智慧型分群演算法—模糊C-回歸模型

An Algorithm for Data Clustering—Fuzzy C-Regression Models

龔宗鈞
大同大學
電機工程研究所
教授

張碩傑
大同大學
電機工程研究所

古宏麒
大同大學
電機工程研究所

蘇瑞堯
工研院機械所
智慧系統工程技術組
綠能智慧化系統部

關鍵詞

- 模糊 c-回歸模型分群演算法　FCRM
- 群集驗證準則　Cluster Validity Criterion

摘要

複雜的機械設備系統，其輸入與輸出的關係，往往不是單一關係式所能描述的，因此對於複雜的系統，必須先對輸入與輸出資料分成數個子群，對各子群分別進行系統辨証。本文提出一種對資料數據分群的方法，並有判斷法則，可決定最小的分群數目。模糊 c-回歸模型分群演算法(Fuzzy C-Regression Models Clustering Algorithm, FCRM Clustering Algorithm)是假設所有的輸入/輸出樣本資料屬於 c 個不同的函數模型。再根據各輸入/輸出資料點與各群集表示式(clustering representative)的符合程度來進行分群演算，因此特別適合於處理輸入/輸出資料具有函數關係時的分群問題。此外，模糊 c-回歸模型分群演算法需要一套有效可靠的群集驗證準則(cluster validity criterion)來決定最適當的群集數目。本文將介紹適用於模糊 c-回歸模型分群演算法之群集驗證準則，以決定模糊 c-回歸模型分群演算法中最適當的 FCRM 群集數目 c。同時，將對此群集驗證準則進行極限分析(limit analysis)，以研究該群集驗證準則之可靠度。

For a modern mechanical system, it is inadequate to describe the complicated behavior by a single mathematical model. Usually, the input-output data have to be clustered into several different groups before performing the system identification. This paper introduces a novel data clustering method and the criteria for determining the proper regression models. The Fuzzy C-Regression Models (FCRM) clustering algorithm assumes that the data space could be drawn

from c different regression models. The distance of object data vector to some representatives of cluster as the measure of goodness of fit is then replaced by the fitness of different regression models to each output. Although FCRM clustering algorithm is useful in fuzzy model identification, the reliable cluster validity criterion to guide the FCRM clustering algorithm is still lacking. Hence, it will be an important issue to design a cluster validity criterion suitable for FCRM clustering algorithm in order to choose the optimal number of clusters. In this article, we will introduce a cluster validity criterion suitable for FCRM, and take the limit analysis on it to understand its behavior as the weighting exponent $m \to 1^+$ and $m \to \infty$.

前言

複雜的機械設備系統中，常包含多個次系統或關鍵組件，甚至每個次系統或關鍵組件中，其數學模型(mathematical model)複雜，而無法以單一組控制方程式(governing equations)表示，也不容易從基本物理定律推導出其動態方程式。因此從實際系統運行時之實驗數據建立模型結構與參數的方法稱爲系統辨証(system identification)，並逐漸在各領域中廣爲使用。然而對於複雜的機械設備系統，其輸入與輸出的關係，往往不是單一關係式所能描述的，因此對於複雜的系統，必需先對輸入與輸出資料分成數個子群，對各子群分別進行系統辨證。本文提出一種對資料數據分群的方法，並有判斷法則，可決定最小的分群數目。

近年來，模糊分群演算法(fuzzy clustering algorithm)[1]-[4]理論被廣泛應用在複雜的非線性系統辨証上。模糊分群演算法的基本運作原理在於“相似度”與“差異性”：屬於同一群集的資料樣本必須彼此之間相似度愈高愈好，而屬於不同群集的資料樣本則彼此的差異性愈明顯愈好。其中，模糊 c-means 演算法(Fuzzy C-Means Algorithm, FCM Algorithm)[5]-[9]是最常見的模糊分群演算法。FCM 演算法是一種以「點的觀點」(point-wise)來進行分群(clustering)的演算法，此演算法首先計算資料樣本中每一點和各個群集中心點(cluster center)之間的幾何距離，進而計算出分群結果的誤差表現，最後分割出具有「超球體型式」 (hyper-spherical-shaped)的分類群集(cluster)。依此觀點，兩個幾何距離相近的資料樣本自然會被 FCM 演算法歸類在同一個分類群集中，所以十分適合處理圖形辨識與分割區塊的分群問題上。但是，以 FCM 演算法來進行系統輸入/輸出資料之間的函數關聯分群時卻不盡然合理，尤其當系統的輸入與輸出混雜有數種函數關係時，距離相近的輸入樣本並不見得會以相同的函數來產生輸出樣本。所以即使 FCM 演算法將兩組幾何距離相近的輸入/輸出資料樣本分在同一個分類群集中，卻不保證此兩組輸入/輸出資料具有同一個函數關係。

模糊 C-回歸模型分群演算法(Fuzzy C-Regression Models Clustering Algorithm, FCRM Clustering Algorithm)[10], [11]假設所有的輸入/輸出樣本資料屬於 c 個不同的函數模型，並根據各輸入/輸出資料點與各群集表示式(clustering representative)的符合程度來進行分群演算，因此特別適合於處理輸入/輸出資料具有函數關係時的分群問題。相較於 FCM 演算法以「點的觀點」產生 c 組「超球體型式」的群集中心，FCRM 分群演算法則是根據「函數關係」建構出 c 個具「超平面型式」（hyper-plane-shaped）的群集表示式。

以圖一來說明，$\mathbf{x}_1$、$\mathbf{x}_2$、$\mathbf{x}_3$、$\mathbf{x}_4$、$\mathbf{x}_5$ 與 $\mathbf{x}_6$ 代表須被分群的六筆資料，FCM 演算法以「點的觀點」進行分群演算，$\mathbf{x}_1$、$\mathbf{x}_2$ 與 $\mathbf{x}_3$ 由於幾何距離較相近，會被歸類爲“以 $\mathbf{v}_1$ 爲群集中心點”的群集，而 $\mathbf{x}_4$、$\mathbf{x}_5$ 與

更完整的內容
請參考【機械工業雜誌】328 期・99 年 7 月號
每期 220 元・一年 12 期 2200 元
劃撥帳號：07188562 工業技術研究院機械所
訂書專線：03-591-9342
傳真訂購：03-582-2011