a new approach to represent membership functions as points in Cartesian space is also proposed [6]. Under this representation, a rule set can be viewed as mappings between antecedent and consequent spaces, and the interpolation problem becomes finding an appropriate image in the consequent space for the antecedent observation. This new representation also allows regions of normal and abnormal membership functions to be readily characterized and incorporated in the analysis. On the other hand, relatively fewer results are available in the extraction of sparse rule base. No formal method for extraction is available so far. In [7], various issues and difficulties in generating sparse rule base are discussed. A case study using polynomial Lagrange interpolation is included to highlight the discussions.

Independently, fuzzy clustering has also attracted a lot of interest in recent years. The goal here is to determine a few points, so-called cluster centers, which would best represent a given set of points. More notably among the works in this area is the mountain method by Yager and Filev [8]. The method assigns a certain potential function for the given points and selects the cluster centers according to the magnitude of the overall potential. In a modified version, Chiu [9] introduced a formulation to identify the outputs models via least square error fitting of the given data. He also fuzzified the stopping rule of the original scheme.

This paper combines the Cartesian representation of [6] with the clustering technique of [9] as possible mean to extracting sparse fuzzy rule base from a given set. The former converts the given rule base to a set of points in Cartesian space, and the latter is applied to extract the cluster centers from these points. Same as [9], output models will be obtained via a least square error minimization process. Both the zero order and first order models of [9] will be considered. For ease of comparing to existing results, the example in [7] will be used as case study.

2. The Cartesian Representation

Let F be a membership function of fuzzy variable f comprised of n characteristic points. The work [6] proposed to represent F as a point $\vec{F}$ in a n-dimensional Cartesian space $\mathcal{R}^n$, with coordinates

$$\vec{F} = \begin{bmatrix} f^{(1)} \\ f^{(2)} \\ \vdots \\ f^{(n)} \end{bmatrix} \quad \in \quad \mathcal{R}^n \tag{1}$$

where $f^{(i)}$ is the value of f at which the ith characteristic point of the membership function F occurs. Under this representation, a fuzzy set of q rules: If $A_i \Rightarrow B_i$, $i = 1, \ldots, q$, where the membership functions A_i and B_i have n and m characteristic points, respectively, can be viewed as mappings between points A_i in the antecedent space $\mathcal{R}^n$ and B_i in the consequent space $\mathcal{R}^m$. This is depicted in Fig. 1, where the representing points A_1, A_2, …, and A_q in $\mathcal{R}^n$ are mapped to B_1, B_2, …, and B_q in $\mathcal{R}^m$. As such, the interpolation problem now amounts to determining an image B^* in $\mathcal{R}^m$ for the given observation A^*.

3. The Clustering Method

Given N data points $z_i = (x_i, y_i) \in \mathcal{R}^{n+m}$, $i = 1, \ldots, N$, where $x \in \mathcal{R}^n$ is the input and $y \in \mathcal{R}^m$ is the output. We focus on the method proposed by Yager and Filev [8] and modified by Chiu [9] to identify, say c, cluster centers from these N data points.

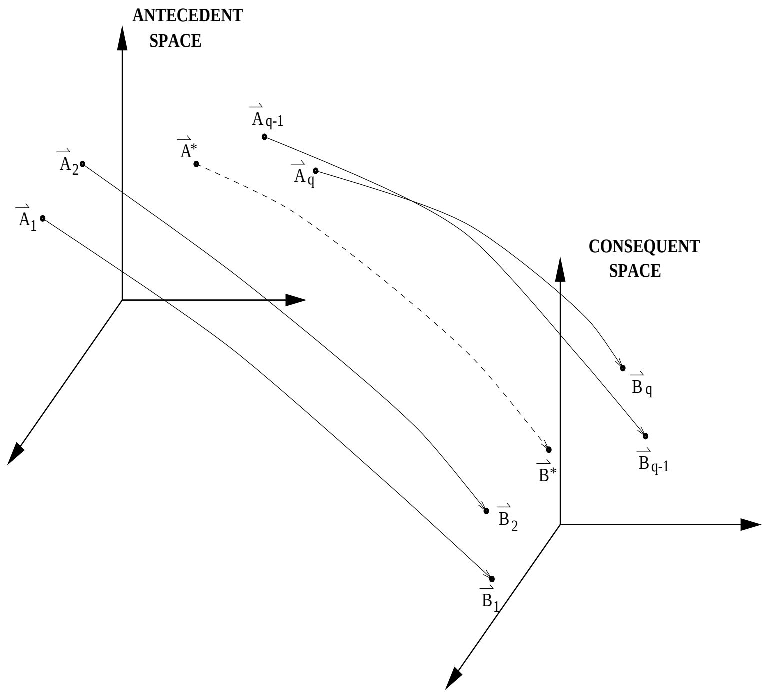

Figure 1: Fuzzy rules as mappings between antecedent and consequent spaces

3.1. Determination of cluster centers

A cluster center, by definition, should have many surrounding points. A natural approach is hence to introduce a numerical index to this effect, $P_1(z_i)$, which measures the potential ability of data point z_i as cluster center. The greater $P_1(z_i)$, the more likely that z_i is the cluster center. Implicit in $P_1(z_i)$ is the Euclidean distance from z_i to other points,

$$P_1(z_i) = \Phi(d(z_i, z_j), \text{ for } j = 1, 2, \ldots, N) \tag{2}$$

where $\Phi : (R^+)^{n+m} \to R^+$ such that Φ is decreasing in each argument. Yager and Filev [8] took

$$P_1(z_i) = \sum_{j=1}^{n} e^{-\alpha d(z_i, z_j)} \tag{3}$$

for some $\alpha > 0$. Chiu [9] took

$$P_1(z_i) = \sum_{j=1}^{n} e^{-\alpha d(z_i, z_j)^2} \tag{4}$$

Other function Φ can also be chosen. A discussion on the optimal choices of Φ can be found in [10]. The first center, say, z_1^*, can now be found as the one such that $P_1(z_1^*) = \max_{1 \le j \le n} P_1(z_j)$. Next, in order to find the second center, the effect of z_1^* is eliminated by revising the function $P_1(z_i)$ to become $P_2(z_i)$,

$$P_2(z_i) = P_1(z_i) - P_1(z_1^*) e^{\beta d(z_i, z_1^*)}, \tag{5}$$

or, in Chiu's case,

$$P_2(z_i) = P_1(z_i) - P_1(z_1^*)e^{\beta d(z_i, z_1^*)^2}, \tag{6}$$

where β is another chosen constant satisfying $\alpha > \beta > 0$. The revising process put zero potential on z_1^* and reduces potentials in the cluster around z_1^* more than far away points, thus somewhat equivalent to physically remove z_i^* and its cluster. Upon determining the second cluster center, the process can be repeated to yield other cluster centers.

3.2. The Stopping Rule

Yager and Filev proposed to stop the process at $k-1$ clusters if

$$\frac{P_k^*}{P_1^*} < \delta \tag{7}$$

where P_k^* is the maximum potential at step k, i.e., $P_k^* = P_k(z_k^*)$. The rationale is that should P_k^* constitute only a small given fraction, say, 15%, or $\delta = 0.15$, of the original maximum potential P_1^*, then there are only few data points remaining and the process can be stopped.

In Chiu's work, instead of a "crisp" stopping rule with threshold δ as in Yager and Filev's, a "fuzzified" rule with various degrees of stopping is adopted. Let two fractions $\bar{\epsilon}$ and $\underline{\epsilon}$, $\underline{\epsilon} < \bar{\epsilon}$, be given, Chiu's stopping rules are:

1. if $\frac{P_k^*}{P_1^*} > \bar{\epsilon}$, accept z_k^* as cluster center.
2. if $\frac{P_k^*}{P_1^*} < \underline{\epsilon}$, reject z_k^* and stop (resulting in $k-1$ centers).
3. if $\underline{\epsilon} < \frac{P_k^*}{P_1^*} > \bar{\epsilon}$, let $d_{min} = \min\limits_{1 \le i \le k-1} ||z_k^* - z_i^*||$, then
   (a) If $\frac{d_{min}}{r_a} + \frac{P_k^*}{P_1^*} \geq 1$: accept z_k^* as new center,
   (b) If $\frac{d_{min}}{r_a} + \frac{P_k^*}{P_1^*} < 1$: reject z_k^* and set $P_k(z_k^*) = 0$. Select the next maximum potential and re-test.

3.3. The Least Square Error Fitting

Let $z_1^*, \ldots, z_c^*$ be c the cluster centers identified above. Chiu utilized the formula

$$y(x) = \frac{\sum\limits_{j=1}^{c} \mu_j y_j^*}{\sum\limits_{j=1}^{c} \mu_j} \tag{8}$$

to generate the output y of any input x, where

$$\mu_j = e^{-\alpha d(x, x_j^*)^2}, \tag{9}$$

and x_i^* and y_i^* are the input and output at the cluster center $z_j^* = (x_j^*, y_j^*)$. Note that (9) utilizes the same α as (4). Equation (8) can be interpreted as the output of a product-sum-gravity-based inference system of c fuzzy rules: *If* $x_i^* \Rightarrow y_i^*$, $i = 1, \ldots, c$, with antecedent membership functions $A_i(x) = e^{-\gamma d(x, x_i^*)^2}$ and singleton output membership functions at y_i^*. In [6], (8) is also interpreted as an interpolation formula, with $A_i(x)$ being the extensibility function for the interpolated rule.

In addition, Chiu allowed the output for the ith rule, y_i^*, $i = 1, \ldots, c$, to be replaced by a linear function of the input variables,

$$y_i^* = G_i x + h_i \tag{10}$$

where G_i is an $m \times n$ constant matrix, and h_i a constant m-vector. Then, defining

$$\rho_i = \frac{\mu_i}{\sum_{j=1}^{c} \mu_j}, \tag{11}$$

(8) can be expressed as

$$y^T(x) = \begin{bmatrix} \rho_1 x^T & \rho_1 & \cdots & \rho_c x^T & \rho_c \end{bmatrix} \begin{bmatrix} G_1^T \\ h_1^T \\ \vdots \\ G_c^T \\ h_c^T \end{bmatrix}. \tag{12}$$

Specifically, we desire (12) to reproduce, closely, the outputs y_i for all input points x_i, $i = 1, \ldots, N$, i.e.,

$$\begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix} = \begin{bmatrix} \rho_{1,1} x_1^T & \rho_{1,1} & \cdots & \rho_{c,1} x_c^T & \rho_{c,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{1,N} x_N^T & \rho_{1,N} & \cdots & \rho_{c,N} x_c^T & \rho_{c,N} \end{bmatrix} \begin{bmatrix} G_1^T \\ h_1^T \\ \vdots \\ G_c^T \\ h_c^T \end{bmatrix}. \tag{13}$$

where $\rho_{i,j}$ denotes ρ_i evaluated at $x = x_j$ (see (9) and (11)). Equation (13) constitutes a least square estimation (LSE) problem whereby the quantities G_i and h_i can be determined. The model $y_i^* = G_i x + h_i$ employed in (10) is termed "first order model" by Chiu. Another type, with $y = h_i$ only, is the "zero order model", which leads to a simpler LSE equation than (13).

4. A Method To Extract Sparse Rule Base

The present work combines the Cartesian representation and clustering technique together as formal procedures to extract sparse rule based from a given set. Upon given a fuzzy rule base of: If $A_i \Rightarrow B_i$, $i = 1, \ldots, q$, the procedures include

1. Convert the membership functions to points in Cartesian spaces. The antecedent space is $\mathcal{R}^n$ and the consequent space $\mathcal{R}^m$, where n and m are numbers of characteristic points for A_i and B_i, respectively.

2. The points A_i now play the role of the input x_i, and B_i that of y_i. There are now $N = q$ data points $z_i = (x_i, y_i) = (A_i, B_i)$, i.e., each fuzzy rule is represented as a point in $\mathcal{R}^{(n+m)}$.

3. Apply the clustering method to the data point z_i, resulting in c cluster centers.

4. Solve the LSE problem (13) to obtain parameters (G_i and h_i) of the output models.

5. The sparse rule base is hence given by: If $A_i^* \Rightarrow B_i^* = G_i^T A + h_i^T$, for $i = 1, \ldots, c$.

6. One can apply (8) to generate outputs at the antecedent points A_i for comparison with the given B_i to check how well the original rule set can be recovered from the sparse rule base.

We use the example in [7] to demonstrate the effectiveness of the present approach. The example is given in terms of 11 antecedent membership functions in the form of overlapping isosceles triangles with centers ranging from 0 to 10, and 7 consequent membership functions also in the form of overlapping isosceles triangles with centers ranging from 0 to 7. Denote these membership functions as A_1 to A_{11}, and B_1 to B_7. The given rule set is:

$$\begin{array}{l} \text{R1: If } a \text{ is } A_1\text{, then } b \text{ is } B_1 \\ \text{R2: If } a \text{ is } A_2\text{, then } b \text{ is } B_3 \\ \text{R3: If } a \text{ is } A_3\text{, then } b \text{ is } B_4 \\ \text{R4: If } a \text{ is } A_4\text{, then } b \text{ is } B_5 \\ \text{R5: If } a \text{ is } A_5\text{, then } b \text{ is } B_6 \\ \text{R6: If } a \text{ is } A_6\text{, then } b \text{ is } B_6 \\ \text{R7: If } a \text{ is } A_7\text{, then } b \text{ is } B_6 \\ \text{R8: If } a \text{ is } A_8\text{, then } b \text{ is } B_6 \\ \text{R9: If } a \text{ is } A_9\text{, then } b \text{ is } B_5 \\ \text{R10: If } a \text{ is } A_10\text{, then } b \text{ is } B_4 \\ \text{R11: If } a \text{ is } A_11\text{, then } b \text{ is } B_3 \end{array} \tag{14}$$

In [7], argument was made that the problem can be tackled by considering the 1st, 2nd, and the 3rd characteristic points of the membership functions independently. And since the characteristic points are shifted by the same amount for all membership functions, the problem can be solved with the solution to just any one of the characteristic points. The example picks the 2nd one, or the central locations of the membership functions, for solution. The rule set (14) is thus graphically represented by 11 data points, one for each rule, in a XY plot, with the X-axis denoting the central locations of A_i and Y-axis that of B_j. It is then desired that a few points, or rules, be found such that their interpolation constitutes a close approximation to all data points. A maximum error bound of $\epsilon = 0.5$ is set so that the original information can be recovered by truncating the interpolated values to the nearest integers. Assuming polynomial Lagrange interpolation, the work [7] presents $\{R3, R6, R9\}$ and $\{R1, R6, R11\}$ as possible selections which leads to satisfaction of the error bound requirement.

The same example is now treated using the present approach. First, the membership functions are converted to points in antecedent and consequent spaces. This is shown in Figs. 2 and 3, which depict A_1 to A_{11}, and B_1 to B_6, respectively. The rule set (14) hence in this case gives rise to 11 data points z_i in $\mathcal{R}^6$. However, with A_i and B_i lying on straight lines within their respective space, the problem can be treated considering as input space the straight line A_1A_{11}, and as output space the straight line B_1B_6. Hence, in its own way, the present approach also reduces the example to a two dimensional problem with linear input and output spaces.

Extraction of cluster center can thus be carried out in $\mathcal{R}^2$. We utilize the same parameters as in Chiu's work [9]. Specifically, the data points are first scaled to lie within a unit square. Then selection of cluster centers is carried out using $\alpha = 4/r_a^2$, $\beta = 4/r_b^2$, with $r_a = 0.25$ and $r_b = 1.5r_a$. The parameter r_a plays the role of a radius of influence. Figure 4 shows the potential functions $P_1(z_i)$, $P_2(z_i)$, and $P_3(z_i)$ against the rule number i. The cluster centers derived from these potentials are R7, R3, and R10. It can be observed from the figure that the cluster around

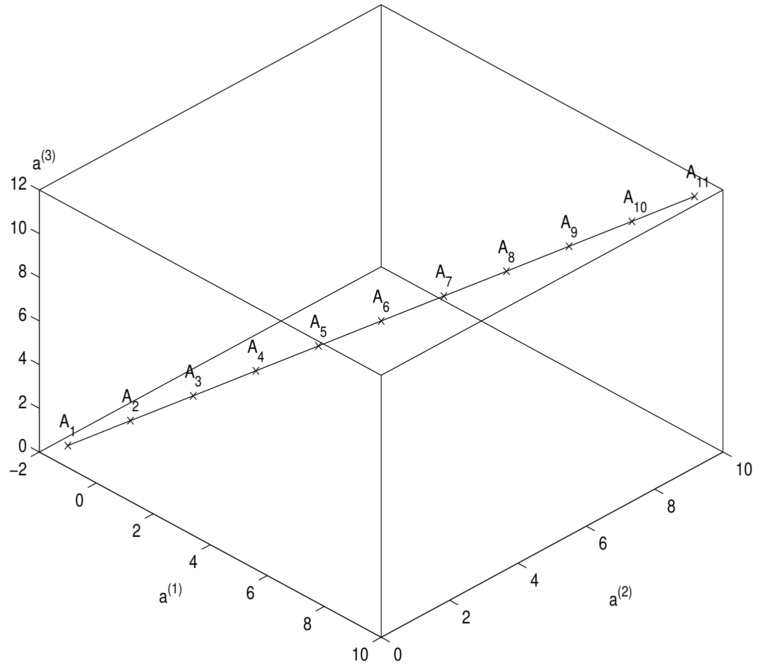

Figure 2: Antecedent space containing representing points of antecedent membership functions

R7 in $P_1(z_i)$ is removed in $P_2(z_i)$, and that the cluster around R3 in $P_2(z_i)$ is removed in $P_3(z_i)$, and so on.

The LSE process to identify the output models for R7, R3, and R10 is carried out according to (10)-(13). In the reduced $\mathcal{R}^2$ space, the quantities G_i and h_i, $i = 1, 2, 3$, are scalars. Figure 5 compares the outputs of the given rules with the interpolated outputs using first order output models. For convenience, the figure uses the central location of the antecedent membership functions as horizontal axis and the central locations of the consequent membership functions as vertical axis to present the rule base. Same as [7], we are interested to see how well the interpolation based on cluster centers would faithfully reproduce the original rules of R1 to R11. In this case the maximum error induced between the two is 0.4, which falls below $\epsilon = 0.5$. The original rules can hence be fully recovered by rounding off to the nearest integer.

On the other hand, Fig. 6 shows the interpolated results using zero order models. In this case, the error induced is over 0.5 and hence full recovery of the original rules from the ensuing sparse rule base is not possible.

Several points are worth mentioning here regarding the approach and case study:

- The rules R6 and R7 have the same potential $P_1(z_i)$ in Fig. 4, and R7 was arbitrarily picked over R6. Should R6 be chosen instead, the other two cluster centers, R3 and R10, would not be affected. The performance of the first order models in recovering the original rules would also be similar. R6 and R7 actually belong to the same cluster.

- The sparse rule base generated in this case is:

$$\begin{array}{l} \text{R1}^*\text{: If } A_7 \Rightarrow B_1^* = 0.168A + 0.897[1\ 1\ 1]^T \\ \text{R2}^*\text{: If } A_3 \Rightarrow B_2^* = 2.541A + 0.066[1\ 1\ 1]^T \end{array} \quad (15)$$

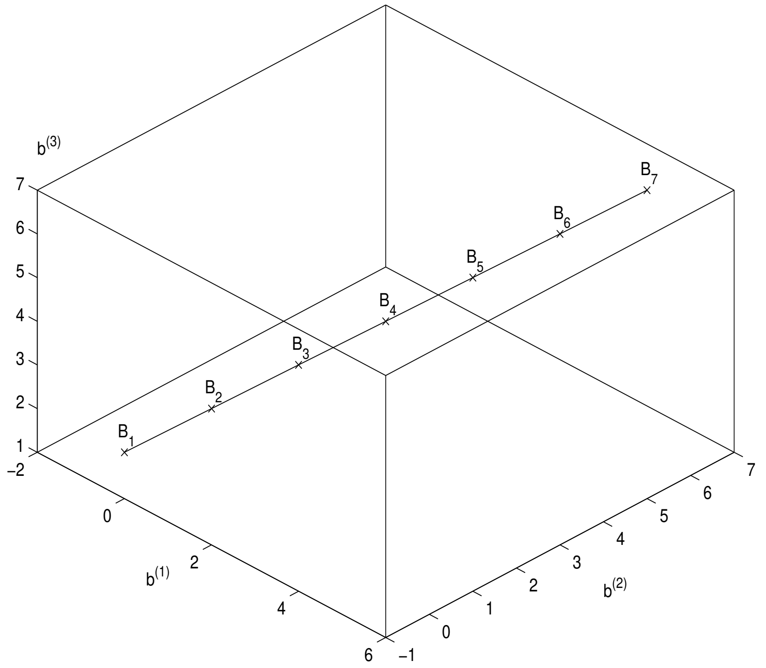

Figure 3: Consequent space containing representing points of consequent membership functions

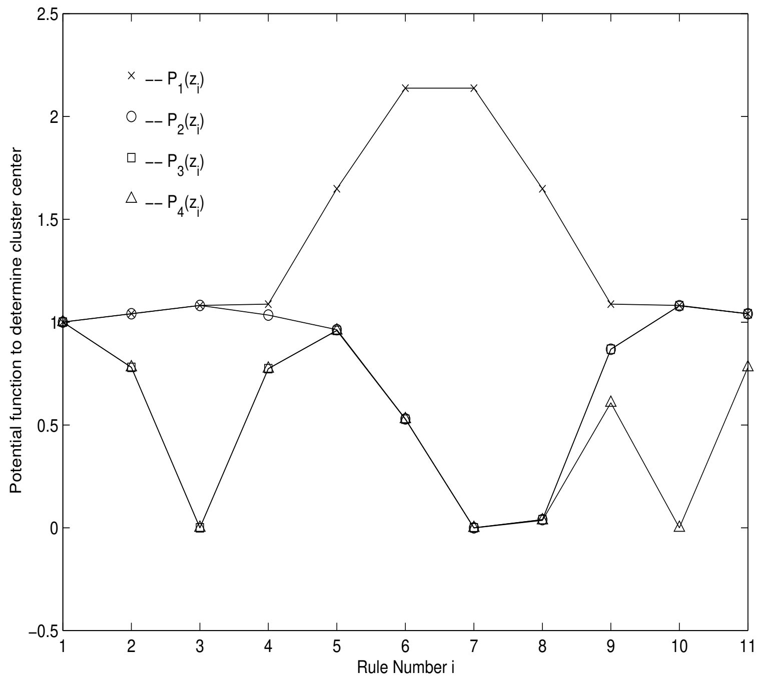

Figure 4: Potential $P_1(z_i)$, $P_2(z_i)$, $P_3(z_i)$, and $P_4(z_i)$ as a function rule number i

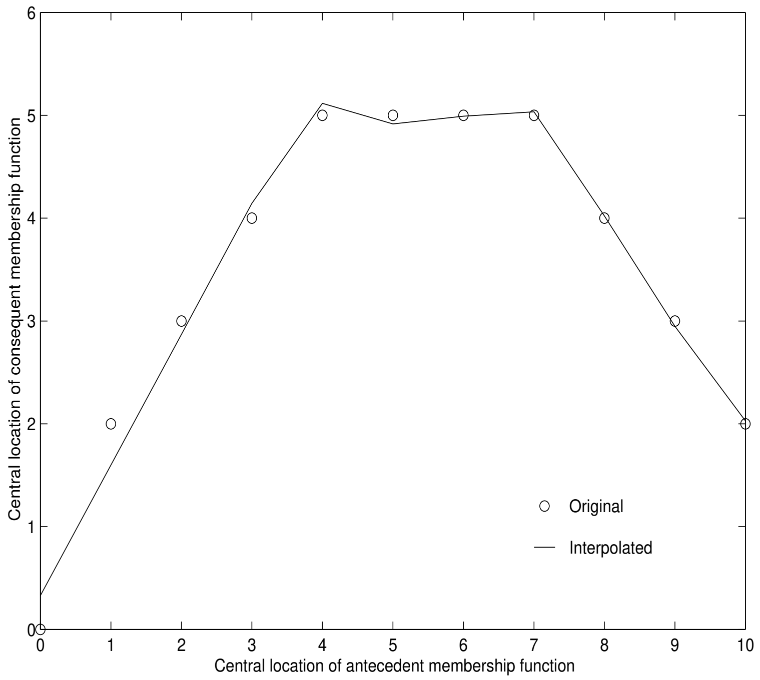

Figure 5: Rule base represented by central locations of membership functions: original (o) and interpolation from first order model (solid line)

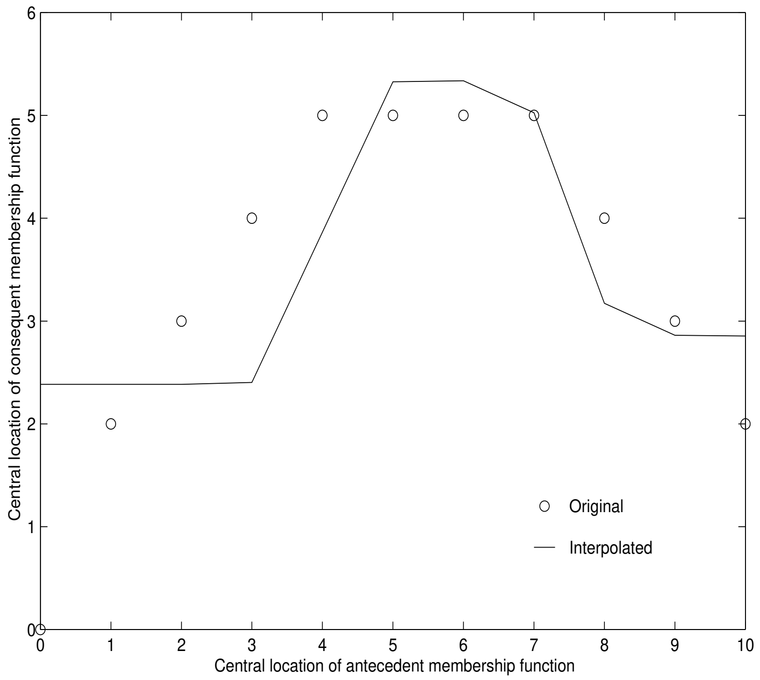

Figure 6: Rule base represented by central locations of membership functions: original (o) and interpolation from zero order model (solid line)

$$R3^*\text{: If } A_{10} \Rightarrow B_3^*\text{=-1.828}A\text{+2.233}[1\ 1\ 1]^T$$

- The present approach constitutes a systematic and standardized method to generate sparse rule base. This contrasts the exhaustive search approach discussed in [7], which requires appreciably more computational efforts. In comparison, it is interesting to note that the cluster centers of (R3, R7, R10) (or (R3, R6, R10) if R6 is picked as first cluster center) from the present approach are quite similar to the result (R3, R6, R9) in [7].

- As pointed out in [7], the exhaustive search approach is further complicated by the fact that while a selected number of rules may not be performing satisfactorily for a given specification ϵ, a reduced number from this selected set may, on the other hand, achieve satisfactory performance. In other words, the fact that a set of rules not performing satisfactorily does not eliminate from further search various subsets of the same set. In contrast, the LSE formulation embedded in the present approach guarantees that performance will be enhanced with the addition of more cluster centers. One can hence gradually increase the number of cluster centers and stop if satisfactory performance has been attained.

- The stopping rules in Chiu's method were not applied in the example. We stopped adding new cluster center when satisfactory performance is attained.

- Instead of being scalars, the parameters G_i and h_i correspond to a 3×3 matrix and a 3×1 vector when the problem is treated in $\mathcal{R}^6$. In the present example, however, G_i turns out to be a constant times identity matrix, and h_i a column vector of the same element, and the LSE problem can be decomposed into three identical reduced order ones in $\mathcal{R}^2$. This would not be true, of course, if the membership functions are of more general type.

- The work [6] classifies two types of interpolation, one within the antecedent spanning set and the other outside. The former is one where the antecedent observation lies within the space spanned by the antecedents of the available rules. In the present example, since the antecedents all lying on the straight line formed by $A_3A_7A_{11}$, the interpolation involved is solely within the antecedent spanning set. Again, the situation will be different if the membership functions are of more general nature.

5. Conclusions

This paper investigates a possible methodology to extract sparse rule base, by converting a given rule set into Cartesian points and then applying clustering technique to generate cluster centers. The work readily adopts Chiu's approach of the mountain method and, in particular, his formulation to generate output models which best fit the given data in a least square error sense. The LSE process ensures that performance in recovering the original rule set enhances with the addition of more cluster centers. An example previously discussed in another work is included as case study. It is demonstrated that the present approach offers a systematic and efficient way of extracting sparse rule base with performance satisfactory to within a specified error bound. The example in this study utilizes standard isosceles triangular-type membership functions and the problem is reducible to a two-dimensional one. More general cases will be studied in the future.

References

[1] L. T. Kóczy and K. Hirota, "Approximate reasoning by linear rule interpolation and general approximation," *International Journal of Approximate Reasoning*, **9**, 1993, pp. 197-225.

[2] L. T. Kóczy and K. Hirota, "Interpolative reasoning with insufficient evidence in sparse fuzzy rule bases," *Information Sciences*, **71**, 1993, pp. 169-201.

[3] P. Baranyi, T.D. Gedeon, and L.T. Kóczy, "A general interpolation technique in fuzzy rule bases with arbitrary membership functions," *IEEE International Conference on Systems, Man, and Cybernetics*, Beijing, 1996, pp.510-515.

[4] T.D. Gedeon and L. T. Kóczy, "Conservation of fuzziness in rule interpolation," *Symposium on New Trends in the Control of Large Scale Systems*, Vol. I., Herlany, Slovakia, 1996, pp. 13-19.

[5] L. T. Kóczy, K. Hirota, and T.D. Gedeon, "Fuzzy rule interpolation by the conservation of relative fuzziness," Technical Report *TR 97-2*, Hirota Laboratory, Department of Intelligent Computing and Systems Science, Toyko Institute of Technology, Yokohama, 1997.

[6] Y. Yam and L.T. Kóczy, "Representing Membership Functions as Points in High Dimensional Spaces for Fuzzy Interpolation and Extrapolation," Technical Report CUHK-MAE-97-03, 1997, Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong.

[7] L. T. Kóczy and K. Hirota, "Size reduction by interpolation in fuzzy rule bases," *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, 1997, Vol. 27, No. 1, pp. 1–13.

[8] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method", *IEEE Trans. Systems, Man and Cybernetics*, 1994, Vol. 24, No. 8, pp. 1279–1284.

[9] S. Chiu, "Fuzzy model identification based on cluster estimation", *J. of Intelligent and Fuzzy Systems*, 1994, Vol. 2, No. 3, pp. 267–278.

[10] V. Kreinovich, H. T. Nguyen, and Y. Yam, "Optimal choices of potential functions in fuzzy clustering", *Technical Report CUHK-MAE-98-01*, 1998, The Chinese University of Hong Kong.