2009 American Control Conference
Hyatt Regency Riverfront, St. Louis, MO, USA
June 10-12, 2009

A novel rule antecedent structure and its identification for fuzzy models

Ming Su and R. Russell Rhinehart

Abstract—**The existing combinatorial antecedent structure in fuzzy models makes them suffer from "the curse of dimensionality". In this work, a novel rule antecedent structure is proposed to design an efficient fuzzy model by using fewer rules. The new rule antecedent only uses nonlinear variables. Additionally, the proposed rule antecedents are expressed as ellipsoids covering the underlying local regions, which make spatial coverage more efficient.**

I. Introduction

Fuzzy models describe complex systems using human-like cause-and-effect rules. A fuzzy system employs a divide-and-conquer concept. Each rule is responsible for a local behavior, but all rules work together for a complete description. For a nonlinear mapping, $y=f(x_1,\ldots,x_n)$, to be expressed by a fuzzy model, a rule is defined by

$$\textbf{IF}\ (x_1 \text{ is } A_1\ \textbf{AND}\cdots\textbf{AND}\ x_n \text{ is } A_n)\ \textbf{THEN}\ y_r = f_r(x_1,\cdots,x_n) \qquad (1)$$

where, f_r is a simpler function than f and represents local system behavior. The subscript r indicates that it is the r^{th} rule. The determining factors for complexity are the dimension, n and the number of linguistic categories for each. Permutations in the combinatorial structure in the antecedent lead to many rule possibilities.

In this work, the objective is to propose a flexible antecedent rule structure, which also reduces dimensionality.

II. A Flexible Antecedent Structure

This approach has two separate ideas. One is to only place variables with a nonlinear impact in the antecedent. The other is to include variable interaction in antecedent structure.

A. Antecedent variables

As an example, Equation (2) represents a nonlinear dynamic model with three regressors, [y(t-1) y(t-2) u(t-1)].

$$y(t) = y(t-1)\left[y(t-2)+2.5\right] + y^2(t-1)u(t-1) \qquad (2)$$

Using the rule Structure (1), the rule antecedent is expressed as [y(t-1) is A_1 **AND** y(t-2) is B_1 **AND** u(t-1) is C_1)]. The antecedent dimension is 3. Assuming that each regressor has 5 fuzzy sets, the standard combinatorial construction will then generate 125 possible rules.

Equation (2) could be represented in a linear format in Equation (3) using functional parameters [1]

$$y(t) = a_1(t)\,y(t-1) + a_2(t)\,y(t-2) + b_1(t)\,u(t-1) \qquad (3)$$

with $a_1(t) = 2.5, \quad a_2(t) = y(t-1), \quad b_1(t) = y^2(t-1)$

where, model parameters a_2 and b_1 are functions of y(t-1). It indicates that the model can be expressed linearly in all variables except y(t-1). Therefore, the regressor, y(t-1) should be the only one included in the antecedent. The simplified rule is then defined by $\textbf{IF}\ y(t-1) \text{ is } A_i \textbf{THEN}\ y = f_i(x_1,\cdots,x_n)$. The antecedent dimension is reduced to 1. The possible rules are reduced from 125 to 5.

Extending the representation, a nonlinear model is expressed by $y = a_1(c_1,\cdots,c_m)x_1 + \cdots + a_n(c_1,\cdots,c_m)x_n$. It then results in a rule with only m variables in the antecedent.

$$\textbf{IF}\ (c_1 \text{ is } A_1\ \textbf{AND}\cdots\textbf{AND}\ c_m \text{ is } A_m)\textbf{THEN}\ y_r = f_r(x_1,\cdots,x_n) \qquad (4)$$

where c_i and x_i are termed as antecedent and consequent variables, representing u and **y** in Equations (2) and (3). The rule Structure (4) has advantages when m is less than n.

B. Antecedent structure

Knowing the antecedent dimension, users might resort to the techniques in [2] to have a more compact fuzzy model. Given a two dimensional antecedent (c_1 is A_1 and c_2 is B_1), if Gaussian membership functions are assumed and the product operator is used for the **AND** conjunction, the antecedent is then evaluated by the truth of antecedent (TA)

$$TA = e^{-\left(\frac{c_1-o_1}{\sigma_1}\right)^2-\left(\frac{c_2-o_2}{\sigma_2}\right)^2} \qquad (5)$$

where TA is an ellipsoid centering at (o_1,o_2) with width by σ_1 and σ_2. A contour plot of TA is shown in Fig. 1, where the highest value of TA =1 is reached at the centroid. The further out is the contour, the smaller the *TA value*. The value of TA can be interpreted as the belongingness of a data point to a local region. A fuzzy model has several rules. Given a two-dimensional antecedent with equal number of fuzzy sets for each antecedent variable, a typical combinatorial antecedent space partition is shown in Fig. 2(a).

This work was supported in part by the Edward E. and Helen Turner Bartlett Foundation.

M. Su is with the School of Chemical Engineering, Oklahoma State University, Stillwater, OK 74078-5021 USA (ming.su@okstate.edu).

R. Russell Rhinehart is with the School of Chemical Engineering, Oklahoma State University, Stillwater, OK 74078-5021 USA (rrr@okstate.edu).

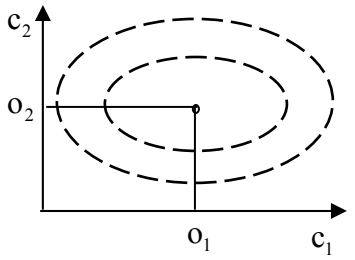

Fig 1. The ellipsoid contour of TA

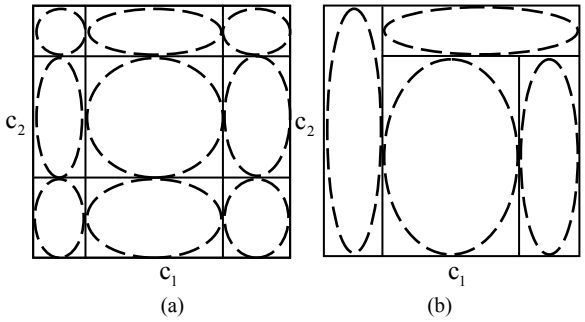

Fig 2. Antecedent space partition and representation

where , 9 rules are resulted by the exhaustive combinations of 3 fuzzy sets for each antecedent variable. The number of rules can be reduced by merging some regions that exhibit similar local behavior. Fig. 2(b) shows a possible simplified partition after merging some regions. The partition in Fig. 2(b) will also become inefficient as shown in Fig. 3, where neither horizontal nor vertical ellipsoid provides an efficient representation of the underlying local region represented by a rotated rectangle.

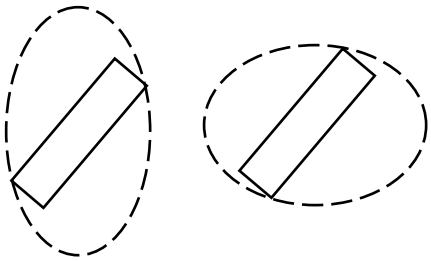

Fig 3. A rotated local region covered by a horizontal or vertical ellipsoid

The proposed solution is to have the ellipsoid rotated to match the underlying shapes of local regions. The concept of rotated ellipsoids is appealing. First, it offers one more degree of freedom, angles, to design ellipsoids. The following TA function with a centroid $\mathbf{o} \in R^m$ and a shape matrix $\mathbf{P} \in R^m \times R^m$ defines a rotated ellipsoid given the antecedent dimension is m .

$$TA = \mathrm{e}^{-(\mathbf{c}-\mathbf{o})^T \mathbf{P}(c-\mathbf{o})} \tag{6}$$

Second, it reduces the antecedent to one element defining truth.

III. IDENTIFICATION

A. Methodology

A data-driven procedure could be used to estimate parameters in Equation (6), if the number of regions is known

In this work, a regression tree [3] is used to determine the number and shapes of regions by a progressive binary partition of the antecedent space. A typical partition is shown in Fig. 4, where 5 regions (t_4, t_5, t_6, t_8, t_9) are obtained by 4 separations (s_1, s_2, s_3, s_4).

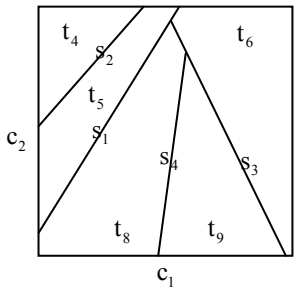

Fig 4. The resultant antecedent space partition

B. Analysis of the splitting and regression problem

The fundamental step is to solve a splitting and regression problem (SRP). Assuming that there are N data points, m antecedent variables and n consequent variables, the linear separation boundary equation is normally defined by

$$s_i = \theta_1 c_{i,1} + \cdots + \theta_m c_{i,m} \tag{7}$$

Where i refers to the i^{th} boundary. To determine the belongingness of data to two categories A and B, use

$$\varphi_i = \left(1 + \exp\left(-s_i / t\right)\right)^{-1} \tag{8}$$

where, t adjusts the 'sharpness' of the separation. At the limiting case, Equation (8) becomes a two-value (0,1) indicator function when t approaches 0.

$$\varphi_i = \begin{cases} 0, & s_i < 0 \\ 1, & s_i \geq 0 \end{cases} \tag{9}$$

and then two local linear models $\boldsymbol{A}$ and $\boldsymbol{B}$ for the segregated data groups are defined by

$$\begin{aligned} \hat{y}_{A,i} &= a_1 x_{i,1} + \cdots + a_n x_{i,n} \\ \hat{y}_{B,i} &= b_1 x_{i,1} + \cdots + b_n x_{i,n} \end{aligned} \tag{10}$$

which is combined with Equation (8) or (9) to compute the output.

$$\hat{y}_i = (1-\varphi_i)\hat{y}_{A,i} + \varphi_i \hat{y}_{B,i} \tag{11}$$

Parameters including **a** and **b** in Equation (10), and **θ** in Equation (7) are adjusted to minimize the criterion

$$J = \frac{1}{2}\sum_{i=1}^{N} \varepsilon_i^2 \tag{12}$$

where, $\varepsilon_i = y_i - \hat{y}_i$ is the residual.

The derivatives of J with respect to **θ**, **a** and **b** can be analytically expressed as below if Equation (8) is used in Equation (11) since it is continuous and differentiable.

$$\frac{\partial J}{\partial a_k} = -\sum_{i=1}^{N} \varepsilon_i (1-\varphi_i) x_{i,k} \tag{13}$$

$$\frac{\partial J}{\partial b_k} = -\sum_{i=1}^{N} \varepsilon_i \varphi_i x_{i,k} \tag{14}$$

$$\frac{\partial J}{\partial \theta_k} = \sum_{i=1}^{N} \varepsilon_i \varphi_i (1-\varphi_i) \frac{c_{i,k}}{t} w_i \tag{15}$$

where $w_i = \hat{y}_{a,i} - \hat{y}_{b,i}$ is the difference between two local models. The Hessian matrix can also be analytically expressed

$$\frac{\partial^2 J}{\partial a_k \partial a_l} = \sum_{i=1}^{N} (1-\varphi_i)^2 x_{i,k} x_{i,l} \tag{16}$$

$$\frac{\partial^2 J}{\partial a_k \partial b_l} = \sum_{i=1}^{N} \varphi_i (1-\varphi_i) x_{i,k} x_{i,l} \tag{17}$$

$$\frac{\partial^2 J}{\partial a_k \partial \theta_l} = -\sum_{i=1}^{N} \varphi_i (1-\varphi_i) \frac{c_{i,l}}{t} x_{i,k} w_i + \sum_{i=1}^{N} \varphi_i^2 (1-\varphi_i) \frac{c_{i,l}}{t} x_{i,k} w_i + \sum_{i=1}^{N} \varepsilon_i \varphi_i (1-\varphi_i) \frac{c_{i,l}}{t} x_{i,k} \tag{18}$$

$$\frac{\partial^2 J}{\partial b_k \partial b_l} = \sum_{i=1}^{N} \varphi_i^2 x_{i,k} x_{i,l} \tag{19}$$

$$\frac{\partial^2 J}{\partial b_k \partial \theta_l} = -\sum_{i=1}^{N} \varphi_i^2 (1-\varphi_i) \frac{c_{i,l}}{t} x_{i,k} w_i - \sum_{i=1}^{N} \varepsilon_i \varphi_i (1-\varphi_i) \frac{c_{i,l}}{t} x_{i,k} \tag{20}$$

$$\frac{\partial^2 J}{\partial \theta_k \partial \theta_l} = \sum_{i=1}^{N} \varphi_i^2 (1-\varphi_i)^2 \frac{c_{i,k} c_{i,l}}{t^2} w_i^2 + \sum_{i=1}^{N} \varepsilon_i \varphi_i (1-\varphi_i)^2 \frac{c_{i,k} c_{i,l}}{t^2} w_i - \sum_{i=1}^{N} \varepsilon_i \varphi_i^2 (1-\varphi_i) \frac{c_{i,k} c_{i,l}}{t^2} w_i \tag{21}$$

Once the gradients and Hessian matrix are obtained, it is then possible to analyze equilibrium solutions. Only the limiting case in Equation (9) is considered for its simplicity. Equilibrium solutions are defined if Equations (13-15) are zero. One solution is to have all $\varphi_i = 0$ (or $\varphi_i = 1$), which results in the following Hessian matrix.

$$\mathbf{H} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & 0 \\ 0 & 0 \end{bmatrix}; \text{ with } \mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,n} \end{bmatrix} \tag{22}$$

where, $\mathbf{X}^T\mathbf{X}$ is positive semidefinite if $\mathbf{X}$ has linear independent columns, which is a reasonable assumption for a linear regression model. **H** is hence positive semi-definite and the solution at $\varphi_i = 0$ is stable. Otherwise, the equilibrium condition is expressed as below

$$\sum_{i=1}^{N_A} \varepsilon_i x_{i,k} = 0 \quad \mathbf{c}_i \in A; \qquad \sum_{j=1}^{N_B} \varepsilon_j x_{j,k} = 0 \quad \mathbf{c}_j \in B \tag{23}$$

where, N_A and N_B are the number of data for model ***A*** and ***B***. Equation (15) is zero since φ_i is either 0 or 1.

Equations (23) are satisfied if **a** and **b** are obtained by least square estimation. The Hessian matrix then becomes

$$\mathbf{H} = \begin{bmatrix} \mathbf{X}_A^T\mathbf{X}_A & 0 & 0 \\ 0 & \mathbf{X}_B^T\mathbf{X}_B & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{24}$$

where $\mathbf{X}_A$ and $\mathbf{X}_B$ are similarly defined as in Equation (22) for model *A* and *B* respectively. The Hessian matrix (24) is also positive semidefinite and indicates a stable solution.

It then provides a two-step procedure to reach an equilibrium solution starting from an arbitrary separation followed by least square estimation. A solution is trivial if the separation is outside the antecedent space. Otherwise, two local models are obtained. Every solution is stable so that a gradient based optimization is easily trapped.

C. Solving the splitting and regression problem

In order to avoid the difficulty in using gradients for parameter tuning, the SRP is solved in this work by a heuristic suboptimal approach, which starts by guessing **a** and **b**, and optimizes the belongingness, φ. The optimal φ will be projected to a linear separation boundary to obtain the boundary parameters **θ**, which is then used to update **a** and **b**.

Guessing **a** and **b** is the first step. Given two linear models in Equations (10), it would be reasonable to assume that $\mathbf{y}_A$ and $\mathbf{y}_B$ are two distinct distributions. Finding two models for these two distributions could be viewed as separating one from the other. A threshold variable, v, is introduced to conduct the separation in y.

$$y_A = \{y | y < v\}, \qquad y_B = \{y | y \geq v\} \tag{25}$$

The optimal v is obtained by a linear search to minimize the following criterion S defined as the weighted sum of variance for two distributions $\mathbf{y}_A$ and $\mathbf{y}_B$.

$$S = N_A \hat{\sigma}_A^2 + N_B \hat{\sigma}_B^2,$$

with $\hat{\sigma}_A^2 = \frac{1}{N_A}\sum_{i=1}^{N_A}\left(y_A(i)-\bar{y}_A\right)^2, \quad \bar{y}_A = \frac{1}{N_A}\sum_{i=1}^{N_A} y_A(i)$ (26)

The separation in y determines also the separation of $\mathbf{X}$ into $\mathbf{X}_A$ and $\mathbf{X}_B$ that is used to estimate $\mathbf{a}$ and $\mathbf{b}$, and output prediction.

$$\hat{\mathbf{a}} = \left(\mathbf{X}_A^T\mathbf{X}_A\right)^{-1}\mathbf{X}_A^T\mathbf{y}_A; \quad \hat{\mathbf{b}} = \left(\mathbf{X}_B^T\mathbf{X}_B\right)^{-1}\mathbf{X}_B^T\mathbf{y}_B$$
$$\hat{\mathbf{y}}_A = \mathbf{X}_A\hat{\mathbf{a}}; \qquad \hat{\mathbf{y}}_B = \mathbf{X}_B\hat{\mathbf{b}} \tag{27}$$

The optimal φ_i is then found by solving the problem

$$\text{minimize} \quad \sum_{i=1}^{N}\left(y_i - (1-\varphi_i)\hat{y}_{A,i} - \varphi_i\hat{y}_{B,i}\right)^2$$
$$\text{subject to} \quad \varphi_i > 0,\ \varphi_i < 1 \tag{28}$$

where, φ_i is bounded between 0 and 1. The solution is used to obtain s by inverting Equation (8) with $t = 1$.

$$s_i = -\ln\left(\varphi_i^{-1} - 1\right) \tag{29}$$

which then results in the following $\boldsymbol{\theta}$ estimate

$$\hat{\boldsymbol{\theta}} = (\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{s}; \qquad \text{with} \quad \mathbf{C} = \begin{bmatrix} c_{1,1} & \cdots & c_{1,m} \\ \vdots & \ddots & \vdots \\ c_{N,1} & \cdots & c_{N,m} \end{bmatrix} \tag{30}$$

The estimated $\boldsymbol{\theta}$ is used in Equation (7) and (9) to define a new separation, φ_{new}. A trivial separation is obtained if φ_{new} is constant (0 or 1). In this case, the following problem instead of the one in Equation (8) is solved, where two extra constraints are introduced to ensure that each local model contains sufficient number of data. The variable β is linear searched to minimize the objective function.

$$\text{minimize} \quad \sum_{i=1}^{N}\left(y_i - (1-\varphi_i)\hat{y}_{A,i} - \varphi_i\hat{y}_{B,i}\right)^2$$
$$\text{subject to} \quad \varphi_i > 0, \quad \varphi_i < 1$$
$$\sum_{i=1}^{N}\varphi_i \ge \beta, \qquad \sum_{i=1}^{N}(1-\varphi_i) \ge \beta, \tag{31}$$

A non-trivial φ_{new} defines a new separation and the procedure then returns back to Equation (27) if two consecutive φ_{new} values are significantly different.

The SRP solving procedure is used to grow a tree. Tree growth and trim procedures could be found in [3].

D. Rule antecedent identification

The tree growth procedure generates a number of separation boundaries that partition the antecedent space. Given a partitioned antecedent space, there are many views on recognizing a local region.

In this work, in using data points to identify an ellipsoid, the quality of each data point is also considered. The quality is related to the residual. The solid dots in Fig. 5 have small residuals. The circles represent data points with larger residuals.

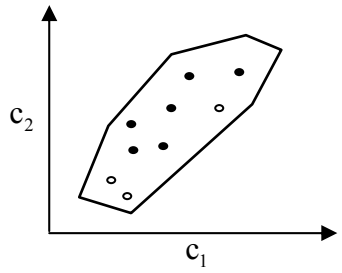

Fig 5. A local region in an antecedent space

A rule antecedent in fact represents the region where the consequent model is accurate. It is then reasonable to use only data samples with smaller residuals to estimate the antecedent parameters. Extending this concept, each data point is then associated with a weight that reflects its influence on ellipsoid estimation. The weight used in this work is defined by

$$\beta_{r,i} = e^{-N_r\frac{\varepsilon_{r,i}^2}{\boldsymbol{\varepsilon}_r^T\boldsymbol{\varepsilon}_r}} \tag{32}$$

where N_r is the number of data points in region r. The script (r,i) represents the i^{th} data in the r^{th} region. $\beta_{r,\text{i}}$ reaches the highest value at 1 when $\varepsilon_{r,\text{i}}$ is zero. High $\varepsilon_{r,\text{i}}$ values mean small $\beta_{r,\text{i}}$. Note: Equation (32) is the author's choice for exploring weighting. It is effective in this work, but may not be an essential part. Reasonable alternates would be to use problem-dependent knowledge or heuristics, or to ignore weighting by letting all $\beta_{r,\text{i}} = 1$. The centroid $\mathbf{o}_r$ is estimated by

$$\mathbf{o}_r = \sum_{i=1}^{N_r}\beta_{r,i}\mathbf{c}_{r,i} \Big/ \sum_{i=1}^{N_r}\beta_{r,i} \tag{33}$$

and the matrix $\mathbf{P}_r$ is defined by its inverse

$$\mathbf{P}_r^{-1} = \sum_{i=1}^{N_r}\beta_{r,i}\left(\mathbf{c}_{r,i} - \mathbf{o}_r\right)\left(\mathbf{c}_{r,i} - \mathbf{o}_r\right)^T \Big/ \sum_{i=1}^{N_r}\beta_{r,i} \tag{34}$$

E. Deffuzificaion

Modeling and prediction using the fuzzy model consisting of the above identified rules is conducted by a conventional weighted average approach.

$$\hat{y} = \sum_{r=1}^{R} TA_r y_r \Big/ \sum_{r=1}^{R} TA_r \tag{35}$$

where y_r is the prediction by the consequent model of the r^{th} rule. They are blended by the TAs to give an overall

prediction.

IV. Testing and Discussion

A. Testing models

Model 1 :

$$y = \sin(x)x, \quad x \in [-2, 8]$$

Illustrated in Fig. 6, Model 1 is a simple function with only one variable, which should occur in both antecedent and consequent parts. The vertical lines represent the linear separation boundaries. Between each pair of adjacent boundaries is a local regime labeled by a number that is based upon the sequence of tree nodes being generated. The boundaries are placed where the function curves 'most'. In a local regime, between boundaries, the function could be reasonably approximated by a linear function.

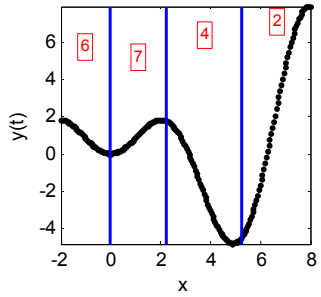

Fig 6. The boundaries for Model 1 by the suboptimal approach

The estimated antecedent parameters **o** and **P**, and local model parameters are summarized in Table 1.

TABLE 1
RULES IDENTIFIED FOR THE MODEL 1

Rule	Antecedent		Consequent
	o	**P**	
2	6.4738	1.7993	-32.9778 + 5.2878 x
4	3.4960	1.4483	8.0097 – 2.6437 x
6	-1.0589	3.4511	-0.2281 – 1.0943 x
7	1.0050	2.7209	-0.1636 + 1.0165 x

The four consequent linear models provide reasonable piece wise linear approximation of the function in each regime. If only a piece wise linear model is used, the model is however discontinuous. By using the deffuzification (35), the sequential linear models become a smoothed function as shown in Fig. 7 (the dashed line, labeled yhat). The mean squared error (MSE) is 0.1285.

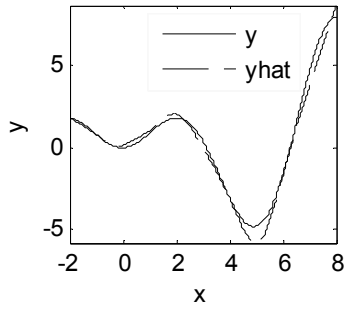

Fig 7. Function approximation by the fuzzy model

Model 2 [4]:

$$y(t) = 0.3y(t-1) + 0.6y(t-2) + 0.6\sin(\pi u(t-1)) + 0.3\sin(3\pi u(t-1)) + 0.1\sin(5\pi u(t-1)) + e(t)$$

Model 2 is a nonlinear dynamic model. The nonlinearity is due to the Sine function, which affects u(t-1). There are three regressors [y(t-1) y(t-2) u(t-1)], which should be included in consequent models. However, there is not a generic method for discovering nonlinear regressors. One practical approach might be to try each regressor in the antecedent. The difference will then hint which regressor is nonlinear. The experiment for this example is summarized in Table 2. The performance is evaluated by the sum of square error (SSE) between the output y and its prediction. Without any splitting, the SSE is 414.8067. Reduction of SSE is observed for each trial and the largest reduction corresponds to $u(t\text{-}1)$. The other two due to $y(t\text{-}1)$ or $y(t\text{-}2)$ barely improves the SSE. It then indicates that $u(t\text{-}1)$ should be included in antecedent and also hints that $u(t\text{-}1)$ has a nonlinear effect on y(t), which is true in this case.

TABLE 2
TRIALS OF ANTECEDENT VARIABLES FOR MODEL 2

Antecedent	Number of rules	SSE
$u(t\text{-}1)$	7	250.7703
$y(t\text{-}1)$	5	397.8031
$y(t\text{-}2)$	5	401.3424

Having $u(t\text{-}1)$ in the antecedent, the resultant separation boundary is shown in Fig. 8.

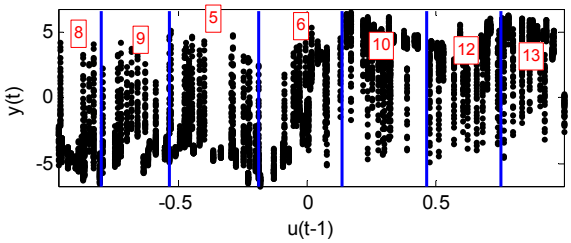

Fig 8. The separation boundaries for Model 2

Fig. 9 shows the separations in the nonlinear part of Model 2, $g(u(t\text{-}1))$, the sum of three Sine functions of $u(t\text{-}1)$. As with

the Model 1 test, Model 2 behaves relatively linearly in the seven local regions.

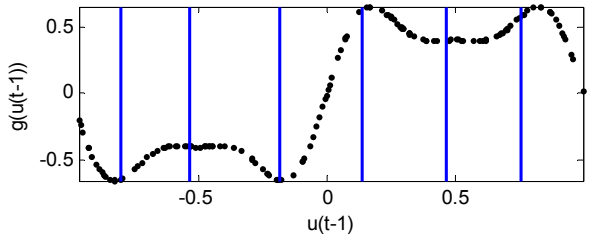

Fig 9. The separation boundaries for the nonlinear part in Model 2

Model 3 [4]:

$$y(t)=\frac{y(t-1)y(t-2)\left(y(t-1)+2.5\right)}{1+y(t-1)^2+y(t-2)^2}+u(t-1)+e(t)$$

Model 3 also has regressors [y(t-1) y(t-2) u(t-1)]. The same procedure is applied to try each regressor in antecedents. The result is summarized in Table 3

TABLE 3
TRIALS OF ANTECEDENT VARIABLES FOR MODEL 3

Antecedent	Number of rules	SSE
$u(t\text{-}1)$	5	197.3187
$y(t\text{-}1)$	6	137.3870
$y(t\text{-}2)$	6	131.6964

The SSE due to having $y(t\text{-}1)$ or $y(t\text{-}2)$ in the antecedent is lower than that due to $u(t\text{-}1)$. This observation is consistent with the Model 3, where $u(t\text{-}1)$ has no nonlinear effect on $y(t)$. The further investigation is to see if both $y(t\text{-}1)$ and $y(t\text{-}2)$ have interacting effect on $y(t)$. Fig. 10 shows the result by having both $y(t\text{-}1)$ and $y(t\text{-}2)$ in antecedent. The separation boundaries are neither horizontal nor vertical, which indicates that there is interacting effect of $y(t\text{-}1)$ and $y(t\text{-}2)$ on $y(t)$. To validate if it is necessary to include such interaction in antecedent, it then needs to check the SSE, which is 137.2988. The SSE is higher than that due to $y(t\text{-}2)$. It then could conclude that the best antecedent in terms of SSE is $y(t\text{-}2)$.

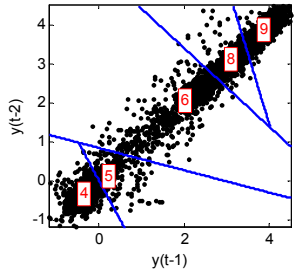

Fig 10. The separation boundaries for Model 3 with $y(t\text{-}1)$ and $y(t\text{-}2)$ in the antecedent

If preference is set to fewer local models, the antecedent might include both $y(t\text{-}1)$ and $y(t\text{-}2)$, which results in 5 local models only.

V. CONCLUSION

A procedure for reducing the curse of dimensionality in fuzzy models was described and demonstrated on simple cases.

The proposed rule antecedent rule structure is able to control the complexity in a fuzzy model. As shown in the last test, only 5 rules are generated even when the antecedent dimension increased from 1 to 2.

Instead of directly estimating model parameters, the proposed approach solves a series of splitting and regression problems to partition the antecedent space as well as compute the antecedent and consequent parameters. As shown in Fig.s 6 and 9, the resultant antecedent partition is meaningful. The placed boundaries divide an antecedent space into regions where function or system behaves relatively linearly.

The interpretability for individual antecedent variables will be lost due to the additional degree of freedom that combines all antecedent variables. However, the interpretability of the antecedent as a whole is still possible. A rule antecedent can be interpreted as a function that defines active region for the rules consequent model.

REFERENCES

[1] P. Young, "Time variable and state dependent modeling of non-stationary and nonlinear time series", in Developments in time series analysis, T.S Rao, Ed. New York: Chapman and Hall, 1993, pp 374-413.

[2] J. Yen and W. Liang, "Simplifying fuzzy rule-based models using orthogonal transformation methods", IEEE Transactions on Systems, Man, and Cybernetics, Part B, Feb, 1999, pp 13-24.

[3] L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone, "Classification and regression trees", Belmont: Wadsworth International Group, 1984.

[4] K.S. Narendra and K. Parthasarathy, "Identification and control of dynamical systems using neural networks", IEEE transactions on neural network, vol. 1, no.1, March 1990, pp 4-27