強健式模糊模式在系統行為曲線建模的應用

謝鴻琳 [1]
聖約翰科技大學 電機系 副教授
e-mail:shieh@mail.sju.edu.tw

張博綸 [2]
龍華科技大學 電機系 講師
e-mail:dirty@mail.lhu.edu.tw

摘要

近年來，在科學應用上，對一個未知的系統行為函數建模(modeling)一直是非常重要的研究課題，而模糊理論(fuzzy theory)常被用在取得的系統知識不完整、口語化的知識表達或是以專家知識描述系統的操作行為等等的建模上，並逐漸發展成為系統建模的有效工具。但是，在實際應用系統的取樣資料中，由於各種因素影響，常常使得取樣資料中存有一般雜訊(noise)或大誤差(outlier)的資料。若將這些雜訊或大誤差資料直接使用在系統行為函數的建模上，往往使得所建立的系統模式與實際的系統行為產生很大的誤差，這種現象稱為「過度適應」。為了解決這個問題，本文提出一個非監督模糊建模方式，可直接由具有雜訊及大誤差的輸入-輸出取樣資料中擷取模糊規則，以建立系統行為的近似函數。本文中利用2個實驗證明所提方法可以用在不同的資料領域上，且在效能上，比其他方法好。

關鍵詞：強健式學習、模糊建模、非監督模式、大誤差

1. 前言

近年來，在科學應用上，對一個未知的系統使用不明確知識建模(modeling)一直是非常重要的研究課題[1]，而模糊理論(fuzzy theory)常常是這些不明確知識建模，最常被使用的方法。但在實際的應用系統的取樣資料中，由於系統的環境影響、使用取樣的儀器或取樣的方法等因素，常常使得取樣資料中存有一般雜訊(noise)或大誤差(outlier)的資料。若將這些雜訊或大誤差的資料直接使用在系統的行為函數的建模上，往往使得所建立的系統模式與實際的系統行為產生很大的誤差，這種現象稱為「過度適應」(overfitting)[2][3][4][5][6]。因此，在系統的建模過程中，如何將雜訊或大誤差資料的影響降至最低，誠屬系統建模中重要關鍵因素之一。

在系統建模的領域中，強健式學習(robust learning)便是為了將雜訊或大誤差的資料的影響降到最低的方法之一[7]。目前大部分的強健式學習的演算法在學習的過程中，都對目標函數(objective function)加入調整函數(tuning function)(例如 loss function)，使得在系統的學習過程中對雜訊或大誤差的資料的影響降低。但此種方法最大的問題是不知道那些資料點是正確的資料點，那些點是具有大誤差的資料點，使得強健式學習的結果產生誤差[8]。

利用模糊建模方法以解決非線性系統建模的研究一直是頗受重視的方法，各種不同的方法分別被提出，例如，以數值型訓練資料(numberical training data)分析為主的研究[9][10][11][12]。這些研究的優點在於方法簡單且快速，缺點是必須事先指定(predefine)歸屬函數(membership function) 及 規則的個數。也有以智慧型資料分析及模糊類神經網路(fuzzy nerual networks，FNNs)為主的研究，例如： Hollatz [13]配合模糊群聚分析架構的RBF(Radial Basis Function)類神經網路建模及Roger Jang [14][15]架構一個 ANFIS 類神經網路建模。使用 ANFIS 建模可經由試誤法(trial-and-error)或專家知識(expert knowledge)事先指定歸屬函數(membership function) 及規則(rules) 的個數[16][17]。

另外，以模糊建模的另一個問題就是模式的參數初值的設定問題。因此，本文希望提出一種新的模糊建模方式，藉以建立系統行為的近似函數。此種建模方式乃基於資料分析(data analysis)的技巧，在不需要指定大誤差的的資料點下，將資料的雜訊及大誤差的的資料點降低其影響程度，並可以不需事先指定歸屬函數及規則的個數前提下，即能將資料依照其分佈的特徵自動分成不同的群聚，並以此群聚為基礎，建立系統的行為模式。

2. 模糊群聚分析

在模糊系統的群聚分析中，FCM 是最常被使用的方法之一。但是它仍然存在兩個主要的缺點：首先，它必須事先設定群聚的個數 c，但在某些應用上，我們並無法事先知道資料到底有多少的群聚，第二，群聚中心初值的設定會影響到它的演算法最後的結果。尤其當系統的取樣資料中若存在著雜訊及大誤差資料時，FCM 所建立的模式受到這些有雜訊的資料或大誤差資料的影響，而產生「過度適應」的現象。因此，本節提出一個新的演算法，基於距離計算的方式，針對這些有雜訊的資料或大誤差資料進行資料的群聚分析，以找出足以代表真正資料的群聚中心。

假設一組 n 個資料的系統取樣資料點，X= $\{x_i, y_i\}_{i=1}^{n}$，其中 $x_i = \{x_{i1}, x_{i2} \cdots x_{id}\}, x_i \in \Re^d$，$x_{ij}$ 代表系統的第 j 個輸入變數的第 i 個取樣資料，y_i 為系統的第 i 個取樣資料的輸出。模糊群聚的資料分析主要的目的，就是要將這些系統的取樣資料依照其特徵分成不同的群組，使得具有距離接近且相類似特徵的取樣資料點分成同一個群組。

本研究以資料點距離為基礎，利用疊代(iteration)方式計算群聚的中心點。因此開始的時候，是將每個資料點當作一個群聚的中心，而每一次疊代時，根據群聚之間距離的遠近來計算新的群聚中心。因此，任兩個群聚之間的距離計算如(1)式所示：

$$r_{ij} = \exp(-\frac{\|v_i - v_j\|^2}{2\sigma^2}) \quad i=1,2..,n, \quad j=1,2..,n \quad (1)$$

其中 v_i 及 v_j 代表第 i 個及第 j 個群聚的中心，$\|v_i - v_j\|$ 代表第 i 個及第 j 個群聚的中心 v_i 及 v_j 之間的歐基里德距離 (Euclidean distance)，σ 代表高斯函數(Gaussian function)的標準差(standard deviation)。

在(1)式中，當兩個群聚之間中心點 v_i 及 v_j 的距離愈大時，$\|v_i - v_j\|$ 的值就愈大，使得 r_{ij} 的值愈小，而當兩個群聚之間中心點 v_i 及 v_j 的距離愈小時，$\|v_i - v_j\|$ 的值就愈小，使得 r_{ij} 的值就愈大，且 $0<r_{ij}\leq 1$。所以 r_{ij} 可以代表兩個群聚之間中心點 v_i 及 v_j 的距離。換句話說，r_{ij} 的值愈大，代表這兩個群聚之間中心點 v_i 及 v_j 的距離很小，因此兩個群聚之間的特徵是相接近的。所以，當以 v_i 作為一個參考的群聚時，若群聚 v_j 與 v_i 之間的特徵是相接近而要產生新的群聚中心時，v_i 的新的群聚中心點 v_i' 可以用 (2)式表示為：

$$v_i' = \frac{\sum_{j=1}^{n} r_{ij}^m v_j}{\sum_{j=1}^{n} r_{ij}^m} \qquad j = 1,2,..., n, \quad m \geq 1 \qquad (2)$$

其中 $m\geq 1$，代表距離 r_{ij} 在計算新的群聚中心 v_i' 時的重要性的權重因子(weight factor)。因為 r_{ij} 的值小於或等於 1，若與參考群聚 v_i 相隔距離較遠的群聚 v_j，則 r_{ij} 會較小。當 m 的值較大時，r_{ij}^m 會使得在計算新的群聚

中心 v_i' 時，群聚 j 的中心點 v_j 對 v_i' 影響變得更小。由(1)式，原來的參考群聚中心點 v_i，由接近於 v_i，且有較相似特徵的群聚中心點的加權平均(weighted-average) v_i' 來取代。也就是說，新的群聚中心會較接近於真正資料的群聚中心點。

3. 強健式模糊 c-means 演算法

本節中提出一個 FCM 演算法的改進版本－強健式-FCM(Robust-F c-means，RFCM)，以改進在第 2 節開始所述 FCM 缺點。RFCM 主要的核心在於使用一組資料的群聚中心來表示系統的行為函數或所建立模式的真正曲線。假設 $X=\{x_1,x_2,\ldots x_n\} \subseteq \Re^d$ 為一組在 d-維空間，具有雜訊及大誤差的系統取樣資料，且 $v=\{v_1,v_2,\ldots,v_c\}$，$v_k \in \Re^d$ 為一個群聚的中心向量。圖 1(a)為一個原來的系統行為曲線。圖 1(b)中，因為雜訊或大誤差，使得非監督群聚演算法所產生的 3 個相鄰群聚中心，v_{i-1} ,v_i ,v_{i+1} 造成真正系統行為曲線與模式的群聚中心產生誤差。如圖 1(b)所示，點 v_i 因為受到大誤差資料點的影響，使得 v_i 不在真正的系統行為曲線上。如果我們可以將 v_i 以點 v_i^* 來取代，則由 v_{i-1} ， v_i^* , v_{i+1} 所構成的曲線比 v_{i-1} , v_i, v_{i+1} 所構成的曲線還接近真正的系統行為曲線如圖 1(c)所示。若進一步將 v_{i-1} ， v_i^* , v_{i+1} 取代 v_{i-1} , v_i, v_{i+1} 所得的平滑曲線，其系統群聚中心的分佈情形，如圖 1(d)中所顯示，以 v_i^* 來取代 v_i 之後的曲線，比原先的群聚中心更接近真正的系統行為曲線。

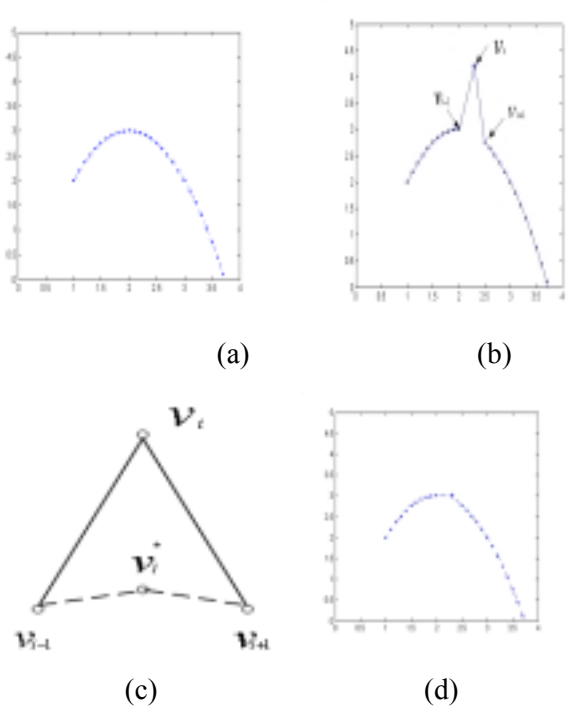


圖 1　一個不平滑群聚中心的例子

由三角形規則得知，v_{i-1} ,v_i 之間的距離與 v_i, v_{i+1} 之間的距離的和,大於 v_{i-1} ， v_i^* 之間的距離與 v_i^*， v_{i+1} 之間的距離的和，因此可得下面的(3)式：

$$\| v_i - v_{i-1} \|^2 \quad + \quad \| v_{i+1} - v_i \|^2 \quad > \quad \| v_i^* - v_{i-1} \|^2 + \| v_{i+1} - v_i^* \|^2 \qquad (3)$$

另外，由非監督群聚演算法的分析理論，對第 i 個群聚的群聚中心 v_i 及所有屬於這個群聚的每一個資料點 $x_{ki}(k = 1,2\ldots m_i$，其中 m_i 代表第 i 個群聚的資料點數)之間的距離和，比起在此群聚的其他任何一點的位置與所有屬於這個群聚的每一個資料點之間的距離和都小。因此，當第 i 個群聚的群聚中心 v_i 被 v_i^* 取代之後，自然的， v_i^* 與所有屬於這個群聚的

每一個資料點 $x_{ki}(k = 1,2\ldots m_i)$之間的距離和會比原先的 v_i 及所有屬於這個群聚的每一個資料點 $x_{ki}(k = 1,2\ldots m_i)$之間的距離和還要大，將這個關係以 (4)式表示為：

$$\sum_{k=1}^{m_i} \| x_{ki} - v_i^* \| \geq \sum_{k=1}^{m_i} \| x_{ki} - v_i \| \qquad (4)$$

由以上的分析，當點 v_i 以點 v_i^* 取代後，系統行為函數或曲線，比原先的群聚中心更為平滑。因此，根據(3)式，與點 v_i^* 相鄰的群聚中心之間距離的和會變小，但是由(4)式，當 v_i 移動到 v_i^* 時，又會使得屬於這個群聚的資料與 v_i^* 之間的距離和會變大。利用 Lagrange 定理[18]將(3)式與(4)式合成一個方程式，可得下面的(5)式，為了方便敘述，在(5)式中，v_i^* 以 v_k 來表示：

$$J_{RFCM}(\mu, v) = \sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik})^m \| x_i - v_k \|^2 + \alpha(\| v_1 - v_2 \|^2 + \sum_{k=2}^{c-1}(\| v_k - v_{k-1} \|^2 + \| v_{k+1} - v_k \|^2) + \| v_c - v_{c-1} \|^2)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{c}(\mu_{ik})^m \| x_i - v_k \|^2 + \lambda \sum_{k=1}^{c-1} \| v_{k+1} - v_k \|^2 \qquad (5)$$

其中 λ 稱為 Lagrange 乘積子(Lagrange multiplier)。為找出(5)式最佳的平滑點，以 v_k 對(5)式 $J_{RFCM}(\mu, v)$微分，以求最小的 $J_{RFCM}(\mu, v)$ 值，以式(6)式表示為：

$$\frac{\partial J_{RFCM}(\mu, v)}{\partial v_k} = 0 = \sum_{i=1}^{n} \mu_{ik}^m (x_i - v_k) + \lambda(v_{k+1} - v_k) \qquad (6)$$

所以 v_k 的值可以表示成(7)式：

$$v_k = \frac{\lambda v_{k+1} + \sum_{i=1}^{n} \mu_{ik}^m x_i}{\lambda + \sum_{i=1}^{n} u_{ik}^m} \qquad (7)$$

為表示任一群聚中心 v_k 與資料點 x_i 之間距離的關係，本文使用高斯型態(Gaussian type)的模糊歸屬函數(membership function)，以表示 v_k 與 x_i 之間距離愈小時，其關係就愈大，此關係以(8)式表示為：

$$\mu_{ik} = \exp(-\frac{\| x_i - v_k \|^2}{2\sigma^2}) \qquad (8)$$

此處的 μ_{ik} 代表第 i 個資料點屬於第 k 個群聚的歸屬度(membership grade)，v_k 是第 k 個群聚的中心點，$\|x_i - v_k\|$ 表示 x_i 與 v_k 之間的歐基里德距離，σ 表示高斯函數的標準差。

RFCM 與 FCM 有幾個不同點：(1) RFCM 採用 2 節所提的非監督式群聚演算法計算群聚中心點的初值，而 FCM 則採用隨機(randomly)方式來選擇群聚中心的初值，而這種隨機選擇的方法，可能因選擇不同的初值而使得所得的系統模式會有不同，繼而可能產生較差結果。(2)因為 RFCM 採用非監督式群聚演算法來計算群聚的初值，所以其分割成的群聚數量依非監督式群聚演算法產生，有別於 FCM 分割的群聚數量必須事先指定的方式。(3)對於 μ_{ik} 值的疊代計算，本文以(8)式的方式來取代原來 FCM 的 μ_{ik} 計算，原先 FCM 所建立的歸數函數是非凸集合特性的，但(8)式的歸屬函數值則具有凸集合的特性，而這種凸集合特性在模糊系統中相當的重要[19]。(4)本文在 μ_{ik} 計算比 FCM 在 μ_{ik} 需計算所有的資料點 x_i 與所有群聚中心點距離的和，節省很多時間。本文所提的 RFCM 可以綜合歸納如下：

Algorithm 1: RFCM algorithm

Step 1) Set value of m satisfying $1<m<\infty$. Use fuzzy clustering algorithm proposed in section 2 to calculate initial cluster centers v_k, k=1,2…c.

Step 2) Calculate the membership matrix of μ by Equation (8).

Step 3) Update fuzzy cluster centers using Equation (7).

Step 4) Calculate the objective function value by Equation (5). Stop if either the value is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

Step 5) Go to step 2.

4.以資料分佈為基礎的群聚分割演算法

非線性近似指將系統輸出-輸入的關係以一個非線性方程式描述。一個非線性系統可以被分割成多個線性區域的組合近似。所謂線性組合近似是指將系統分割成多個區域，每個區域的輸出-輸入關係近似於一個線性的關係，而每個區域以一個線性方程式描述，例如，對於一個 d 個輸入變數 x_i(i=1,2..d)及單一個輸出 y 的系統，如果將系統切割成 c 個近似線性區域的組合，每個區域以一個線性方程式描述時，則系統可以表示為：$y_j = a_{0,j} + \sum_{i=1}^{d} a_{i,j}x_{i,j}$ ，j = 1,2,…c。因此，對一個非線性系統，當由系統中取得 n 筆的資料時，由這 n 筆資料分佈的趨勢，可找出系統的局部最大及局部最小點，這些點代表系統的轉折特徵，利用這些局部最大及局部最小的特徵點，可將系統分割成多個區域的組合。為了方便說明起見，以單一個輸入變數 x 及單一個輸出變數 y 的非線性系統 $y=f(x)$ 為例。如果從此系統所擷取的取樣資料集合為 $\Omega=\{(x_i,y_i)\mid i=1,2...n\}$, $x_i\in\overline{X}$，$y_i\in\overline{Y}$，其中 $\overline{Y}$、$\overline{X}$ 分別表示 y 與 x 的定義域。尋找這些取樣資料的相依(dependent)變數 y 的局部最大點、最小點，這些最大、最小點若以獨立變數 x 的升冪(ascending order)排列，順序為 $m_1,m_2,...m_{c+1}$，其中 $m_j\in\overline{X}, j=1,2...,c+1$。如圖 2 所示，此系統的 4 個局部最大、最小點分別為 (x_j,y_j) = {(1,5) , (8,25) , (12,13) , (16,27)}，在圖中分別以 ⊗ 表示，所以 m_1=1，m_2=8，m_3= 12，m_4= 16，但是 m_1 及 m_4 兩個點為系統的邊界點，所以，在系統分割時不予考慮。因此，考慮圖 2，以 m_2，m_3 點做為系統的分割點，將系統分割成三個近似線性區域的組合，然後每個區域以一個線性的方程式來表示。圖 2 中，每個區域的線性方程式以一條直線表示。在一個多輸入單一輸出的非線性系統中，如同圖 2 所示的單一輸入單一輸出系統一樣，我們也可以找出在多維系統取樣資料的局部最大及最小點，並以這些局部最大及最小點作為系統分割的依據，將系統分割成多個區域的組合。

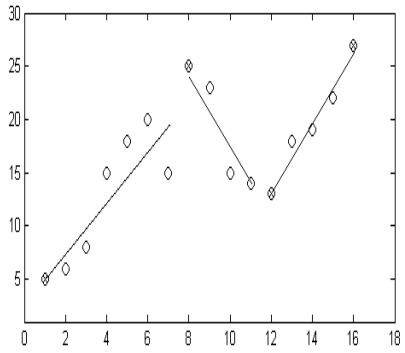
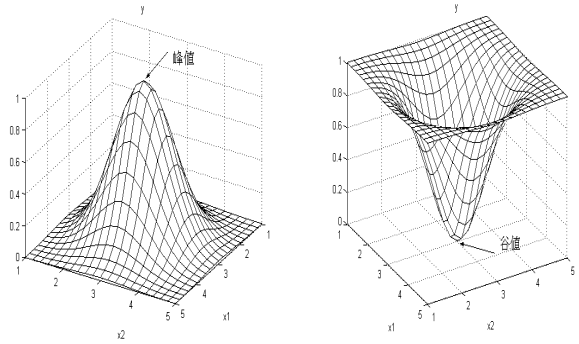


圖 2 一個可由三個線性區間組合而成的非線性系統

系統行為函數建模中，為找出系統轉折特徵點，本節提出一種具有模糊型樣辨識(fuzzy pattern recognition)能力的模糊資料篩選器(fuzzy-based data sifter；FDS)，搜尋未知目標系統取樣資料點的各種轉折特徵點。利用這些特徵點將未知的目標系統切割成多個子群聚(subclusters)的組合，並配合片段線性迴歸分析演算法(piecewise linear regression algorithm)計算代表每一個子群聚的迴歸參數，每一個子群聚的迴歸參數即代表 TS 模糊模式的後件部參數。

FDS 主要的工作，是一個資料篩檢系統，以一組轉折點篩選網路(turn-point sift network；TPSN)把資料點的局部最大、最小特徵點篩檢出來。FDS 除了可以尋找局部最大、最小點之外，也可以依據預先指定的資料分佈特徵做篩檢。如圖 3 所示，轉折特徵包括轉折特徵非常明顯的局部最大值、局部最小值，可以分別稱為：峰、谷。圖 3 (a)為峰值特徵型樣，圖 3 (b)為谷值特徵型樣。在圖 3 (a)的峰值特徵型樣中，縱軸的輸出點在峰值點上有最大值，在圖 3 (b)的谷值特徵型樣中，縱軸的輸出點在谷值點上有最小值。

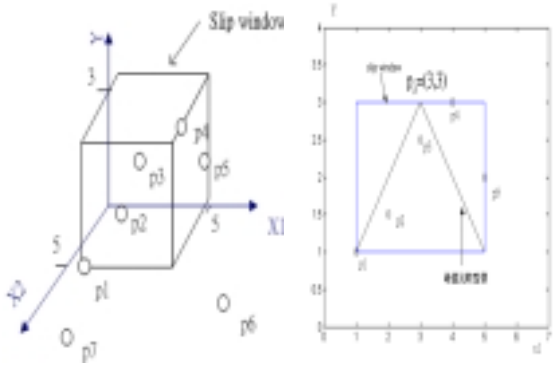


(a) (b)

圖 3 2 個輸入(x_1,x_2) 一個輸出(y) 的模糊特徵型樣(a) 峰值特徵型樣 (b) 谷值特徵型樣

FDS 從 *n* 個資料流中，依序每次取 *m* 個資料點與峰值及谷值兩個特徵型樣比較，以計算這 *m* 個點對兩個特徵型樣的模糊轉折點吻合值，經由計算後，系統的每一個取樣資料點對個別的峰值及谷值特徵型樣分別有相對應的模糊轉折點吻合值，因此在 *n* 個取樣資料點中，模糊轉折點吻合值愈高，代表吻合特徵型樣的程度愈高，該點可做為系統分割的基礎(base)點。

對取樣資料點中任一個資料點 p_j，*j*=1,2…*n*，每次取空間中距離 p_j 最近的 *m*-1(加上 p_j 共 *m* 個點)個點進行計算。這 *m* 個資料點代表系統的局部輸出與輸入關係，因此資料分布情形不一定都會完全符合式峰值與谷值比較型樣的定義。例如，圖 4(a)為兩個輸入，單一輸出的取樣資料點 p1~p7，以 p3 為中心，取得距離 p3 最近的 p1,p2,p4,p5 四個取樣資料點。



(a) (b)

圖 4 取樣資料點與峰值特徵型樣

圖 4 將這 5 個取樣資料點(含 p3)映射到 x_1 軸及 y 軸，並以 p3 為中心，將峰值特徵型樣套在這 5 個取樣資料點上，其中 p_s=(3,3)，代表峰值特徵型樣的最高點。所擷取 5 個取樣資料點，圖 4(a)及圖 4(b)中以「o」表示。這 5 個點利用峰值特徵型樣做比對，結果除了 p_1 剛好在峰值特徵型樣上外，其餘 4 個點均不在峰值特徵型樣，表示這 5 個點不是完全符合峰值特徵型樣，但也不是完全不符合。因此我們利用模糊理論的觀念定義這 5 個取樣資料點滿足峰值特徵型樣的程度。假設在取樣資料點中的 m 個點分別為$\{(x_{j1},x_{j2}\ldots x_{jk},y_j)|1\le j\le m\}$，對第 j 點$(1\le j\le m)$的取樣資料點，計算該點在第 i 個輸入變數 $x_{ji}(1\le i\le d,1\le j\le m)$的峰值比較型樣值為 $Y_j^{cp}=\{y_{ji}^{cp},y_{j2}^{cp}\ldots y_{jd}^{cp}\}$，$1\le i\le d,1\le j\le m$，所以 y_{ji}^{cp} 代表在 m 個資料點中，第 j 個取樣資料點的輸入變數 x_i(i=1,2,…d)所對應的峰值特徵型樣值。同理，對於谷值特徵型樣，對第 j 個取樣資料點，計算該點輸入變數 $x_{ji}(1\le i\le d)$ 的的峰值比較型樣值為 $Y_j^{cv}=\{y_{j1}^{cv},y_{j2}^{cv}\ldots y_{jd}^{cv}\}$，其中 $y_{ji}^{cv}=T_v(x_{ji})$，$1\le i\le d$；$1\le j\le m$，所以 y_{ji}^{cv} 代表在 m 個資料點中的第 j 個取樣點的輸入變數 x_i(i=1,2,…d)所對應到的谷值特徵型樣值。

但前面提到，第 j 點的取樣資料點的 y_j 值不一定與對應的峰值特徵型樣的 y_{ji}^{cp} (或谷值特徵型樣的值 y_{ji}^{cv})相同，所以對於第 y_j 點滿足峰值特徵型樣的 y_{ji}^{cp} 的程度(degree)，可以模糊歸屬函數定義如下：

$$\mu_{j1}(y_j)=\exp\left(-\frac{1}{2}\frac{\sqrt{\sum_{i=1}^{d}(y_j-y_{ji}^{cp})^2}}{y_{\max}-y_{\min}}\right) \qquad (9)$$

其中 y_{max} 及 y_{min} 分別為 m 個資料中最大及最小 y 值。

同理，對於第 y_j 點滿足谷值特徵型樣的 y_{ji}^{cv} 的程度(degree)，以模糊歸屬函數定義如下：

$$\mu_{j2}(y_j)=\exp\left(-\frac{1}{2}\frac{\sqrt{\sum_{i=1}^{d}(y_j-y_{ji}^{cv})^2}}{y_{\max}-y_{\min}}\right) \qquad (10)$$

對 n 個取樣資料點 $p_j=\{(x_{j1},x_{j2}\ldots x_{ji},..x_{jd},y_j)|1\le j\le n\}$，利用模糊轉折點特徵型樣對每個取樣資料點 p_j 做比對，以計算資料點 p_j 符合模糊轉折點特徵型樣的程度。滑下面的(11)式對 m 個取樣資料點符合值加以計算便會得到以 p_s 為中心的 m 個取樣資料點相對應模糊轉折點特徵型樣的整體吻合度(match degree)κ_{qs}，q=1,2，s=1,2…n。其中 s 表示以第 s 個取樣資料點為主的 m 筆資料，q=1 表示以第 s 點為主的 m 筆資料中所有資料點滿足峰值特徵型樣的程度，q=2 表示以 s 點

為主的 m 筆資料中滿足谷值特徵型樣的程度，其中 w_j(j=1,2…m)代表轉折型樣各點的權重值，w_1 為主點 p_s 的權重，w_2 為 p_{s-1},w_3 為 p_{s+1} 的權重，以此類推，且 $w_1 \geq w_2 \geq w_3$,>…$\geq w_m$，即距離點 p_s 愈遠的點，其權重愈低。

$$\kappa_{qs}=\frac{\sum_{j=1}^{m} w_j \mu_{jq}(y_j)}{\sum_{j=1}^{m} w_j}=\frac{w_1\mu_{1q}(y_1)+w_2\mu_{2q}(y_2)+\cdots+w_m\mu_{mq}(y_m)}{w_1+w_2+\cdots+w_m} \quad (11)$$

其中 $\frac{1}{\sum_{1}^{m} w_j}$ 為正規化因子(normalized factor)。且 q=1,2;s=1,2…n。

κ_{qs} 代表以 p_s 為中心的 m 個資料點滿足比較型樣的程度，當 κ_{qs} 愈大，表示取樣資料點 p_s 符合某一種特徵型樣的程度愈高，亦即符合此種轉折特徵的特性愈大，當群聚分析時，即可以依 κ_{qs} 大小的順序作切割。

5.線性迴歸

線性迴歸常常被用來求一個未知系統的輸入與輸出之間的線性關係式。假設 x_1,x_2,…… x_d 為 d 個自變數(independent)，且因變數為 Y。以系統的觀點來看，這 d 個自變數視為系統的輸入變數，因變數 Y 視為系統的輸出變數。對 d 個獨立(independent)輸入變數， Y| (x_1,x_2,…… x_d)的平均值以(12)式表示[20]：

$$M_{Y|(x_1,x_2\ldots.x_d)}=\alpha_0+\alpha_1x_1+\alpha_2x_2+\cdots+\alpha_dx_d \quad (12)$$

其中 α_i ，i=0,1,2,3…d 表示線性迴歸參數。對於一組系統的 n 個取樣資料 $\{(x_{i1},x_{i2}\cdots x_{id},y_i),i=1,2,\cdots n\}$ 且 n>d，其中 y_i 代表系統第 i 筆取樣資料的輸出。對於一個未知系統的輸出 y_i，可以用(13)式來表示估測 y_i 值：

$$y_i^{'}=b_0+b_1x_{i1}+b_2x_{i2}+\cdots+b_dx_{id} \quad (13)$$

其中 b_i 為 α_i ，i=0,1,2,3…d 的估計值。

$y_i^{'}$ 表示當輸入為 $\{(x_{i1},x_{i2}\cdots x_{id}),i=1,2,\cdots n\}$ 時，由估測值 b_i，i=0,1,2,3…,d 計算得到的輸出估測。假設輸入訓練資料為 $\{(x_{i1},x_{i2}\cdots x_{id},y_i),i=1,2,\cdots n\}$，其中 x_{ij}，i=1,2…n，j=1,2….d 表示第 i 個訓練資料的第 j 個輸入變數，y_i 則為輸出，由 b_i，i=0,1,2,3…d 所計算得到的估測值 $y_i^{'}$ 與真正輸出值 y_i 之間存在著誤差 $\varepsilon_i = y_i - y_i^{'}$，即 $y_i = y_i^{'}+\varepsilon_i = b_0+b_1x_{i1}+b_2x_{i2}+\cdots+b_dx_{id}+\varepsilon_i$。為使 ε_i 最小，定義此式的平方和殘差(Sum of Squares of the Error ,SSE)，S_r 如(14)式：

$$\text{SSE} = S_r = \sum_{i=1}^{n}\varepsilon^2 = \sum_{i=1}^{n}(y_i-(b_0+b_1x_{i1}+b_2x_{i2}+\cdots+b_dx_{id}))^2 \quad (14)$$

將 S_r 分別對 b_i(i=0,1,2,3…d)微分並令其等於 0，可得下列 d+1 個方程式[20]：

$$nb_0+b_1\sum_{i=1}^{n}x_{i1}+b_2\sum_{i=1}^{n}x_{i2}+\ldots+b_d\sum_{i=1}^{n}x_{id}=\sum_{i=1}^{n}y_i$$

$$b_0\sum_{i=1}^{n}x_{i1}+b_1\sum_{i=1}^{n}x_{i1}^2+b_2\sum_{i=1}^{n}x_{i1}x_{i2}+\ldots+\sum_{i=1}^{n}x_{i1}x_{id}=\sum_{i=1}^{n}x_{i1}y_i$$

$$\vdots$$

$$\vdots$$

$$b_0\sum_{i=1}^{n}x_{id}+b_1\sum_{i=1}^{n}x_{id}x_{i1}+b_2\sum_{i=1}^{n}x_{id}x_{i2}$$

$$+\ldots.+b_d\sum_{i=1}^{n} x_{id}^2=\sum_{i=1}^{n} x_{id}y_i$$

如果我們使用陣列(array)表示，我們可將上式表示為：

$$\overline{\overline{X}}\,\overline{B} = \overline{Y}\ , \tag{15}$$

其中：

$$\overline{\overline{X}} = \begin{bmatrix} n & \sum_{i=1}^{n}x_{i1} & \sum_{i=1}^{n}x_{i2} & \cdots & \sum_{i=1}^{n}x_{id} \\ \sum_{i=1}^{n}x_{i1} & \sum_{i=1}^{n}x_{i1}^2 & \sum_{i=1}^{n}x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n}x_{i1}x_{id} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \sum_{i=1}^{n}x_{id} & \sum_{i=1}^{n}x_{id}x_{i1} & \sum_{i=1}^{n}x_{id}x_{i2} & \cdots & \sum_{i=1}^{n}x_{id}^2 \end{bmatrix} \tag{16}$$

$$\overline{B} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ b_d \end{bmatrix}$$

$$\overline{Y} = \begin{bmatrix} \sum_{i=1}^{n}y_i \\ \sum_{i=1}^{n}x_{i1}y_i \\ \vdots \\ \sum_{i=1}^{n}x_{id}y_i \end{bmatrix} \tag{17}$$

$$\therefore\ \overline{B} = (\overline{\overline{X}})^{-1}\overline{Y} \tag{18}$$

上式 $\overline{\overline{X}}$ 、$\overline{Y}$ 可由訓練資料計算得到，若 $|\overline{\overline{X}}|$不為 0，即可求得向量(Vector) $\overline{B}$，且$\overline{B}$代表線性迴歸參數。

6. TS 模糊模式

模糊系統依照知識庫的結構可分為語言式的模糊模式(linguistic fuzzy models)、模糊關係模式(fuzzy relational models)及 Takagi-Sugeno 模糊模式(TS model)三種。其中 TS 模糊模式將系統的輸入空間分割成多個子空間(subspace)，這些子空間被視為一個描述線性或非線性系統的群聚。一階 Sugeno 模糊模式的規則可以寫為：

$$R^i : \text{IF } x_1 \text{ is } A_1^i \text{ and } x_2 \text{ is } A_2^i \ldots \text{and } x_d \text{ is } A_d^i$$

$$\text{THEN } y^i = a_0^i + a_1^i x_1 + a_2^i x_2 + \ldots + a_d^i x_d$$

$$= \sum_{j=0}^{d} a_j^i x_j \ ,\ \text{其中 } x_0=1 \tag{19}$$

其中 R^i (i=1,2…c)表示第 i 規則 ， x_j (j=1,2,….d)表示第 j 個輸入變數。 $y^i= f(x_j)$ 表示第 i 條規則的輸出，a_j^i 為常數 A_j^i 表示第 j 個輸入變數的語言項歸屬函數。本文以一個雙邊高斯函數(Two Side Gaussian Function，TSFG)[15]來描述 A_j^i 。 A_j^i 的定義如(20)式。

$$\text{TSGF}(x;\varphi_1,\sigma_1,\varphi_2,\sigma_2) = \begin{cases} \exp\left[-\frac{1}{2}\left(\frac{x-\varphi_1}{\sigma_1}\right)^2\right], & x \le \varphi_1 \\ 1, & \varphi_1 \le x \le \varphi_2 \\ \exp\left[-\frac{1}{2}\left(\frac{x-\varphi_2}{\sigma_2}\right)^2\right], & \varphi_2 \le x \end{cases} \tag{20}$$

當系統的取樣資料具有雜訊或大誤差資料，在建立系統行為函數模式時，可先使用本文第 2 節所提「非監督式群聚演算法」，產生初始群聚中心，並使用 RFCM 以降低雜訊及大誤差資料的影響，使所得的群聚中心更接近系統的行為模式。再依群聚中心分佈特性找出

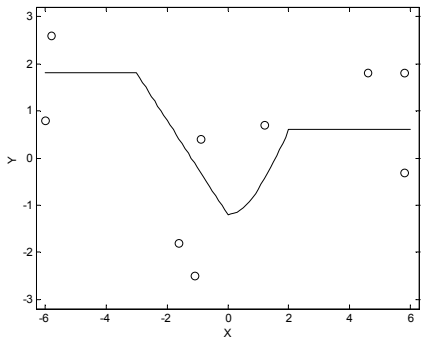
轉折點，即可將系統行為函數分割成線性子群聚的組合，每一個線性子群聚以一個 TS 模糊模式規則來表示，後件部參數以線性迴歸分析求得，前件部則如(20)式所示。最後再以梯度下降法來細調 TS 模糊模式的參數。

7. 實驗結果

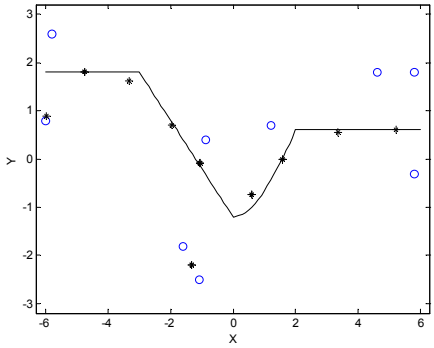
實驗一：本實驗目的是將所提的演算法用在一個具有常數值(constant value)的函數趨近，該函數為[21]:

$$y = f(x) = \begin{cases} 1.8 & x < -3 \\ -x - 1.2 & -3 \le x < 0 \\ 3e^{-0.1x}\sin(0.2x^2) - 1.2 & 0 \le x < 2 \\ 0.6 & 2 \le x \end{cases} \quad (21)$$

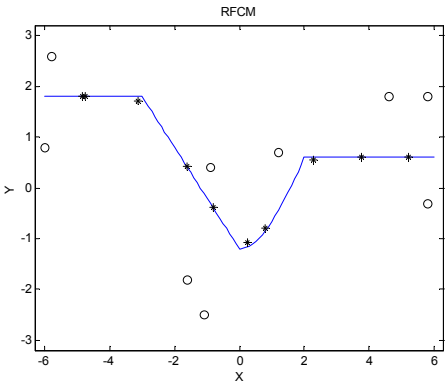
其中 x∈[-6,6]。如右圖 5(a)所示，由(21)式的 $f(x)$取 121 個取樣資料，並加上 9 個大誤差資料，其中實線代表原來的函數曲線，「。」表示 3 個大誤差資料點。圖 5(b)是經過非監督式群聚演算法所得到的資料點群聚中心，其中 m 與 σ^2 的值分別為 2 及 0.02，「*」代表群聚中心，由圖 5(b)中可以看出，由於大誤差點的關係，初始群聚中心包括大誤差點的群聚中心點，為消除大誤差點，利用 RFCM 對這些群聚中心作計算，最大疊代次數設為 10，可得圖 5(c)群聚中心結果。使用 FDS 對圖 5(c)的群聚中心計算，獲得的 6 個轉折點，圖 5(d)是扣除第一點及最後一點，剩下的 4 個轉折點以「□」來表示，因此，系統被分成 5 個線性子群聚的組合，每一子群聚以一條 TS 模糊規則表達。圖 5(e) 為 TS 模糊模式經過梯度下降法細調後所得輸出與原來函數曲線的比較，其中實線為原來函數曲線，點線為本文所提細調後的 TS 模糊模式輸出。圖 5(f)為本文所提模式輸出與 SVM 方法輸出的比較，其中實線為原來函數曲線，點線為本文所提的 TS 模糊模式輸出，虛線為 SVM 輸出，由圖中可以看出，本文的所提的 TS 模式輸出較 SVM 更 接近原來的函數曲線。另外，由(21)式產生 121 個取樣資料，利用本文所提方法進行測試，所得的 RMSE 值為 0.0374。



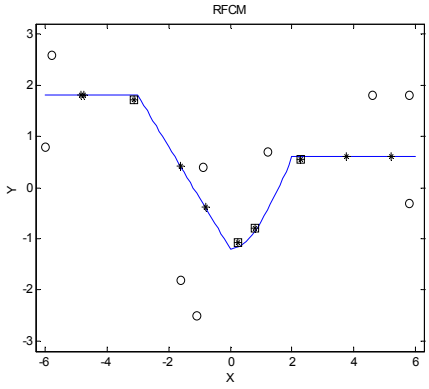
(a) 原來的函數曲線與9個大誤差點



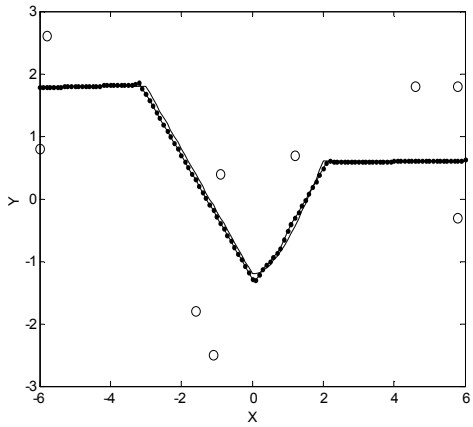
(b) 初始的群聚中心點



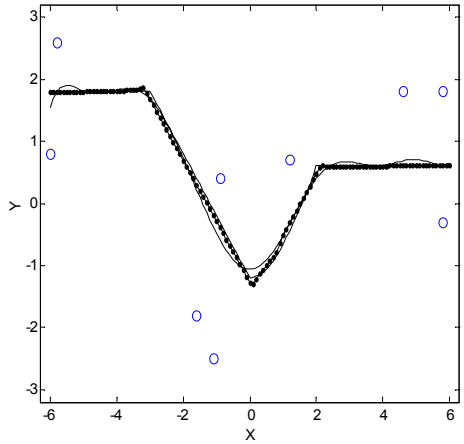
(c) 經過RFCM產生的群聚中心



(d) 轉折點計算，「□」代表轉折點



(e) 細調後的TS模糊模式輸出與原來函數曲線的比較。其中實線為原來函數曲線點線為本實驗模糊模式輸出



(f) 與SVM的比較

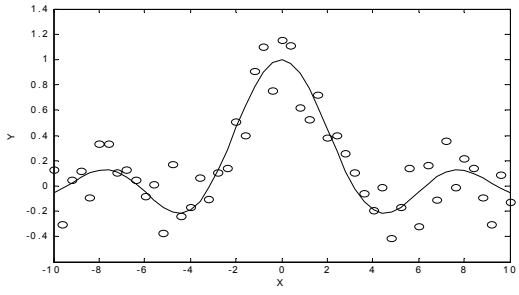
實線: 原來函數曲線 ,點線為本實驗模糊模式輸出,虛線為SVM模式輸出

圖5 具有常數值(constant value)的函數

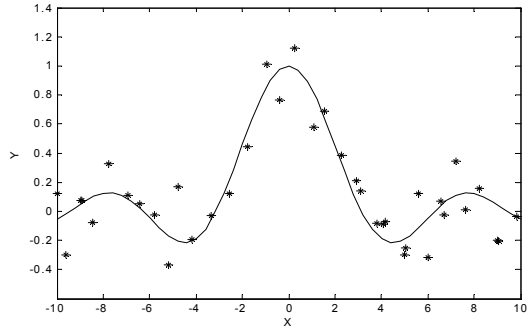
實驗二：本實驗主要測試所提演算法在具有雜訊取樣資料時的建模效能。考慮(22)式的一個*sinc*(*x*)函數[4]：

$$y = \frac{\sin x}{x} \quad , x \in [-10,10] \qquad (22)$$

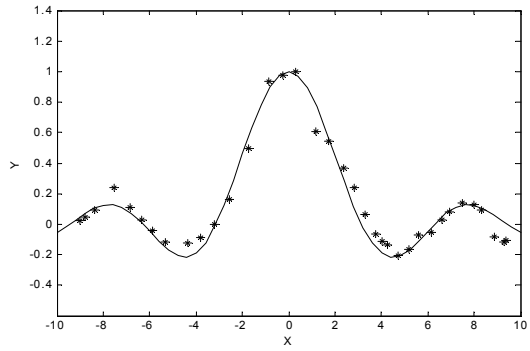
由(22)式產生 51 個具N(0,0.03)高斯雜訊的取樣資料，如圖 6(a)所示，其中實線代表原來的函數曲線，圓圈表示取樣資料點。圖 6(b)是經過非監督式群聚演算法所得到的資料點群聚中心，其中 m 與 σ^2 的值分別為 2 及 0.05，「*」代表群聚中心。利用 RFCM 對這些群聚中心作計算，最大疊代次數設為 10，可得圖 6(c)群聚中心結果。使用 FDS 對圖 6(c)的群聚中心計算，獲得的 9 個轉折點，圖 6(d)是扣除第一點及最後一點，剩下的 7 個轉折點以「□」來表示。因此，系統被分成 8 個線性子群聚的組合，每一子群聚以一條 TS 模糊規則表達。圖 6(e)為 TS 模糊模式經過梯度下降法細調後所得輸出與原來函數曲線的比較，其中實線為原來函數曲線，點線為本文所提細調後的 TS 模糊模式輸出。圖 6(f)為將梯度下降法用在群聚中心對模式細調的 RMSE 值遞減情形，經過 400 次的細調後，再將 51 筆正確的函數曲線取樣資料與模式輸出值作比較，RMSE 值為 0.0505.。與[4]所用的 ARRBFN 方法比較，該篇論文產生 12 條規則(12 個隱藏節點)及 RMSE 值為 0.0560，本文只需 8 條規則，且 RMSE 值為比較低，顯示本文所用方法在規則數及效能上優於該論文。



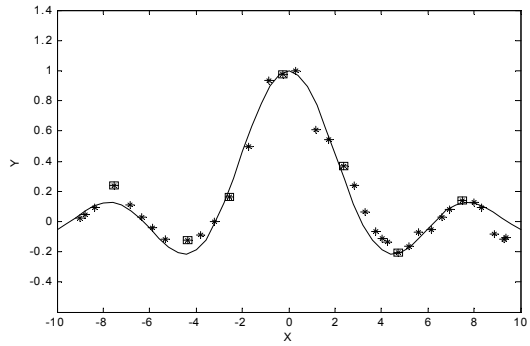
(a) 具N(0,0.03)雜訊的*sinc*(*x*)



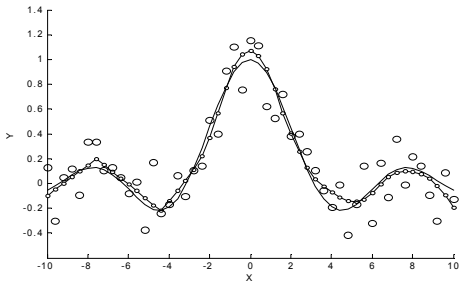
(b) 初始的群聚中心點



(c) 經過RFCM產生的群聚中心

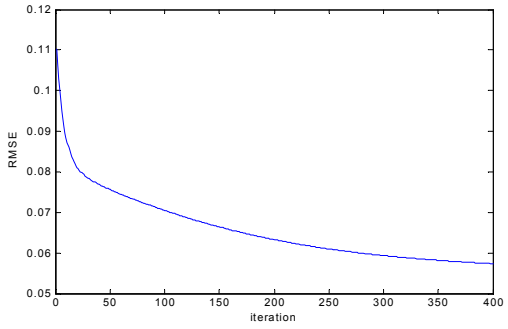


(d) 轉折點計算，「□」代表轉折點



(e) 細調後的TS模糊模式輸出與原來函數曲線的比較

實線: 原來函數曲線　點線: TS模糊模式輸出



(f) 細調的RMSE值遞減情形

圖 6 實驗三 具雜訊的函數趨近

8. 結論

本論文提出一個非監督模式模糊建模方式，利用強健式模糊群聚演算法對一群具有雜訊及大誤差的系統取樣資料建立系統模式，所提模式在不需要事先指定群聚數目下，產生模糊群聚。並使用強健式FCM 方式建立系統的行為曲，最後使用兩個實驗證明，本論文所提方式具相當不錯的效能。

參考書目

1. R. Babuska and H.B. Verbruggen, “Constructing Fuzzy Models by Product Space Clustering, ” *Fuzzy Model Identification. Springer-Verlag*, 1997.
2. Chen-Chia Chuang, Shun-Feng S*u,* Jin-Tsong Jeng and Chih-Ching Hsiao, “Robust Support Vector Regression Networks for Function Approximation With Outliers, ” *IEEE trans. On neural networks*, 13(6), pp1322~1330, 2002.
3. D. S. Chen and R. C. Jain, “A robust back-propagation learning algorithm for function approximation,” *IEEE Trans. on Neural Network*s,5, pp467–479, 1994.
4. C.-C. Chuang, S.-F. Su and C.-C. Hsiao, “The annealing robust backpropagation (ARBP) learning algorithm,” *IEEE Trans. on Neural Network*s, 11, pp1067–1077, 2000.
5. F. E. H. Tay and L. Cao, “Application of support vector machines in financial time series forecasting,” *Int. J. Manage.*, pp309–3172001,.
6. J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, “Weighted least squares support vector machines: Robustness and sparse approximation,” Neurocomputing, 48, pp85-105, 2002.
7. V. David Sanchez, “Robustization of a learning method for RBF network,” Neurocomputing 9, pp85~94,1995.
8. 莊鎮嘉, Robust Approaches of modeling under outlier, 台科大博士論文,2000
9. L.X.Wang and J.M. Mendel,"Generating fuzzy rules by learning from examples," *IEEE Trans. on Syst. Man, Cybern.*, 22, pp1414-1427,1992.
10. T. P. Hong and C.Y. Lee, “Induction of fuzzy rules and membership functions from training examples,” *Fuzzy Sets and System*, 84, pp33-47, 1996.
11. K. Nozaki, H. Ishibuchi,and H.Tanaka ”A simple but powerful heuristic method for generating fuzzy rules from numerical data,” *Fuzzy Sets and System*, 86, pp251-270, 1997.
12. R. Thawonmas and S. Abe,“Function approximation based on fuzzy rules extracted from partioned numerical data,” *IEEE Trans. on Syst. Man, Cybern.* B, 29, pp525-534, 1999.
13. J. Hollatz, “Fuzzy Identification Using Methods of Intelligent Data Analysis,“ *Fuzzy Model Identification. Springer-Verlag*. 1997.
14. J.–S. Roger Jang, “ANFIS: Adaptive-Network-based Fuzzy Inference Systems,” *IEEE Trans. on Syst. Man, Cybern.* 23(03), pp 665-685,1993.
15. R. J.–S. Jang, C. –T. Sun and E. Mizutani ,*Neuro-fuzzy and Soft Computing*, Prentice Hall,1997
16. Shiqian Wu, Meng Joo Er, and Yang Gao “ A Fast Approach for Automatic Generation of Fuzzy Rules by Generalized Dynamic Fuzzy Neural Network,"*IEEE Trans. on Fuzzy Systems*, 9 (4), pp578-594,2001.
17. Y.H. Lin and G. A. Cunningham , “A new approach to fuzzy-neural system modeling," *IEEE Trans. on Fuzzy Systems*, 3, pp190-197,1995.
18. C. R. Wylie and A. Barrett, *Advanced Engineering Mathematics*, pp840~pp846, McGraw-Hill, New York, 1982..
19. Frank Hoppner, Frank Klwawonn, Rudolf Kruse and Thomas Runkler, *Fuzzy Cluster*

Analysis, John Wiley & Sons, Inc., New York, 1999.

20. 顏月珠，*現代統計學*, 三民書局，民國 82 年 2 月.
21. Chen-Chia Chuang, Jin-Tsong Jengb, Pao-Tsun Lin, “Annealing robust radial basis function networks,” *Neurocomputing*, 56, pp123-139, 2004