灰聚類演算法之設計及其在模糊建模及控制上之應用

Design of grey clustering algorithm and its application on the fuzzy modeling and control

計畫編號： NSC 89-2213-E-032-027
執行期間： 自 88 年 8 月 1 日 至 89 年 7 月 31 日
主 持 人： 翁慶昌 淡江大學 電機系

一、中文摘要

本研究提出以灰關聯測度為基礎的灰聚類演算法來對資料進行分類，並探討如何在只知系統輸出入資料的情況下，將所提的灰聚類演算法應用於模糊系統建模及控制的設計上。首先，灰聚類演算法將資料點間的灰色關聯度量測值視為各資料點間相似程度的代表值，並依此相似值的大小來將資料分群。接著並應用此聚類演算法於系統輸出入資料的分析結果，做為系統結構的鑑別方法，另外再藉由最佳化方法來完成系統鑑別的參數鑑別部分，來建立一個適當合理的模糊系統。最後，以非線性系統的建模及倒單擺系統的控制問題來說明所提設計方法的可行性及有效性。

關鍵詞：聚類分析演算法；灰色關聯度分析；模糊建模；模糊控制

Abstract

In this project, a grey clustering algorithm based on the grey relational analysis is proposed to clustering data and applied to the design of the fuzzy modeling and control with only some input-output data of the system. First, the grey relational grades between data are regarded as the similarities of them, and then the grey clustering algorithm is utilized to group data according to the magnitude of similarity. Next, the algorithm is applied to analysis the input-output data of the considered system, and used the result of clustering to construct a rough model of the system. An optimization approach proceeds to finish the parameter identification part of the system identification. A nonlinear system modeling and an inverse pendulum control issue are illustrated the feasibility and effectiveness of the proposed design approach.

Keywords：Clustering Analysis; Grey Relational Analysis; Fuzzy Modeling; Fuzzy Control

二、計劃緣由與目的：

本研究計畫主要針對目前常用的聚類演算法中所存在的一些問題進行方法上的改善，並根據在灰色理論領域中所做的研究，將灰色理論裡有關分析資料序列的灰關聯分析方法應用在資料聚類演算的工作上。並將所提之灰聚類演算法成功的應用於模糊建模與模糊控制的問題上，做為系統鑑別中結構鑑別的主要方法。在常見的聚類演算法(如 Fuzzy c-means 聚類演算法)中，往往存在一些無可避免的人為設定，而這些人為的設定也通常會對最後資料聚類的結果造成影響。例如在使用 Fuzzy c-means 聚類演算法來將資料聚類時，我們就必須對初始的群聚數目、初始的群聚中心位置以及資料的距離量測方式做一個初始的設定及選擇。一般來說，對於低維度且群聚明確的資料來說或許可以很容易由視覺的反應來做出適當的設定與選擇，但事實上一旦所要分析的資料量很大，或是對於一些高維度資料的聚類問題，我們並無法根據所看到的來決定初始值的選取時，這些設定便會成為我們對資料進行聚類分析時的困擾。而且往往不同的初始設定值又很容易對最後資料的聚類結果造成影響。有鑑於此，我們希望所提出的聚類演算法能在參數選擇上以及資料聚類方法上

進行改善原有的聚類方法的缺失，並能夠正確且有效的對任何型態的資料進行正確的聚類分析，於是提出了一個以灰關聯分析為基礎的聚類演算法。考慮以灰色理論中專門在計算序列資料的關聯性的灰關聯分析來取代大多數聚類分析方法中的距離量測方法來做為資料間相似程度的代表值，其主要的原因在於灰關聯分析的結果是一個整體性的考量，並不會因為各個資料維度的大小不同而有所影響。而在距離量測中，對於各維度分佈範圍大小不同的資料而言，決定資料點之間距離的往往是由大範圍的維度資訊來主導。另一個以灰關聯分析來做為相似度量測的優點是灰關聯度的量測值永遠介於[0, 1]之間，這對於在處理各種不同資料的聚類問題時，對於參數的選擇相較於無法限定其範圍的距離量測來說會有較大的幫助。而且以灰色關聯分析來做為資料間相似度的測量也可以免於考慮因為不同的資料分佈型態而需要選擇不同的距離量測方式(例如，歐幾里得距離量測方式或橢圓距離距離量測方式)來處理聚類問題。此外，我們也利用了一個很直覺式的概念來進行資料的聚類動作，也就是希望相似程度高的資料能夠聚類在相同的群聚裡。首先，我們以一些可移動的向量來代表所有的資料點，然後再根據兩兩向量間的相似度找出與某一向量的相似度高於一某閥值的所有向量(也就是鄰近區域裡的向量資料)，並以這些相似度高之向量的平均位置來做為此一可移動向量位置的更新值。如此一來，每一個可移動的向量便會受到鄰近區域裡之高相似度的向量牽動而慢慢有靠攏聚集的狀況發生，經由不斷的疊代會將相互之間具有高相似度特值的向量聚集在一起，最後會將所有的向量收斂成為少數幾個向量。我們稱這些具有相同收斂位置的向量其原來所代表的那些資料為相同的群聚，而最後的收斂位置為該群聚的中心位置。而對於其中閥值的決定便是影響到此一聚類方法的主要關鍵，但很直觀的我們可以發現，當所選取的閥值越大時，對於相似度高的定義越行嚴格，向量之間會相互影響而相互牽引最後收斂的向量分佈範圍較小，所以會形成較多的群聚數；反過來說，當選取較小的閥值時，對於相似度高的定義較鬆，向量間相互影響的範圍大，因此就算相似度較低(距離較遠)的兩個向量也會因為相互的牽引而最後收斂於相同的位置上，所以會有較少的群聚產生。根據這些直觀的判斷，以及對於所有資料間灰關聯度量測值均落在[0, 1]之間的特性，我們其實很容易就可以掌握到要如何來決定這個閥值的選取，而且除此之外，這個聚類方法並不需要其他的設定便能將資料聚類。雖然如此，我們還是提出了一個效能指標的判斷式來幫助使用者選取適當的閥值來將資料聚類。這個效能指標的設定是以希望聚類所得到的群聚要有較高的緊密度為出發點來選取較適當的閥值。所以最終的灰聚類演算法便只需要在最初決定一個較小的初始閥值(也可以從 0 開始)及閥值的增量後，演算法即能幫我們找到一個適合的閥值來將資料聚類。所提出的方法可以具有以下的特性：(1)可自動找出群聚數目;(2)可找出極為精確的群聚中心位置;(3)群聚中心的初始位置不用須先設定;(4)與所欲分類的資料群的形狀無關。相信此方法的提出將使資料的聚類分析將更形有效且適用性更廣。此外，本計畫更結合所提的聚類演算法於模糊建模與模糊控制的問題探討上，利用這個聚類方法來對只知輸入輸出資料的系統進行結構鑑別的工作，再利用一最佳化方法來選取精確的參數來達到模糊建模與控制問題的解決。

三、研究方法及成果：

3.1 灰聚類演算法

灰關聯度為一個描序資料序列間關係大小的測度方法。對於灰關聯空間裡的因子集 $P(X)$ 中一序列 $x_i=(x_i(1),x_i(2),\ldots,x_i(m))\in X$，$i=1,2,\ldots,n$，$x_i(k)$ 代表此序列中的任一元素，其中 $k=1,2,\ldots,m$，首先定義灰關聯係數為：

令任一序列 x_i 為參考序列，x_j 為比較序列，$j=1,2,\ldots,n$，則 $x_i(k)$ 與 $x_j(k)$ 的灰關聯係數為

$$\gamma(x_i(k),x_j(k))=\frac{\Delta_{\min}+\zeta\Delta_{\max}}{\Delta_{ij}(k)+\zeta\Delta_{\max}} \qquad (1)$$

其中，$\Delta_{ij}(k)=\left|x_i(k)-x_j(k)\right|$ 為 $x_i(k)$ 與 $x_j(k)$ 相差的絕對值。

$$\Delta_{\min}=\min_{\forall i,\forall j}\min_{\forall k}\left|x_i(k)-x_j(k)\right| \qquad (2)$$

$$\Delta_{\max}=\max_{\forall i,\forall j}\max_{\forall k}\left|x_i(k)-x_j(k)\right| \qquad (3)$$

當灰關聯係數產生之後，我們可以根據下式得到 x_i 與 x_j 兩序列的灰關聯度量測值：

$$\gamma(x_i, x_j) = \sum_{k=1}^{m} \beta_k \gamma(x_i(k), x_j(k)) \tag{4}$$

其中，β_k 為權重值，$\sum_{k=1}^{m} \beta_k = 1$。

灰關聯度的量測是一種具有整體性的測度方法。利用這個特性，希望能將灰關聯的測度方法應用於複雜度較高的資料分群的工作上。這裡所指的複雜度高不單是指資料的維度高，且當資料在每個維度的範圍大小不相同時，易造成以歐式距離做為資料間相似量測的聚類分析方法所形成在較小範圍維度的資料中容易被忽略的問題，以致於在聚類結果會有錯誤分類的情形。

由於使用方法不同，本研究也發展出一套有系統的方法，以很簡單的概念、直接的方式對資料做聚類處理。灰聚類演算法所使用的概念簡單地說就是：我們希望具有相同特性的資料，能經由測量與疊代之後，收斂到相同的一點。而對於這些能夠收斂到相同位置的資料，我們稱之為同一群聚。關於特性，在各方法中會對聚類所依據的特性有不同定義，在本文的方法中，特性即是以資料間的灰關聯度量測來代表資料間相關程度的特性。詳細的灰聚類演算法敘述如下：

對於資料集 $X = \{\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n\}$，每筆資料 $\underline{x}_i = (x_i(1), x_i(2), \cdots, x_i(m))$ 可視為一向量，$\mathbf{R}^m$，每個向量中包含有 m 個元素分別代表 m 個不同的特徵來描述問題。所提演算法的目的即是要將這 n 筆資料，經由灰色關聯度的量測所得，再依據各資料間相似程度的大小，將資料加以聚類成數個群聚。首先，我們指定 n 個可移動的向量來代表原始的 n 筆資料向量，並將這些向量視為可比較的序列，然後依次指定每一向量為參考序列並測量與其他各比較序列的灰色關聯度之後，每一參考序列都找出與自己有較高關聯度的幾個比較序列，並以所得的序列中元素的平均值來更新自己。在這裡所選取較高關聯度的比較序列來更新參考序列的原因是不希望參考序列受到關聯度低的比較序列影響。當所有序列完成更新，即為一次疊代的結束。經由疊代動作，相互之間關聯度較高的幾個序列便會慢慢聚集起來，到所有的序列均收斂成為少數幾個不同值序列，且不再改變時，我們便稱這些收斂在一起的序列所代表的原始向量具有相似的特性，故可視為相同的群聚。

此外，對於相似度高的認定我們是以一個閥值的設定來做判別。也就是在資料間的灰關聯度求出之後，我們再以一閥值 ω 來選取所謂相似度高的資料。對於灰關聯度大於 ω 值的資料我們才稱為是具有較高相似度的資料。而 ω 值的大小會影響到參考向量更新值的決定，當然也會影響到最終的聚類結果。有鑑於此，我們在聚類的過程中提出了一個效能指標(performance index, PI)的設定來協助做閥值的選取。整個聚類演算法的實行步驟如下所述：

步驟一：以一可移動的向量集合 $V = \{\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n\}$ 代表原始的資料向量，即令 $\underline{v}_i = \underline{x}_i$，$i = 1,2,\ldots,n$；設定初始的閥值 ω 及閥值增量 $\Delta\omega$。

步驟二：以灰關聯測度方法來計算資料序列 $\underline{v}_i$ ($i = 1,2,\ldots,n$) 和 $\underline{v}_j$ ($j = 1,2,\ldots,n$) 間的灰關聯度 $r_{ij} = r(\underline{v}_i, \underline{v}_j)$ 做為兩資料序列間相似程度的代表值。

步驟三：更新向量 $\underline{v}_i = (v_i^*(1), v_i^*(2), \ldots, v_i^*(m))$，其中

$$v_i^*(k) = \frac{\sum_{j=1}^{n} n_{ij} v_j(k)}{\sum_{j=1}^{n} n_{ij}}, \quad k = 1,2,\ldots,m \tag{5}$$

且

$$n_{ij} = \begin{cases} 1 & \text{if } r_{ij} \geq \omega \\ 0 & \text{if } r_{ij} < \omega \end{cases} \tag{6}$$

步驟四：若 $\underline{v}_i$，$i = 1,2,\ldots,n$ 的值不再變動，則繼續下一步驟；否則，跳至步驟二繼續執行。

步驟五：計算效能指標(PI)

$$PI = \sum_{i=1}^{c} \bar{d}_i = \sum_{i=1}^{c} \left(\frac{1}{\#(G_i)} \sum_{x_j \in G_i} \left\| \underline{x}_j - cc_i \right\| \right) \tag{7}$$

其中 c 為聚類完成後的群聚數目，G_i，$i = 1,2,\ldots,c$ 為歸屬於第 i 個群聚的資料所構成的集合，$\#(G_i)$ 為集合的元素個數，cc_i 為第 i 個群聚中心位置。

步驟六：若 $\omega < 1$，令 $\omega = \omega + \Delta\omega$，以及 $V = X$ 並回到步驟二。

選擇能使得 PI 值為最小的最大閥值所得之聚類結果為最終結果。令最終收斂點個數為群聚數目，且視收斂於相同收斂位置的資料歸

屬於相同群聚。

3. 應用灰聚類演算法於模糊系統之設計

考慮一個多輸入單輸出的系統，由系統所收集到的輸入輸出資料對中即可以定義出系統的輸入輸出變數。假設以具有 m 個維度的向量來做為系統的輸入變數，即系統為 m 輸入單輸出的系統，其第 i 個模糊規則 R_i, $i=1,2,\ldots,c$，可以表示成：

R_i: IF x_1 is A_{i1} and…and x_m is A_{im}

THEN y is $y_i = a_{i0} + a_{i1}x_1 + \cdots + a_{im}x_m$ (8)

其中 $X=(x_1,x_2,\ldots,x_m)$ 為系統的輸入項，y 為系統的輸出項，y_i 是一實數值，為輸入變數的線性組合。而 A_{ik} 為變數 x_k 的模糊集合，其歸屬函數表示如下：

$$A_{ik}(x_k) = \exp\left(-\frac{(x_k - v_{ik})^2}{\delta_{ik}^2}\right),\ k=1,2,\ldots,m \tag{9}$$

其中 v_{ik} 為鐘形歸屬函數的中心點，δ_{ik} 為其寬度，而整個系統最後的明確輸出值 y 為：

$$y = \frac{\sum_{i=1}^{c} y_i \tau_i}{\sum_{i=1}^{c} \tau_i} \tag{10}$$

其中

$$\begin{aligned}\tau_i &= \prod_{k=1}^{m} A_{ik} = \prod_{k=1}^{m}\left[\exp\left(-\frac{(x_k - v_{ik})^2}{\delta_{ik}^2}\right)\right] \\ &= \exp\left[-\sum_{k=1}^{m}\left(\frac{x_k - v_{ik}}{\delta_{ik}}\right)^2\right]\end{aligned} \tag{11}$$

以群聚中心位置來決定歸屬函數的中心點 v_{ik}，而歸屬函數的寬度，我們是利用每個群聚的範圍大小來決定歸屬函數的 δ_{ik} 值，

$$\delta_{ik} = \sqrt{\frac{-(x_{ik}^{*} - v_{ik})^2}{\ln(\alpha)}} \tag{12}$$

其中，x_{ik}^{*} 為第 i 群中離中心位置最遠資料點的第 k 個維度資料，α 為在 x_{ik}^{*} 這個位置上歸屬值的指定值。

到目前為止，我們已經架構出一個粗略的模糊模型來描述系統，至於更精確的系統描述則須再經由細部調整規則中的參數值來達到目的。我們以最陡坡降法來調整剛所建立的粗略模糊模型之參數。首先定義一誤差函數做為調整這些參數的依據：

$$E = \frac{1}{2}(y - y^d)^2 \tag{13}$$

其中 y 為粗略模糊模型的輸出值，y^d 為期望的輸出值。所要調整的參數共包含 v_{ik}, δ_{ik} 及 a_{ij}，其中 $i=1,2,\ldots,c$, $k=1,2,\ldots,m$, $j=0,1,\ldots,m$. 我們應用最陡坡降法來調整這些參數直到誤差小於容忍值，如此我們就可得到一更精確的模糊系統。

四、結論與討論：

本研究如上所述已成功的提出一灰聚類演算法來改善一些聚類演算法在資料之群聚分析上之缺點，此外在只知道系統輸入輸出資料之前提下提出一有效建立模糊系統的方法並將所提方法應用於一些非線性系統的建模及倒單擺系統的控制上，從其結果可說明所提的設計方法確實有效，相信所提方法可提供研究聚類演算法、模糊建模與控制的學者參考。

References:

[1] Berenji H. R. and Khedkar P. S., “Clustering in product space for fuzzy inference,” *in Proc. of the Second IEEE Inter. Conf. on Fuzzy Systems*, pp. 1402-1407, 1993.

[2] Deng J., “Control problems of grey system,” *System & Control Letters*, vol. 1, no. 5, pp. 288-294, 1982.

[3] Deng J., “Introduction to grey system theory,” *The Journal of Grey System*, vol. 1, pp. 1-24, 1989.

[4] Sugeno M. and Kang G. T., “Structure identification of fuzzy model,” *Fuzzy Sets and Systems*, vol. 28, pp. 15-33, 1988.

[5] Takagaki T. and Sugeno M., “Fuzzy identification of systems and its applications to modeling and control,” *IEEE Trans. Syst. Man Cybern.*, vol. SMC-15, pp. 116–132, 1985.

[6] Wong C. C. and Chen C. C., “Data clustering by grey relational analysis,” *The Journal of Grey System*, vol. 10, no. 4, pp. 281-288, 1998.

[7] Wong C.C. and Lai H.R., “A grey clustering algorithm,” *National Conference on Grey Theory and Applications*, 2000. (In Chinese)

[8] Yoshinari Y., Pedrycz W. and Hirota K., “Construction of fuzzy models through clustering techniques,” *Fuzzy Sets and Systems*, vol. 54, pp. 7-13, 1993.