以資料探勘技術建置多準則決策

黃光宇
嶺東科大資管系 教授
e-mail：kyhuang@teamail.ltu.edu.tw

林威廷
嶺東科大資管系 研究生
e-mail：a0m5001@stumail.ltu.edu.tw

摘要

隨著電腦資訊發展，以及大量的資料有如爆炸性的散布，讓人想從中找出有用的資訊越來越不容易。要想在其中找出有用的資訊，可以運用資料挖掘、機器學習等技術來獲得。其中如決策者在做決策時，就得以知識為基礎所建構的知識系統，來得到最佳決策。然而現在的決策環境日趨複雜，所需考量的不再只是單一目標決策方法可以處理，須藉由多目標決策方法來解決。因此，在本研究中將結合變精度粗集合理論（Variable Precision Rough Set，VPRS）、叢集效度指標函數(Cluster Validity Index method，CVI)、模糊集理論(Fuzzy Set Theory，FST)等理論方法來建置一個有效的多決策屬性分類方法，供後續多目標決策支援系統使用。

關鍵字：多準則決策、模糊理論、變精度粗集合、叢集效度指標、資料探勘

Abstract

In recent years, the decision makers required to resolve the issues of multiple objective decision-making(MODM) system rather than those of single objective decision-making (SODM) systems in the real world. In the study, we will therefore presents a novel approach to classifying datasets with multi-decision attributes. The hybrid scheme incorporates fuzzy set (FS) theory, Variable Precision Rough Set (VPRS) theory and a modified form of the PBMF index function. When the approach is built, it can be efficiently used for a MODM system in the future. This approach can help decision makers to derive suitable and the most appropriate decisions.

Keywords: multiple objective decision-making, Rough Set, PBMF index function, fuzzy set theory, Data Mining

1.前言

在過去數十年中，隨著電腦科技進步和收集資料的能力快速地發展，為了能夠提供決策者做出最佳的決策，就發展出由知識建構的知識系統，並廣泛地應用在許多真實的系統中。而大部份的系統資料量過於龐大，所以在處理這些資料時，就會透過機器學習、資料挖掘等技術來發現隱含知識，以獲得有用的知識。而面對日趨複雜難解的決策環境，決策問題之考量可能不再是單一目標決策方法可以處理，需藉由多個目標決策方法來解決，例如：民眾在購買房子時，以往考量的可能只是價格低，但現在考量到的有價格低、是否在學區、是否在捷運站旁、安全性等多重目標考量。

而多準則決策（Multiple Criteria Decision Making，MCDM）國內有學者亦稱為多水準決策或多評準決策，此種決策方法適用於經濟、科學、社會、管理等方面。依據[1]之分類，多準則決策又可分為多目標決策(multiple objective decision-making, MODM)及多屬性決策 (multiple attribute decision-making, MADM)。其中，多目標決策之主要方法，是

透過數學規劃之模式，以求得決策之替選方案；而多屬性決策之主要方法，則是利用評估各屬性之相對重要性，以找出各替選方案之最佳方案。至此，已有多位學者 [2、14-17] 曾對於多準則決策進行相關研究，例如（1）多屬性決策如何透過如加權法（Weighted Sum Model， WSM）、層級分析法（Analytic Hierarchy Process，簡稱 AHP）等，『讓決策者評估和選擇適當方案』；或（2）多目標決策如何利用多目標線性規劃（Multi-Criteria Linear Programming，MCLP；Multi-Objective Linear Programming，MOLP），『建立在一套最佳化的演算程序，決策者扮演選擇角色，無法直接設計可行方案』，但是這些方法得考慮到決策者偏好或專家意見。另外，以往學者對於因果(模型)如何產生的，卻未多加著墨。對於每一個目標決策與輸入資料間的關係，可以利用不同分類器的訓練加以獲得。然而，對於一個多輸入(條件屬性)及多輸出(決策屬性)的資訊系統，屬性間(條件對條件、條件對決策以及決策對決策)複雜關係，以及其所衍生的資料間不準確問題，(舉例：資料間不準確的情況為相同的條件屬性對應的決策屬性卻是不相同的狀況下)，皆未加以考量。且目前大部分的多目標決策未過濾到不確定性的資料，將會使的決策規則可能產生互相排斥的現象。而變精度粗集合它的優點可以處理資料不確定性問題，所以我們將利用變精度粗集合來解決這個問題，但是它在分類時的前提是各屬性分群是已知的，如此將會影響到分群的結果，所以以前所提出研究[7、10、11]即想到結合叢集效度指標即可解決此一問題，以達到分群及分類之最佳化方法。然而，此一方法是針對單一決策，為了處理多目標決策的問題，本研究將再結合模糊理論中的運算技巧，以拓展原先單一決策屬性分群暨分類指標方法的應用。

有鑑於此，我們將結合變精度粗集合、叢集效度指標方法及模糊理論等相關方法來處理多條件既多決策屬性的不確定性資料，以解決資料間複雜的因果關係，及資料不確定性間所產生不準確的問題。解決了這些問題，即可建立出有效的多決策屬性分類方法。

2.文獻探討

2.1 多目標決策

多目標決策是對多個相互矛盾的目標進行科學、合理的評估選擇，然後作出決策的理論和方法。它是 20 世紀 70 年代後迅速發展起來的一種決策方法。多目標決策與一般決策有所不同，它並不是從許多可行的方案中只考慮到單一一個目標選出的最佳方案。在多目標決策中，要同時考量到多種目標，而這些目標往往是很難比較的，甚至是彼此矛盾的；一般很難使每個目標都達到最優，作出各方面都很滿意的決策。因此多目標決策實質上是在各種目標之間和各種限制之間求得一種合理的妥協，這就是多目標最優化的過程。只有使這些相互聯系和相互制約的因素都能得到最佳的協調、配合和滿足，才是最優的決策。

如多目標決策中的多目標線性規劃文獻方法有(1) 針對山西壽陽旱農試驗區農業生態系統結構進行優化[15]。由於中國內陸北方旱地，農業綜合發展歷史悠久，是該國主要的中低產地區和重要的農業生態經濟區之一。但該地區水資源嚴重短缺、土壤侵蝕、植被稀疏、災害頻繁、生態環境脆弱。且農業生產面臨的最大問題是農業生態系統結構不合理，如糧食"一頭傾"、果園"一頭熱”等問題。所以中國大陸將於重點科技項目上，在系統辨識和明確結構優化的基本策略基礎上，採用多目標線性規劃(MOLP) 方法，以農業生產純收入、蛋白質產量和降水利用率最高作為目標函數，研究結果表明:引種飼草作物小黑麥，大力發展畜牧

業，與種植業有機結合，形成農牧系統良性循環結構，將提高旱地經濟效益、社會效益和生態效益，促進農業持續發展。
(2) 利用參數規劃技術求出多目標證券決策的有效證券組合集作為組合證券投資決策定量分析的依據[16]。該研究採用風險證券的收益率為預期收益的度量指標, 每年實際收益率與各年平均收益率偏差的絕對值之和為投資風險的度量指標建立存在無風險投資時, 證券組合投資決策的多目標線性規劃模型應用偏好係數加權轉化為單目標線性規劃模型。

而多屬性決策中的文獻方法有(1)AHP 應用於醫療服務選擇[2]，它可以有效地幫助病患，透過系統以評估他自己需要的醫療服務。(2) 另一研究運用多屬性決策方法(Multiple Attribute Decision Making, MADM) 之簡單加權和法(Simple Additive Weighting Method, SAW)、選擇法(Elimination et Choice Translating Reality, ELECTRE)、理想點法(Technique for Order Preference by similarity to Ideal solution, TOPSIS) 與灰關聯分析(Grey Relational Analysis, GRA)等技術的輔助來評估出較優良之觀光旅館業[17]，透過多屬性決策亦可以進一步的探討在不同的決策者喜好之下，各家觀光旅館整體表現之不同點。如此進行觀光旅館業的企業聲望評價與績效評估，期能為觀光旅館業者提供標竿學習之重要指標、投資者評估出較適合投資之標的、消費者評選旅遊住宿旅館的方法，或是提供與即將踏入社會之新鮮人作為評估就業目標之參考依據之一。

然而就上述運用的多準則決策方法中，可以了解到現今大部分的多準則決策得仰賴專家意見，或決策者偏好來得出最適當的決策。所以本研究將提出新的多目標決策分類方法以供後續多目標決策支援系統所使用。

2.2 變精準粗集合（VPRS）理論[13、19]

原粗集合理論採用精確集合概念定義上、下近似集合，描述數據之間的相關性。由於對數據的要求過於嚴格，導致存在一些不足之處，主要體現在：缺乏對雜訊數據的適應能力，抗干擾能力差；分類只有嚴格的“包含”和“不包含”，缺乏容錯能力(部分分類正確的資料集)；對於邊緣區域，不能區分等價類與集合的重疊度，沒有體現程度上的差別等。

所以原粗集合模型建立在絕對精確的包含關係之上，產生了極大的限制。因此 Ziarko 就提出了變精度粗集合模型，允許上近似和下近似存在一定的分類誤差，對原粗集合理論進行了擴展，主要分析了屬性間統計意義上的數據模式，或者存在概率上的不確定關係時的分類問題，而不是嚴格意義上的屬性函數依賴關係，增强了粗集合模型的數據分析能力。

2.3 模糊理論（FST）

2.3.1 模糊聚群演算法 (Fuzzy c-Means，FCM)演算法[18]

Mitra et al. [3]提到，一般而言，模糊集合（Fuzzy Set）適合用來處理不確定性的資料，其主要優點為能夠快速地提供推論結果。模糊集合在資料探勘應用非常廣泛包括分群、發現關聯規則（Association Rule）以及影像處理（Image Processing）等。Miyamoto [4]提到模糊分群演算法分為階層式模糊分群演算法（Hierarchical Fuzzy Clustering Algorithm）和非階層式模糊分群演算法（Non-Hierarchical Fuzzy Clustering Algorithm）。在非階層式模糊分群演算法中，最有名和最常使用的就是 目標函數基礎的模糊聚群演算法 (FCM)演算法，最早由 Dunn [5]所提出，隨後 Bezdek [6] 將 Dunn 所提出的 FCM 演算法一般化，加入模糊度參數的概念。FCM 其主要是針對 k-means 演算法進行改良，FCM 就如同 k-means

一樣，利用目標函數(objective function methods)的概念，對於每一個可能的分割進行評估，再藉由最佳化目標函數，使之小於設定之容忍誤差，達到最佳的分群。

但 FCM 與 K-means 不同的是歸屬矩陣，不再是二位元矩陣。由於傳統分群分析屬於明確群集(crisp clustering)，不同類別間有著明確的分割區域，同一個樣本，只能屬於一個類別。而在現實中很多都是無法明確界定的，許多介於不同類別間模糊地帶的樣本，常常無法明確定義其歸屬哪個特定類別。FCM 加入了模糊理論的概念，使得每一輸入項不再僅歸屬於某一特定的群集，而以歸屬程度來表現屬於各群集的程度，歸屬程度的數學化表示便是利用歸屬函數式。

2.3.2 模糊理論的運算關係[21]

模糊邏輯推論的創造概念類似人類的思維方式，它是利用語意變數(linguistic variable)來代表以往的經驗或專家建議，模擬人類的行為構成條件式的模糊規則，再經由模糊推論機構(fuzzy inference machine)模仿人類下決策之近似推理(approximate reasoning)模式，將此條件式模糊規則轉化為決策策略，以得出推論。

而其中模糊推論引擎是模糊系統中負責模糊推理運算的核心單元，其是依據近似推理的概念發展而來，應用於模糊系統中的模糊推論方法有許多種，例如：Mandani 的 min-min-max 模糊推論法、Larsen 之 min-product-max 模糊推論法、Tsukamoto 之模糊推論法和 Takagi 與 Sugeno 之模糊推論法等。由於有上述的優點及方法，所以本研究將予以納入。

2.4 叢集效度指標 (CVI)

叢集效度指標可大分為兩大類，一為針對 FCM 分群法來進行效度評估，例如 PBM(F)-index[7、8]，這些指標都是需要使用到 FCM 分群法所獨有的模糊權重矩陣 U，因為 FCM 是使用一個模糊權重矩陣 U，來判斷資料點應歸屬於哪一群，所以這些指標皆會使用到 U 矩陣中的值；另一為對任意分割式分群法皆適用的效度指標，有 Bcri t index、SV index 等，上敘這些方法都已被廣泛運用在處理無監督式的分群問題上。

而在過去的叢集效度指標方法（Cluster Validity Index）用於處理連續值資料的最佳化分群方法，只考慮到對資料本身的分群問題（其分群的對象是資料本身），並未對存在資料中各種屬性以及彼此間關聯性加以考量。對於一個特徵明顯的簡單訊息系統，或許這樣的分群方法所獲的之分群結果可以使用。然而，對於一個真正的訊息系統，每筆資料必須利用到很多個屬性才可以描述其特徵，對每一個屬性而言，應視為訊息系統中的獨立參數。因此，在使用傳統分群方法（對資料加以分群）來描述此一真實訊息系統時，往往因為條件屬性（自變數）與決策屬性（因變數）的複雜關係，若利用傳統叢集效度指標分群方法進行分群，其結果僅能提供分群結果的真實性（realness）,但卻無法提供分群結果的品質（伴隨產生之分類結果）好壞（goodness）。因此，在建構資料的分類演算法時，若能考量對屬性加以分群，或許可以更精準地描述資料屬性間複雜關係，對於獲得更多有效的分群知識有極大的幫助。

3.研究方法

3.1 模糊理論（Fuzzy Set Theory）

3.1.1FCM

FCM 屬於一種無監督式的分群運算法則，其應用範圍很廣。從特徵分析、分群甚至

到分類設計都有。Dunn 在 1973 中首先提出，以傳統的 k-means 結合模糊理論而成。除了引入模糊隸屬度的概念之外，並以目標函數來判斷元素與中心點的相似度，而且在尋找最佳相似度時，採用最小目標函數的準則；目標函數事實上就是糊模集群分析轉換古典數學中的最小平方法（method of least squares）的誤差平方和函數（squared error function）而來的。隨後 Bezdek 在 1981 加以改良， FCM 演算法的基本概念及演算法則如下：

$$\sum_{j=1}^{p}\mu_j(x_i)=1,\quad i=1,2,3...,k \qquad (1)$$

其中 p 是分群的群數，k 是資料筆數， x_i 是第 i 筆的資料，而 $\mu_j(x_i)$ 則代表 x_i 在第 j 群的歸屬函數值，也就是 x_i 隸屬於第 j 群的程度。由方程式可以很清楚地看出，某筆資料中任何一個屬性，其對應的所有分群歸屬函數總和一定要等於 1。FCM 的目的將目標函數最小化，也就是

$$J=\sum_{i=1}^{p}J_i=\sum_{j=1}^{p}\sum_{i=1}^{n}\left[\mu_j(x_i)\right]^{m'}\left\|x_i-c_j\right\|^2,\quad 1<m'<\infty \qquad (2)$$

其中 p 是分群的群數， n 是資料筆數，而 $\mu_j(x_i)$ 則代表 x_i 在第 j 群的歸屬函數值，x_i 是 第 i 筆的資料， m' 則是模糊化參數，而 c_j 則是第 j 群的群心。

為了符合所有分群歸屬函數總和等於1的必要條件，根據上述方程式可推導出下列方程式：

$$c_j=\frac{\sum_i\left[\mu_j(x_i)\right]^{m'}x_i}{\sum_i\left[\mu_j(x_i)\right]^{m'}}\ for\quad 1\le j\le p \qquad (3)$$

$$\mu_j(x_i)=\frac{\left(\frac{1}{d_{ji}}\right)^{\frac{1}{m'-1}}}{\sum_{k=1}^{p}\left(\frac{1}{d_{ki}}\right)^{\frac{1}{m'-1}}}=\frac{1}{\sum_{k=1}^{p}\left(\frac{d_{ji}}{d_{ki}}\right)^{\frac{1}{m'-1}}}\ for\quad 1\le j\le p,\quad 1\le i\le n \qquad (4)$$

3.1.2 模糊理論的運算關係

對模糊理論的運算係針對歸屬函數而言。舉例而言，假設資料集中每一筆資料有 M 個決策屬性，又假設第 l 個決策屬性（ d_l ）及第 k 個決策屬性（ d_k ）被分為 p_j 及 p_k 個群數， ${}_{d_l}\mu_{ji}$ 及 ${}_{d_k}\mu_{ji}$ 各代表資料集中第 j 筆資料 x_j 的第 l 個決策屬性（ d_l ）及第 k 個決策屬性（ d_k ），屬於第 i 個 DACV 分群向量的歸屬函數值。在本研究中可能採用模糊理論的運算關係有多種選擇，來獲得每一筆資料的整併歸屬函數，包括有 minimize, maximize, product 等運算關係。若以 minimize（最小化）運算關係而言，其定義如下：

$$ {}_d\overline{\mu}_{ji}=\min({}_{d_1}\mu_{ji},{}_{d_2}\mu_{ji},\cdots{}_{d_M}\mu_{ji}) \qquad (5)$$
$$={}_{d_1}\mu_{ji}\wedge{}_{d_2}\mu_{ji}\wedge\ldots\wedge{}_{d_M}\mu_{ji}$$

其中 ${}_d\overline{\mu}_{ji}$ 是第 j 筆資料 x_j、第 i 個 DACV 分群向量的歸屬函數值的乘積。舉例而言，假設決策屬性的個數是 3，且每一個決策屬性被分為 2 群；且假設第一筆資料（ x_1 ）中每一個決策屬性的分群指標（每一個決策屬性被歸屬於某一分群）分別為 $C_{d_1}(x_1)$=2、 $C_{d_2}(x_1)$=1，以及 $C_{d_3}(x_1)$=2。在此，先定義決策屬性分群指標向量（cluster vector of decision attributes，簡稱 DACV）為 $C_d(x_1)=\{2,1,2\}=c$ 。又假設伴隨第一筆資料各個決策屬性的最大歸屬函數值，各為 $\mu(C_{d_1}(x_1))=0.62$ 、 $\mu(C_{d_2}(x_1))=0.95$，以及 $\mu(C_{d_3}(x_1))=0.73$。因此，透過模糊理論的最小化、最大化及乘積運算關係，第一筆資料之 DACV 即為 $\mu_{\min}(C_d(x_1))={}_d\overline{\mu}_{jc}$ =min(0.62,0.95,0.73)=0.62 ， $\mu_{\max}(C_d(x_1))=\max((0.62,0.95,0.73)=0.95$，及 $\mu_{prod}(C_d(x_1))=0.62\times0.95\times0.73=0.43$。

3.2 變精度粗集合（VPRS）理論

集合理論是在 1982 年 Pawlak [9] 提出，是一種能從關聯式資料表抽取其中的知識法則的一種理論。在粗集合理論中，知識被認為是一種對抽象或現實的對象進行分類的能力，根據所討論對象的特徵差異，其某種分類的結果均可看作是某種知識的呈現。粗集合主要針對不確定性的資料進行分析，Pawlak 在粗集合中所提出的決策分析方法，是用來找出物件的關鍵屬性，建立物件集合的上下近似集合。

但是由於原來的粗集合容錯能力低，包含度過於精準，如果資料獲取不易，將使得最後得到的可用知識太少，所以本研究將利用Ziarko學者所提出了變精準粗集合，是粗集合的擴充，藉由精度參數β的引進，可以處理集合中，僅有部分分類正確的資料集合問題。其中的β值在本研究中代表著某一條件屬性皆相同之特定集合中的元素，屬於某一決策屬性集合的正確分類比例，因此，本研究將變精準粗集合分類精度的技巧納入到原先的叢集效度指標函數[18]中，建立一種新的分類方法。

3.2.1 資訊系統

一個資訊系統可被描述為$S=(U,A,V_q,f_q)$，其中U是一個非空集合所組成的有限物件集合，而A則是一個非空集合所組成的有限屬性集合。假定，在集合A的屬性可以分成條件屬性$C\neq\phi$，以及決策屬性$D\neq\phi$，其中$A=C\bigcup D$並且$C\bigcap D=\phi$。如果將所有屬性（包含條件及決策）以表列出，則此一表格稱為決策表。若某一屬性$q\in A$，V_q代表屬性q的值域，也就是$V=\bigcup V_q$。最後，$f_q:U\times A\to V$為訊息函數，代表著就$\forall q\in A$以及$\forall x\in U$，可以獲得$f(x,q)\in V_q$的關係。

對於C的每個子集$P\neq\phi$，表示為$P\subseteq C$，相等關係可由以下的式子來表示$I=\{(x,y)\in U\times U: f(x,q)=f(y,q)\forall q\in P\}$。而其所對應的相等集合（equivalence class）則表示為$U/I(P)$，代表具有相同條件屬性的集合元素，完全隸屬於相同決策屬性的集合。

3.2.2 β-下近似和β-上近似集合[20]

假設訊息系統中的決策表（decision table）可表示為$S=(U,A,V_q,f_q)$，其中$X\subseteq U$且$R\subseteq A$，其中$A=C\bigcup D$、$X\subseteq U$、$P\subseteq C$。在利用VPRS（$0.5<\beta\leq 1$）模型處理一個資訊系統S時，資料分析程序與二個基本觀念密不可分，也就是β-下近似和β-上近似集合。β-下近似集合可以表示成：

$$\underline{R}_P^{\beta}(X)=\bigcup\left\{I(P):\frac{|I(P)\bigcap X|}{|I(P)|}\geq\beta\right\} \qquad (6)$$

同樣地，β-上近似集合可以表示成：

$$\overline{R}_P^{\beta}(X)=\bigcup\left\{I(P):\frac{|I(P)\bigcap X|}{|I(P)|}>1-\beta\right\} \qquad (7)$$

當$\beta=1$時，$\underline{R}_P^{\beta}(X)$和$\overline{R}_P^{\beta}(X)$則分別變為RS理論的下近似和上近似集合，變精準粗集合模型則轉變為最初的粗集合模型。

Beynon[12]則改變一種表達方式，以β-negative區域和β-boundary區域來定義：

$$NEG_P^{\beta}(X)=\bigcup\left\{I(P):\frac{|I(P)\bigcap X|}{|I(P)|}\leq 1-\beta\right\} (8)$$

$$BND_P^{\beta}(X)=\bigcup\left\{I(P):1-\beta<\frac{|I(P)\bigcap X|}{|I(P)|}\leq\beta\right\} (9)$$

其中$|I(P)|$代表集合中元素的個數。

在確定β-上集合、β-下近似集合後，可獲得近似分類準確度（Accuracy of approximation）。準確度主要是在解釋以R等價類分配的近似集合，集合中x元素是否被正確的分類。如果α_c^{β}接近1代表分類與決策表是相當一致的，屬性集合R運用得當，近似分類準度被定義為：

$$\alpha_c^{\beta}=\left|\underline{R}_P^{\beta}(X)\right|\Big/\left|\overline{R}_P^{\beta}(X)\right| \qquad (10)$$

其中$X=\{x:C_d(x)=c,\forall x\in U\}$，$\left|\underline{R}_P^{\beta}(X)\right|$和$\left|\overline{R}_P^{\beta}(X)\right|$代表決策屬性屬於第$c$分群所形成的下近似集合和上近似集合的元素個數。$\alpha_c^{\beta}$代表決策屬性屬於第$c$分群的分類準度（$\beta$-下近似集合和$\beta$-上近似集合元素個數的比值）。

3.3 結合RS理論的分群暨分類指標方法

3.3.1Huang-index 方法與參數

Huang-index 方法[7]中，將粗集合理論納入原先的PBMF-index方法中，因此可以同時

針對分群群數以及分類準度進行最佳化處理。假設U是資料集中的元素集合， 而R代表近似集合，則 RS 問題可以表示如下：

$X \subseteq U$ is ：$(\underline{R}_P(X), \overline{R}_P(X)), BND_P(X)$，

其中X 是資料集，而U/I_P 被稱為 U中的不可辨識子集；I_P是關於R屬性集的不可辨識關係； ϕ 代表空集合；R 是 X 資料集的屬性集，此一屬性集包含條件屬性集(C)以及決策屬性集(D)；$P \subseteq C$；$\underline{R}_P(X)$ 代表 X 資料集中的下近似集合；$\overline{R}_P(X)$ 代表 X 資料集中的上近似集合；而 $BND_P(X)$稱為 X 資料集中的邊界集合。

Huang-index 方法所提出的分群暨分類模型中，資料集中的每筆資料 (X_i)有n個條件屬性($C_1 \sim C_n$)和一個決策屬性 (d)。與以往的 PBMF-index 分群模型相比，Huang-index 方法分群的對象是資料集中的屬性，而 PBMF-index 方法分群的對象則是資料集中的資料。

為擴大 PBMF-index 方法的適用性，我們發展出 Huang-index 方法。Huang-index 公式為

$$H(C, \alpha_C) = \left(\frac{1}{C} \times \frac{E_1'}{F_C'} \times D_C' \right) \tag{11}$$

而其參數定義如下：

第一項參數中的C是決策屬性的分群數目。

第二項參數中的F_C'值，則取決於所有分群E_c'的總和，其公式為$F_C' = \sum_{c=1}^{C} E_c'$；其中的$E_c'$應該被定義為

$$E_c' = \sum_{j=1}^{n} \overline{\mu}_{cj}^{m'}\left(x_j(d)\right) \left\| x_j - z_c' \right\| \Big/ \alpha_c \tag{12}$$

有關於上述等式右邊中的$\overline{\mu}_{cj}\left(x_j(d)\right)$代表資料$x_i$中決策屬性$d$的第$c$個歸屬函數值、而$z_c'$則代表決策屬性$d$ 第c個分群的群心（將所有屬於決策屬性d 第c個分群之下近似資料集加以平均，即可獲得此一群心）、α_c則是決策屬性d 第c個分群之分類準度、而n則是在數據集中的資料總數。

最後第三項參數，D_C'代表根據決策屬性加以分群所得到不同兩組群心之距離中的最大值，公式記作

$$D_C' = \max_{i,j=1}^{C} \left\| z_i' - z_j' \right\| \tag{13}$$

在 Huang-index 公式中第二項參數中的F_C'（由E_c'所組成）與 PBMF-index 方法介紹中所定義的$J_{m'}$有一些不同的地方。

在這裡，我們使用一個明顯地的例子可以很容易並清楚地解釋E_c'如何影響其所組成的F_C'。舉例而言，假設每筆資料中有三個屬性，包括二個條件屬性和一個決策屬性，並且每個屬性皆被劃分成 3 群。第一筆資料x_1決策屬性的歸屬函數值假設分別為$\overline{\mu}_{11}(x_1(d)) = 0.78$，$\overline{\mu}_{21}(x_1(d)) = 0.05$，以及 $\overline{\mu}_{31}(x_1(d)) = 0.13$。假設第一筆資料$x_1$的資料值為$x_1(2.5,10.4,7.2)$，並且假設屬於決策屬性第二分群的下近似集合群心是$z_2'(3.5,9.9,6.2)$。 因此，可以獲得$|x_1(a_1) - z_2'(a_1)| = |2.5 - 3.5| = 1.0$， $|x_1(a_2) - z_2'(a_2)| = |10.4 - 9.9| = 0.5$，以及 $|x_1(a_3) - z_2'(a_3)| = |7.2 - 6.2| = 1.0$。所以，$x_{12} = x_1 - z_2'$ 向量被表示為$[x_{12}(a_1), x_{12}(a_2), x_{12}(a_3)] = [1.0, 0.5, 1.0]$其所對應的 norm $\|x_{12}\| = \sqrt{x_{12}(a_1)^2 + x_{12}(a_2)^2 + x_{12}(a_3)^2}$ $= \sqrt{1.0^2 + 0.5^2 + 1.0^2} = 1.5$ 。

3.3.2 Huang-index 函數與 PBMF-index 函數的差異點

Huang-index 函數是從 PMBF-index 函數所延伸出來的一種索引函數，其公式為$H(C, \alpha_C) = \left(\frac{1}{C} \times \frac{E_1'}{F_C} \times D_C' \right)$。這個索引函數包括三項參數，即，$1/C$、$E_1'/F_C'$和$D_C'$。二種函數之間的差別列於表 1 中。如表 1 所顯示，第一項參數與決定屬性分群數目C呈反比，隨著決定屬性分群數目的增加，此一 Huang-index 值將會減少。第二項參數包含兩個變量E_1'和F_C'，其值由兩者的比值所組成。

其中 E_1' 對一特定的資料集而言是個常數，與 PBMF-index 方法中的 E_1 值相同。F_C 由 E_c' 所組成，而 E_c' 則由決策屬性 d 第 c 個分群之分類準度所組成。通常 E_c' 會隨著決定屬性分群數目 C 增加而減少。因此，當 E_C' 減少時 Huang-index 的值會增加。第三項參數，D_C'，根據決策屬性 d 加以分群後，所得到的群心中，測量不同兩組群心的最大分離距離。此一參數值當資料集中的每筆資料各為一獨立分群時，兩筆資料間的最大分離距離值為其限制值。

由 Huang-index 函數的表示式可以看出，此一方法可以同時對分群數目及分類準度提供最佳的解決方法。

3.3.3 多決策屬性的指標函數

多決策屬性的指標函數是從 Huang-index 函數所延伸出來的一種指標方法，其公式為 $MD(C)=\left(\frac{1}{C}\times\frac{E_1}{\overline{F}_C}\times\overline{D}_{N_I}\right)$，其中的參數 C 與 E_1' 與 Huang-index 函數是一樣。C 是決策屬性的分群數目；E_1 對一特定的資料集而言是個常數，與 PBMF-index 方法中的 E_1 值相同。

而 $\overline{F}_C$ 與 $\overline{D}_C$ 則與將模糊理論中的運算關係納入 VPRS 理論後，與所獲得的 β-下近似集合結果有關，將與 Huang-index 函數中的 F_C' 及 D_C' 有所不同。$\overline{F}_C=\sum_{i=1}^{N_I}E_i'$，其中 N_I 是所有決策屬性分群向量（Decision Attribute Cluster Vector，簡稱 DACV) 的數目，$E_i'=\sum_{j=1}^{n}{}_d\overline{\mu}_{ji}^{m'}\left\|x_j-z_i'\right\|\Big/\alpha_i^{\beta}$。$E_i'$ 與資料到各群心的距離有關；其中 α_i^{β} 是分類精準度；${}_d\overline{\mu}_{ji}$ 是所有決策屬性，透過糊糊運算後所得到的合成歸屬函數值；m' 是模糊係數；n 則是資料及所有資料筆數。$\overline{D}_C$ 則代表不同 β-下近似集合群心間的最大距離，即 $\overline{D}_{N_I}=\max_{i,j=1}^{N_I}\left\|z_i'-z_j'\right\|$。

3.4 研究步驟

3.4.1 決策屬性分群向量（DACV) 及整併歸屬函數的建立

目前對於具連續值多目標決策分析，多數僅就『決策面』考量，例如多屬性決策『讓決策者評估和選擇適當方案』；或多目標決策『建立在一套最佳化的演算程序，決策者扮演選擇角色，無法直接設計可行方案』。這些方法大部分都沒有考量到『影響因素與決策之間的關係』。因此，本研究模型針對多決策屬性連續值訊息系統，如何以系統化的方式獲得決策屬性分群向量（Decision Attribute Cluster Vector，簡稱 DACV) 及整併歸屬函數的步驟如下：

（1）利用 FCM 方法將連續值訊息系統中的條件屬性及決策屬性模糊化。

（2）根據各分群之歸屬函數值律定各屬性的分群指標，並根據某個屬性各模糊分群的歸屬函數值中之最大值者，定義為該屬性的所屬分群。

（3）定義多決策屬性之分群指標向量（cluster vector of decision attributes，簡稱 DACV）：假設決策屬性的個數是 3，且每一個決策屬性被分為 2 群；且假設第一筆資料（x_1）中每一個決策屬性的分群指標分別為 $C_{d_1}(x_1)$=2、$C_{d_2}(x_1)$=1，以及 $C_{d_3}(x_1)$=2，則決策屬性分群指標向量（DACV）為 $C_d(x_1)=\{2,1,2\}=c$，其中 c 代表所有 DACV 第 c 種決策屬性分群指標的組合。

（4）利用模糊理論運算關係找出每一筆連續值資料多決策屬性的整併歸屬函數值。

根據本研究中擬採用模糊理論的最小化運算關係，獲得每一筆資料決策屬性的整併歸屬函數值。假設資料集中每一筆資料有 m 個條件屬性及 M 個決策屬性；就每筆資料所有決策屬性的合併歸屬函數值而言，假設第 l 個決策屬性（d_l）及第 k 個決策屬性（d_k）被分為 p_j 及 p_k 個群數，${}_{d_l}\mu_{ji}$ 及 ${}_{d_k}\mu_{ji}$ 各代表資料集中第 j 筆資料 x_j 的第 l 個決策屬性（d_l）及第 k 個決策屬性（d_k），屬於第 i 個 DACV 決策屬性分群向量的歸屬函數值。首先就決策屬性的歸屬函數值加以整併，該運算關係定義如下：

$$_{d}\overline{\mu}_{ji}=\min({}_{d_1}\mu_{ji},{}_{d_2}\mu_{ji},\ldots{}_{d_M}\mu_{ji})$$
$$={}_{d_1}\mu_{ji}\wedge{}_{d_2}\mu_{ji}\wedge\ldots\wedge{}_{d_M}\mu_{ji}$$

其中 $_{d}\overline{\mu}_{ji}$ 是第 j 筆資料 x_j、第 i 個 DACV 決策屬性分群向量的歸屬函數值的乘積。

（5）決定 VPRS 的 β-下近似集合(β-lower approximation)。

透過 VPRS 的決策表，可以決定資料是屬於 β-下近似集合、β-上近似集合或是邊界集合，$\underline{R}_P(X)$ 、 $\overline{R}_P(X)$ 和 $BND_P(X)$。

3.4.2 具連續值多決策屬性分群暨分類指標方法的建立

本研究將 FCM 方法、模糊理論運算關係、變精度粗集合、叢集效度指標函數加以整合，產生一個最適於具連續值多決策屬性之分群暨分類指標方法，其步驟以圖 1 表示，而詳細的步驟則如下所示：

（1）律訂每個連續值屬性分群群數的範圍為 $[2,\ N_{max}]$。

（2）利用 FCM 方法將訊息系統內屬性加以模糊分群。

一般而言，一個連續值資訊系統可以利用模糊分群技術，將其轉換成一個對等的模糊資訊系統。舉例來說，假設訊息系統內屬性的分群群數為 2，而符號 $\mu(a_1)=\{\tilde{A}_{11},\tilde{A}_{12}\}$ 代表屬性某筆資料的第一個屬性 a_1，隸屬於第一分群的程度為 $\tilde{A}_{11}$，而隸屬於第二分群的程度為 $\tilde{A}_{12}$。而在利用分群暨分類指標方法進行 FCM 模糊分群的過程中，假設所有條件及決策屬性之屬性值的範圍介於 $[\alpha,\beta]$ 之間且被分為 p_l 模糊分群。因此，此一具有連續資料的資訊系統 (U,A,V_q,f_q) 將被轉變為一模糊資訊系統 (U,A,Φ,d)；其中的 $\Phi=\{\tilde{A}_{lj}\mid l\le m,j\le p_l\}$、$\tilde{A}_{lj}=\mu_j(x_i(a_l))$（代表第 i 筆資料之第 l 條件屬性 a_l 的歸屬函數）。

（3）對每筆連續值資料中之條件及決策屬性給予正確的屬性分群。

根據指標函數 $C_{a_l}(x_i)=I_{\max}(\mu_j(x_i(a_l)))=Index(\max(\mu_j(x_i)))$ for $1\le l\le m, 1\le i\le n$ 的定義，將某個屬性各模糊分群的歸屬函數值中之最大值者，定義為該條件屬性或決策屬性的所屬分群。

（4）根據 DACV 計算具連續值多目標決策屬性之整併歸屬函數值，並確定 VPRS 中的 β-近似集合。

將上述步驟每筆連續值資料所獲得之每個屬性決策屬性分群指標，將多決策屬性之決策屬性分群向量（decision attribute cluster vector，簡稱 DACV)，利用模糊理論中運算關係將向量中每一個歸屬函數值，整併為單一的歸屬函數值。再根據『研究方法』中「近似集合」理論，可以獲得在決策屬性分群向量中的第 c 個分群向量的下近似、上近似及邊界集合的訊息。因此，透過對多決策屬性資料集下近似集合內元素與上近似集合內元素的比值的計算，伴隨多決策屬性中的第 c 個分群向量的 RS 分類準度即可獲得。

（5）計算隸屬於連續值多決策屬性每一分群向量之下近似集合的群心。

假設連續值信息系統的屬性個數不只一個，因此伴隨於多決策屬性每一分群向量之下近似集合的群心，必須計算該集合下所有資料之每個屬性的平均值(包含條件及決策屬性)。

（6）計算連續值多決策屬性集之叢集效度指標。

當下近似集合的分類準度及群心獲得後，透過連續值多決策屬性集之叢集效度指標方法，即可獲得分群數目及分類準度的結果。

（7）檢查終止條件是否滿足。

當計算完某一特定分群數目 N 下的叢集效度指標值，檢查此一分群數目 N 是否超過分群數目之最大值 N_{max}。若 N 尚未超過分群數目之最大值，將分群數目 N 增加 1 後，到第一步驟，重新進行 FCM、模糊理論運算關係、VPRS 及叢集效度指標的計算。直到終止條件滿足時，終止到第一步驟的遞增流程，並且進入最後一個步驟。

（8）確定連續值多決策屬性集之叢集效度指標值

一但終止條件滿足時，將所有 N = 2~ N_{max} 的叢集效度指標值加以比較，所得的最大叢集效度指標值即作為多決策屬性集之叢集效度指標值，此一數值代表每一屬性分群數目以及資料集整體分類準度的最佳化結果。

4.例子資料分類暨分群說明

（1）假設有 10 筆資料，每筆資料中有 2 個條件屬性（c_1及c_2）、2 個決策屬性（d_1及d_2），在此假設，每個條件屬性及決策屬性分類組數分群群數設定值皆為 2，自定義資料集的資料內容如表 2。

（2）利用 FCM 方法將訊息系統內屬性加以模糊分群

當每個條件屬性被設定分為兩群時，所有 10 筆資料之每個屬性的歸屬函數值如表 3。舉例來說，${}_j\mu_{12}$為第 j 筆資料x_j第 1 個屬性被歸為第 2 分群的歸屬函數值。因此，${}_1\mu_{12}$ $=0.0077$。

（3）根據指標函數

$C_{a_l}(x_i)=I_{\max}(\mu_j(x_i(a_l)))=Index(\max(\mu_j(x_i)))$ for$1\le l\le m, 1\le i\le n$的定義，將某個屬性各模糊分群的歸屬函數值中之最大值者，定義為該條件屬性或決策屬性的所屬分群，如表 4 所示。其中括號中的數值為每筆資料中選出條件屬性值的最大值。舉例而言，第 1 筆資料x_1之第一及第二個屬性的歸屬函數值（${}_1\mu_{11}$及${}_1\mu_{12}$）分別為 0.9923 及 0.0077。因此，

$C_{a_1}(x_1)=I_{\max}(\mu_1(x_1(a_1)),\mu_2(x_1(a_1)))$ = $I_{\max}(0.9923, 0.0077)=1$。

（4）(a)根據 DACV 計算具連續值多目標決策屬性之整併歸屬函數值，並確定 VPRS 中的β-近似集合。

由於模糊理論的最小化運算關係，獲得每一筆資料決策屬性的整併歸屬函數值。依自訂資料集，每一筆資料有 2 個條件屬性及 2 個決策屬性。

(b)由表 3 之決策屬性的歸屬函數值可以看出，舉例而言第 1 筆資料x_1來說明，第 1 個決策屬性（d_1）及第 2 個決策屬性（d_2）都被分為 2 個群數，${}_{d_1}\mu_{11}$及${}_{d_2}\mu_{11}$各代表第 1 個決策屬性（d_1）及第 2 個決策屬性（d_2）屬於第 1 個 DACV 決策屬性分群向量（[1,2]）的歸屬函數值，分別為 0.9290 及 0.9815；而${}_{d_1}\mu_{12}$及${}_{d_2}\mu_{12}$各代表第 1 個決策屬性（d_1）及第 2 個決策屬性（d_2）屬於第 2 個 DACV 決策屬性分群向量（[2,1]）的歸屬函數值，分別為 0.0710 及 0.0185。該運算關係定義如下：

${}_d\overline{\mu}_{j1}$ = $\min({}_{d_1}\mu_{j1}, {}_{d_2}\mu_{j1})$ =min(0.9290, 0.9815)=0.9290。

${}_d\overline{\mu}_{j2}$ = $\min({}_{d_1}\mu_{j2}, {}_{d_2}\mu_{j2})$ =min(0.0710, 0.0185)= 0.0185。

其餘資料的計算與上述相似，結果請參考表 5。

(c)其次，本步驟需確定 VPRS 中的β-近似集合，為節省篇幅，省略不同 DACV 決策屬性分群向量的β值的計算。第 1 個 DACV 決策屬性分群向量（[1,2]）的β值為 0.9119；第 2 個 DACV 決策屬性分群向量（[1,2]）的β值為 0.6958。

(d)觀察表 4 分群指標所形成的決策表，{x_1、x_2、x_3}具有相同的條件屬性分群向量（[1,1]）及決策屬性分群向量（[1,2]），$\alpha_1^\beta=\frac{3}{3}$=1.000＞0.9119，所以{$x_1$、$x_2$、$x_3$}屬於$\beta$-下近似集合。而{$x_4$、$x_5$、$x_6$、$x_7$、$x_8$、$x_9$、$x_{10}$}具有相同的條件屬性分群向量（[2,2]），但僅{x_4、x_5、x_6、x_7、x_8、x_9}具有相同的決策屬性分群向量（[2,1]）。就 RS 而言，這些資料屬於邊界集合，但就 VPRS 而言，因為$\alpha_2^\beta=\frac{6}{7}$≒0.8571＞0.6958，所以這些資料亦被納入β-下近似集合中。

（5）計算隸屬於連續值多決策屬性每一分群向量之下近似集合的群心。

(a) z_1' = $\frac{1}{3}\sum_{i=1}^{3}x_i$ = $\frac{1}{3}[(2.0+2.1+2.2), (4.1+4.2+4.3), (8.5+9.0+10.3), (1.7+2.2+2.3)]$ = (2.1,4.2, 9.27,2.07)；

(b) $z_2' = \frac{1}{7}\sum_{i=4}^{10} x_i = \frac{1}{3}[(2.0+2.1+2.2),$
$(4.1+4.2+4.3),(8.5+9.0+10.3),(1.7+2.2+2.3)]$
$=(3.20,2.30,5.73,3.83)$

（6）計算連續值多決策屬性集之叢集效度指標值當資料分為一群時，

(a)其群心

$z_1 = \frac{1}{10}\sum_{i=1}^{10} x_i$ =(2.87,2.87,6.79,3.30)；

(b)因此，$E_1 = \sum_{i=1}^{10}\|x_i - z_1\|$ =26.6840，其中 $\|x_i - z_1\|$ 的資料如表 6 所示。

(c)其次，$\overline{F}_2 = \sum_{i=1}^{2} E_i'$，其中 $E_i' = \sum_{j=1}^{n} {}_d\overline{\mu}_{ji}^{m'}\|x_j - z_i'\| \Big/ \alpha_i^{\beta}$。

(d)在此，舉 E_1' 而言，$E_1' = \sum_{j=1}^{n} {}_d\overline{\mu}_{j1}^{m'}\|x_j - z_1'\| \Big/ \alpha_1^{\beta}$，$\overline{\mu}_{j1}$ 的資料可以由表 5 獲得；$\alpha_1^{\beta} = \frac{3}{3}$ =1.000；$\|x_j - z_1'\|$ 的值計算在表 7。m' =1.5；

(e)同樣地，根據相同的步驟，E_2' =0.074；

(f) $\overline{F}_2 = \sum_{i=1}^{2} E_i' = E_1' + E_2'$ =4.500；

(g) $\overline{D}_2 = \max_{i,j=1}^{2}\|z_i' - z_j'\| = \|z_2' - z_1'\|$ =4.521。

(h)將這些參數值帶入 $MD(2) = \left(\frac{1}{2} \times \frac{E_1}{\overline{F}_2} \times \overline{D}_2\right)$，即可獲得此自訂資料集在分群數為 2 的條件下的指標值 1.658。

5.結論與建議

由於在以往探討文獻中，發現大部分的多準則決策方法得需要專家意見，或決策者偏好來建立。所以本研究將利用變精度粗集合它具有的容錯能力，使資料獲得不易時得以保留下來，及可以處理資料不確定性問題的優點，再結合叢集效度指標，以達到分群及分類之最佳化方法。然而，此一方法是針對單一決策，所以本研究將再結合模糊理論中的運算技巧(歸屬函數)，以拓展原先單一決策屬性分群暨分類指標方法，得出一個新的多目標決策分類之方法，以供往後研究的多目標決策支援系統使用，而未來研究中，將會應用於實證研究上，如商業、社會、科學等地方，提供有效的多目標決策準則。

參考文獻

[1] Hwang, C.L. and Yoon K., “Multiple Attribute Decision Making Methods and Applications:A State-of-the-Art Survey” ***Springer-Verlag***, 1981.

[2] Tseng, C.C., Hong, C.F. and Chang, H.L, “Multiple Attributes Decision-Making Model for Medical Service Selection: An AHP Approach” ***Journal of Quality***, Vol. 15, No.2, pp. 155-165, 2008.

[3] Mitra, S., Pal, S.K., and Mitra, P., “Data Mining in Soft Computing Framework: A Survey” ***IEEE Transactions on Neural Networks***, Vol. 13, No.1, 2002.

[4] Miyamoto, S.,“An Overview and New Methods in Fuzzy Clustering” ***Second International Conference on Knowledge-Based Intelligent Electronic Systems***, Vol. 1, pp.33-40, 1998.

[5] Dunn, J.C., “A Fuzzy Relative of the ISODATA Process and Its Use in Detecting CompactWell-Seperated Clusters” ***Journal of Cybernetics***, 1973.

[6] Bezdek, J.C., “Some New Index of Cluster Validity,” ***IEEE Transaction on Systems ,Man And Cybernetics -Part B: Cybernetics***, Vol. 28, No.3, 1998.

[7] Huang, K.Y., “Applications of an Enhanced Cluster Validity Index method based on the Fuzzy C-means and Rough Set Theories to Partition and Classification,” ***IEEE Transaction on Systems , Expert Systems with Applications***, Vol. 37, No.12, pp. 8757-8769, 2010.

[8] Zadeh, L.A., “Fuzzy sets,” IEEE Transaction on Systems , ***Information and Control*** , pp.338-353, 1965.

[9] Pawlak, Z., “Rough sets” ***International Journal of Information and Computer***

Sciences11, pp.341-356, 1982.

[10] Huang, K.Y. ,"An Enhanced Classification Method Comprising a Genetic Algorithm, Rough Set Theory and a Modified PBMF-Index Function," ***Applied Soft Computing***, pp. 46-63, 2012.

[11] Huang, K.Y, "A Heuristic Approach to Classifying Labeled/Unlabeled Data Sets," ***Journal of the Operational Research Society***, pp. 1248-1257, 2012.

[12] Beynon M, "Reducts within the variable precision rough sets model: a further investigation," ***European Journal of Operational Research***, pp. 592-605, 2001.

[13] Huang, K.Y.," Application of VPRS model with enhanced threshold parameter selection mechanism to automatic stock market forecasting and portfolio selection"，***Expert Systems with Applications***，pp.11652-11661，2009

[14] 王俊凱、蕭文龍、許昌齡，" Meta-Analysis of AHP in expert and information management for Knowledge Management"，***知識社群與資訊安全學術研討會***，2008

[15] 高德明、陈丽娟、梅旭荣、崔玉亭，" 晋东豫西旱地农业生态系统结构优化研究"，***中国农业大学学报***，pp.33-40，1997

[16] 徐大江，"证券投资决策的多目标线性规划方法"，***系统工程理论与实***，pp.46-52，1995。

[17] 林士彥，"應用多屬性決策評價觀光旅館業聲望"， ***觀光研究學報***，pp.101-123，2004

[18] 黃美玲、施啟翔，" 以模糊分群法結合邏輯斯迴歸於醫療輔助診斷"， ***第 17 屆模糊理論及其應用研討會***，pp.138-143，2009

[19] 王加阳、陈松乔、罗安，"可变精度粗集模型研究"，***計算機與數字工程***，pp.53-55，2005

[20] 鄢勝偉，" 叢集效度指標為基礎之分類方法的建立"， ***嶺東科技大學資訊管理與應用研究所碩士論文***， 2010

[21] 施宏政、葉榮木、蔡俊明，"應用模糊推論之人臉膚色光線補償"， ***Symposium on Digital Life and Internet Technologies***，2004

[22] 李秀琴、林孟郁、黃木榮，" 運用智慧型代理人與分析階層程序（AHP）於商品選購策略－以旅遊行程規劃為例"， ***資訊、科技與社會學報***，pp.1-12，2002

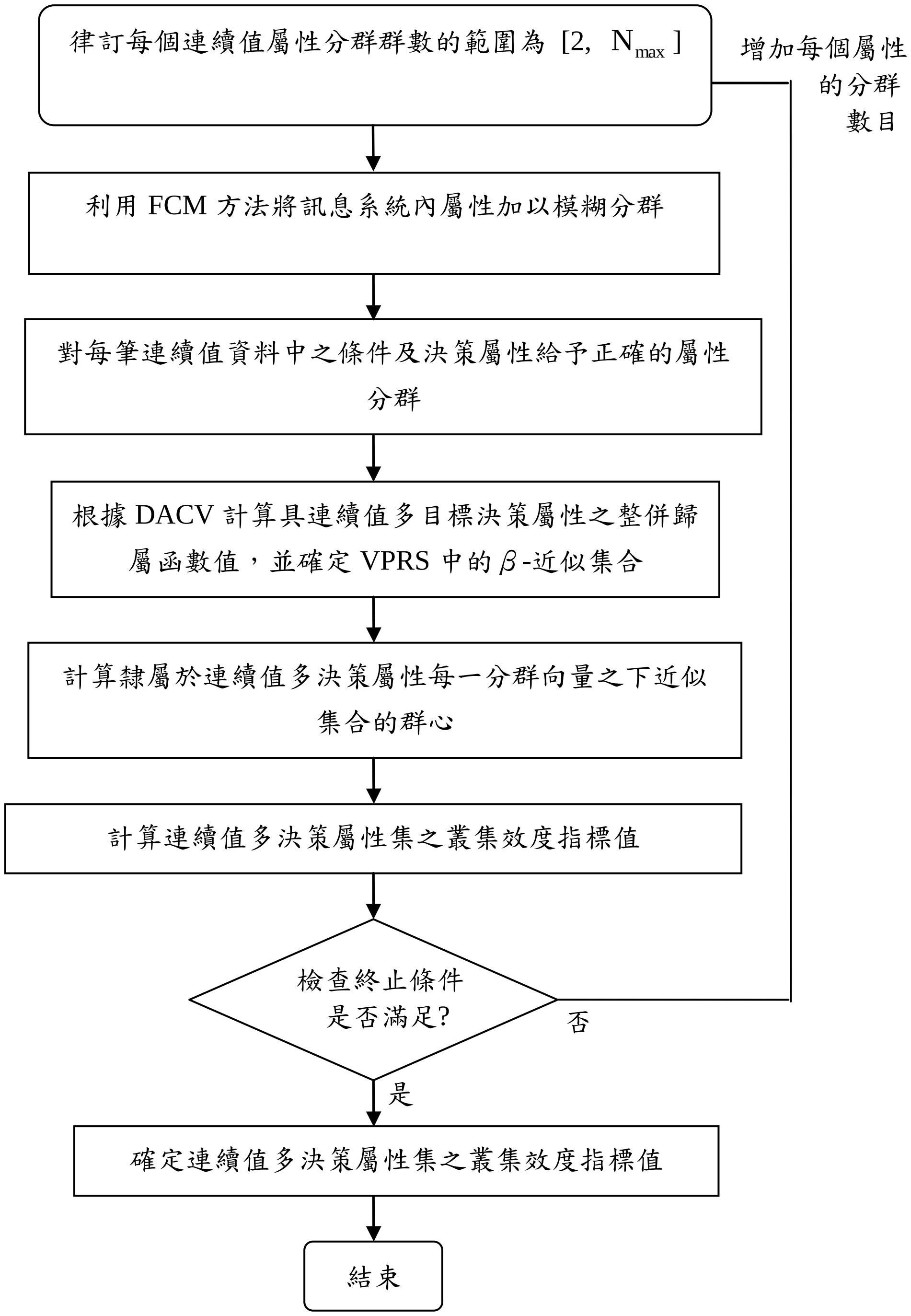

圖 1 多決策屬性集分類方法的流程圖

表 1. Huang-index 和 PBMF-index 兩種指標之間的差異

方法	Huang-index	PBMF-index
公式	$H(C,\alpha_c)=\left(\frac{1}{C}\times\frac{E'_1}{F'_C}\times D'_C\right)$	$PBMF(K)=\left(\frac{1}{K}\times\frac{E_1}{J_{m'}}\times D_K\right)$
如何對資料分群	對資料中的所有屬性加以分群	對資料集中的資料加以分群
	C代表決策屬性的分群數目	K代表資料集的分群群數
	$F'_C=\sum_{c=1}^{C}E'_c\,,E'_c=\sum_{j=1}^{n}\overline{\mu}_{cj}^{m'}\left(x_j(d)\right)\Vert x_j-z'_c\Vert\Big/\alpha_c$ (1) $\overline{\mu}_{cj}\left(x_j(d)\right)$代表資料$x_i$中決策屬性$d$的第$c$個歸屬函數值。 (2) z'_c則代表決策屬性d之第c個分群的群心。 (3.1) $\Vert x_j-z'_c\Vert$則是第 x_j 筆資料到第c個群心z'_c之間的距離。 (3.2) $E'_c=\sum_{j=1}^{n}\Vert x_{jc}\Vert$代表所有資料到第$c$個群心$z'_c$之間的距離與其對應歸屬函數乘積的總和，其中$\Vert x_{jc}\Vert=\overline{\mu}_{cj}^{m'}\left(x_j(d)\right)\Vert x_j-z'_c\Vert$ 。 (3.3) α_c則是決策屬性d第c個分群之分類準度（下近似集合與上近似集合中資料個數之比值）d	$J_{m'}=\sum_{k=1}^{K}E_k\,,E_k=\sum_{j=1}^{n}\mu_{kj}^{m'}\Vert x_j-z_k\Vert$ (1) μ_{kj} 為第j筆資料的歸屬函數值。 (2) z_k是利用 FCM 模糊分群所獲得之決策屬性第k分群的群心。 (3,1) $\Vert x_j-z_k\Vert$ 則是第 x_j 筆資料到第k個群心z_k之間的距離。 (3.2) $E_k=\sum_{j=1}^{n}\Vert x_{jk}\Vert$代表所有資料到第$k$個群心$z_k$之間的距離與其對應歸屬函數乘積的總和，其中$\Vert x_{jk}\Vert=\mu_{kj}^{m'}\Vert x_j-z_k\Vert$。
	$D'_C=\max_{i,j=1}^{C}\Vert z'_i-z'_j\Vert$ 根據決定屬性d加以分群所得到不同兩組群心之距離中的最大值	$D_K=\max_{i,j=1}^{K}\Vert z_i-z_j\Vert$ 根據資料集所獲得之任何兩組不同群心之距離中的最大值

表 2 自定義資料集

	$c_1(=a_1$，條件屬性 1)	$c_2(=a_2$，條件屬性 2)	$d_1(=a_3$，決策屬性 1)	$d_2(=a_4$，決策屬性 2)
x_1	2.0	4.1	8.5	1.7
x_2	2.1	4.2	9.0	2.2
x_3	2.2	4.3	10.3	2.3
x_4	2.9	2.0	4.4	3.4
x_5	3.0	2.1	4.5	3.5
x_6	3.1	2.2	5.2	4.2
x_7	3.2	2.3	5.3	4.3
x_8	3.3	2.4	6.0	5.1
x_9	3.4	2.5	4.7	4.5
x_{10}	3.5	2.6	10.0	1.8

表 3 資料歸屬函數值

	$c_1\ (=a_1)$		$c_2(=a_2)$		$d_1(=a_3)$		$d_2(=a_4)$	
x_j	${}_j\mu_{11}$	${}_j\mu_{12}$	${}_j\mu_{21}$	${}_j\mu_{22}$	${}_j\mu_{31}$	${}_j\mu_{32}$	${}_j\mu_{41}$	${}_j\mu_{42}$
x_1	0.9923	0.0077	0.9970	0.0030	0.9290	0.0710	0.0185	0.9815
x_2	1.0000	0.0000	1.0000	0.0000	0.9865	0.0135	0.0050	0.9950
x_3	0.9915	0.0085	0.9975	0.0025	0.9759	0.0241	0.0155	0.9845
x_4	0.1321	0.8679	0.0181	0.9819	0.0138	0.9862	0.7055	0.2945
x_5	0.0522	0.9478	0.0088	0.9912	0.0100	0.9900	0.7792	0.2208
x_6	0.0120	0.9880	0.0024	0.9976	0.0022	0.9978	0.9989	0.0011
x_7	0.0001	0.9999	0.0000	1.0000	0.0052	0.9948	0.9998	0.0002
x_8	0.0057	0.9943	0.0032	0.9968	0.0768	0.9232	0.9310	0.0690
x_9	0.0212	0.9788	0.0139	0.9861	0.0039	0.9961	0.9913	0.0087
x_{10}	0.0416	0.9584	0.0344	0.9656	0.9888	0.0112	0.0104	0.9896

表 4 分群指標及其對應最大屬性值

x_i	$C_{a_1}(x_i)$	$C_{a_2}(x_i)$	$C_{a_3}(x_i)$	$C_{a_4}(x_i)$
x_1	1（0.9923）	1（0.9970）	1（0.9290）	2（0.9815）
x_2	1（1.0000）	1（1.0000）	1（0.9865）	2（0.9950）
x_3	1（0.9915）	1（0.9975）	1（0.9759）	2（0.9845）
x_4	2（0.8679）	2（0.9819）	2（0.9862）	1（0.7055）
x_5	2（0.9478）	2（0.9912）	2（0.9900）	1（0.7792）
x_6	2（0.9880）	2（0.9976）	2（0.9978）	1（0.9989）
x_7	2（0.9999）	2（1.0000）	2（0.9948）	1（0.9998）
x_8	2（0.9943）	2（0.9968）	2（0.9232）	1（0.9310）
x_9	2（0.9788）	2（0.9861）	2（0.9961）	1（0.9913）
x_{10}	2（0.9584）	2（0.9656）	1（0.9888）	2（0.9896）

表 5 決策屬性整併歸屬函數值

x_j	${}_d\overline{\mu}_{j1}$（DACV=[1,2]）	${}_d\overline{\mu}_{j2}$（DACV=[2,1]）
x_1	0.9290	0.0185
x_2	0.9865	0.0050
x_3	0.9759	0.0155
x_4	0.0138	0.7055
x_5	0.0100	0.7792
x_6	0.0011	0.9978
x_7	0.0002	0.9948
x_8	0.0690	0.9232
x_9	0.0039	0.9913
x_{10}	0.9888	0.0104

表 6 E_1 的資料值

x_i	$x_i - z_1$	$\lVert x_i - z_1 \rVert$
x_1	(-0.870,1.230,1.710,-1.600)	2.785
x_2	(-0.770,1.330,2.210,-1.100)	2.908
x_3	(-0.670,1.430, 3.510,-1.000)	3.977
x_4	(0.030, -0.870,-2.390,0.100)	2.546
x_5	(0.130,-0.770,-2.290,0.200)	2.428
x_6	(0.230,-0.670,-1.590,0.900)	1.960
x_7	(0.330, -0.570,-1.490,1.000)	1.912
x_8	(0.430, -0.470,-0.790,1.800)	2.066
x_9	(0.530, -0.370,-2.090,1.200)	2.495
x_{10}	(0.630,-0.270,3.210,-1.500)	3.609

表 7 E_1' 的資料值

	$x_j - z_1'$	$\lVert x_j - z_1' \rVert$	$\overline{\mu}_{j1}^{m'} \lVert x_j - z_1' \rVert$
x_1	(-0.1000,-0.1000,-0.7667,-0.3667)	0.8615	0.7714
x_2	(0.0000,0.0000,-2.6667,0.1333)	0.2981	0.2921
x_3	(0.1000,0.1000,1.0333,0.2333)	1.0687	1.0303
x_4	(0.8000,-2.2000,-4.8667,1.3333)	5.5626	0.0090
x_5	(0.9000,-2.1000,-4.7667,1.4333)	5.4768	0.0055
x_6	(1.0000,-2.0000,-4.0667,2.1333)	5.1077	0.0002
x_7	(1.1000,-1.9000,-3.9667,2.2333)	5.0539	0.0000
x_8	(1.2000,-1.8000,-3.2667,3.0333)	4.9550	0.0898
x_9	(1.3000,-1.7000,-4.5667,2.4333)	5.5996	0.0014
x_{10}	(1.4000,-1.6000,0.7333,-0.2667)	2.2647	2.2268
$E_1' = \sum_{j=1}^{10} {}_d\, \overline{\mu}_{j1}^{m'} \lVert x_j - z_1' \rVert \Big/ \alpha_1^{\beta}$ =4.427			