

結合生物多樣性概念之群智慧混合演算法應用於資料分群

王欲韶

國立高雄第一科技大學 資訊管理所

u9924814@nkfust.edu.tw

黃承龍

國立高雄第一科技大學 資訊管理所

clhuang@nkfust.edu.tw

摘要

傳統群體智慧演算法在運算過程中，易落入區域最佳解，本研究提出結合生物多樣性及子群體之混合系統，其作法是混搭不同的演算法，以增加混合模式結構的多樣性，希求在求解過程中，不致於過早收斂及落入區域最佳解。

在本研究中實作兩種結合多樣性及群智慧演算法的模型，設計生物多樣性程度不同的兩種模型，並與生物多樣性程度較低之一般混合模型進行比較。應用在資料探勘中的分群領域，以多個公開資料集進行分群實驗。透過實驗結果可得知，本研究所提出的研究模型均能提高尋優的能力，表示生物多樣性的程度高低，的確會影響找尋優的能力。

關鍵詞：生物多樣性、子群體、群智慧演算法、分群、混合系統

壹、緒論

在傳統的尋優問題上，大多利用貪婪法(Greedy Method)進行求解，但該方式在處理問題時，有極大可能會在搜尋過程中落入區域最佳解，導致無法找到最佳解。因此在這樣的環境下，有學者提出來所謂的群體智慧(Swarm Intelligence)的方法，即透過觀察自然界的動物的行為所仿照的演算法，如：螞蟻的覓食行為。且近年來，群體智慧常被應用在解決各種最佳化問題上。而群體智慧中最常見的蟻群演算法(Ant Colony Optimization; ACO) (Dorigo et al. 1997)和粒子群演算法(Particle Swarm Optimization, PSO) (Eberhart and Kennedy 1995)。

ACO 可分為離散型最佳化問題的蟻群演算法(Mullen et al. 2009)及連續型最佳化問題的蟻群演算法ACO_r(Socha and Dorigo 2008)，其運作模式是根據螞蟻在尋找食物過程中，會在搜尋的路徑佈上費洛蒙(Pheromone)，運算時根據費洛蒙的濃度，不斷找尋新的路徑，經過迭代(Iteration)的計算來求得最佳解。

PSO 則是藉著粒子(Particle)在空間的移動來找尋最佳的位置，每個粒子根據本身所找尋的最佳值、整體粒子的最佳值和上一回合的加速度，三者產生出新的加速度，來改變目前粒子的位置，迭代運算之後會留下表現最佳的粒子而找到最佳解。

在傳統群體智慧演算法在迭代運算的過程中，其解集合內的個體會變得很相近，容易落入區域最佳解，研究者無不對這個議題，提出許多模型與方法以提高演算法的求解的效果。改良的方式通常有良好的模型參數設定、子群體(Sub-Populations)策略、混合系統(Hybrid Systems)等方式。近年來許多混和型式的演算法被應用在各個領域，是個頗為熱門的研究議題。

對於群智慧混合系統的研究，本研究提出一個思考的方向：如何讓具有多樣性特色的演算法混和在一起？

在研究中認為，如同自然界的運作方式，一個生態系統內蘊含了各種不同物種，而即使是相同類別的物種，也是由具不同特色的個體所組成，而這樣的狀態即是生物多樣性(Biodiversity)。若將此概念應用群體智慧演算法之改良，並設計出一個具有多樣性特色的混合演算法，則在求解複雜問題時可能會有較佳的能力。

而子群體(Spears 1994 ; Lovbjerg et al. 2001)是及混合演算法(Lozano and García-Martínez 2010)亦是提升解之多樣性的一種可行方式。故在此研究中亦將利用子群體及混合群智慧演算法方式，並加入基因演算法(Genetic Algorithms, GA)(Holland 1975)，以求解更具有多樣性。

綜合上述說明，本研究提出具有多樣性特性之群智慧混合系統，經由多種不同演算法混合運算，並結合生物多樣性的概念，透過遷移及子群體策略，藉此提高搜尋的效能，避免落入區域最佳解。最後經由資料分群(Data Clustering)方式，藉以測試本研究所提出方法，和傳統分群方式及單一群智慧分群演算法比較，確實可提升求解表的多樣性，進而獲得最佳解。

貳、相關文獻

(一)、資料分群

資料分群是主要是將資料分成數個群集(Clusters)，其作法為在搜尋資料中相似度高的群集，並找出各群聚的群心，透過群心將相似度大的資料分為同一群集，來達到資料分類成群的目標。

(二)、群體智慧

群體智慧屬於人工智慧的一種，此類演算法基於群體的行為(Swarm Behavior)。自然界中許多動物間覓食就是依照著群體的行為，如：鳥群、魚群、蟻群等。群體智慧演算法就是觀察這些自然界的動物的行為所仿照的演算法，近年來，群體智慧常被應用在解決各種最佳化問題上，如應用在網路的入侵偵測系統(Intrusion Detection System)上(Kolias et al. 2011)。

(三)、生物多樣性及子群體

生物多樣性(Wilson 1988)是指物種的豐富度(Species Richness)。根據生物多樣性公約(Convention on Biological Diversity)之定義(CBD 1992)，生物多樣性的組成可以分為基因多樣性(Genetic Diversity)、物種多樣性(Species Diversity)和生態系多樣性(Ecosystem Diversity)三種層次。

1. 基因多樣性：存在於物種內或物種間的基因多樣性。若物種的基因多樣性變異越豐富，則該物種對環境變動的適應力也就越強。
2. 物種多樣性：一個區域之內物種數目的多寡以及其數目組成。若區域內物種的多樣性越高，其好處為在因應大自然的演化過程中，物種之間產生了一種合作又競爭的現象，使得物種彼此之間共生又共榮。
3. 生態多樣性：在生態系內部由於環境差異和生態系組成的多種多樣，造就成豐富的系統。且在同一個區域，可能包含著多個生態系。地球上有許多生態系，例如草原生態系、沙漠生態系、海洋生態系等。

而在生物多樣性的概念中提到，自然界的運行不僅是追求生存，更重要的是生態系中的生態均衡，以求達到永續經營。同樣的，在尋優問題上，如何達到「平衡」將是探討重點。

亦有學者提到在尋優問題中，提高良好解的方法，在於如何達到「Exploration」及「Exploitation」的平衡(Gen and Cheng 1997)，後續學者更進一步提到 Exploration 及 Exploitation 又可被稱為是「Diversification」和「Intensification」(Clerc 2008)。其中 Intensification 是指在現有的良好區域上，搜尋鄰近區域是否有更好的解，反之 Diversification 則是透過跳脫現有的搜尋區域，往外尋找更好的解(Thangaraj 2011)。學者(Lozano and Garcia-Martínez 2010)再對 Diversification 有更深入的解釋，是指在搜尋最佳解過程中，擁有拜訪更多及不同搜尋空間的能力，藉以使得解不易掉入局部最佳解中。

因此，不少的研究學者提出了各種增加多樣性的方法。(1)如有相關學者提出了子群體來提升母體的多樣性(Spears 1994)。子群體是由全域群體(Global Population)切割成於多個子群體，每個切割出來的子群體可以示為獨立且不重疊之探索位置(Belding 1995)。(2)亦有學者提出和子群體概念相近的方式，利用可調式多樣性門檻值方法在於運算過程中，檢視母體的多樣性的密度是否介於門檻值容許的範圍內，若過低則導入外部的人工

染色體，藉以增加母體的多樣性(Chang et al. 2010)。(3)亦有利用搜尋不同的解空間達到多樣性的作用(Chang and Chen 2009)。

而本研究所指的多樣性跳脫在單一解空間求解之較狹義定義，而提出更為廣義的方式，是為具有物種多樣性(或是生態多樣性)特色之混和演算法：擁有多種混和演算法以「物種」模式(Species)之設計。即是在文獻中所提到的演算法及相關方法加以搭配，讓具有不同物種的群體共同運作。

(四)、遷徙

透過自然界中遷移作用，使得物種之間的交流頻繁，得就生物群日益增大。相同的在尋優問題上，亦可採用此方式，有學者提出島嶼基因演算法(Parallel Island GAs)(Cantu-Paz 1998)方式增加解的多樣性，遷移率(Migration Rate)為島嶼間每次資訊交換的比率，即遷移出的個體數量除以粒子群體的大小，因而求得島嶼間溝通交換的成員比例。透過遷移溝通行為，使得在探索解空間上增加多樣性，不易陷入區域最佳解，進而利用此優勢搜尋至全域最佳解，

(五)、混合系統

混合系統的概念是透過不同的演算法及概念所建構而成的系統。近年來，透過單一演算法處理問題，發展至今似乎有難已突破的瓶頸，且有學者提到單一演算法會比混合演算法差(Blum 2005)，故有學者提出混合方式(Thangaraj et al. 2011)，並證明該方式為提高演算法性能的不可缺少觀點，該方式可改善原本的演算法、並可得到品質更佳解以及混合模型可當成整體系統的一部分。

在其他方面，群體智慧結合其他演算法的應用也有非常多，有學者提出 PSO/ACO 的混合方法(Holden et al. 2007; Niknam 2009; Meng et al. 2007)。

而粒子群在分群上亦有廣泛的應用，有學者提出兩個新方法應用於資料分群，標準的 PSO 及結合 k-means 找尋新的粒子的新式 PSO(Merwe and Englbrecht 2003)。結果顯示結合 k-means 的混合演算效果較為良好。之後學者皆陸續發表出結合 k-means 的粒子群演算法(Zhenkui et al. 2008; Ahmadyfard and Modares 2008)，以及結合模糊適應粒子群演算法(Fuzzy PSO)、ACO 和 k-means 的演算法。亦有 PSO 結合其他的方式，如結合蒙地卡羅模擬法(Monte Carlo Method)(Gao et al. 2011)應用在電力系統上。亦有學者提出使用 ACO/K-means 方式解決高光譜影像(Chandrasekhar et al. 2011)。

GA 也有和群體智慧結合的應用，有學者將之應用在處理 DNA 問題上，利用 PSO 結合 GA 的方式，並搭配 FUZZY 的專家系統，處理醫學上的種種問題(Ganesh-Kumar et al. 2012)。群體智慧中最常見的兩種演算法—ACO 和 PSO，及亦導入基因演算法，詳述如以下小節。

(六)、連續型蟻群演算法

連續型蟻群演算法是用來解連續型數值最佳化，以往的作法大部份都是將連續型數值離散化後，再由離散型蟻群演算法處理，但此方式會因為離散化的細緻程度而影響結果。

ACO_R 演算法是利用一費洛蒙表(Pheromone Table)進行運算，此表有 k 組解，每一組解由 n 個決策變數所決定。連續型蟻群演算法主要運作如下：

1. 隨機亂數產生初始的 k 組解，並求出每一組解的目標函數(Fitness Value)。
2. 依目標函數的值將費洛蒙表排序，算出每一個解的權重 ω ，如公式(1)：

$$\omega = \frac{1}{qk\sqrt{2\pi}} e^{-\frac{(l-1)^2}{2q^2k^2}} \quad (1)$$

q =學習率，其值介於 0-1； k =費洛蒙的大小，代表有 k 組解

l =利用目標函數順序後的順序， $\omega_1 \geq \omega_2 \geq \dots \geq \omega_l \geq \dots \geq \omega_k$

3. 利用輪盤法方式，計算權重 ω_l 的輪盤比例 P_l 如公式(2)。輪盤上的機率分配代表費洛蒙表上解被選中之機率，被選中解用以建構下一代的解。

$$P_l = \frac{\omega_l}{\sum_{r=1}^k \omega_r} \quad (2)$$

4. 螞蟻產生多個新解，依照上述步驟(3)選出第 l 個解的第 i 個變數值當作 μ_i^i ，依照公式(3)計算標準差 σ_i^i ，接著使用常態分配 Normal(μ_i^i , σ_i^i) 方式產生新的解。

$$\sigma_i^i = \xi \sum_{e=1}^k \frac{|s_e^i - s_l^i|}{k-1} \quad (3)$$

ξ =費洛蒙的蒸發率，其值大於 0； l =費洛蒙表排名的順序

5. 費洛蒙表的更新是依照新解的目標函數與連續型費洛蒙表比較，如果新解的目標函數較好，則該新解取代費洛蒙表中較差的解。

(七)、粒子群演算法

粒子群演算法是模擬鳥類、魚的覓食的概念，每一隻鳥或魚的個體在粒子群演算法中稱為粒子(Particle)，每個粒子會隨著一個加速度去移動位置如公式(4)，而加速度會受到上一回合加速度、過去本身所找尋最佳解(P-Best)和群體目前為止所找尋最佳解(G-Best)所影響，且之後學者加入權重 ω 來平衡 global 搜尋和 local 搜尋如公式(5)。

$$X_i^{(t+1)} = X_i^{(t)} + V_i^{(t+1)} \quad (4)$$

$X_i^{(t+1)}$ = 粒子第 $t+1$ 的位置 $X_i^{(t)}$ = 粒子第 t 代的位置 $V_i^{(t+1)}$ = 加速度

$$V_i^{(t+1)} = \omega \cdot V_i^{(t)} + c_1 \cdot rand_1() \cdot (Pbest_i - X_i^{(t)}) + c_2 \cdot rand_2() \cdot (Gbest_i - X_i^{(t)}) \quad (5)$$

$V_i^{(t+1)}$ = 加速度; ω = 慣性權重; c_1 、 c_2 = 學習因子; $rand_1()$ 、 $rand_2()$ = 介於 0 和 1 之間的隨機變數; $Pbest_i$ = 過去本身所找尋到的最佳解; $Gbest_i$ = 群體目前為止所找尋最佳解
 $X_i^{(t)}$ = 粒子第 t 代的位置

(八)、基因演算法

基因演算法是由學者 John Holland 於 1975 年所提出，其流程是利用基因演算法中三項運算「選擇(Selection)」、「交配(Crossover)」及「突變(Mutation)」。在利用基因演算法運算前，必須將問題的需求做一番詳細的規劃，以定義出合適的適應函數(Fitness Function)及染色體(Chromosome)個數，而每條染色體所代表之意思即是該問題的最佳解。目前常用的編碼方式有整數編碼(Integer Encoding)、實數編碼(Real Value Encoding)及二進制編碼(Binary Encoding)等，在本研究中使用實數編碼。

基因演算法使用的範圍相當廣泛，(1)如有學者利用在找尋產業界中可合作的對象，並促使成為策略同盟(Martínez-Torres and Toral- Marín 2010)。(2)亦有學者提出改進方法，利用將染色體分段(Segment)方式，使用在樣式辨認(Pattern Recognition)上(Yang et al. 2011)。(3)或是使用在物理學上找尋粒子的能量分配亦可使用基因演算法(Ji et al. 2011)。(4)或是使用在探討排程問題中的模糊時間(Alcan and Başlıgil 2012)。

參、實驗架構

本研究的重點是具有物種多樣性(或是生態多樣性)特色之混和演算法：擁有多種混和演算法以「物種」模式(Species)之設計。即是在文獻中所提到的演算法及相關方法加以搭配，讓具有不同物種的群體共同運作。

(一)具生態多樣性與子群體特色之混合模式設計理念

在本研究中對於物種的定義更為廣義，可以是相同的物種，例如：居住在鄰近區域不同巢穴的螞蟻等；也可以是相同的物種的不同族群，例如：紅火蟻、熱帶大頭蟻等；也可以是完全不同的物種共同生存在一個生態系統中，例如：螞蟻群、鳥群等。

此處廣義的定義即是可以讓混和系統有更為靈活與彈性的結合再一起，在本研究中之設計的重點包括「各物種混合運作的架構」及「各物種之合作與競爭的策略」。在此指的合作與競爭策略，是透過不同的演算法混搭使用及解的比較，實作多樣化特色的混合系統，據以瞭解並比較物種多樣性(或是生態多樣性)，對於求解最佳化問題之效果。因此，本研究跳脫傳統的混和系統只有具備狹義的多樣性，設計一個具有廣義多樣化混和系統。

(二)系統流程說明

在本研究中實作了兩組具多樣性的模型(Hybrid-I & Hybrid-II)，Hybrid-I 和 Hybrid-II 的差別在於使用的種族(演算法)數量的不同。分別展示兩模型的示意圖，見圖 1 及圖 2，及在此利用 Hybrid-II 模型進行流程說明，以下為其步驟。

步驟一 初始適應解組合表

利用均勻分配方式亂數產生「適應解組合表」上的數值，其大小共為 k 組解。將

該運行表分割成三個子表，並且每一個子適應解組合表分配一組子族群演算法。

步驟二 分別處理三個子適應解組合表

將分割的三個子適應解組合表，使用三種子族群演算法分別平行處理。

1. 子族群演算法-1(PSO/ACO_R)

(Step-1) 依據式(4)計算「子適應解組合表-1」的加速度。

(Step-2) 依據式(5)計算「子適應解組合表-1」的新位置。

(Step-3) 將更新過後的「子適應解組合表-1」當作ACO_R的費洛蒙表，執行ACO_R。並計算「子適應解組合表-1」的績效值、權重 ω_i 值如公式(1)和比例 P_i 式(2)。

(Step-4) 利用(2)所計算出的值並搭配輪盤法的方式，選擇出「子適應解組合表-1」上的值，當作常態分配 $Normal(\mu_i^i, \sigma_i^i)$ 公式中的平均數 μ_i^i 。

(Step-5) 計算產生標準差 σ_i^i 並產生新解，如式(3)。

(Step-6) 更新「子適應解組合表-1」。如果新解的分群績效值比原來的分群績效值小，則該解取代原來的「子適應解組合表-1」。

2. 子族群演算法-2(ACO_R/GA)

(Step-1) 先計算「子適應解組合表-2」中的ACO_R演算法所需參數值。

(Step-2) 利用(2)所計算出的值並搭配輪盤法的方式，選擇出「子適應解組合-2」上的值當作常態分配 $Normal(\mu_i^i, \sigma_i^i)$ 公式中的平均數 μ_i^i 。

(Step-3) 產生標準差 σ_i^i 並產生新解。

(Step-4) 更新「子適應解組合表-2」。如果新解的分群績效值比原來的分群績效值小，則該解取代原來的「子適應解組合表-2」。

(Step-5) 計算「子適應解組合表-2」內 GA 演算法所需用之適應值。

(Step-6) 利用輪盤法選擇父母兩代，並隨機選取一交配點進行交配。

(Step-7) 選取交配點後交換動作，並產生新解。

(Step-8) 進行突變動作，並評估後代基因之適應值及更新「子適應解組合-2」。

3. 子族群演算法-3(GA/PSO)

(Step-1) 計算「子適應解組合表-3」內 GA 演算法所需用之適應值。

(Step-2) 利用輪盤法選擇父母兩代，並隨機選取一交配點進行交配。

(Step-3) 選取交配點後交換動作，進行交配並產生新解。

(Step-4) 進行突變動作，並評估後代基因之適應值及更新「子適應解組合表-3」。

(Step-5) 根據公式(4)計算「子適應解組合表-3」的加速度。

(Step-6) 根據公式(5)計算「子適應解組合表-3」的新位置，並更新「子適應解組合表-3」。

步驟三 遷移動作

1. 在每一子適應解組合表中隨機選擇 25%的解，並利用循序交換方式，交換彼此的解。且交換後，對每一個子適應解組合表和原本未交換前之表內的解進行比較。

2. 更新並移除每個子演算表中較差的解。

步驟四 判斷是否到達停止條件

判斷是否到設定的迭代數，如果尚未達到則從步驟二開始執行。

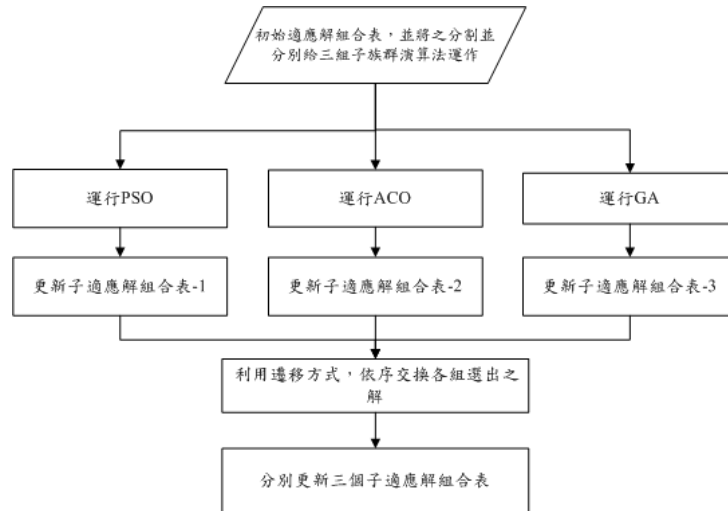


圖 1 Hybrid-I 模型流程圖

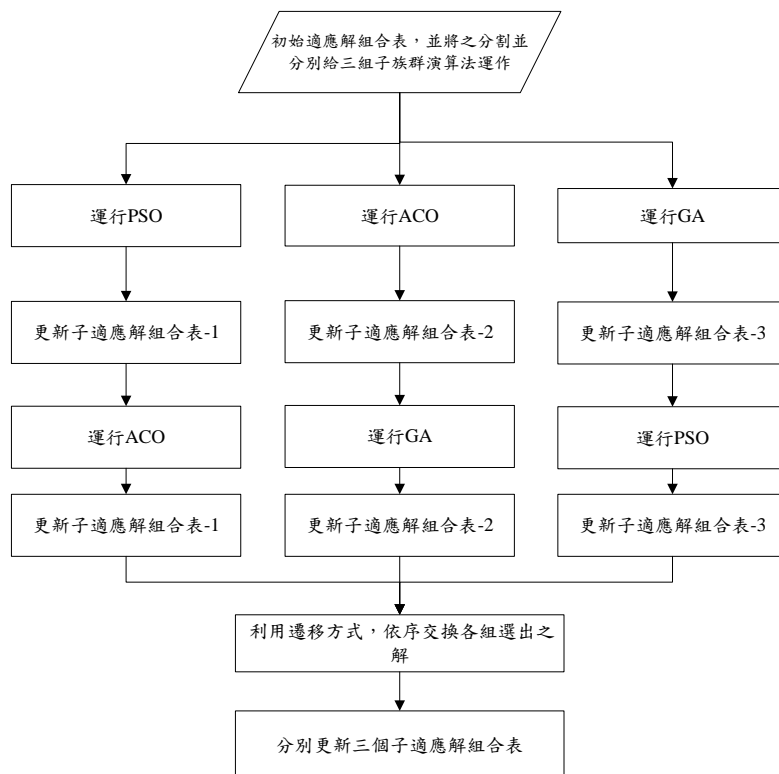


圖 2 Hybrid-II 模型流程圖

肆、實驗展示

(一) 測試資料集

在本研究中，共使用五組 UCI Machine Learning Repository 網站上之真實資料，包含二組醫學資料及三組一般期刊論文中常用資料集，如表 1 所示。

表 1 實驗中所使用的測試資料集

測試資料集名稱	屬性個數	資料筆數
Abalone	8	4177
Breast Cancer Wisconsin	10	683
Iris	4	150
Liver Disorders	6	345
Wine	13	178

(二) 比較模型設計

在本研究中將比較生物多樣性的程度多寡，為了證實的確可提升解的效能。分別比較(1)生物多樣性程度低的單一群智慧演算法及遺傳演算法-PSO、 ACO_R 及 GA，以及(2)生物多樣性程度中等的三組混合演算法-PSO/ ACO_R 、PSO/GA 及 GA/ ACO_R ，(3)最後再與本研究中所提出的兩組生物多樣性程度高之實驗模型-Hybrid-I 及 Hybrid-II 比較。

(三) 各實驗模型參數設定

1. 分群效度指標

當使用者不確定測試資料集最後可分成幾群時，則必須透過分群效度(Cluster Validity)的分析，以便找出資料最佳分群數。

在本研究中，為了比較分群的效果，採用 XBI (Xie and Beni's Index) (Xie and Beni 1991)指標來做評估。XBI 指標以各資料群與群中心的距離計算離目前群心之最短距離，來評估各組之間的總變異，該指標數值愈低表示分群效果愈佳，其公式(11)如下。

$$XBI = \frac{\sum_{i=1}^c \sum_{j=1}^N \mu_{ij}^2 \|x_j - v_i\|^2}{N \min_{i,j} \|\bar{x}_i - \bar{x}_j\|^2} \quad (11)$$

c :分群數; N :資料集的筆數 ; x_j :第 j 筆資料值; v_i :第 i 群的群心

μ_{ij} :第 j 筆資料值對 i 群的歸屬值

2. 資料正規化

避免資料集內數值影響分群的結果，因此利用正規化方式轉變換成設定的上下限之間。轉換後的屬性值將會介於 0 到 1 的區間內，如公式(12)所示。

$$x' = \frac{x - \min_i}{\max_i - \min_i} \quad (12)$$

x =未正規化前資料數值 ; x' =正規化後的資料數值

\min_i =原始資料中第 i 個屬性值的最小值;

\max_i =原始資料中第 i 個屬性值的最大值

3. 每種演算法之初始值設定

本研究對於每一種演算法的參數設定上，均進行了大量的微調實驗，但在本研究模型上，因採用多種混合系統，相同演算法之相同參數若用在不同的組合中，仍存在著不確定的因素而影響結果。實驗參數設定如表 2。

表 2 測試模型所使用的初始參數

參數名稱	定義	設定值
學習率 $c1$	每個粒子自身的學習率	1.5
學習率 $c2$	每個粒子自身的學習率	1.5
學習率 q	連續型螞蟻演算法的學習率	0.05
蒸發率 ζ	連續型螞蟻演算法的蒸發率	0.05
母體個數	基因演算法中的染色體個數	30
螞蟻數量	連續型螞蟻演算法中使用螞蟻數量	5
迭代次數	實作模型的運算次數	500
交配率	染色體的交配率	60%
突變率	染色體的突變率	5%
族群遷移率	遷移時的百分比	25%
族群遷移代數	每隔多少代進行遷移動作	100

(三)實驗結果

將本研究所提出的兩種方式及其他比較的演算法，分別對五種測試資料集進行實驗評估，每種演算法均進行 2-10 群運算，其平均績效如表 3 所示，實驗結果顯示生物多樣性最高的 Hybrid-II 均優於其他演算法，次高的 Hybrid-I 亦優於大部份演算法。生物多樣性程度中等的三種混合演算法，亦優於生物多樣性低的單一演算法，如表 4 所示。證明具結合多樣性之群智慧演算法及遺傳演算法，的確可提高求解過程中的能力。

表 3 各種演算法對五個資料集之平均績效

資料集名稱	K-means	PSO	ACO_R	GA	PSO/ ACO_R	ACO_R / GA	GA/ PSO	Hybrid -I	Hybrid- II
Abalone	5.4098	1.1354	1.0924	1.1614	1.0930	1.0863	1.0974	1.0827	1.0807
B.C.W.	2.7636	1.2731	1.2019	1.3271	1.2142	1.1998	1.2474	1.1912	1.1853
Iris	3.1601	1.2075	1.1275	1.2745	1.1060	1.1244	1.1295	1.1096	1.1035
L.D.	2.8466	1.3274	1.2247	1.3732	1.2175	1.2335	1.2451	1.2074	1.2056
Wine	1.8234	1.1480	1.0975	1.1819	1.0941	1.1107	1.0950	1.0846	1.0822

※B.C.W. : Breast Cancer Wisconsin; L.D. : Liver Disorders

表 4 生物多樣性程度對測試資料集之平均績效

K-means	生物多樣性程度 -低	生物多樣性程度 -中	生物多樣性程度 -次高	生物多樣性程度 -最高
3.2007	1.2102	1.1529	1.1351	1.1314

※生物多樣性程度-低之演算法：PSO、 ACO_R 、GA;

※生物多樣性程度-中：PSO/ ACO_R 、 ACO_R /GA、GA/PSO

※次高之演算法：Hybrid-I; 最高之演算法：Hybrid-II

伍、結論

本研究提出了具多樣性特色及子群體的混合系統，藉由混搭不同的演算法並使用遷移概念，以獲得更好的解。

在本實驗採用五個資料集作為測試分群績效，根據平均分群績效顯示，Hybrid-II 皆會比傳統 k-means、PSO、ACO_R、GA 及多種混合演算法的平均績效來的良好，而 Hybrid-I 亦比大部份的演算法佳，因為混合模式不僅可以讓求解表中的解在迭代過後能保留多樣化，可以快速的找到最佳解。

實驗結果證明具多樣性之群智慧混合模式的確可提升分群能力，其原因為透過不同演算法之混合運作，提供比單一演算法更具多樣化的解，如此可避免落入區域最佳解。

陸、參考文獻

1. Ahmadyfard, A., and Modares, H. "Combining PSO and K-means to Enhance Data Clustering," *IEEE International Symposium on Telecommunications*, Tehran, 2008, pp: 688-691.
2. Alcan, P., and Başlıgil, H. "A genetic algorithm application using fuzzy processing times in non-identical parallel machine scheduling problem," *Advances in Engineering Software*(45:1), 2012, pp. 272-280.
3. Belding, T. C. "The Distributed Genetic Algorithm Revisited," *In Proceedings of the Sixth International Conference on Genetic Algorithms*, 1995, pp. 114-121.
4. Blum, C. "Ant colony optimization: Introduction and recent trends," *Physics of Life Reviews* (2:4), 2005, pp.353-373.
5. Cantú-Paz E. "A survey of Parallel Genetic Algorithms," *Calculateurs parallèles Réseaux et Systèmes Répartis* (10:2), 1998, pp. 141-171.
6. Chandrasekhar, U., and Naga, P.R.P. "Recent trends in Ant Colony Optimization and data clustering: A brief survey," *Intelligent Agent and Multi-Agent Systems (IAMA)*, 2011, pp.32-36.
7. Chang, P. C., Huang, W. H., and Ting, C. J. "Dynamic diversity control in genetic algorithm for mining unsearched solution space in TSP problems," *Expert Systems with Applications* (37:3), 2010, pp. 1863-1878.
8. Chang, P. C., and Chen, S. H. "The development of a sub-population genetic algorithm II (SPGA II) for multi-objective combinatorial problems," *Applied Soft Computing* (9:1), 2009, pp. 173-181.
9. Clerc, M. "The mythical balance or when Particle Swarm Optimization does not exploit," 2008, From http://clerc.maurice.free.fr/ps0/Balanced_PSO/Balanced_PSO.pdf
10. Convention on Biological Diversity, CBD (1992) from <http://www.cbd.int/convention/text/>.
11. Dorigo, M., and Gambardella, L. M. "Ant colony system: A cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary*

- Computation* (1:1), 1997, pp. 53-66.
12. Eberhart, R., and Kennedy, J. "A new optimizer using particle swarm theory," *Proceedings Sixth International Symposium on Micro Machine and Human Science*, Nagoya, Japan, 1995, pp. 39-43.
 13. Ganesh Kumar, P., Aruldoss Albert Victoire, T., Renukadevi, P., and Devaraj, D. "Design of fuzzy expert system for microarray data classification using a novel Genetic Swarm Algorithm," *Expert Systems with Applications*(39:2), 2012, pp. 1811-1821.
 14. Gao, H., and Xu, W. "Particle swarm algorithm with hybrid mutation strategy," *Applied Soft Computing* (11:8), 2011, pp. 5129-5142.
 15. Gen, M., Cheng, R. Genetic algorithms and engineering design, Wiley, New York, 1997.
 16. Holden, N., and Freitas, A. A. "A Hybrid Particle Swarm/Ant Colony Algorithm for the Classification of Hierarchical Biological Data," *IEEE Swarm Intelligence Symposium*, Pasadena, California, USA, 2005.
 17. Holland, J. H. Adaptation in Natural and Artificial Systems, University of Michigan Press, 1975.
 18. Ji, Z., Li, Z., and Ji, Z. "Research on Genetic Algorithms based on Real Coded and Data Information for Global Optimization," *Procedia Engineering* (23), 2011, pp. 161-166.
 19. Koliass, C., Kambourakis, G., and Maragoudakis, M. "Swarm intelligence in intrusion detection: A survey," *Computers & Security* (30:8), 2011, pp.625-642.
 20. Lovbjerg M., Rasmussen, T. K., and Krink T. "Hybrid Particle Swarm Optimiser with Breeding and Subpopulations," *Proceedings of the Genetic and Evolutionary Computation Conference*, 2001.
 21. Lozano M., and Garcia-Martinez C. "Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: Overview and progress report," *Computers & Operations Research* (37:3), 2010, pp.481-497.
 22. Martinez-Torres, M.R., and Toral-Marin, S.L. "Strategic group identification using evolutionary computation," *Expert Systems with Applications* (37:7), 2010, pp. 4948-4954.
 23. Meng, Y., Kazeem, O., and Muller J. C. "A Hybrid ACO/PSO Control Algorithm for Distributed Swarm Robots," *Proceedings of the IEEE Swarm Intelligence Symposium*, 2007.
 24. Merwe, D. W., and Engelbrecht, AP. "Data Clustering using Particle Swarm Optimization," *Proceedings of IEEE Congress on Evolutionary Computation*(1), 2003, pp. 215-220.
 25. Mullen, R., Monekosso, D., Barman, S., and Remagnino, P. "A review of ant algorithms," *Expert Systems with Applications* (38), 2009, pp. 9608-9617.
 26. Niknam, T. "An efficient hybrid evolutionary algorithm based on PSO and ACO for distribution feeder reconfiguration," *European Transactions on Electrical Power*, 2009.

27. Socha, k., and Dorigo, M. "Ant colony optimization for continuous domains," *European Journal of Operational Research* (185:3), 2008, pp. 1155-1173.
28. Spears, W. M. "Simple Subpopulation Schemes," *Proceedings of the 1994 Evolutionary Programming Conference World Scientific*, 1994, pp. 296-307
29. Thangaraj, R., Pant, M., Abraham, A., and Bouvry, P. "Particle swarm optimization: Hybridization perspectives and experimental illustrations," *Applied Mathematics and Computation* (217:12), 2011, pp. 5208-5226.
30. Wilson, E. O. "The current state of biological diversity," National Academy Press , Washington, D.C., , 1988.
31. Xie, X.L., Beni, G. "A validity measure for fuzzy clustering," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* (13:8), 1991, pp.841-847.
32. Yang, W., Li, D., and Zhu, L. "An improved genetic algorithm for optimal feature subset selection from multi-character feature set," *Expert Systems with Applications* (38:3), 2011, pp. 2733-2740.
33. Zhenkui, P., Xia H., and Jinfeng H. "The Clustering Alogorithm Based on Particle Swarm Optimization Algorithm," *International Conference on Intelligent Computation Technology and Automation* (1), 2008, pp. 148-151.

Hybrid Swarm Intelligence Algorithms with Biodiversity Applied to

Data Clustering

Yu-Shao Wang

Department of information management

National Kaohsiung First University of Science and Technology

u9924814@nkfust.edu.tw

Cheng-Lung Huang

Department of information management

National Kaohsiung First University of Science and Technology

clhuang@nkfust.edu.tw

Abstract

The traditional swarm algorithms are easy to find the local optimal due to some factors in the computing process. We propose the innovative hybrid models with the diversity and subpopulation, which mix and match different algorithms, and prevent early convergence and local optimal in the computing process.

We propose two models through different levels of biodiversity to implemented the combination with biodiversity and swarm intelligence algorithms, and compare the general hybrid models of lower level of biodiversity.

The proposed hybrid models were applied to data clustering using the UCI public datasets. Experimental results show that the proposed hybrid systems with high biodiversity improve the performance of data clustering.

Keywords: Biodiversity, Subpopulation, Swarm Intelligence, Clustering, Hybrid Systems