

逢 甲 大 學
交通工程與管理學系碩士班
碩士論文



應用螞蟻族群系統構建群落分析演算法

Ant Colony System Based Clustering Algorithms

指導教授：邱裕鈞

研 究 生：周世章

中 華 民 國 九 十 三 年 六 月

誌謝

轉眼之間兩年研究所生活接近盡頭，回首這兩年的研究所生活，首先感激恩師 邱裕鈞老師於論文起草時的啟蒙，讓我能找到研究方向與方法，研究過程中的細心指導更令學生在論文撰寫過程更加嚴謹，在程式語言方面有老師的指導與指正，平時在研究進度上對我們的鞭策，讓我能夠順利完成論文，並且在生活態度上也給予許多寶貴的意見，讓學生不僅在學問上受益匪淺，在做人處事上更有深一層的體認；感謝系上老師們的傾囊相授，使學生如沐春風、受益良多，此外感謝系上助教在平日的照顧與幫忙。論文撰寫期間，感謝胡大瀛老師與林大傑老師在程式與模式上的啟發與指教；論文口試期間，感謝馮教授正民與魏教授健宏的寶貴意見，使得本論文能更趨完善。特別感謝馮老師對於研究方法與圖形的指正，同時點出論文中的缺失；感謝魏老師在模式驗證方面的意見，使論文中的模式驗證更加完善

這兩年來感謝同門守潔、秋美、心怡在各方面的協助，班上同學蔚明、靖峰、延祥、凱斌、芳誼、佳穎、國樑、政霖、婉郁、世淵、沛儒、亮勳、伯鴻、俊享、振嘉、宗泓不論在生活與學業上的互相鼓勵與陪伴，令我研究所生活更加豐富；另外高中知己在我研究瓶頸與情緒低潮時的激勵，當然還有令我在低潮時感動不已的垃圾話之旅，讓在異鄉的我堅強地面對許多挑戰；於國外工作的唯豪，兩年來在英文方面的指導與討論，令我受益良多；還有在國外求學的盈方，不時的加油打氣，無形中給予許多幫助；從小到大的朋友凱熙，在生活與求學的態度，那種麻吉感覺讓我在多年的求學生活中都有盞明燈；汶璇等學弟妹在平時的鼓勵與幫助，更讓我感到窩心；雅惠表姊與姊夫，在我六年的逢甲求學生涯中提供了不少幫助，讓我在台中也有家的溫暖，還有兩位小朋友郁喬、郁家帶給我不少歡笑；大甲的外婆、舅舅、舅媽和沙鹿的小阿姨、姨丈，在生活上的關心與支持，讓在异地求學的我仍有家人親情的照料。

最後，最感激我的父母還有妹妹佳穎，從小到大的栽培和給予我一個美滿的家庭，讓我沒有後顧之憂，能夠全心全力投入研究中，雖然有時的關心讓我感到不悅但卻又令我特別懷念，這感覺不是筆墨三兩句可以完全形容的；當然還有許多我的貴人朋友這些日子來的陪伴與鼓勵，在此實難一一提到，特此，將此成果與我的家人還有關心我的朋友們共同分享。

周世章 謹誌于逢甲

中華民國九十三年七月

摘要

群落分析 (Cluster analysis) 是一種傳統的多變量統計分類方法，用以探討如何將所蒐集的資料或物件予以分群 (cluster) 或是分類 (classification)，使得群內個體間具有高度的同質性，以及不同群體間具有高度的異質性。群落分析的用途甚廣，如圖樣辨識、資料分析、影像處理及市場研究等。但由於求解方法的組合特性，當問題規模的增加造成群落分析問題在計算上的困難度快速攀升，而且群落分析問題也被證明在分群數超過三個時，就會變成 NP-hard 問題。甚至連為了特定目標式 (複雜度如 $O(N^3 \log N)$ 或 $O(N^3)$) 所發展出的最佳演算法，都有大幅的改善空間。因此，許多啟發式演算法漸漸被用來處理群落分析問題。螞蟻族群系統 (ant colony system) 為 Dorigo 等人於 1996 年模仿螞蟻外出覓食的行為發展而成的演算法。經過驗證其對於 NP-hard 或 NP-complete 的組合最佳化問題，如：旅行銷售員問題 (traveling salesman problem)、車輛途程問題 (vehicle routing problem)、二次分派問題 (quadratic assignment problem) 等，皆具有相當優良的求解能力。

基此，本研究嘗試提出與驗證以螞蟻族群系統為基礎的分群演算法—螞蟻分群演算法 (ant-based clustering algorithm, ACA)。在模式驗證部分，本研究隨機產生三種不同規模大小的二維樣本包括 10 組、50 組及 100 組等小、中及大規模樣本。並分別與統計分群之凝聚法、K-means 法，以及遺傳分群演算法 (genetic clustering algorithm, GCA) 進行比較。結果顯示，在小規模樣本時 (10 筆樣本資料)，除凝聚法以外，其他分群方法均可求得整體最佳解 (以窮舉法求得)。在中規模樣本及大規模樣本時 (分別為各 30 組之 50 筆及 100 筆樣本資料)，在不同的分群組數下 (分別測試 3、5、7、9 群)，各分群模式之表現，經檢定以 ACA 表現最佳，GCA 次之，K-means 再次之，而凝聚法表現最差，而 ACA 對於其他分群演算法所求的結果有 1.04%~53.42% 的改善率。不過，各分群模式的求解穩定度 (robustness)，透過各分群演算法所求得結果之標準差進行分析，均無顯著差異。在實例應用與驗證方面，總共挑選 100 筆筆事鑑定資料，並利用卡方檢定挑選出 6 個與責任鑑定結果具有顯著影響的變數，作為進行群落分析之準則變數，而結果顯示，ACA 於分群規模為 3 群與 5 群時，仍有較其他分群演算法良好的分群結果。

關鍵字：群落分析、螞蟻分群演算法、遺傳分群演算法、凝聚法、K-means

Abstract

Cluster analysis is a traditional method of multivariate statistic classification. Cluster analysis is mainly to group all objects into several mutually exclusive clusters in order to make the degree of homogeneity within cluster and the degree of heterogeneity among clusters as high as possible. Cluster analysis is widely applied to many fields, such as pattern recognition, data analysis, image processing and market research. However, Cluster analysis is rapidly becoming computationally intractable as problem scale increases, because of the combinatorial character of the method. It has been proven that cluster analysis becomes an NP-hard problem when the number of clusters exceeds 3. Even the best algorithms developed for some specific objective functions, exhibit complexities of $O(N^3 \log N)$ or $O(N^3)$, leaving much room for improvement. Therefore, lots of heuristic algorithms have been proposed for cluster analysis. The performance of ant colony system developed by Dorigo *et al.* in 1996 based on the behaviors of nature ants out-searching for food has been proven in solving NP-hard and NP-complete combinatorial optimization problems, such as traveling salesman problem, vehicle routing problem, and quadratic assignment problem.

This study attempts to propose and validate a clustering algorithm based on ant colony system, which is called ant-based clustering algorithm (ACA). For validating the performance of proposed algorithm in different scale of problems, three different scales of two-dimension data sets have been produced randomly, including small scale (10 samples), medium scale (50 samples) and large scale (100 samples). The comparison is also conducted by comparing its performance with that of agglomerative method, k-means method, and genetic clustering algorithm (GCA). In small scale problem, in addition to agglomerative method, all other three clustering algorithms can solve the optimum solution which is solved by the total enumeration method. In the medium and large scale problems with different number of clusters (3, 5, 7, 9 clusters), ACA statistically significantly outperforms than any other algorithms by 1.04% ~ 53.42%. GCA performs better than two statistic cluster analysis methods and agglomerative method have worst performance. However, no remarkable difference in the robustness, represented by standard error, has been observed for these four methods. In the case study, a total of 100 accident records data sets have been selected and 6 clustering variables which have significant influence on determining accident responsibility are selected by chi-square test. The results show that ACA still have the best performance in clustering this accidents data into 3 and 5 clusters.

Key words : cluster analysis, ant-based clustering algorithm, genetic clustering algorithm , agglomerative method, k-means method

目錄

誌謝.....	
中文摘要.....	
英文摘要.....	
目錄.....	
圖目錄.....	
表目錄.....	
第一章 緒論	1
1.1 研究背景與動機.....	1
1.2 研究目的	2
1.3 研究內容與流程.....	3
第二章 文獻回顧	6
2.1 螞蟻理論	6
2.2 其他分群演算法.....	11
第三章 研究方法	16
3.1 螞蟻群落系統 (Ant Colony System)	16
3.2 統計群落分析方法 (Statistical Cluster Analysis)	20
3.2.1 凝聚法	20
3.2.2 K 平均值法.....	21
3.3 遺傳演算法 (Genetic Algorithm)	22
3.4 小結.....	25
第四章 模式構建	26
4.1 演算法變數定義.....	26
4.2 演算法運算流程.....	28
第五章 模式驗證與比較分析	33
5.1 小規模問題之驗證與比較.....	33
5.1.1 參數設定.....	33
5.1.2 模式驗證.....	34
5.1.3 不同交換樣本方式的求解差異.....	36

5.1.4 小規模問題之運算過程.....	42
5.1.5 不同分群規模之比較.....	47
5.2 中規模問題之比較.....	55
5.2.1 參數設定.....	55
5.2.2 不同分群規模之比較.....	56
5.3 大規模問題之比較.....	63
5.3.1 參數設定.....	63
5.3.2 不同分群規模之比較.....	64
第六章 實例應用分析	73
6.1 敘述統計分析.....	73
6.2 分群結果之分析.....	78
第七章 結論與建議	86
7.1 結論	86
7.2 建議	87
參考文獻	88



圖目錄

圖 1.1	研究流程圖	3
圖 3.1	螞蟻系統的基本觀念	16
圖 3.2	蒙地卡羅轉盤篩選概念	23
圖 3.3	染色體交配方式示意圖	24
圖 3.4	染色體突變方式示意圖	24
圖 4.1	ACA 運算流程圖	28
圖 4.2	隨機將螞蟻放置在資料節點上	29
圖 4.3	螞蟻初步完成分群之結果	30
圖 4.4	挑選出欲作交換資料節點	31
圖 4.5	完成一個回合之較佳分群結果	31
圖 5.1	決定 α 、 β 之曲線圖(小規模問題)	34
圖 5.2	演算法驗證之資料	35
圖 5.3	窮舉法所得之最佳分群結果	35
圖 5.4	ACA-1 在分群規模 5 群之分群結果	39
圖 5.5	ACA-2 在分群規模為 5 群之分群結果	39
圖 5.6	ACA-3 在分群規模為 5 群之分群結果	42
圖 5.7	不同分群規模下不同交換樣本方式求解結果趨勢圖	42
圖 5.8	小規模簡例	43
圖 5.9	ACA 分群結果	44
圖 5.10	k-means 分群結果	44
圖 5.11	凝聚法分群結果	45
圖 5.12	GCA 分群結果	45
圖 5.13	20 筆資料下 ACA 分群結果(5 群)	53
圖 5.14	20 筆資料下 ACA 分群結果(7 群)	54
圖 5.15	20 筆資料下 ACA 分群結果(9 群)	54
圖 5.16	不同分群規模各分群演算法求解結果趨勢圖(20 筆資料)	55
圖 5.17	決定 α 、 β 之曲線圖(中規模問題)	56
圖 5.18	50 筆資料下 ACA 分群結果(3 群)	61
圖 5.19	50 筆資料下 ACA 分群結果(5 群)	62
圖 5.20	50 筆資料下 ACA 分群結果(7 群)	62
圖 5.21	50 筆資料下 ACA 分群結果(9 群)	63
圖 5.22	不同分群規模各分群演算法求解結果趨勢圖(50 筆資料)	63
圖 5.23	決定 α 、 β 之曲線圖(大規模問題)	64

圖 5.24	100 筆資料下 ACA 分群結果(3 群).....	69
圖 5.25	100 筆資料下 ACA 分群結果(5 群).....	70
圖 5.26	100 筆資料下 ACA 分群結果(7 群).....	70
圖 5.27	100 筆資料下 ACA 分群結果(9 群).....	71
圖 5.28	不同分群規模各分群演算法求解結果趨勢圖(100 筆資料).....	71



表目錄

表 2.1	螞蟻理論相關文獻彙整表	9
表 2.2	求解群落分析問題的相關文獻彙整表	14
表 5.1	驗證資料.....	34
表 5.2	各不同分群演算法之分群結果.....	36
表 5.3	ACA-1 與 ACA-2 在不同分群規模之 F 檢定結果.....	37
表 5.4	ACA-1 與 ACA-2 在不同分群規模之 t 檢定結果.....	38
表 5.5	ACA-2 與 ACA-3 在不同分群規模之 F 檢定結果.....	40
表 5.6	ACA-2 與 ACA-3 在不同分群規模之 t 檢定結果.....	41
表 5.7	不同分群演算法之分群結果.....	45
表 5.8	三十組資料分析結果.....	46
表 5.9	ACA 與凝聚法分群結果之 F 檢定(20 筆資料).....	48
表 5.10	ACA 與凝聚法分群結果之 t 檢定(20 筆資料).....	49
表 5.11	ACA 與 k-means 分群結果之 F 檢定(20 筆資料).....	50
表 5.12	ACA 與 k-means 分群結果之 t 檢定(20 筆資料).....	50
表 5.13	ACA 與 GCA 分群結果之 F 檢定(20 筆資料).....	51
表 5.14	ACA 與 GCA 分群結果之 t 檢定(20 筆資料).....	52
表 5.15	ACA 與凝聚法分群結果之 F 檢定(50 筆資料).....	56
表 5.16	ACA 與凝聚法分群結果之 t 檢定(50 筆資料).....	57
表 5.17	ACA 與 k-means 分群結果之 F 檢定(50 筆資料).....	58
表 5.18	ACA 與 k-means 分群結果之 t 檢定(50 筆資料).....	59
表 5.19	ACA 與 GCA 分群結果之 F 檢定(50 筆資料).....	60
表 5.20	ACA 與 GCA 分群結果之 t 檢定(50 筆資料).....	60
表 5.21	ACA 與凝聚法分群結果之 F 檢定(100 筆資料).....	65
表 5.22	ACA 與凝聚法分群結果之 t 檢定(100 筆資料).....	65
表 5.23	ACA 與 k-means 分群結果之 F 檢定(100 筆資料).....	66
表 5.24	ACA 與 k-means 分群結果之 t 檢定(100 筆資料).....	67
表 5.25	ACA 與 GCA 分群結果之 F 檢定(100 筆資料).....	68
表 5.26	ACA 與 GCA 分群結果之 t 檢定(100 筆資料).....	68
表 5.27	ACA 對於其他分群演算法求解結果的改善率.....	72
表 6.1	影響變數之代號與說明.....	73
表 6.2	筆事責任與影響變數間之卡方檢定表.....	77
表 6.3	各分群演算法之求解結果.....	78
表 6.4	區別變數在每一分群之情況(3 群).....	78

表 6.5	區別變數在每一分群之情況(5群).....	79
表 6.6	飲酒變數在每一分群之情況(3群).....	79
表 6.7	飲酒變數在每一分群之情況(5群).....	80
表 6.8	天色變數在每一分群之情況(3群).....	80
表 6.9	天色變數在每一分群之情況(5群).....	81
表 6.10	道路型態變數在每一分群之情況(3群).....	81
表 6.11	道路型態變數在每一分群之情況(5群).....	82
表 6.12	行向變數在每一分群之情況(3群).....	82
表 6.13	行向變數在每一分群之情況(5群).....	83
表 6.14	路權變數在每一分群之情況(3群).....	83
表 6.15	路權變數在每一分群之情況(5群).....	83
表 6.16	各分群中變數分布之情形(3群).....	84
表 6.17	各分群中變數分布之情形(5群).....	85



第一章 緒論

1.1 研究背景與動機

群落分析 (cluster analysis) 為一種傳統的多變量統計分類方法，其主要是探討如何將資料或物件予以分群 (cluster) 或是分類 (classify)，主要目的就是依照收集的樣本將它分成數個分群數，使得群內個體間高度的同質性，以及不同群體高度的異質性。而群落分析屬於 NP-hard 問題，其求解難度會隨著問題的規模大小快速攀升，其中最重要的兩個問題必須解決：決定正確的群聚數目與如何採用適當的相似度量測，運用多變量分析的特點來求解問題的近似解。

在文獻中，近似解演算法 (approximation algorithm) 大抵上可以概分成兩類：第一類為經驗式法則 (ad hoc heuristics)，即針對問題特性，設計簡單的決策方法；在群落分析問題上如凝聚法 (agglomerative)、分散法 (divisive)，以及 K 平均值法 (k-means) 等，均在統計上廣被採用。第二類則是一些通用型概念 (meta-heuristics)，可適用於多種類型問題，而這些解題策略通常取材自各種自然或科學等領域，如遺傳演算法 (genetic algorithms, GA)、模擬退火法 (simulated annealing, SA)、禁忌搜尋法 (tabu search, TS) 皆為此類；在群落分析問題上，則以遺傳演算法最多 (稱為遺傳分群演算法)，其分群效果已證明優於經驗式法則。

螞蟻演算法為一新近發展的求解近似解演算法，由 Dorigo 等學者於 1996 年所提出 (Dorigo *et al.*, 1996)，此演算法利用自然界螞蟻出外覓食時，會在行經巢穴與食物間留下一種化學物質，稱為費洛蒙 (pheromone)，之後經過的螞蟻會遵循先前螞蟻所留下較濃的費洛蒙濃度路徑前進，進而尋得最佳解，其精神在於善用既有的經驗，並在可能的解答空間中探索新的可能解答 (exploration) 以及改善既有解答 (exploitation)，以反覆的自我改善過程逼近最佳解。對於解決一些困難的組合最佳化問題 (combinatorial optimization problems) 如：旅行銷售員問題 (traveling salesman problem, TSP) (Dorigo *et al.*, 1997)、二次分派問題 (quadratic assignment problem, QAP) (Maniezzo *et al.*, 1994) 等，已被證明有相當好的成效。與諸多以自然界現象為譬喻的演算法如：模擬退火法、遺傳演算法等相同，螞蟻演算法也是一成功的自然演算法，並吸引了相當多學者的注意與關切。而這股銳不可當的應用趨勢，當然得歸功於進步神速的電腦軟體效能，藉著目前電腦的強大運算能力，才使得以往只能紙上談兵的人工智慧理論得以實地的演練，俾能展現出較傳統演算法則更為優秀的解題能力。

螞蟻演算法與傳統演算法最為不同的地方是具有極高的問題獨立性，以往演算法大都是伴隨著某一特定的問題模式，有著相當多的實用限制，然而螞蟻演算法無此考量，只需要略為調整一下就可以應用在各式各樣最佳化的問題中。這種彈性也是傳統演算法所難望其項背的。同時，螞蟻演算法和類神經網路一樣。有著能夠平行處理的特性。這種平行處理的能力並不是如電腦硬體架構般顯而易見的，而是巧妙的隱藏在其演算法中，靠著在空間中散佈的各點同時探索不同的區域。

過去甚少文獻針對利用螞蟻演算法來作群落分析，如今擁有了螞蟻演算法這種與傳統演算法截然不同的工具，使我們能夠以嶄新而無拘束的觀點來針對螞蟻演算法在群落分析的效度（effectiveness）上的不同作一分析。並且利用另一種啟發式演算法如遺傳演算法來做比較，進而分析各演算法間孰者為優，期能對於啟發式演算法之運用有所貢獻。

1.2 研究目的

基於以上的動機，本研究之研究目的如下：

1. 彙析群落分析求解方法與應用領域之相關文獻。
2. 應用螞蟻族群系統之理論，建立螞蟻分群演算法（ant clustering algorithm, 以下稱之為 ACA）。
3. 利用程式語言（Fortran）編撰本研究所建立之 ACA。
4. 利用隨機產生的小規模問題驗證 ACA 之精確度。
5. 設計不同資料規模問題，進行不同求解方法之效度比較，包括本研究所建立之 ACA、遺傳分群演算法（genetic cluster algorithm, 以下稱之為 GCA）及傳統統計方法如 k 平均值法（以下稱之為 k-means）與凝聚法。
6. 在交通運輸運用方面，考量肇事案件於群落分析問題可依據所訂定群數來分析各分群的資料趨向，故本研究採用肇事鑑定資料作為實例驗證的資料來源，並與其他分群演算法所求出的分群結果比較，藉此來驗證 ACA 於實例中仍有相當程度的準確性。

1.3 研究內容與流程

本研究之研究流程如圖 1.1 所示：

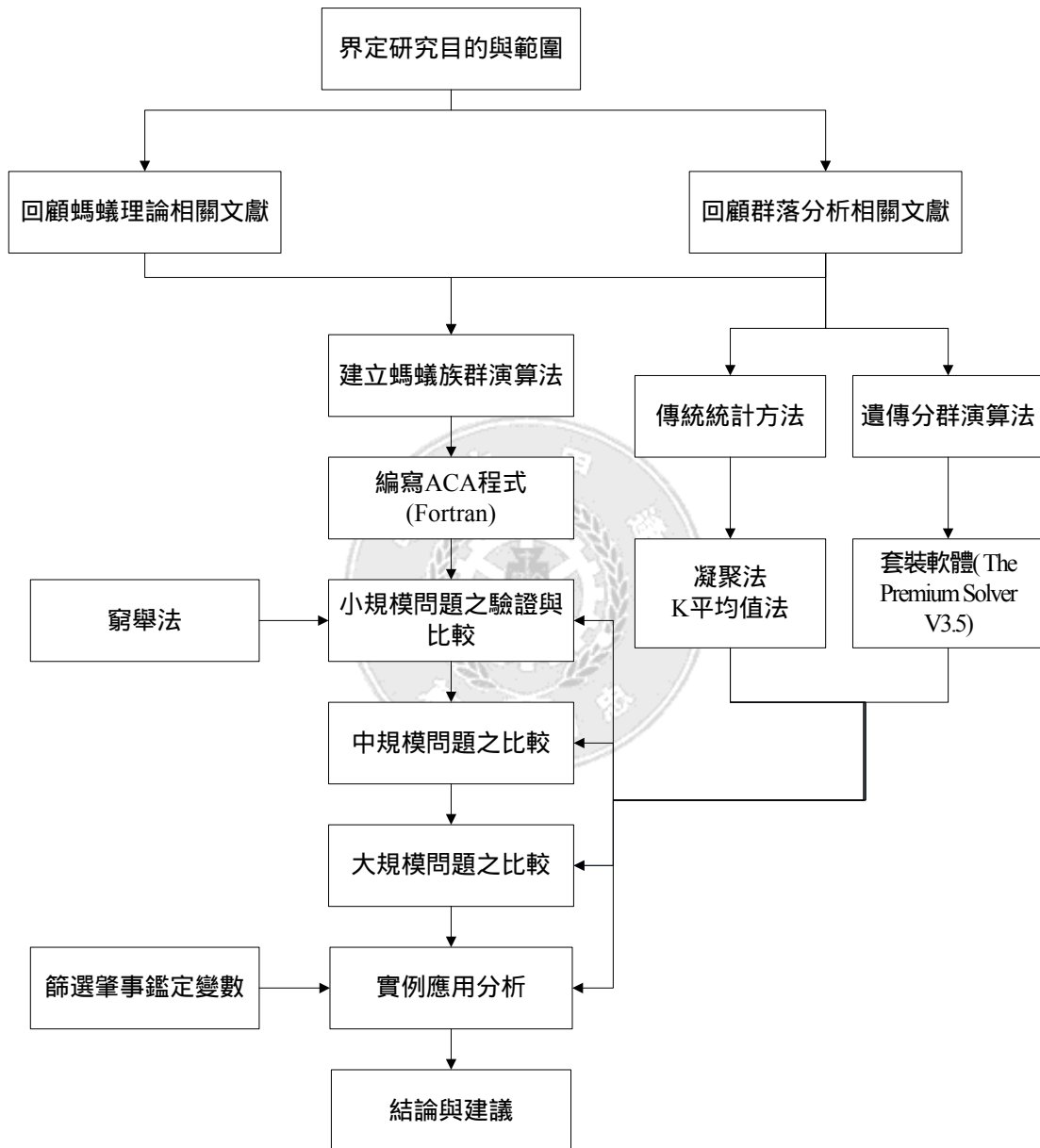


圖 1.1 研究流程圖

根據本研究之研究流程圖，本研究之研究內容可分述如下：

1. 界定研究目的與範圍

本研究為利用以螞蟻族群系統為基礎建立一個群落分析演算法，並以此演算法與其他啟發式演算法為基礎的群落分析演算法透過隨機產生的樣本，在不同規模的分群數下，分析各演算法之求解效度，並比較其優劣。

2. 回顧相關文獻

回顧國內外有關利用啟發式演算法為基礎的群落分析方法之文獻，以及對螞蟻族群系統的發展過程進行研究，並蒐集啟發式演算法所運用的範疇，藉此探討以啟發式演算法為基礎的群落分析方法之發展與應用現況，並確立本研究中 ACA 的架構。

3. 建立螞蟻分群演算法

根據回顧相關文獻之整理，界定本研究所要使用的螞蟻族群系統之範圍，並且根據其他啟發式分群演算法的精神，建立本研究的分群演算法—螞蟻分群演算法（ACA）。

4. 小規模問題之驗證

利用隨機產生的樣本，建立一個小規模的資料庫比較不同交換樣本方式的 ACA 來挑選適當的交換樣本方式，並透過與其它傳統分群演算法比較，進而驗證 ACA 的精準性。

5. 其他啟發式分群演算法

根據回顧相關文獻之整理，挑選一個適當的啟發式分群演算法，在不同規模問題之比較中，作為一個比較的對象。

6. 不同規模問題之比較

此部分為採用隨機產生的樣本，建立不同資料規模的分群問題，將本研究所發展出的 ACA 與其他分群演算法如凝聚法、k-means、GCA 分別作求解的動作，並針對各演算法在求解過程中求得的最佳解之效度作一比較分析。

7. 實例應用分析

以實際資料如肇事鑑定資料、問卷資料等，運用 ACA 作求解的動作，對於其效度作一結果分析，並且分析與傳統演算法間的優劣差異。

8. 結論與建議

根據前述的分析結果提出本研究的結論，並對往後相關研究與發展提出建議，以期獲得更完備的理論架構。



第二章 文獻回顧

本研究旨在運用螞蟻族群系統 (ant colony system, ACS) 的特性，構建出一種求解群落分析的演算法—螞蟻分群演算法 (ant clustering algorithm, ACA)，過去國內外有關利用啟發式解法之群落分析的文獻多以遺傳演算法 (genetic algorithm, GA) 為主，在九零年代初期，螞蟻系統發表之後，國內外相關文獻孕育而生，本研究就螞蟻理論與其他分群演算法為基礎的群落分析相關文獻做進行回顧，分述如下：

2.1 螞蟻理論

Dorigo *et al.* (1996) 首次提出螞蟻系統 (ant system, AS)。文章中描述螞蟻理論的基本定義與演算步驟，在研究模式中提出螞蟻理論求解問題最佳化的方法，從模擬試驗中，螞蟻在行走過的路徑上會留下一種化學物質費洛蒙 (pheromone) 當眾多螞蟻通過後會在最短路徑上留下濃度最高的費洛蒙路徑即為最佳解。而要把此螞蟻族群間這種自我組織而能搜尋出最短路徑的能力轉換進入電腦系統時，加入部份設定特性轉換成人工螞蟻的概念。在研究結果中利用螞蟻系統與其他啟發式演算法如模擬退火法 (simulated annealing, SA)、禁忌搜尋法 (tabu search, TS) 比較三者間的效率與效度，結果皆為螞蟻理論的表現比其他兩者來的優良。

Dorigo and Gambardella (1997) 針對 AS 作了三處改良，包括加入轉換規則 (transition rule)、改變費洛蒙路徑更新方法及增加局部更新法 (local updating)，而發展出螞蟻族群系統 (ant colony system, ACS) 演算法。在研究驗證部份，首先設計一小規模 (50 個城市) 旅行銷售員問題 (travelling salesman problem, TSP) 得到的結果與其他啟發式演算法比較雖有較佳的效率與效度，但是差異不大；該研究利用過去 TSPLIB 國際題庫的大規模問題 (城市數大於 100) 驗證資料，得到的結果顯示 ACS 具有較佳的效率與效度。

Gambardella and Dorigo (1996) 利用 ACS 在求解 TSP 問題中，對於路徑分成最佳路徑 (best edges, BE)、可測試路徑 (testable edges, TE) 與無利益路徑 (uninteresting edges, UE)，依據費洛蒙的特性螞蟻對於適當的路徑會累積較高的濃度，所以在一開始所有路徑皆為最佳路徑，但是隨著時間的增加，較佳路徑的費洛蒙濃度相對其他路徑來得高，因此依據螞蟻選取費洛蒙濃度較高的路徑行動可以得到我們所要的最短路徑，並以不同問題規模來驗證演算法的效率。在該研究結論中提及對於如何避免陷入局部最小值的陷阱和加速搜尋速度提出利用增加人工螞蟻族

群，而一些新的螞蟻會正確地決定局部最佳解，以此來改善局部最佳解的最佳路徑。

Dorigo and Di Caro (1999) 對於螞蟻族群最佳化做一完整詳細的介紹，而螞蟻族群最佳化對於解決 NP-hard 問題的應用在過去相關文獻有不同方面之見解，本研究針對過去各種方法做一整合的動作，該研究之目的在於整合過去有關螞蟻族群最佳化的模式，並對於此演算法運用於 TSP 驗證效率與效度的方式做一完整的介紹與分析。

Dorigo *et al.* (2000) 對於整體螞蟻演算法的理論基礎與應用範疇提出一整理性的研究，並且明確定義螞蟻理論屬於 Stigmergy 形式的一種，Stigmergy 一詞為法國為生物學家 Pierre-Paul Grasse 提出，其精神在於模擬社群昆蟲自主組織的行為，經由直接或非直接的方式達到分群分類之目的，而實際運用在人類日常生活上最明顯則是網際網路，而螞蟻利用費洛蒙濃度達到找尋食物、分辨最短路徑等皆是屬於 Stigmergy 重要的部分，該研究主要目的在於闡述定義螞蟻演算法與利用一些簡例說明螞蟻演算法的效率與效度。

McMullen (2001) 運用螞蟻族群最佳化來處理即時生產系統 (just-in-time, JIT) 問題，即時生產系統之目的在於穩定的保持最小投資，以求能夠在生產管理上保持較大的彈性並且保持在市場上的競爭力，在研究模式中利用二維與三維資料並且針對條件的不同來設計不同形式的 ACO 演算法，並以這些 ACO 演算法與其他啟發式演算法比較其效率與效度，在研究結果中得到 ACO 演算法的表現令人滿意。

Gómez and Puerta (2002) 運用螞蟻族群系統來解決貝氏網路問題，在研究模式中，透過對於螞蟻族群系統的不同定義衍生出有些微差異的螞蟻族群系統，在利用這些演算法與相同方式定義出的遺傳演算法式的演算法，做一比較分析，在研究結果中得知螞蟻族群演算法有較優秀的效率與效度，而在增加電腦 CPU 運算時間到 120 秒其結果仍是螞蟻族群演算法有較佳的效率與效度。

Randall and Lewis (2002) 利用螞蟻族群最佳化演算法中平行處理 (parallelization) 的特性來針對所設計的問題中每一步驟有眾多解的情形下，如何求得最佳解；由於定義平行處理的效率是有困難的，研究模式是利用一些通用型法則的基本要素來構建一平行處理的機制，發展出平行獨立螞蟻族群系統、平行交互作用螞蟻族群系統、平行處理的螞蟻系統、平行求解要素評價和結合螞蟻與求解要素評價的平行處理。在研究結果部分，實例驗證結果顯示在大樣本 (樣本數大於 200) 可以得到理想的效率與效度，而平行處理的效率可以經由演算法的轉變進而轉換成能更同時處理更多樣本的機制。

蕭宗勝（民 91 年）對螞蟻族群最佳化進行深入之研究。除了詳實探討螞蟻族群演算法的特性外，也針對新的組合問題提出了螞蟻族群演算法，以擴大螞蟻族群最佳化的應用範圍；這包括了：最小節點覆蓋問題、基地台分配問題、遠端擴張樹問題與演化樹建構問題。在本研究之實驗與驗證下證實，螞蟻族群演算法實是一成功的解題策略，對於許多困難的組合問題均能在短時間內取得優良之近似解，即使相較於其他近似解法也毫不遜色。本研究在一些問題的實驗上說明了只要能找到適當的圖形模式，螞蟻族群演算法均能提供優良之解答。

藍坤銘（民 91 年）基於過去眾多文獻中，幾乎只能看到螞蟻理論的求解品質，對於螞蟻理論求解效率卻鮮少著墨，本研究提出一種新的啟發式方法 ARACS 解決旅行者推銷員問題，經由多個 TSPLIB 國際題庫之求解，證實本研究提出之方法，結合順序樣式法於螞蟻理論之啟發式方法比傳統螞蟻理論收斂速度更快，而同樣能保證得到相同最佳解品質。ARACS 所採用的方法為擷取螞蟻前端搜尋所得之路徑順序為資料，進一步採用資料挖掘的順序樣式法則，找出各城市之間的關聯性，建立費洛蒙素動態矩陣，避免費洛蒙素更改時陷入局部最佳解。

朱文正（民 92 年）由於螞蟻族群演算法求解 VRP 問題之績效已獲驗證且可求解多目標問題等特性，本研究乃利用螞蟻族群演算法求解研究模式，同時修改螞蟻族群法之狀態轉換法，使得螞蟻選擇路徑時，考量時窗可靠度之影響，可有效選擇最佳路徑。

周仕雄（民 92 年）透過螞蟻釋放費洛蒙的概念提出兩種新的資料挖掘之集群分析方法：Ant system-based clustering algorithm (ASCA) 及 Ant k-means(AK)演算法。ASCA 為一種能自動分群之集群方法，而 AK 為一種必須輸入群數以及其中心點的集群演算法，在驗證部份此文獻透過蒙地卡羅模擬的方式模擬出所需的資料集合，藉以評估上述兩種演算法以及 Self-organizing feature map (SOM) 與 SOM+GA-based clustering method (S+G) 的效能，其結果顯示 ASCA 為評估出之最佳自動分群方式，本研究與此文獻的差異在於兩者局部搜尋步驟的不同，本研究在分群過程中的局部搜尋 (local search) 步驟，為在螞蟻蒐集所屬分群後，把分群中不適合的資料節點取出跟其他分群作交換資料的動作，以此來加速達到最佳分群結果；在蒐集資料節點過程中，本研究依螞蟻理論的轉換規則與更新費洛蒙濃度來進行分群動作，與該文獻依路徑上的費洛蒙濃度來進行分群動作有所不同。

茲將上述有關螞蟻理論相關研究彙整如表 2.1 所示。

表 2.1 螞蟻理論相關文獻彙整表

作者	運用方法	應用範圍	驗證資料來源	比較之演算法	驗證後效率與效度的表現
Dorigo <i>et al.</i> (1996)	AS	求解 TSP、QAP、JSP	Oliver30	TS、SA、GA、ES	AS 較佳
Dorigo and Gambardella (1997)	ACS	求解 TSP	1. 設計一個 50 城市數規模的 TSP 問題 2. TSPLIB 國際題庫	SA、NNs、EN、SOM、EC、GA、EP、AG、FI	在小規模(50)問題各演算法差異不大，大規模(100以上)ACS 較佳
Gambardella and Dorigo (1996)	ACS	求解對稱與非對稱 TSP 問題	TSPLIB 國際題庫	—	ACS 較佳
Dorigo and Di Caro (1999)	ACO	求解 TSP	—	—	—
Dorigo <i>et al.</i> (2000)	AS 相關	整理相關運用範圍	—	—	—
McMullen (2001)	ACO	求解 JIT 問題	設計三種不同需求規模的問題	GA、SA、ANN、TS	ACO 較佳
Gómez and Puerta (2002)	ACS	貝式網路 (Bayesian Networks)	<i>Medianus</i> (Kjoerulff, 1990)	GA	ACS > GA

註：TS：Tabu Search；SA：Simulated Annealing；GA：Genetic Algorithm；
ES：Evolution Strategy；NNs：Neural Nets；EN：Elastic Net；SOM：Self Organizing Map；EC：Evolutionary Computation；EP：Evolutionary Programming；AG：a combination of SA and GA；FI：Farthest Insertion；
ANN：Artificial Neural Network

表 2.1 (續) 螞蟻理論相關文獻彙整表

作者	運用方法	應用範圍	驗證資料來源	比較之演算法	驗證後效率與效度的表現
Randall and Lewis (2002)	ACO	求解 TSP 問題	設計多個不同城市規模的 TSP 問題	—	在大樣本 (200) ACO 較佳
蕭宗勝 (民 91 年)	ACO	基地台分配問題、最小節點覆蓋問題、遠端擴張樹問題、演化樹建構問題	針對各個問題設計適當的問題	基地台分配問題：H1、H2、H3 最小節點覆蓋問題：GW、GD、REP、GM、MGM 遠端擴張樹問題：HEU、OPT 演化樹建構問題：GM、MM、UM、HM	ACO 較佳
藍坤銘 (民 91 年)	ARACS	求解 TSP 問題	TSPLIB 國際題庫	ACS	ARACS 較佳
朱文正 (民 92 年)	ACS	求解 VRP 問題	假設資料與 Solomon 提出 R111 標竿範例	—	—
周仕雄 (民 92 年)	ASCA、AK	分群問題	蒙地卡羅模擬出之資料	SOM、S+G	ASCA 較佳

註：H1：Heuristic ；H2：Heuristic ；H3：Heuristic ；GW：優先選取權重最小節點之貪婪法則；GD：優先選取分支度最大節點之貪婪法則；REP：隨機選取任意連線上一點之法則；GM：優先選取權重除以分支度最小節點之貪婪法則；MGM：優先選取權重除以分支

度最小節點之貪婪法則，但權重在過程中會更動；HEU：貪婪近似解；OPT：最佳解法；MM：以最小擴張樹為基礎之 MiniMax 建構法；UM：對距離矩陣作修改之 UPGMA 法則；HM：隨機貪婪法則之近似解 Gstep-wR

由表 2.1 可知，有關螞蟻理論的研究多為探討如何去改善在求解旅行銷售員問題的最佳解品質與效率，而運用範圍也漸漸從 TSP、QAP、VRP 到電信領域（貝式網路問題）與即時生產系統（JIT）等，研究範圍漸漸跳脫在理論面上的探討轉而運用於實務方面，且此類研究大部分都會以過去曾經利用的傳統統計方法或啟發式演算法做效率與效度上的驗證比較，基此，本研究在驗證部分亦會以傳統統計方法（如 k-means）在效度上進行比較分析，以驗證螞蟻族群演算法為一個優於過去方法的新興演算法。

2.2 其他分群演算法

過去群落分析問題多以非層級集群方法中的 k-means 法求解，而群落分析問題為一 NP-Hard 問題在求解近似解的過程中，會隨著問題規模的增加，大幅增加求解問題近似解的難度，因此近年來一些以自然現象為基礎發展出的演算法稱之為啟發式演算法，透過其模擬自然界行為的優點，在求解群落分析問題的效率與效度已獲得令人滿意的結果，以下就利用啟發式演算法求解群落分析問題的相關文獻做一介紹：

Yuan and Zhuang (1996) 該研究目的在於發展出一套利用模糊遺傳演算法 (fuzzy genetic algorithm, FGA) 建立模糊分類問題的模糊規則集合，並且改善模糊遺傳演算法的效率，在研究模式中利用遺傳演算法 (genetic algorithm, GA) 來產生模糊分類規則 (fuzzy classification rules)，模糊分類系統以模糊邏輯 (fuzzy logic) 為基礎用來處理認知上的不確定性，但是這些方式通常有搜尋到的為區域最佳解 (local optimal) 而不是絕對最佳解 (global optimal) 的問題，GA 可以解決此問題但過程繁瑣，有鑑於此該研究利用 FGA 產生模糊分類規則並透過一些方法如相對適合度函數 (composite fitness function)、確認可實行性 (viability check) 等方式來改善演算法的效果。

Murthy and Chowdhury (1996) 運用遺傳演算法來找出最適的群集數目。該研究先利用 k-means 進行初步的分群，然後將其解答當作遺傳演算法初始的染色體之一，再利用遺傳演算法來改善 k-means 最後的答案。透過人工模擬與實際的資料來實驗，證明該研究所提出的方式確實可以

改善 k-means 所產生的分群結果，得到較為理想的效率與效度。

Sarkar *et al.* (1997) 等三位學者，以演化程式為基礎 (evolutionary programming-based) 的觀念來解決分群的問題。主要在於改善 k-means 演算法必須要事先知道分群數的問題。因此，透過此理論可以在未知分群數的情況下，能有效的將一群的資料分成最佳的群數。

Jiang *et al.* (1997) 本研究為發展修正遺傳演算法於整數方面上的表現，提出整數遺傳演算法 (integer genetic algorithm, IGA) 並應用於群落分析問題，此演算法為一種發展基本 GA 系統的演算法，其中包括適度規模、修正選擇操作和三個先進基因操作方式。在研究驗證中，利用中國茶葉資料來驗證 IGA 的效率與效度得到比傳統遺傳演算法有較佳的表現。

Krishna and Murty (1999) 利用遺傳演算法結合 k-means 演算法而提出新的演算法稱之為 genetic k-means algorithm (GKA)。主要是利用遺傳演算法尋找在已知分群數情況下，資料點在各群分佈的情況，利用 k-means 的運算當作基因演算法中交配的準則，同時提出一種以距離為基礎的突變方式解決 genetic k-means algorithm 演化過程中所產生的不合理解答。最後利用馬可夫鏈 (Markov chain) 來證明 genetic k-means algorithm 的收斂情況可以達到全域的最佳近似解。

Cowgill and Watson (1999) 該研究所提出的演算法為 COWCLUS 是一種以遺傳演算法為基礎的分群演算法，其中主要是利用變異比標準 (variance ratio criterion) 來定義資料集中正確的分群數目，COWCLUS 搜尋資料的方式為測試分群指派的數目、保持最佳的樣本數、合併或交配存活的樣本、突變或改變小部分樣本。但是此演算法仍無法有效透過數學方式得到有利的佐證，因為此演算法所找出的解並不能保證是一個整體最佳解，該研究只是利用遺傳演算法的特性來修正過去傳統方式的缺點，而且在一些特定的條件限制之下，COWCLUS 仍比傳統的群落分析方法來的有效率，也可較有效的找出整體最佳解。

Maulik and Bandyopadhyay (2000) 利用遺傳演算法加上 k-means 演算法之結合而提出新的演算法稱之為 GA-Clustering。GA-Clustering 最主要的目的是希望利用遺傳演算法廣域的搜尋能力，期望在有限制條件下找出最適合的群集中心點，以改善 k-means 演算法無法妥善處理重疊的資料點。該研究所提出的演算法比傳統的 k-means 法來的有效率，而且是一種運用範圍廣泛的演算法，而遺傳演算法中的分群矩陣是依點跟點之間的絕對歐幾里德距離來作分群的動作，文章提到也可以利用歐幾里德距離平方合作為一個最小的標準，這兩種方式在相關文獻中證明其效率是相同的。

Chiou and Lan (2001) 該研究目的為利用遺傳演算法來解決分群問題，過去利用啟發式演算法在處理分群問題時有兩項重大缺點：第一、分群數目必須事先限定；第二、大部分非 GA 型的演算法在應用上仍被限制，有鑒於這兩項缺失，所以分群方法結合遺傳演算法來克服，該研究之主要目的就是發展 GA 的加碼/解碼技巧來解決同時產生的最佳分群數目，並與一般統計分群方法比較最佳分群結果。研究模式中利用遺傳演算法中不同的編碼解碼技巧建立三種模式分別為：SICM、STCM、CSPM，此三種方法再與傳統分群法 AHCM 透過設計簡例的驗證，得知 SICM、STCM、CSPM 皆較 AHCM 有較佳的效率與效度。

葉承銓（民 91 年）該研究是以適應性遺傳演算法結合 k-means 的方法來解決資料分群的問題，期望透過遺傳演算法的平行處理特性，同時給予多組不同的參考點進行搜尋最佳近似解，以降低 k-means 對於資料量大的分群與重疊資料分群的誤判程度。導入適應性觀念，是期望遺傳演算法在尋找近似最佳解的過程中，可以加快演化的速度，而順利找到近似最佳解。研究結果中，我們發現利用適應性遺傳演算法來解決分群問題時，可以有效的改善 k-means 誤判情況以及找出合適的中心代表點。同時在演化的速度上，確實明顯的優於遺傳演算法，尤其在資料量大的情況下（由實驗結果得知，資料量大於 1500 筆），使用適應性遺傳演算法解決分群的問題，是一個值得考慮方法。

蔡崇煒（民 91 年）由於其演化特性，將傳統 GA 應用於實際問題及工程問題時，其彈性仍略顯不足，同時也成為一個難以解決、相矛盾及多工之問題，因此，修改 GA 的結構、適應性函數、交配及突變的設定以及微調的問題，以改善 GA 之效能，是這類研究中較常見的方式。在該研究中，針對這些方式的效益評估比較，並提出一有效的方法，稱為多重搜尋遺傳演算法 (MSGGA)。此一方法在經由實驗證明，其成效超越傳統 GA、疫苗式(immune) GA 及其它 GA 之方法。

茲將上述有關利用啟發式演算法求解群落分析問題之相關研究彙整如表 2.2 所示。

表 2.2 求解群落分析問題的相關文獻彙整表

作者	運用方法	驗證資料來源	比較之演算法	驗證後效率與效度的表現
Yuan and Zhuang (1996)	FGA	假設資料	—	FGA 表現較佳
Murthy and Chowdhury (1996)	GA + K-Means	德國與英國城鎮資料	ES、EP、GAX、ABF、KMA	GA + K-Means 表現較佳
Jiang <i>et. al.</i> (1997)	IGA	中國茶葉資料	GA	IGA 表現較佳
Sarkar and Khemani (1997)	GA	假設資料	—	—
Krishna and Narasimha Murty (1999)	Genetic K-Means Algorithm	假設資料	K-Means	Genetic K-Means Algorithm 表現較佳
Cowgill, Harvey and Watson (1999)	COWCLUS	限制資料庫與非限制資料庫	Ward's、K-Means	COWCLUS 表現較佳
Maulik, Bandyopadhyay (2000)	GA-Clustering	假設資料與實際資料	K-Means	GA-Clustering 表現較佳
Chiou and Lan (2001)	SICM、STCM、CSPM	假設資料	AHCM	CSPM 具較佳效度，
葉承銓(民 91 年)	適應性 GA	假設資料與實際資料 (Iris data)	K-Means、H-Means	適應性 GA 表現較佳

表 2.2 (續) 求解群落分析問題的相關文獻彙整表

作者	運用方法	驗證資料來源	比較之演算法	驗證後效率與效度的表現
蔡崇煒(民 91 年)	MSGGA	TSPLIB 國際題庫、假設資料	GA	MSGGA 表現較佳

由表 2.2 可以了解到相關文獻大部分皆針對過去傳統多以 k-means 法求解群落分析問題所產生在效率與效度上的缺點透過遺傳演算法來做一修正的動作，就求解的效率與效度而言，過去傳統統計方法僅能有效處理特定形式的問題，在運用方面欠缺彈性，而啟發式解法不論目標函數的特性為何，皆不用修改設計法則，此運算的優點為此類演算法在求解分群問題上廣為採用的主要原因之一，從過去文獻中可以發現到遺傳演算法是最常被利用的，基此，本研究採用傳統統計方法來和本研究所提出的螞蟻分群演算法，在求解的效度上做一比較分析。



第三章 群落分析之研究方法

本研究目的在於利用螞蟻族群系統為基礎提出一個有效求解群落分析問題的演算法—螞蟻分群演算法 (ACA)，在研究驗證方面將會利用其他啟發式演算法求解群落分析問題比較兩者間的效度。本節茲將對群落分析問題的特性與過去傳統求解群落分析方法做一介紹，並且針對本研究中模式構建最重要的螞蟻理論與在模式驗證中所運用比較的遺傳演算法分述如下。

3.1 螞蟻群落系統 (Ant Colony System)

螞蟻系統 (Ant System, AS) 為 Dorigo 於 1996 年於國際文章首次發表，其理論基於自然界螞蟻外出覓食，並將食物有效分類的行為，應用於求解近似解的一種人工智慧啟發式解法。AS 的基本概念：自然界的螞蟻會離開巢穴尋找食物，藉由牠們自己本身所散發出來的特有化學分泌物—費洛蒙 (pheromone) 來幫助螞蟻在找尋食物時，能夠遺留下某些訊息給其他的螞蟻知道其尋找食物時曾經所走過的路徑。而自然界的螞蟻在經由不斷的搜尋食物的過程中，往往會遺留許多的費洛蒙，隨著越來越多的螞蟻走過相同的路徑時，此路徑的費洛蒙也隨之增加；相對的，其餘路徑的費洛蒙則會慢慢蒸發；最後螞蟻利用費洛蒙來找尋出離食物最近的相關路徑。而螞蟻理論就是利用模擬自然界螞蟻的搜尋方式，演化出一種人工螞蟻，進而代替一般的搜尋方法，進而達到搜尋最短路徑，如圖 3.1 所示：

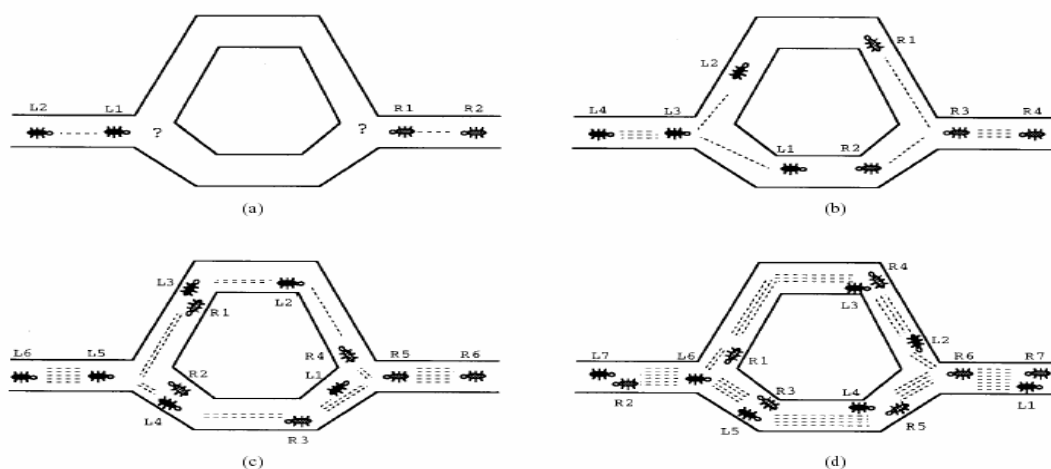


圖 3.1 螞蟻系統的基本觀念

資料來源：Dorigo, M. and Gambardella, L. M., 1997

在將自然界螞蟻轉化進入電腦系統時，我們該如何保留住某些特性，使得人工螞蟻（artificial ants）也有像自然螞蟻般有著優秀的尋優功能，Dorigo 與 Gambardella 在 1997 年提出人工螞蟻模擬自然界螞蟻行為的三大關鍵特性：

1. 螞蟻傾向於選擇具有較高費洛蒙濃度的路徑。
2. 對於較短的路徑，其費洛蒙累積的速度較為快速。
3. 螞蟻透過費洛蒙達到間接溝通的效果。

雖然 AS 的人工螞蟻主要為模仿自然界螞蟻的行為，但為了便於演算法及電腦程式的運作，仍有以下的差異（Dorigo *et al.*, 1991）：

1. 人工螞蟻具有某一程度的記憶。
2. 人工螞蟻不完全是瞎的，即存在著除了費洛蒙外的引導。
3. 人工螞蟻所生存的環境，其時間是非連續性的。

以下就 AS 重要觀念分述如下：

1. 轉換機率（transition probability）

螞蟻系統基本上是一種機率型尋優方法，在螞蟻進行路徑尋優時，係利用一轉換機率進行下一節點的選擇，而考慮的因素為費洛蒙強度及節線長度，其方程式如下：

$$p_{ij}^k = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{u \in J_k(i)} [\tau_{iu}(t)]^\alpha [\eta_{iu}]^\beta} & , \text{ if } j \in J_k(i) \\ 0 & , \text{ otherwise} \end{cases}$$

其中， $\tau_{ij}(t)$ ：代表節線 ij 在時間 t 的費洛蒙濃度。

η_{ij} ：代表期望值，通常設為節線 ij 距離長度的倒數。

$J_k(i)$ ：位於節點 i 的螞蟻 k 尚未拜訪過的鄰近節點集合。

α 、 β ：用以決定費洛蒙濃度與距離間的相對重要性的參數。

若 $\alpha=0$ 則螞蟻選擇距離最近的鄰近節點作為下一個拜訪節點，則 AS 成為貪心解法 (greedy heuristic, GH)。若 $\beta=0$ 則螞蟻選擇下一個拜訪節點時，完全不考慮其間距離長短，而完全以隨機尋找方式進行，故 AS 變成隨機尋優法 (random search, RS)。可見 GH 及 RS 啟發式解法均屬於 AS 的一種特例。當然，一般在操作 AS 時，都會設定 α 、 β 為一不為零的數值。

2. 費洛蒙蒸發 (pheromone evaporation)

與自然界的現象相同，螞蟻留置於地面的費洛蒙會隨著時間而蒸發。螞蟻族群最佳化進行費洛蒙蒸發的目的，在於避免某些路徑上費洛蒙無限量的累積，使得螞蟻有機會探索其他解答；另一方面也是讓人工螞蟻族群可對於曾經取得的可行解，在時間經過後，未再有其他螞蟻經過而獲得費洛蒙累積的情況，能適時的遺忘。設定過高的費洛蒙蒸發率將使得求解經驗無法累積，大幅降低費洛蒙所應有的功用，進而拖垮系統取得解答的效能 (Maniezzo and Coloni, 1999)。決定適當的費洛蒙累積率，將有助於取得解答的效能與效率。

3. 費洛蒙更新 (pheromone updating)

螞蟻在行經的路徑上會留下費洛蒙，而殘留費洛蒙的路徑，稱為費洛蒙路徑 (pheromone trail)，以與沒有螞蟻走過或費洛蒙已揮發完畢的路徑作為區分。各節線費洛蒙的更新方式如下：

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{k=1}^K \Delta\tau_{ij}^k$$

其中， $\tau_{ij}(t)$ 及 $\tau_{ij}(t+1)$ 為節線 ij 在 t 及 $t+1$ 時間的費洛蒙濃度。

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{L^k} & , \text{ if 節線 } ij \in L^k \\ 0 & , \text{ otherwise} \end{cases}$$

L^k 為螞蟻 k 所完成的路徑總長度。

Q 為一參數值，一般設為 100。

ρ 為費洛蒙衰退參數， $0<\rho<1$ 。

1997 年 Dorigo 以 AS 為基礎，加入轉換規則、變更整體更新法與加入局部更新法進一步提出螞蟻族群系統 (ant colony system, ACS)：

1. 轉換規則 (transition rule)

$$j = \begin{cases} \arg \max_{u \in J_k(i)} \{ [\tau_{iu}(t)]^\alpha [\eta_{iu}]^\beta \} & , \text{ if } q \leq q_0 \\ J & , \text{ otherwise} \end{cases}$$

其中， q ：為 $0, 1$ 間呈均勻分配的隨機亂數。

q_0 ：為一參數， $0 \leq q_0 \leq 1$ 。

J 為屬於 $J_k(i)$ 的某一節點 (即 $J \in J_k(i)$)，選擇方式係由 p_{ij}^k 計算轉換機率，再隨機選取之。

依據上式，若隨機亂數值 $q > q_0$ 於則 ACS 的轉換規則與 AS 的轉換機率完全相同。若 $q \leq q_0$ 則直接選取具有最大吸引力的節點 (費洛蒙濃度高、距離長度短)。因此，ACS 螞蟻的選擇行為可進一步區分為「追隨」(exploitation)與「探索」(exploration)兩種。其中，當 $q > q_0$ 時，雖然費洛蒙濃度高及距離長度短的節點被選的機率也較高，但螞蟻係以機率方式進行選擇，未必會選到機率最高的節點，因此，螞蟻還有選到其他節點的機率，故其行為較偏向探索，即類似 AS 螞蟻的行為。但當 $q \leq q_0$ 時，則螞蟻一定要選費洛蒙濃度高及距離長度短的節點，故其行為較偏向追隨。

2. 變更「整體更新法」

針對每回合 (iteration) 取得最佳解答的路徑改變其費洛蒙；在此「回合」是指所有人工螞蟻均取得一可行解的時間。總體更新的精神則在於對於好的解答給予獎賞，以引導螞蟻依據這些路徑進行開發及探索。

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho \cdot \Delta\tau_{ij}$$

$$\text{其中，} \Delta\tau_{ij} = \begin{cases} \frac{Q}{L^+} & , \text{ if 路段 } ij \in T^+ \\ 0 & , \text{ otherwise} \end{cases}$$

3. 加入「局部更新法」

只要有螞蟻走過的路徑，即改變路徑上的費洛蒙。區域更新的主要精神在於避免產生一過於強勢的路徑，吸引所有的螞蟻走上該路徑，如此無法進行適當的探索新路徑動作，而影響所得解答之品質。

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho\tau^0$$

其中， τ^0 ：起始費洛蒙濃度，通常設定為： $\tau_0 = (NL_{mn})^{-1}$ ， N 為節線總數， L_{mn} 為以貪心解法 (greedy heuristic) 所求解的總距離。

3.2 統計群落分析方法 (Statistical Cluster Analysis)

群落分析 (cluster analysis) 是一種一般邏輯程序，它能根據相似性與相異性，客觀地將相似者歸集在同一群落 (cluster) 內，所謂「群落」就是相似事物的集合。群落分析的目的在于辨認在某些特性上相似的事物，並將這些事物按照這些特性劃分成幾個群落，使在同一群落內的事物具有高度的同質性 (homogeneity)，而不同群落間的事物則具有高度的異質性 (heterogeneity)。

在群落分析中最基本的分群概念就是「相似性」，若兩個觀測值很相似，則兩個觀測值歸為同一群，而群落分析最主要就是以兩點間的距離來當作相似性的判斷，距離近的就當作相似性較高。最常是使用歐幾里

得式距離：若 X 和 Y 兩點距離定義為 $d(x, y) = \left[\sum_{i=1}^p |x_i - y_i|^2 \right]^{\frac{1}{2}}$ ，各種距離定義視其分析特性的需要使用。

如何將相似性的資料分群，主要分成兩種，一為層級群集法 (hierarchical clustering) 如凝聚法，另一為非層級群集法 (nonhierarchical clustering) 如 k-means 法，以下就凝聚法與 k-means 法做一介紹：

3.2.1 凝聚法 (Agglomerative Hierarchical Clustering Method)

凝聚法為一種傳統的統計方法，多應用於求解群落分析問題，其觀念為：起初，物件有許多分群，而這些起始分群依據本身的物件價值 (objective value) 來改善其分群狀況，其分群過程如下：

步驟 1 起始有 N 個分群，每個分群包含著單一物件，即 $S_i^{(1)} = \{O_i\}, i = 1, \dots, N$ ，而目標式為 $MF = \{\Delta F_{ij}, i, j = 1, \dots, N, i \neq j\}$ 其中 ΔF_{ij} 為目標值的增加量，即第 i 個分群與第 j 個分群會結合成一單一分群。令 $K = 1$

步驟 2 若 $\Delta F_{vu} = \max\{\Delta F_{ij}, i, j = 1, \dots, N-k, i \neq j\}$ 且 $v > u$,

$$S_u^{(k+1)} = S_u^{(k)} \cup S_v^k, S_1^{(k+1)} = S_1^{(K)}, \dots, S_{u-1}^{(k+1)} = S_{u-1}^k,$$

$$S_{u+1}^{(k+1)} = S_{u+1}^{(k)}, \dots, S_v^{(k+1)} = S_v^{(k)}, \dots, S_{N-v-1}^{(k+1)} = S_{N-v}^{(k+1)}$$

and $S_{N-k}^{(k+1)} = \Phi$, 計算部分目標值 $F(X)^{(K)}$ 。令 $K = K+1$

步驟 3 重覆步驟 2 直到 $K = N-1$, $F(X)^* = \max\{F(X)^{(K)}, K = 1, \dots, N-1\}$

3.2.2 K 平均值法 (K-Means Method)

k-means 是 MacQueen (1967) 提出，其內容為將資料分為 K 群，每群有一中心值 (centroid or seed) 即平均值，將各個觀測值分配給與中心最鄰近的群集，其主要分群技術可由下列三個步驟來說明：

1. 設定 k 個起始分群 (initial clusters) , k 為期望群數。
2. 觀測值分配到距中心值 (mean) 最近的群集，其中距離最小者，即將此觀測值分配與該群，然後再重新計算各群的中心值。
3. 以新的各群中心為基準，重覆步驟 2 之分群過程，直到各群中沒有再分配的觀測值出現。

其敘述如下：

步驟 1 將 n 個事物點 $\{X_1, X_2, \dots, X_n\}$ 隨機分割成 K 個起始族群。

步驟 2 分派事物點 X_i , $i = 1, 2, \dots, n$ 到 C_j 分群， $j \in \{1, 2, \dots, K\}$

$$\|x_i - z_j\| < \|x_i - z_p\|, p = 1, 2, \dots, K \text{ and } j \neq p, \text{ 若相等則從新求解。}$$

步驟 3 計算新的分群中心 $z_1^*, z_2^*, \dots, z_k^*$, $z_i^* = \frac{1}{n_i} \sum_{x_j \in C_j} x_j$, $i = 1, 2, \dots, K$,

其中 n_i 為 C_i 分群中的事物點數量。

步驟 4 若 $z_i^* = z_i$, $i = 1, 2, \dots, K$ 則停止，否則重覆步驟 2。

由上可知 k-means 演算法的觀念是相當簡單，其目的在於使各群中之離差平方和最小化，重覆動作直到其收斂各樣本不需重新指派到其他群集中，便可完成分群，該方法已經成功的應用在許多分群的問題上。但是，k-means 演算法卻存在著諸多的缺失，其缺點如下：

1. 使用 k-means 時，必須要先知道分群數才可進行運算。
2. k-means 的運作必須依賴初始的群集中心點。
3. k-means 在分群後所得到的解答可能陷入局部解。
4. 無法妥善處理重疊的資料點。

3.3 遺傳演算法 (Genetic Algorithm)

遺傳演算法 (genetic algorithm, GA) 的基本理論最早是由 John Holland 於 1975 年首先提倡，為一種以達爾文『進化論』為基礎，運用『物競天擇，適者生存』生物法則構建而成的最佳化搜尋法，它能夠選擇物種中具有較佳特性的上一母代，並且隨機性的相互交換彼此的基因資訊，以期能產生較上一母代更優良的子代，如此不斷重覆而產生適應性最強的最佳物種。

在演化過程中，將每一條染色體中的每一個基因 (gene) 設計成一個二元字串 (binary string) 或者是一個離散字串 (discrete string)，並且經由選擇 (selection)、交配 (crossover) 及突變 (mutation) 此三種機制來完成演化的過程，且此三種機制皆有其不可或缺之功能。以下就此三種機制分述如下 (馮正民、邱裕鈞，民 93)：

1. 選擇 (selection)

選擇是依據每一個染色體 (chromosome) 之適合度 (fitness) 高低，決定其繁衍子代多寡的一種運算過程，適應度高的物種在下一子代中將被大量複製；適應度低的物種在下一子代中則被淘汰，其中適應程度的量測是運用適應函數 (fitness function) 來決定，其中處理方式多以蒙地卡羅轉盤 (Monte Carlo wheel) 之概念加以篩選，即每一染色體被選中進行繁衍後代之機率為該染色體之適合度值占每一染色體之適合度值總合之比例，如圖 3.2 所示：

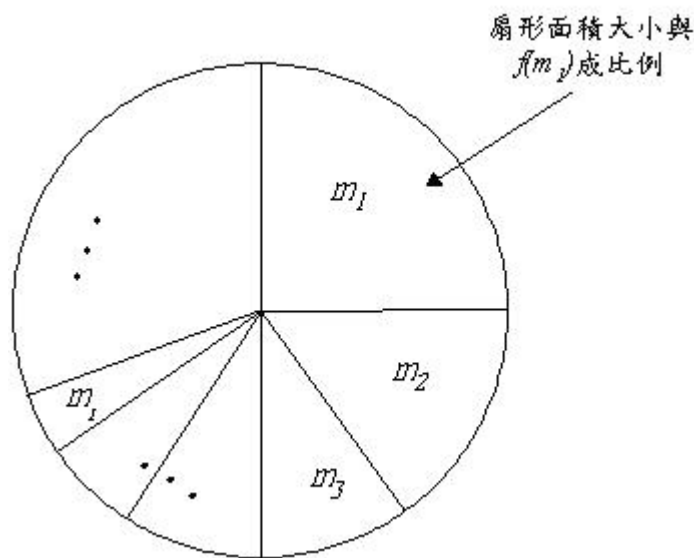


圖 3.2 蒙地卡羅轉盤篩選概念

資料來源：馮正民、邱裕鈞，民 92 年

$$P(m_j) = \frac{f(m_j)}{\sum_{i=1}^n f(m_i)}$$

其中， m_j 表第 j 個染色體， $P(m_j)$ 表第 j 個染色體被選中之機率， $f(m_i)$ 表第 i 個染色體之適合度值， n 為族群數。

因此，適合度高之染色體被選中，用以產生下一代之機率亦較高，使得較「優良」之遺傳因子得以流傳與繼承。

2. 交配 (crossover)

交配乃是將經過複製過程篩選出之成對染色體以一定機率（交配率）決定是否進行配對，再經由彼此間所進行之基因交換行為產生子代。以隨機方式產生的 0-1 均勻分配（uniform distribution）亂數值若低於交配率，則進行交配，以新的子代取代母代。反之，則被挑選的母代不進行交配，直接置入新的族群中。藉由交配機制，子代可同時具有母代雙方的優良基因，合組成更具適應能力之染色體。其處理方式主要有兩步驟：(1) 進行篩選後族群之隨機配對；(2) 隨機產生成對染色體之配對點，再依配對點之位置，進行基因互換。交配的方法亦有數種，較常採用的方法有三：單點交配（one-point crossover）、雙點交配（two-point crossover）以及均勻交配（uniform crossover）。

交配率的高低會影響母代能否繼續生存於下一世代的機率。交配率愈高，母代會被強迫以子代取代。交配率愈低，則產生子代的比率則會較低。一般交配率設定為 0.8~0.9 間。但如果交配的機制是由兩個母代產生兩個子代時，不一定以子代取代母代，而是挑選適合度值較高的兩個置回族群，另外兩個加以剔除，則不必另設交配率。

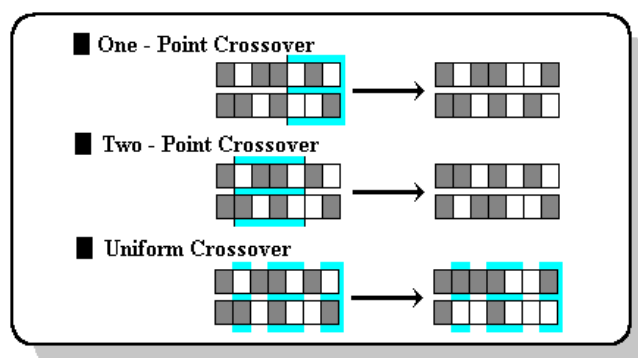


圖 3.3 染色體交配方式示意圖

資料來源：馮正民、邱裕鈞，民 92 年

3. 突變 (mutation)

突變乃是針對子代之遺傳基因以一定機率（通常很低），更動某一基因值（0→1 或 1→0），以防止染色體於複製及交配過程中，遺漏重要訊息或落入局部最佳解，亦即預防族群內各染色體之僵化。以隨機方式產生的 0-1 均勻分配亂數值若低於突變率，則該子代進行突變。反之，則子代不進行突變，可保留其交配結果的基因組合。一般而言，突變的處理方式有二：(1)基因突變 (gene mutation)，即當產生之亂數低於設定之突變率時，即隨機更動某一基因值；(2)移轉突變 (shift mutation)，即變動同一染色體內各基因之位置，而不改變其值。突變率之設定影響尋優甚鉅，該值過小，將無法發揮突變之功能，可能發生族群僵化現象，反之將破壞子代繼承母代之優良基因。一般建議設定突變率等於族群數之倒數，或設定為 0.01~0.03 間。

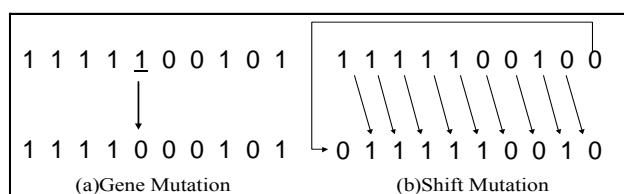


圖 3.4 染色體突變方式示意圖

資料來源：馮正民、邱裕鈞，民 92 年

3.4 小結

螞蟻理論為九零年代新興的啟發式演算法，與過去傳統類似演算法差別在於同時考慮追隨與隨機搜尋，而非只單就一個步驟予以求解，如同其他啟發式演算法，螞蟻理論也是根據自然界現象當中，螞蟻會留下一種自然化學物－費洛蒙來達到間接溝通的觀念，其中費洛蒙更新的觀念使得前一循環當中較佳的求解表現能夠有更顯著的差異，較差的求解表現有較低的顯著值，以此讓下一循環的求解過程當中，能夠參考前一循環的結果進而找尋到最佳的近似解，再加上近來電腦運算技術的大幅提升，對於複雜度高與不確定性高的問題，提供一個更有效的求解近似解的求解介面，在下一個章節本研究將利用螞蟻理論中的轉換規則與費洛蒙更新觀念，針對群落分析問題提出一個求解近似解的分群演算法－螞蟻分群演算法（ant cluster algorithm, ACA），利用一小規模問題驗證演算法的準確性，再透過不同的資料規模與分群規模驗證 ACA 比其他分群演算法是否有較佳的求解穩定度與求解結果，進而驗證 ACA 為一種能夠有效求解群落分析的演算法。



第四章 模式構建

本演算法係依據 Dorigo 等學者修正螞蟻系統 (ant system) 所發展出的螞蟻族群系統 (ant colony system, ACS) (Dorigo *et al.*, 1997) 與周仕雄發表的以螞蟻系統為基礎之分群演算法 (ant system-based clustering algorithm, ASCA) (周仕雄, 2001) 為基礎, 所建構出的分群演算法稱之為螞蟻分群演算法 (ant cluster algorithm, ACA)。

4.1 演算法變數定義

在介紹這個演算法之前, 首先定義一些記號 (notation), 以下所提到的專有名詞跟記號將適用於 ACA 演算法中。

- $\tau_{ij}(t)$: i 與 j 間第 t 個回合 (iteration) 的費洛蒙濃度, 其起始值為利用 k -means 所求解出的 TWCV 值之倒數, 再乘上資料節點的數目。
- η_{ij} : i 與 j 間的期望值 (desirability value), 依據 TSP 可定義為 $\eta_{ij} = (d_{ij})^{-1}$, 其中 d_{ij} 為兩資料節點間的歐基里德距離。
- α : 螞蟻隨機選擇資料時賦予費洛蒙的權重
- β : 螞蟻隨機選擇資料時賦予啟發值的權重
- m : 螞蟻的數目
- n : 資料的數目
- N_i^k : 第 k 隻螞蟻還沒走過的資料集合
- $P_{ij}^k(t)$: 第 k 隻螞蟻在第 t 個回合於資料節點 i 選擇資料節點 j 的機率值
- q_0 : 轉換規則參數, 為螞蟻選擇路徑時, 有 q_0 機率直接選擇往 $P_{ij}^k(t)$ 最大的路徑前進
- \bar{P} : 所有選擇未分群資料的機率平均值
- T_k : 第 k 隻螞蟻所蒐集的資料集合
- $O_{center}(T)$: T 分群中的資料節點之重心, 換言之

$$O_{center}(T) = \frac{1}{n_T} \sum_{O_i \in T} O_i$$

n_T : T 集合內的資料數目

- $D_{mean}(T)$: 為 T 集合內資料與其重心的距離平均值，換言之

$$D_{mean}(T) = \frac{1}{n_T} \sum_{O_i \in T} D(O_i, O_{center}(T))$$

- ρ : 為在分群過程中螞蟻經過路徑所蒸發的費洛蒙比率， $0 < \rho < 1$
- δ : 目標分群次數
- η : 決定是否進行更新費洛蒙濃度的門檻次數
- $\Delta\tau_{ij}^k$: 第 k 隻螞蟻在路段 ij 留下費洛蒙的量
- WCV : Within Cluster Variance 各分群組內變異，換言之

$$WCV = \sum_{i \in k} (O_i, O_{center}(T_k))^2$$

- TWCV : Total Within Cluster Variance 總組內變異，換言之

$$TWCV = \sum_{k=1}^m \sum_{i \in k} (O_i, O_{center}(T_k))^2$$

4.2 演算法運算流程

本節將 ACA 之程序列出，運算的流程如圖 4.1 所示：

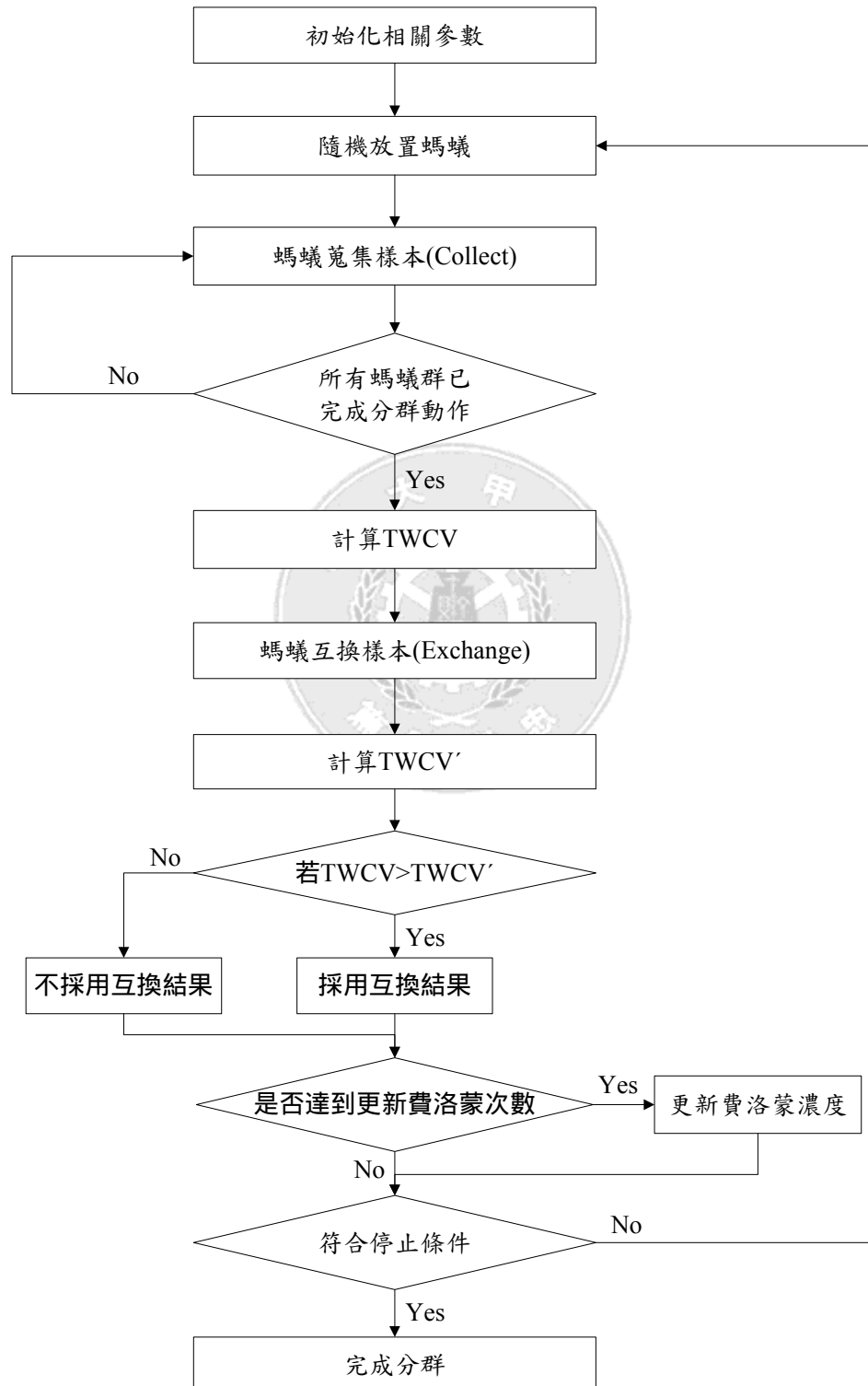


圖 4.1 ACA 運算流程圖

ACA 之分群技術程序內容如下：

步驟一：起始狀態及參數設定：包含螞蟻數目 m 、轉換規則參數 q_0 、費洛蒙蒸發比率 ρ 、目標分群次數 δ 、決定更新費洛蒙濃度的門檻次數 η 、 α 與 β 值。

步驟二：隨機放置螞蟻：將 m 隻螞蟻隨機放置於節點上，開始進行分群的動作如圖 4.2 所示：

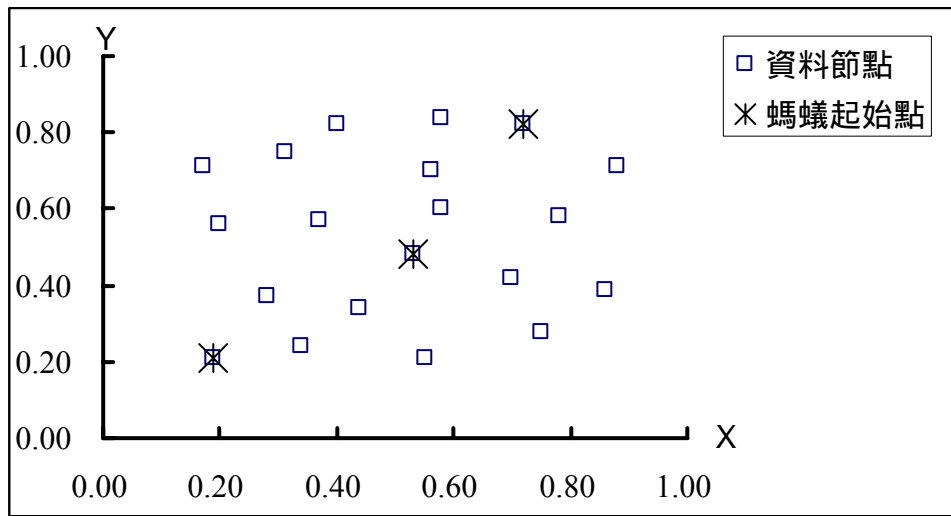


圖 4.2 隨機將螞蟻放置在資料節點上

步驟三：螞蟻蒐集樣本(Collect)：每隻螞蟻隨即依據當時的資料節點蒐集狀況，計算轉換機率，以便選擇下一個所蒐集的資料節點，逐步完成分群動作。依據 ACS 中轉換規則來計算螞蟻移動分群的機率，計算方式如下：

$$P_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}(t)]^\alpha [\eta_{il}]^\beta}, \forall j \in N_i^k, k = 1 \sim m$$

$$q \sim u(0, 1), 0 < q_0 < 1$$

當 $q \leq q_0$ 時：

$$P_{ij}^k(t) = \begin{cases} 1, & \text{if } j = \arg \max \{P_{ij}^k(t) | j \in N_i^k\} \\ 0, & \text{otherwise} \end{cases}$$

當 $q > q_0$ 時，依 P_{ij}^k 來決定被螞蟻蒐集資料節點的機率。

並且 $P_{ij}^k \geq \bar{P}_i^k$ ，則第 k 隻螞蟻蒐集此資料節點。

步驟四：計算 TWCV：計算螞蟻完成初步分群的 TWCV，初步的分群結果如圖 4.3 所示：

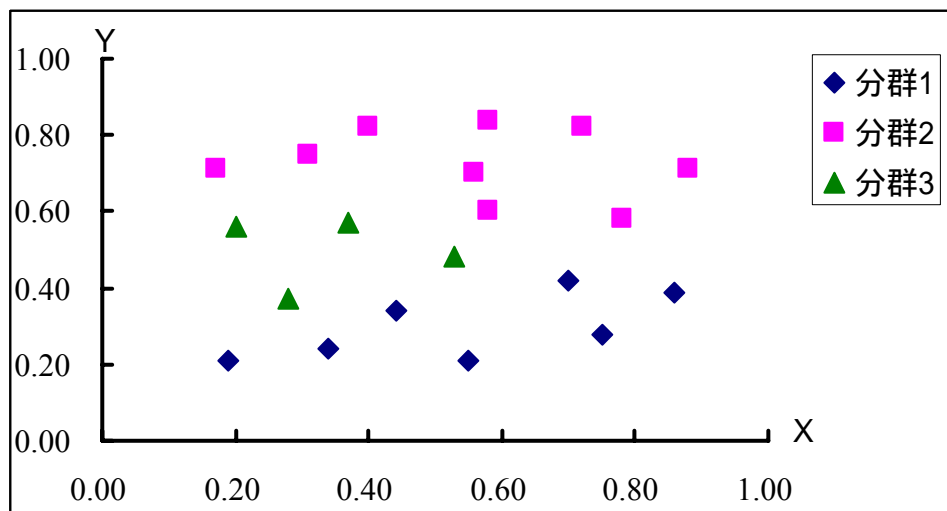


圖 4.3 螞蟻初步完成分群之結果

步驟五：螞蟻互換樣本(Exchange)：在互換樣本部份，為所謂局部搜尋動作，本研究提出三種不同的交換方式：

ACA-1：每一分群內的資料節點依據與其所屬分群重心的距離採比例原則挑選出來做互換動作，即與分群重心距離越遠有越大的被挑選進行交換的機率，而每一分群再採隨機方式分別挑選出一個欲作交換的資料節點；再者，計算交換節點與各分群重心的距離，各分群分別蒐集與本身重心距離最近的節點。

ACA-2：每一分群依據下列條件篩選出可以作為交換的資料節點：

$$D(O_i, O_{center}(T_k)) \geq D_{mean}(T_k)$$

即各分群依據資料節點與其重心的距離來判斷，若距離超過各群內資料節點與其分群重心距離的平均值則可挑選為欲作交換的資料節點；接著計算各分群重心與交換節點的距離，各分群分別蒐集與本身重心距離最近的節點。

ACA-3：演算法求解過程中，不進行互換樣本的動作。

演算法中挑選出欲作交換資料節點方式如圖 4.4 所示：

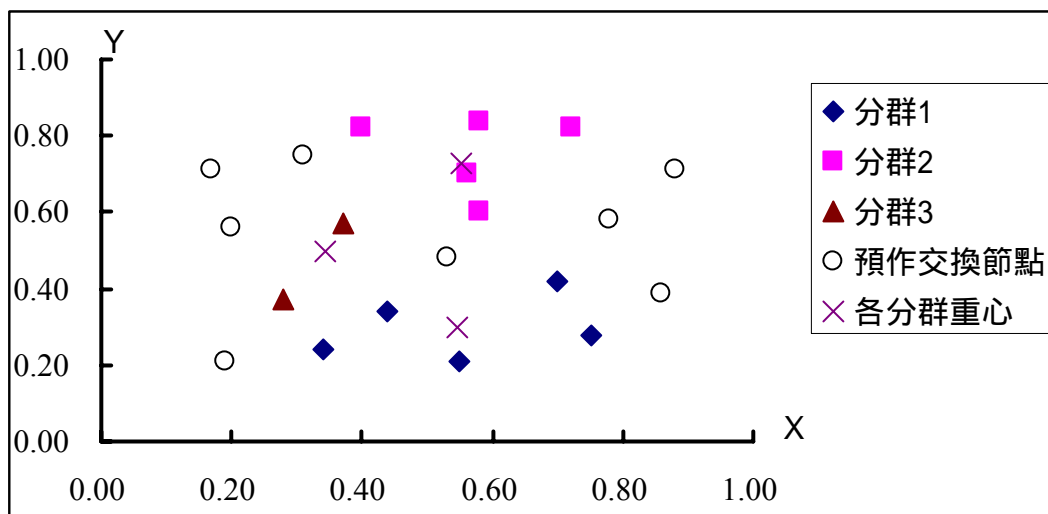


圖 4.4 挑選出欲作交換資料節點

步驟六：計算 $TWCV'$ ：計算互換樣本後的 $TWCV$ 值，與之前尚未進行互換樣本的 $TWCV$ 值做比較，以較小者做為此回合的 $TWCV$ 值，並判斷是否已達到目標分群次數 δ ，若達目標次數則完成分群並挑選出 $TWCV$ 值最小者為分群的結果，此回合較佳的分群結果如圖 4.5 所示；否則進行步驟七。

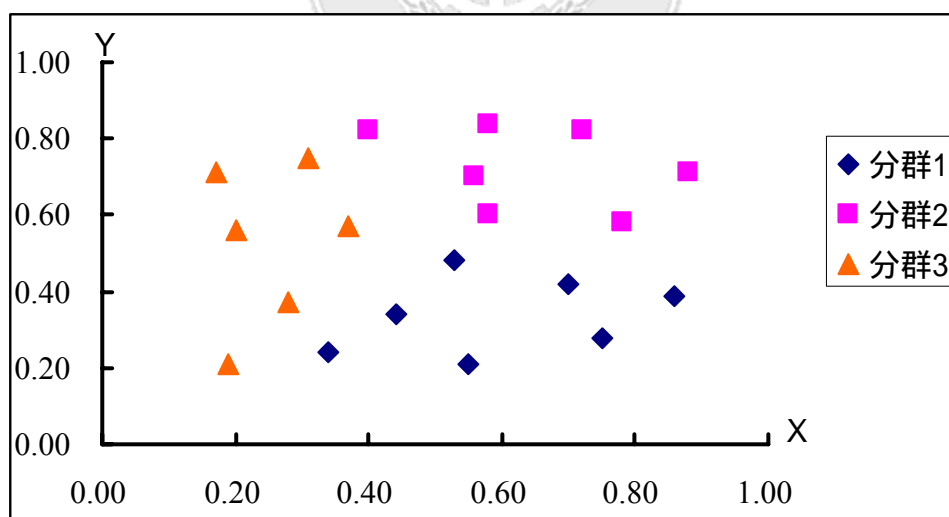


圖 4.5 完成一個回合之較佳分群結果

步驟七：更新費洛蒙濃度：首先判斷已分群次數是否達到更新費洛蒙濃度的門檻次數 η ，若符合條件則進行整體費洛蒙濃度更新，也就是原殘留路段上之費洛蒙會隨時間消逝而逐漸揮發，而螞蟻巡行過之路段則會再留下新的費洛蒙，而使得此路段的費洛蒙量增加。其更新方式如下：

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \frac{1}{WCV_k}$$

其中， WCV_k 為 η 次分群中最小 $TWCV$ 值中所屬分群的 WCV 值，若沒達到更新的門檻次數，則直接進行步驟二。

以上為 ACA 的運算流程，接下來章節將對於演算法的可靠性，透過小規模問題採用窮舉法找出最佳分群結果，進而驗證 ACA 的可靠性；再透過不同資料規模與不同分群規模與其他分群演算法的比較分析，來驗證演算法的適用性。



第五章 模式驗證與比較分析

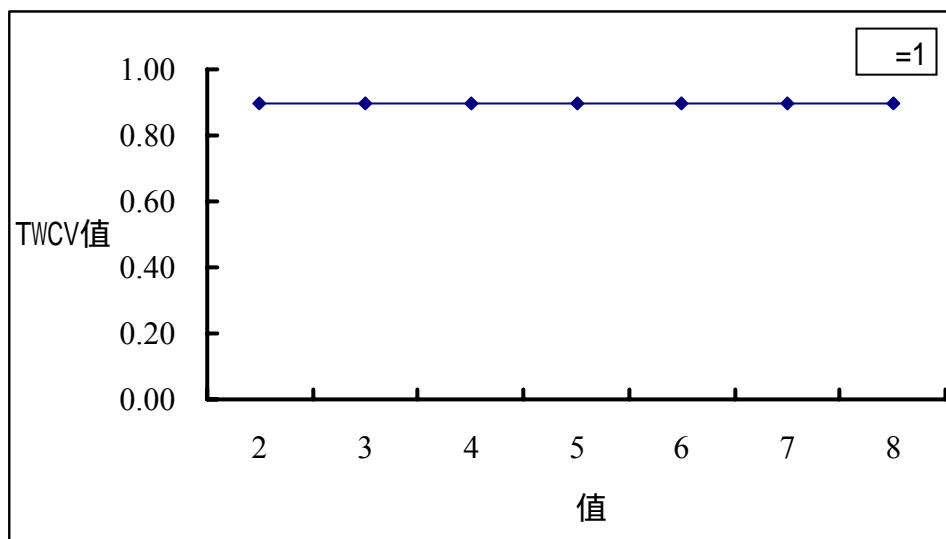
本章節將對於第四章所構建出的三種不同交換樣本方式之 ACA，透過隨機產生符合均等分配（uniform distribution）界於 0 與 1 間之二維資料節點，利用窮舉法來證明 ACA 所求解的分群結果會趨近於最佳解；在小樣本問題中解釋整個演算法的運算步驟，對於不同的交換樣本方式在不同分群規模下的求解結果，利用統計分析方法挑選出適當的交換樣本方式進而與其它分群演算法比較在求解的變異數與效度是否有較佳的表現結果；在中規模與大規模問題中，採用統計分析方法來比較分析各分群演算法的求解變異與分群結果。本研究 ACA 為利用程式語言 Fortran 編寫而成，凝聚法與 k-means 則為利用 SPSS 統計軟體求解，而 GCA 則透過 The Premium Solver 版本 3.5 來求解，分別計算各演算法對設定問題的 TWCV 值，進而對各個分群演算法進行比較分析。

5.1 小規模問題之驗證與比較

本節首先利用隨機產生之資料規模為 10 筆二維資料作為模式驗證的依據，利用窮舉法找出最佳分群結果，再利用本研究所提出的兩種不同交換樣本方式之螞蟻分群演算法來證明螞蟻分群演算法的求解結果會趨近於最佳解以證明其可靠性；再著，運用資料規模為 20 筆二維資料共 30 組樣本與其他分群演算法進行比較分析。

5.1.1 參數設定

ACA 中最重要的兩個參數 α 與 β ，係採用一組隨機產生樣本透過試誤法來決定之，本研究中為固定 $\alpha=1$ ，變動 β （範圍 2~8）來決定不同規模問題時 α 與 β 值，因此小規模問題中，由於 β 值在變動範圍中所呈現的 TWCV 值皆相同（如圖 5.1 所示），表示 α 與 β 值對於此情況下對於求解結果的敏感度低，故小規模問題時參數設定 $\alpha=1$ 、 $\beta=2$ ，其他相關參數如費洛蒙衰退參數 $\rho=0.4$ ，轉換規則參數 $q_0=0.8$ ，而螞蟻數即為分群群數，目標分群次數為 $\delta=1000$ ，更新費洛蒙濃度之門檻次數 $\eta=100$ 。

圖 5.1 決定 α 、 β 之曲線圖（小規模問題）

5.1.2 模式驗證

本節將利用一組資料規模為二維度 10 筆資料，利用窮舉法來找出最佳的分群結果，並利用本研究所提出的兩種不同交換樣本方式之螞蟻分群演算法來證明螞蟻分群演算法的求解結果會趨近於最佳解。表 5.1 為此 10 筆驗證資料，圖 5.2 為所利用之資料節點分布情形：

表 5.1 驗證資料

資料編號	X	Y
1	0.4142	0.9909
2	0.7266	0.8636
3	0.9760	0.4272
4	0.4327	0.0339
5	0.1287	0.6452
6	0.0569	0.4308
7	0.9879	0.0261
8	0.4468	0.2803
9	0.7697	0.3497
10	0.1687	0.8042

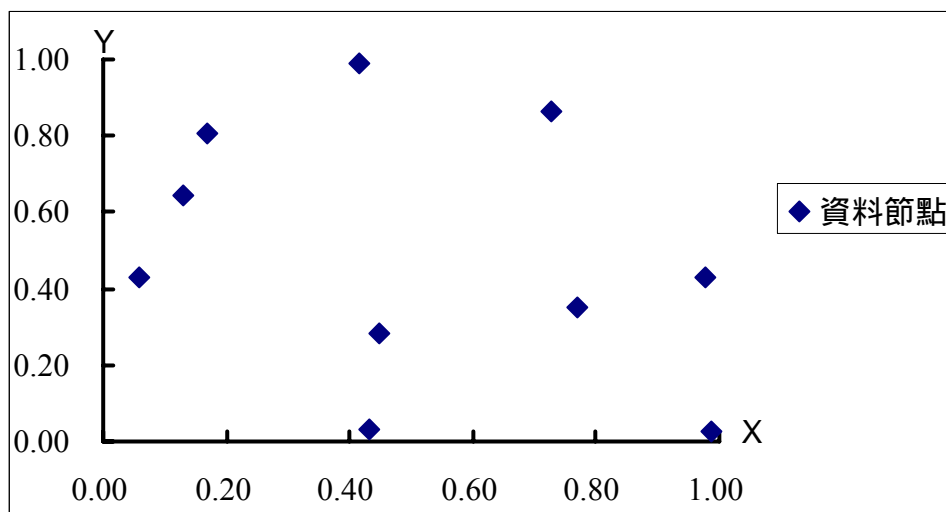


圖 5.2 演算法驗證之資料

模式驗證中分群規模設定為 3 群，經過窮舉法找出最佳分群結果，再與利用 ACA、k-means、凝聚法與 GCA 所求解出的分群結果作一比較，圖 5.3 為窮舉法所找到的最佳分群結果：

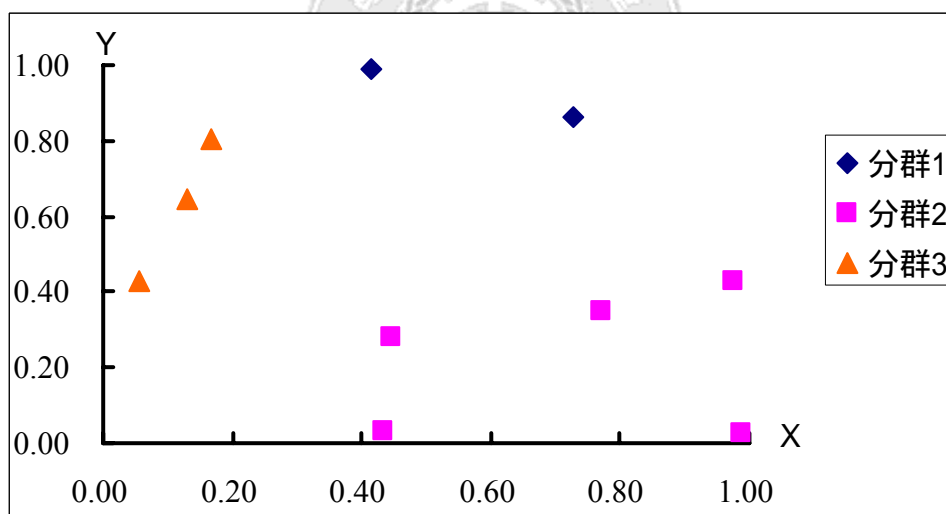


圖 5.3 窮舉法所得之最佳分群結果

經由 ACA、k-means、凝聚法與 GCA 所求解出的分群結果作一比較，三種不同交換樣本方式之 ACA、GCA 凝聚法所求解出結果均與窮舉法所得結果相同，只有 k-means 所求解之分群結果與其他不同，其各分群演算法之結果如表 5.2 所示：

表 5.2 各不同分群演算法之分群結果

資料 編號	各資料節點在不同演算法之所屬分群						
	窮舉法	k-means	凝聚法	GCA	ACA-1	ACA-2	ACA-3
1	1	1	1	1	1	1	1
2	1	3	1	1	1	1	1
3	2	3	2	2	2	2	2
4	2	2	2	2	2	2	2
5	3	1	3	3	3	3	3
6	3	1	3	3	3	3	3
7	2	3	2	2	2	2	2
8	2	2	2	2	2	2	2
9	2	3	2	2	2	2	2
10	3	1	3	3	3	3	3
TWCV	0.566026	0.684868	0.566026	0.566026	0.566026	0.566026	0.566026

計算各分群演算法之 TWCV 值，利用窮舉法所找到之最佳解的 TWCV 值為 0.566026，因 ACA、凝聚法、GCA 所求解出的分群結果與窮舉法相同，故 TWCV 值亦相同，而 k-means 所求解出的 TWCV 值則為 0.684868；由此可知，本研究所提出的 ACA 經過驗證，是一種可以找到最佳分群結果的分群演算法。

5.1.3 不同交換樣本方式的求解差異

本研究提出三種不同的交換樣本方式之 ACA，而交換樣本方式即為局部搜尋的動作，ACA-1 與 ACA-2 皆有考慮到局部搜尋的動作，ACA-3 則不考慮局部搜尋；以下首先針對兩種有考慮交換樣本方式的 ACA 透過統計分析方法來比較兩種交換樣本方式間在不同分群規模下何種有較佳的分群結果。接著，對於交換樣本方式對於求解結果是否有顯著的差異做比較分析，利用前述找出較佳的交換樣本方式與並不考慮交換樣本方式透過統計分析方法來觀察在不同分群規模下是否有顯著的差異，而資料樣本為 30 組資料規模為 20 筆來進行比較依據，表 5.3 為 ACA-1 與 ACA-2 的 F 檢定結果：

表 5.3 ACA-1 與 ACA-2 在不同分群規模之 F 檢定結果

3 群			5 群		
	ACA-1	ACA-2		ACA-1	ACA-2
平均數	0.902345	0.901556	平均數	0.415104	0.396940
變異數	0.01987	0.019767	變異數	0.006625	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.005210		F	1.401810	
P(F<=f) 單尾	0.494467		P(F<=f) 單尾	0.184182	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	
7 群			9 群		
	ACA-1	ACA-2		ACA-1	ACA-2
平均數	0.238611	0.215554	平均數	0.136875	0.120697
變異數	0.002696	0.001701	變異數	0.001079	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.585472		F	1.364651	
P(F<=f) 單尾	0.110291		P(F<=f) 單尾	0.203753	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

ACA-1 與 ACA-2 兩者差異在於互換樣本的方式不甚相同，ACA-1 是以分群內資料節點與其所屬分群重心距離越遠有越大的被交換機率；ACA-2 首先以條件限制有較高相似性的資料節點被提出交換，並挑選出可供交換的資料樣本，被交換節點依據與各分群重心的距離來進行交換動作而完成一個回合的分群動作，由表 5.3 可知，在信賴水準 $\alpha=0.05$ 於不同的分群規模下，兩種交換樣本方式在求解結果穩定度並沒有顯著的差異。

表 5.4 ACA-1 與 ACA-2 在不同分群規模之 t 檢定結果

3 群			5 群		
	ACA-1	ACA-2		ACA-1	ACA-2
平均數	0.902345	0.901556	平均數	0.415104	0.396940
變異數	0.01987	0.019767	變異數	0.006625	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	1.255077		t 統計	3.829573	
P(T<=t) 單尾	0.109735		P(T<=t) 單尾	0.000317	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	
7 群			9 群		
	ACA-1	ACA-2		ACA-1	ACA-2
平均數	0.238611	0.215554	平均數	0.136875	0.120697
變異數	0.002696	0.001701	變異數	0.001079	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	6.301128		t 統計	5.750646	
P(T<=t) 單尾	3.498E-07		P(T<=t) 單尾	1.582E-06	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

表 5.4 為兩種不同交換樣本方式之平均值檢定結果，在分群規模為 3 群時，兩種交換樣本方式的檢定值並不顯著，換言之，兩種分群演算法在分群結果並無顯著的差異；但是隨著問題的複雜度增加，即增加分群規模（5 群、7 群、9 群），透過平均值檢定兩種分群演算法之檢定值有顯著的差異，表示 ACA-2 比較 ACA-1 在資料規模為 20 筆資料時，有較為優良的分群結果，意謂隨者不同的交換樣本方式，在分群結果也會有相當程度的改變，圖 5.4 與圖 5.5 分別為在分群規模為 5 群時，相同資料樣本 ACA-1 與 ACA-2 的分群結果。

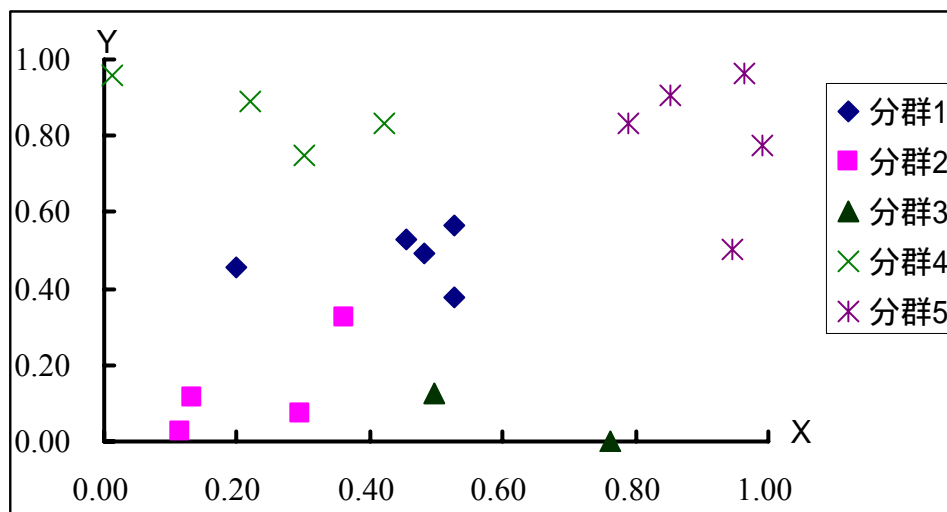


圖 5.4 ACA-1 在分群規模 5 群之分群結果

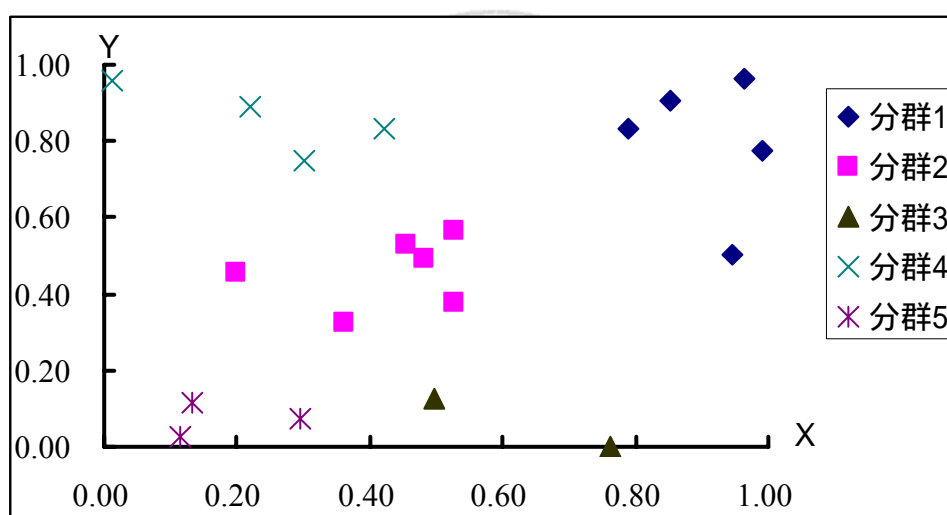


圖 5.5 ACA-2 在分群規模為 5 群之分群結果

接著，對於有無交換樣本方式是否會對求解結果有顯著的差異利用統計分析方法來比較，以下將比較 ACA-2 與 ACA-3 在求解結果的穩定度與效度是否有顯著差異，進而判斷有無交換樣本方式對於求解結果是否有顯著影響。ACA-3 為不考慮交換樣本的步驟即不採用局部搜尋方式，以螞蟻初步完成分群的結果作為一個循環的最終分群結果。

表 5.5 ACA-2 與 ACA-3 在不同分群規模之 F 檢定結果

3 群			5 群		
	ACA-2	ACA-3		ACA-2	ACA-3
平均數	0.901556	0.906337	平均數	0.396940	0.416379
變異數	0.019767	0.019838	變異數	0.004726	0.005935
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	0.996432		F	0.796292	
P(F<=f) 單尾	0.496193		P(F<=f) 單尾	0.271784	
臨界值：單尾	0.537399		臨界值：單尾	0.537399	
7 群			9 群		
	ACA-2	ACA-3		ACA-2	ACA-3
平均數	0.215554	0.242736	平均數	0.120697	0.138621
變異數	0.001701	0.002371	變異數	0.000791	0.001060
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	0.717214		F	0.746275	
P(F<=f) 單尾	0.188005		P(F<=f) 單尾	0.217732	
臨界值：單尾	0.537399		臨界值：單尾	0.537399	

由表 5.5 可知，於資料規模 20 筆不同分群規模下兩者間的 F 檢定值，其檢定結果顯示並無顯著的差異，換句話說兩者間的求解穩定度並沒有顯著的差異。

表 5.6 ACA-2 與 ACA-3 在不同分群規模之 t 檢定結果

3 群			5 群		
	ACA-2	ACA-3		ACA-2	ACA-3
平均數	0.901556	0.906337	平均數	1.528528	1.374884
變異數	0.019767	0.019838	變異數	0.041337	0.019221
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	3.172824		t 統計	-4.540622	
P(T<=t) 單尾	0.001336		P(T<=t) 單尾	4.537E-05	
臨界值：單尾	1.860811		臨界值：單尾	1.699127	
7 群			9 群		
	ACA-2	ACA-3		ACA-2	ACA-3
平均數	0.215554	0.242736	平均數	0.120697	0.138621
變異數	0.001701	0.002371	變異數	0.000791	0.001060
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	-5.426311		t 統計	-5.889052	
P(T<=t) 單尾	3.880E-06		P(T<=t) 單尾	1.080E-06	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

由表 5.6 可知，在不同分群規模下 ACA-2 與 ACA-3 的 t 檢定值皆呈現顯著差異，表示 ACA-2 有較 ACA-3 優良的分群結果，由此可知，有無交換樣本的步驟對於求解結果有顯著的差異，基此，接下來章節將以 ACA-2 於不同資料規模及不同分群規模下，與其他分群演算法透過 F 檢定判斷各分群演算法在求解穩定度是否有顯著差異，利用 t 檢定判斷各分群演算法在求解效度上是否有顯著的差異，以此來驗證 ACA 的適用性。圖 5.6 為 ACA-3 於分群規模 5 群時的分群結果，圖 5.7 為三種交換樣本方式於不同分群規模下的求解結果趨勢圖。

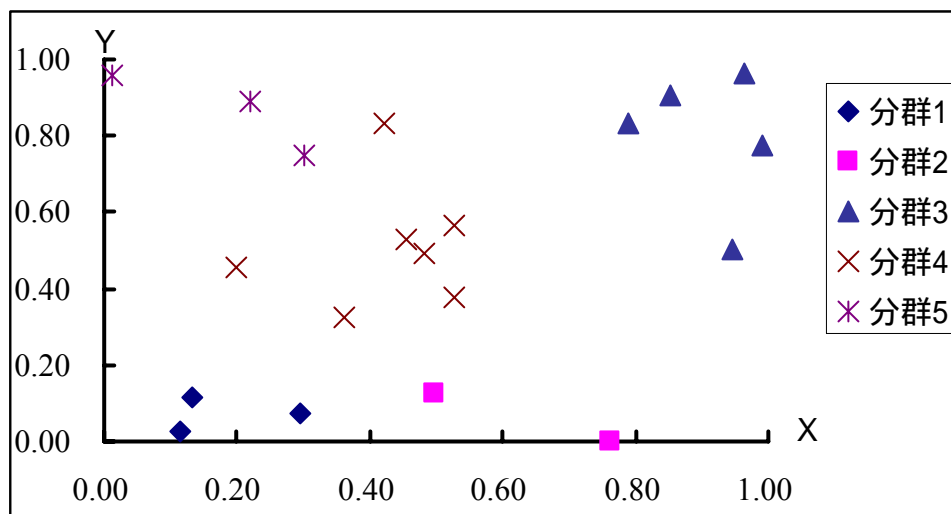


圖 5.6 ACA-3 在分群規模為 5 群之分群結果

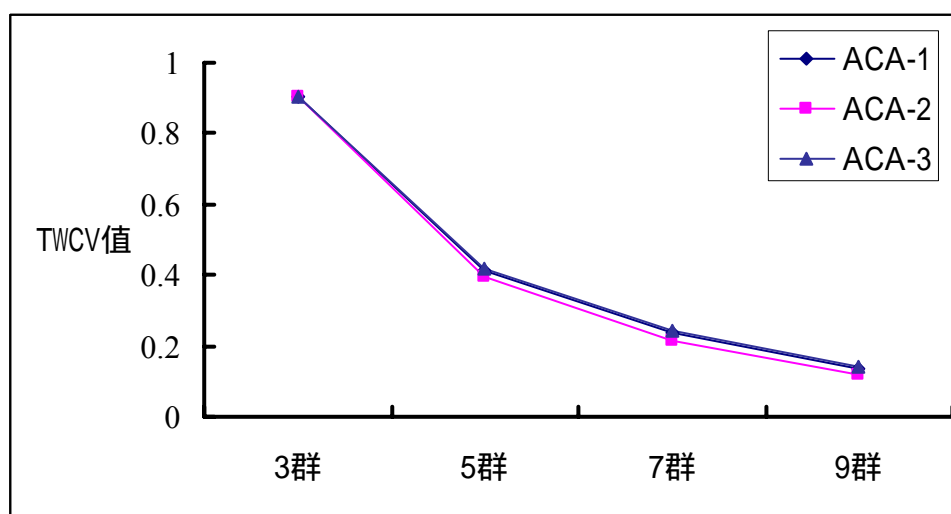


圖 5.7 不同分群規模下不同交換樣本方式求解結果趨勢圖

5.1.4 小規模問題之運算過程

本節將利用一組資料規模為 20 筆資料節點，對於 ACA 做一詳細的運算過程介紹，所採用的資料如圖 5.8 所示：

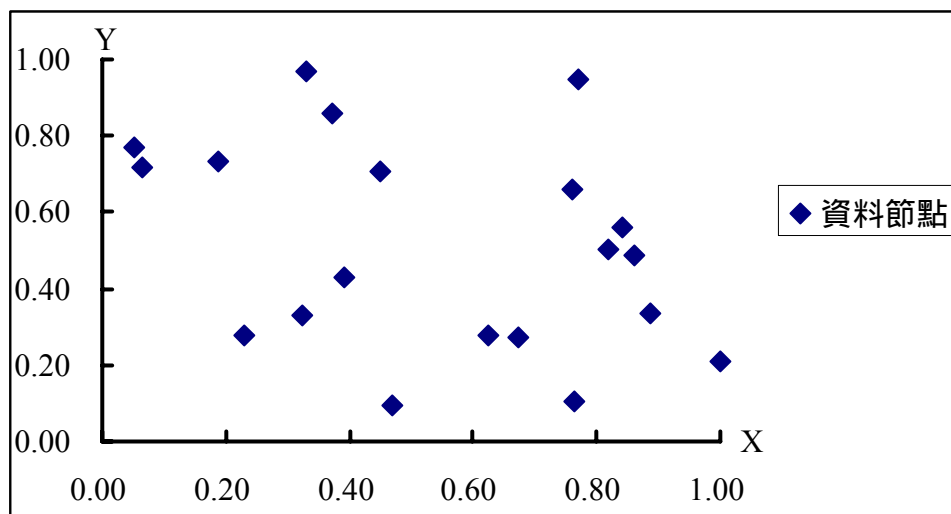


圖 5.8 小規模簡例

以下分述 ACA 操作步驟：

步驟一：起始狀態及參數設定，包含轉換規則參數 (q_0)、費洛蒙起始值 (τ^0)、費洛蒙衰退參數 (ρ)、螞蟻數目 (m)、 α 與 β 值、目標分群次數 (δ)、更新費洛蒙濃度之門檻次數 (η)。在此透過試誤法與相關文獻資料設定 $\alpha=1$ 、 $\beta=2$ 、 $\rho=0.4$ 、 $q_0=0.8$ 、 $m=3$ 、 $\delta=1000$ 、 $\eta=100$ ，而費洛蒙起始值 (τ^0) 為資料數 (n) 乘上以 K 平均值法所求出 TWCV 值倒數，即 $\tau^0 = \frac{n}{TWCV}$ 。

步驟二：隨機放置螞蟻到節點上，並開始進行分群的動作。

步驟三：螞蟻蒐集樣本，螞蟻依據轉換規則來計算移動分群的機率。

步驟四：螞蟻完成初步分群並計算 TWCV 值。

步驟五：計算各分群內資料節點與其重心距離的平均值，依據挑選準則選出各組欲作交換的資料節點；接著計算各分群與交換節點的距離，並依其距離遠近把交換節點分配給適當的分群。

步驟六：完成交換樣本計算 TWCV' 值，與之前尚未交換樣本前的 TWCV 值作一比較，以較小者為此次運算循環的 TWCV 值。

步驟七：若分群次數符合目標分群次數 (δ) 則停止分群動作並挑選出 TWCV 值最小者為分群結果，反之進行步驟八。

步驟八：更新費洛蒙濃度，若分群次數達到更新費洛蒙濃度之門檻次數 (η) 則依據更新準則予以更新之，反之則重複步驟二。

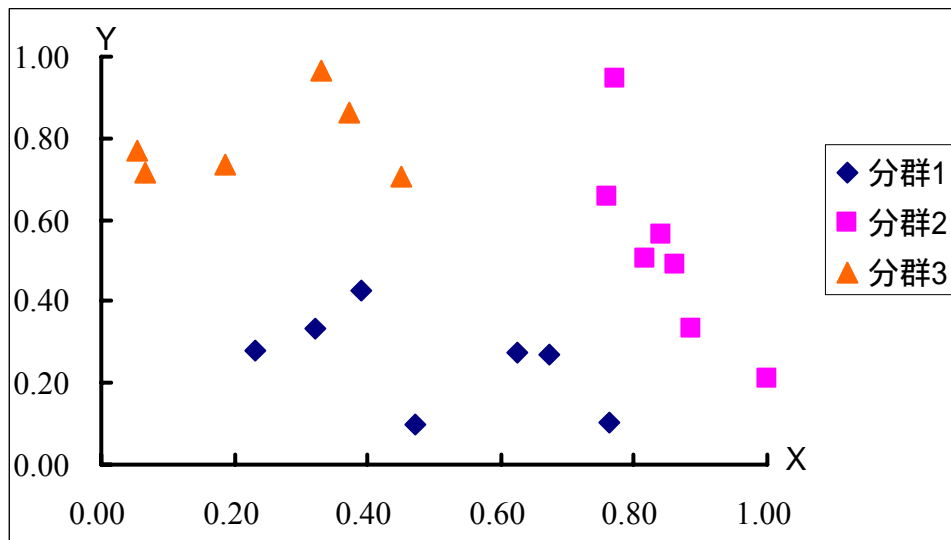


圖 5.9 ACA 分群結果

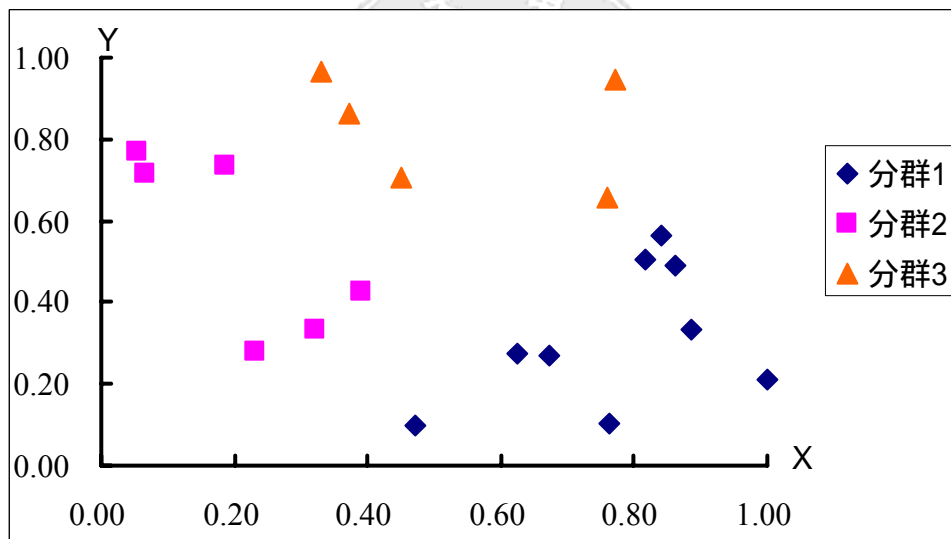


圖 5.10 k-means 分群結果

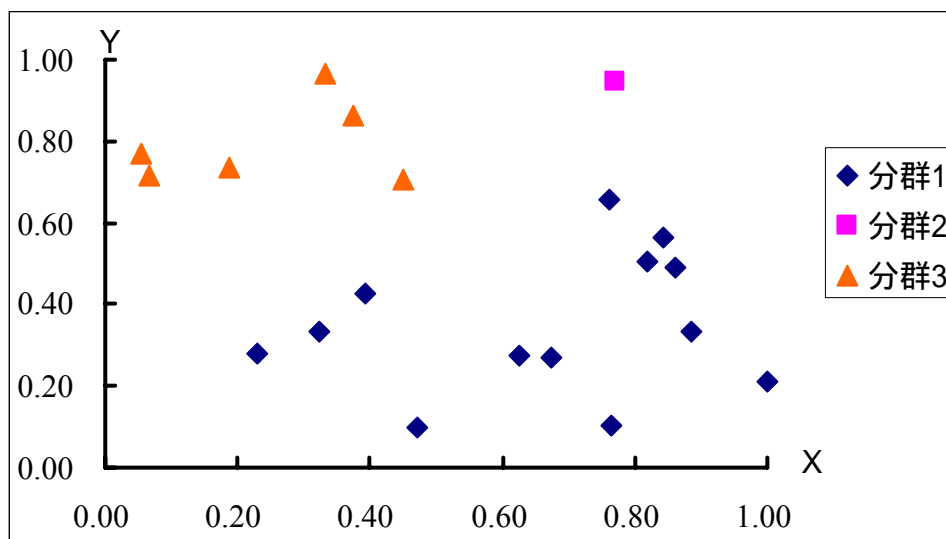


圖 5.11 凝聚法分群結果

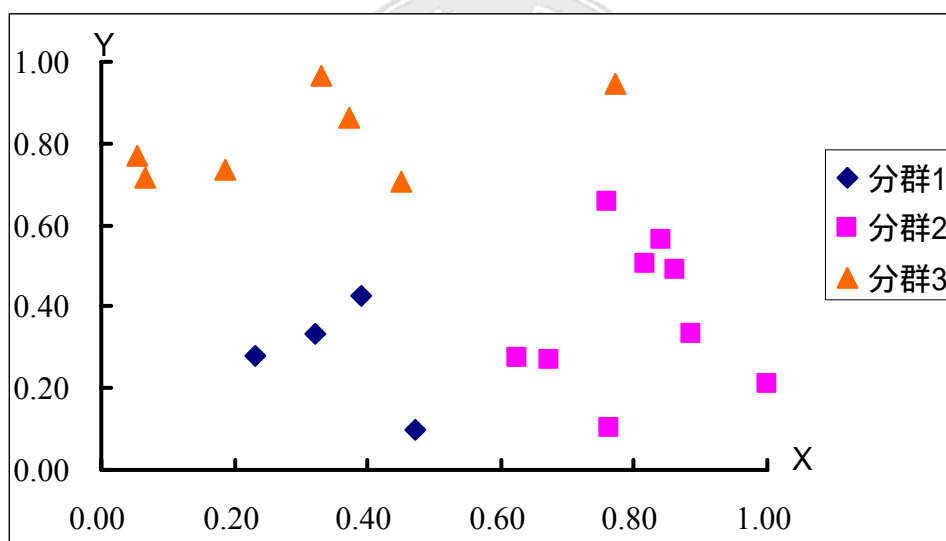


圖 5.12 GCA 分群結果

圖 5.9、5.10、5.11 與 5.12 為在分群規模為 3 群時，各分群演算法於相同的資料樣本的分群結果，並由表 5.7 中各分群法 TWCV 值，概約判斷出 ACA 演算法有較佳的分群結果。

表 5.7 不同分群演算法之分群結果

分群演算法	ACA	k-means	凝聚法	GCA
TWCV 值	0.882794	1.037672	1.235969	0.9104238

為了證明 ACA 分群演算法較其他三種分群演算法有較佳的分群結果，本研究利用隨機產生 30 組資料規模為 20 筆資料的樣本，此 30 組隨機樣本亦符合均等分配，資料樣本於 3 群分群規模下分群結果之 TWCV 值如表 5.8 所示：

表 5.8 三十組資料分析結果

	ACA	k-means	凝聚法	GCA
1	0.895161	0.900735	1.141268	0.895193
2	0.717871	0.717871	0.717871	0.717913
3	0.882794	1.037672	1.235969	1.037615
4	0.968726	0.968726	0.982740	0.968686
5	1.179625	1.179625	1.369426	1.179688
6	0.706837	0.706837	0.706837	0.706837
7	0.976394	0.985207	1.006970	0.976355
8	0.736697	0.786272	0.851331	0.736706
9	0.939820	1.004528	1.049665	0.939809
10	0.741833	0.741833	0.741833	0.741802
11	0.808474	0.862998	1.125641	0.808509
12	0.9005964	0.930472	0.972747	0.900680
13	0.935666	1.097122	1.208515	0.935785
14	0.967657	0.967657	0.967657	0.967671
15	1.094308	1.128075	1.128075	1.094380
16	0.805676	0.829939	0.887031	0.814866
17	1.065096	1.097941	1.602836	1.160268
18	0.916173	0.934838	0.952080	0.916126
19	1.185540	1.236898	1.242216	1.185656

表 5.8 (續) 三十組資料分析結果

	ACA	k-means	凝聚法	GCA
20	0.943840	0.955850	1.368649	0.943767
21	0.880151	0.880151	1.042605	0.880186
22	0.880522	0.952549	1.086715	0.880444
23	0.851554	0.851554	0.940952	0.851600
24	0.885538	0.952673	0.937321	0.885513
25	1.126039	1.182149	1.582442	1.126014
26	0.7203053	0.7395181	0.720305	0.720324
27	0.9665604	0.9711214	0.972266	0.992755
28	0.7782631	0.7782631	0.965091	0.778232
29	0.6207008	0.7281865	0.620701	0.620728
30	0.9682736	0.9682736	1.140110	0.968153

針對各分群演算法所求出的分群結果，往後章節將運用統計分析方法來驗證 ACA 於不同資料規模及不同分群規模是否有較佳的分群結果，藉此來驗證 ACA 是否較其他三種分群演算法有較佳的求解穩定度與求解結果。

5.1.5 不同分群規模之比較

此節將針對資料規模為 20 筆資料，在不同分群規模（3 群、5 群、7 群、9 群）下不同分群演算法是否有顯著不同的分群結果。驗證資料為與前小節相同的隨機產生之 30 組資料，首先針對在不同分群規模下，分群結果的變異數是否有顯著地差別作一分析討論。再利用與前小節所運用的檢定方式進行分析；對於 ACA 與其他三種分群演算法於不同分群規模下進行 F 檢定，其結果如表 5.9 所示：

表 5.9 ACA 與凝聚法分群結果之 F 檢定 (20 筆資料)

3 群			5 群		
	凝聚法	ACA		凝聚法	ACA
平均數	0.916829	0.901556	平均數	0.444613	0.396940
變異數	0.023375	0.019767	變異數	0.011021	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.182536		F	2.332003	
P(F<=f) 單尾	0.327326		P(F<=f) 單尾	0.012936	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	
7 群			9 群		
	凝聚法	ACA		凝聚法	ACA
平均數	0.232881	0.215554	平均數	0.125766	0.120697
變異數	0.002147	0.001701	變異數	0.000801	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.262281		F	1.013215	
P(F<=f) 單尾	0.267273		P(F<=f) 單尾	0.486021	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

由表 5.9 可知，在信賴水準 $\alpha=0.05$ 時，20 筆資料規模僅有 5 群分群規模時，ACA 與凝聚法的 F 檢定值有顯著的差異，表示 ACA 於此情況下較凝聚法有優良的求解穩定度，其於狀況皆顯示 ACA 與凝聚法所求解出的分群結果穩定度間之差異不甚顯著。

表 5.10 ACA 與凝聚法分群結果之 t 檢定 (20 筆資料)

3 群			5 群		
	凝聚法	ACA		凝聚法	ACA
平均數	0.916829	0.901556	平均數	0.444613	0.396940
變異數	0.023375	0.019767	變異數	0.011021	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	2.483037		t 統計	4.126895	
P(T<=t) 單尾	0.009528		P(T<=t) 單尾	0.000142	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	
7 群			9 群		
	凝聚法	ACA		凝聚法	ACA
平均數	0.232881	0.215554	平均數	0.125766	0.120697
變異數	0.002147	0.001701	變異數	0.000801	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	5.203889		t 統計	3.181921	
P(T<=t) 單尾	7.196E-06		P(T<=t) 單尾	0.001738	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

根據表 5.10 可知，在小規模問題中，不同的分群規模下 ACA 與凝聚法的 t 檢定值皆呈現顯著差異的結果，即此情況下 ACA 較凝聚法有優良的分群結果。

對於 ACA 分群演算法與 k-means 的求解結果比較，係利用相同的方式，在信賴水準 $\alpha=0.05$ 透過 F 檢定兩分群方法的求解結果之變異與 t 檢定兩分群方法在求解結果是否有顯著的差異，兩分群方法的 F 檢定如表 5.11 所示：

表 5.11 ACA 與 k-means 分群結果之 F 檢定 (20 筆資料)

3 群			5 群		
	k-means	ACA		k-means	ACA
平均數	0.935851	0.901556	平均數	0.430435	0.396940
變異數	0.021286	0.019767	變異數	0.021286	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.076870		F	1.354546	
P(F<=f) 單尾	0.421650		P(F<=f) 單尾	0.209383	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	
7 群			9 群		
	k-means	ACA		k-means	ACA
平均數	0.243863	0.215554	平均數	0.132132	0.120697
變異數	0.003407	0.001701	變異數	0.000900	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	2.003419		F	1.137767	
P(F<=f) 單尾	0.033113		P(F<=f) 單尾	0.365263	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

由表 5.11 可以得知，大部分情況下 ACA 分群演算法與 k-means 在分群結果之變異數差異並不顯著，但在分群規模為 7 群時，ACA 與 k-means 之變異數差異顯著，表示在此情況下 ACA 有較小的分群結果之變異數。

表 5.12 ACA 與 k-means 分群結果之 t 檢定 (20 筆資料)

3 群			5 群		
	k-means	ACA		k-means	ACA
平均數	0.935851	0.901556	平均數	0.430435	0.396940
變異數	0.021286	0.019767	變異數	0.006401	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	4.270209		t 統計	5.372902	
P(T<=t) 單尾	9.562E-05		P(T<=t) 單尾	4.500E-06	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

表 5.12 (續) ACA 與 k-means 分群結果之 t 檢定 (20 筆資料)

7 群			9 群		
	k-means	ACA		k-means	ACA
平均數	0.243863	0.215554	平均數	0.132132	0.120697
變異數	0.003407	0.001701	變異數	0.000900	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	5.044567		t 統計	3.809611	
P(T<=t) 單尾	1.121E-05		P(T<=t) 單尾	3.346E-04	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

由表 5.12 可知在信賴水準 $\alpha=0.05$ 不同的分群規模下，ACA 分群演算法與 k-means 比較下有顯著的較佳分群結果；對於 ACA 分群演算法與 GCA 的求解結果比較，也是運用相同的方式，在信賴水準 $\alpha=0.05$ 透過 F 檢定兩分群方法的求解結果之變異與 t 檢定兩分群方法在求解結果是否有顯著的差異，兩分群方法的 F 檢定如表 5.13 所示：

表 5.13 ACA 與 GCA 分群結果之 F 檢定 (20 筆資料)

3 群			5 群		
	GCA	ACA		GCA	ACA
平均數	0.911075	0.901556	平均數	0.428213	0.396940
變異數	0.021770	0.019767	變異數	0.005074	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.101357		F	1.073597	
P(F<=f) 單尾	0.398331		P(F<=f) 單尾	0.424830	
臨界值：單尾	1.860812		臨界值：單尾	1.860812	

表 5.13 (續) ACA 與 GCA 分群結果之 F 檢定 (20 筆資料)

7 群			9 群		
	GCA	ACA		GCA	ACA
平均數	0.247275	0.215554	平均數	0.139040	0.120697
變異數	0.002972	0.001701	變異數	0.000955	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.747577		F	1.207928	
P(F<=f) 單尾	0.069336		P(F<=f) 單尾	0.307162	
臨界值：單尾	1.860812		臨界值：單尾	1.860812	

根據表 5.13 可知，兩種演算法在此情況下 F 檢定值並不顯著，意謂兩者間的求解穩定度並沒有顯著差異，以下將以 t 檢定來比較兩者間的求解結果平均值是否有顯著的差異。

表 5.14 ACA 與 GCA 分群結果之 t 檢定 (20 筆資料)

3 群			5 群		
	GCA	ACA		GCA	ACA
平均數	0.911075	0.901556	平均數	0.428213	0.396940
變異數	0.021770	0.019767	變異數	0.005074	0.004726
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	1.592637		t 統計	4.042646	
P(T<=t) 單尾	0.061042		P(T<=t) 單尾	1.781E-04	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

7 群			9 群		
	GCA	ACA		GCA	ACA
平均數	0.247275	0.215554	平均數	0.139040	0.120697
變異數	0.002972	0.001701	變異數	0.000955	0.000791
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	4.941420		t 統計	4.282317	
P(T<=t) 單尾	1.493E-05		P(T<=t) 單尾	9.249E-05	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

由表 5.14 可知，ACA 與 GCA 在分群規模為 3 群時檢定結果並不顯著，而在 5 群、7 群與 9 群其平均值檢定結果顯示兩種演算法間的分群結果有顯著的差異，即 ACA 比 GCA 有較佳的分群結果。

根據本節分析結果，在樣本數為 20 筆資料時，ACA 演算法與凝聚法在求解結果的變異數差異並不顯著，而在求解出的分群結果有較佳的分群表現；ACA 演算法與 k-means 在求解結果的變異數差異並沒有顯著差異，在求解分群問題中則有較佳分群結果；ACA 與 GCA 比較下，在求解結果的變異數差異並沒有顯著差異，在求解分群問題中，除了在分群規模 3 群時結果不顯著外，其他分群規模下皆有較佳分群結果，圖 5.13、5.14、5.15 分別為利用 ACA 分群演算法透過 5.1.3 節的小規模問題，以分群規模 5 群、7 群與 9 群所得的分群結果，圖 5.16 為 20 筆資料規模下，不同分群規模各分群演算法求解結果趨勢圖。

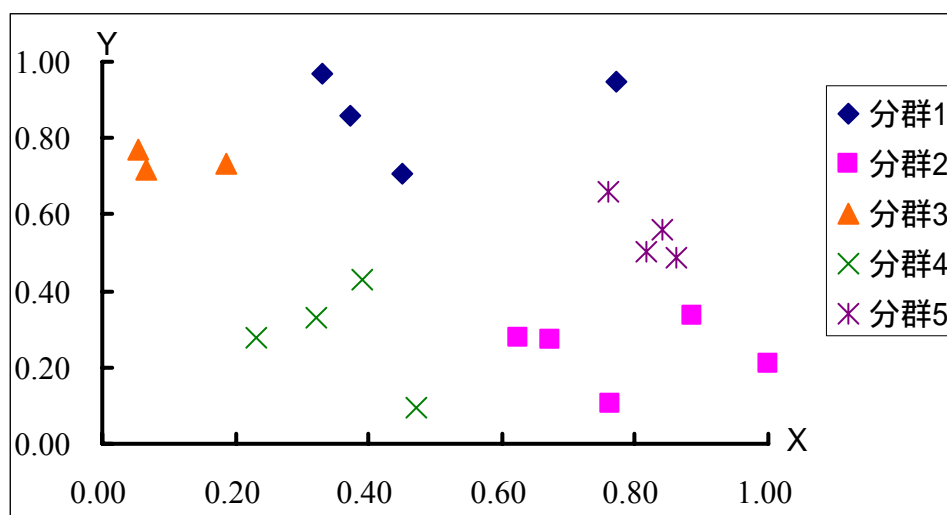


圖 5.13 20 筆資料下 ACA 分群結果 (5 群)

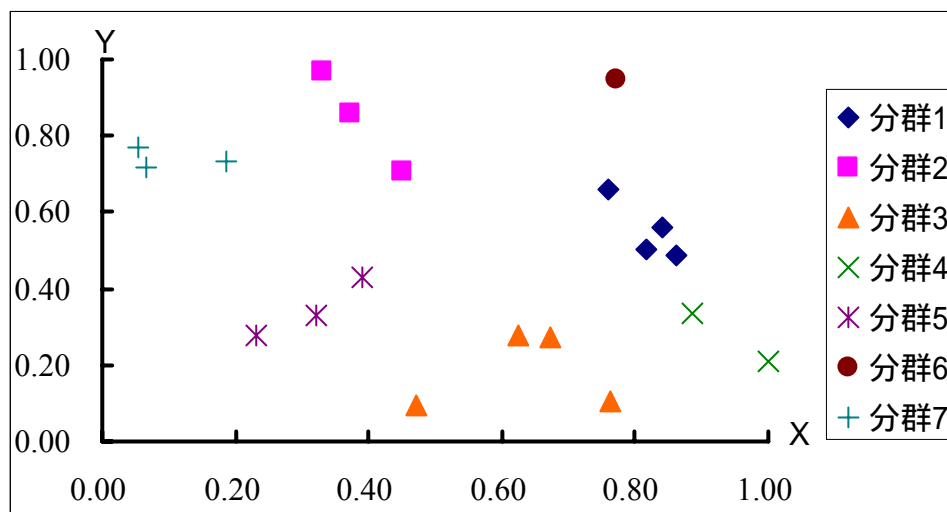


圖 5.14 20 筆資料下 ACA 分群結果 (7 群)

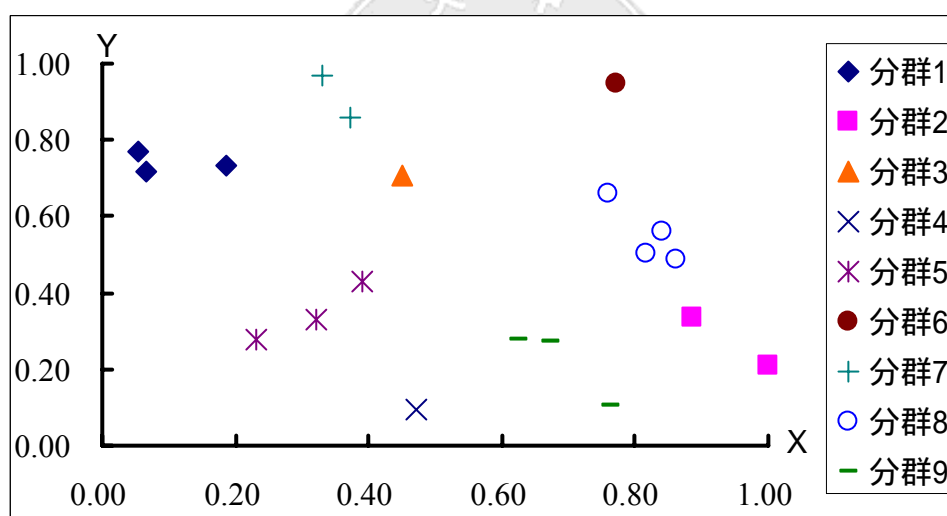


圖 5.15 20 筆資料下 ACA 分群結果 (9 群)

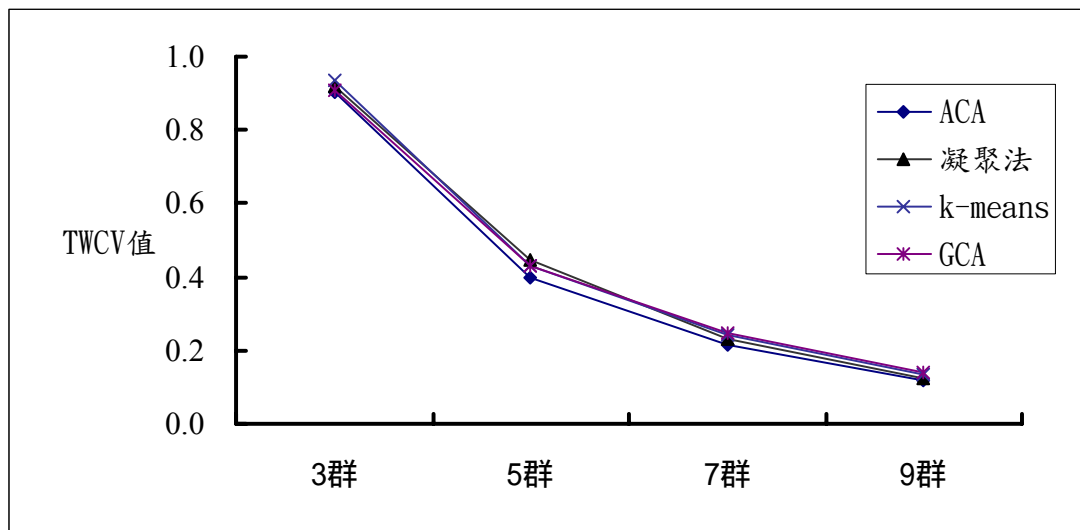


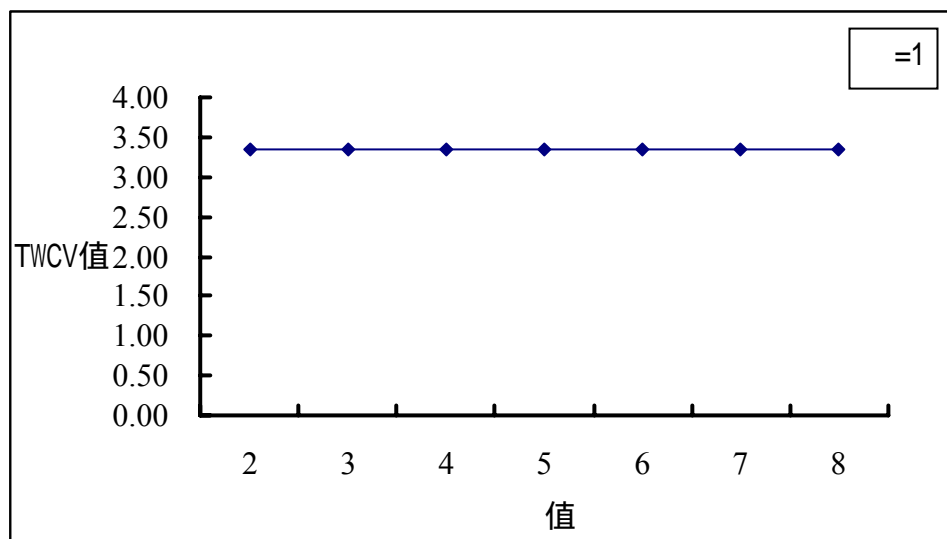
圖 5.16 不同分群規模各分群演算法求解結果趨勢圖 (20 筆資料)

5.2 中規模問題之比較

經過前述分析出 ACA-2 有較佳的分群結果後，本節將 ACA-2 設定為本研究主要的螞蟻分群演算法 (ACA)，並且透過中規模問題來驗證螞蟻分群演算法比其他分群演算法有較佳的分群結果。

5.2.1 參數設定

中規模問題中， β 值在變動範圍中所呈現的 TWCV 值皆相同 (如圖 5.17 所示)，表示 α 與 β 值對於此情況下對於求解結果的敏感度低，故中規模問題時參數設定 $\alpha=1$ 、 $\beta=2$ ，其他相關參數如費洛蒙衰退參數 $\rho=0.4$ ，轉換規則參數 $q_0=0.8$ ，而螞蟻數即為分群群數，目標分群次數為 $\delta=1000$ ，更新費洛蒙濃度之門檻次數 $\eta=100$ 。

圖 5.17 決定 α 、 β 之曲線圖（中規模問題）

5.2.2 不同分群規模之比較

利用隨機方式產生 30 組資料規模為 50 筆資料的樣本，透過統計方法來驗證與其他分群演算法是否有較小的分群變異與較佳的分群結果；利用 F 檢定來針對演算法的變異數作一探討，透過 t 檢定來針對演算法的分群結果與其他分群演算法作一比較分析。以下將針對 ACA 與凝聚法、k-means、GCA 在資料規模為 50 筆資料時不同分群規模下進行一連串的比较分析。

表 5.15 ACA 與凝聚法分群結果之 F 檢定（50 筆資料）

3 群			5 群		
	凝聚法	ACA		凝聚法	ACA
平均數	3.090540	2.731044	平均數	1.528528	1.374884
變異數	0.228300	0.071955	變異數	0.041337	0.019221
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	3.172824		F	2.150599	
P(F<=f) 單尾	0.001336		P(F<=f) 單尾	0.021683	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

表 5.15 (續) ACA 與凝聚法分群結果之 F 檢定 (50 筆資料)

7 群			9 群		
	凝聚法	ACA		凝聚法	ACA
平均數	0.986550	0.890859	平均數	0.671757	0.623296
變異數	0.020491	0.011646	變異數	0.006963	0.005907
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.759471		F	1.178849	
P(F<=f) 單尾	0.066999		P(F<=f) 單尾	0.330336	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

根據表 5.15 可知在分群規模為 3 群和 5 群時，ACA 與凝聚法的檢定值為顯著差異代表兩種演算法間的求解差異是顯著的，在其他狀況下（7 群、9 群）則為不顯著，代表兩種演算法間的求解差異不顯著；接著，對於兩種演算法間進行平均值檢定以了解兩種演算法何者有較佳的分群結果，其結果如表 5.16 所示：

表 5.16 ACA 與凝聚法分群結果之 t 檢定 (50 筆資料)

3 群			5 群		
	凝聚法	ACA		凝聚法	ACA
平均數	3.090540	2.731044	平均數	1.528528	1.374884
變異數	0.228300	0.071955	變異數	0.041337	0.019221
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	5.261116		t 統計	6.191571	
P(T<=t) 單尾	6.138E-06		P(T<=t) 單尾	4.715E-07	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

表 5.16 (續) ACA 與凝聚法分群結果之 t 檢定 (50 筆資料)

7 群			9 群		
	凝聚法	ACA		凝聚法	ACA
平均數	0.986550	0.890859	平均數	0.671757	0.623296
變異數	0.020491	0.011646	變異數	0.006963	0.005907
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	6.660972		t 統計	6.475065	
P(T<=t) 單尾	1.322E-07		P(T<=t) 單尾	2.182E-07	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

根據表 5.16 可知，在資料規模為 50 筆資料時，不同的分群規模下，ACA 皆有較凝聚法優良的分群結果。

表 5.17 ACA 與 k-means 分群結果之 F 檢定 (50 筆資料)

3 群			5 群		
	k-means	ACA		k-means	ACA
平均數	2.856622	2.731044	平均數	1.442647	1.374884
變異數	0.159091	0.071955	變異數	0.022094	0.019221
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	2.210977		F	1.149467	
P(F<=f) 單尾	0.018244		P(F<=f) 單尾	0.355056	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	
7 群			9 群		
	k-means	ACA		k-means	ACA
平均數	0.943636	0.890859	平均數	0.671767	0.623296
變異數	0.015643	0.011646	變異數	0.010026	0.005907
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.343205		F	1.697334	
P(F<=f) 單尾	0.215863		P(F<=f) 單尾	0.080119	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

由表 5.17 可知，在分群規模為 3 群時，ACA 比 k-means 有較小的求解變異，而在其他分群規模下則兩種演算法的求解結果並沒有明顯的差異。

表 5.18 ACA 與 k-means 分群結果之 t 檢定 (50 筆資料)

3 群			5 群		
	k-means	ACA		k-means	ACA
平均數	2.856622	2.731044	平均數	1.442647	1.374884
變異數	0.159091	0.071955	變異數	0.022094	0.019221
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	2.369080		t 統計	4.810887	
P(T<=t) 單尾	0.012354		P(T<=t) 單尾	2.145E-05	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	
7 群			9 群		
	k-means	ACA		k-means	ACA
平均數	0.943636	0.890859	平均數	0.671767	0.623296
變異數	0.015643	0.011646	變異數	0.010026	0.005907
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	3.968956		t 統計	6.258198	
P(T<=t) 單尾	0.000218		P(T<=t) 單尾	3.932E-07	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

表 5.18 顯示在資料規模為 50 筆資料時，各不同分群規模的分群結果平均值檢定，皆顯示 ACA 與 k-means 的分群結果有顯著的差異，即 ACA 有較 k-means 優良的分群結果。

表 5.19 ACA 與 GCA 分群結果之 F 檢定 (50 筆資料)

3 群			5 群		
	GCA	ACA		GCA	ACA
平均數	2.866513	2.731044	平均數	1.521355	1.374884
變異數	0.110211	0.071955	變異數	0.031071	0.019221
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.531665		F	1.616533	
P(F<=f) 單尾	0.128415		P(F<=f) 單尾	0.100964	
臨界值：單尾	1.860812		臨界值：單尾	1.860812	
7 群			9 群		
	GCA	ACA		GCA	ACA
平均數	1.019087	0.890859	平均數	0.785677	0.623296
變異數	0.016734	0.011646	變異數	0.012795	0.005907
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.436900		F	2.166066	
P(F<=f) 單尾	0.167261		P(F<=f) 單尾	0.020743	
臨界值：單尾	1.860812		臨界值：單尾	1.860812	

由表 5.19 可知，在分群規模為 9 群時，因其 P 值小於 0.05 故 ACA 與 GCA 在分群結果的變異有顯著差異，而在 3 群、5 群、7 群時，兩種分群演算法的求解結果變異則不顯著。

表 5.20 ACA 與 GCA 分群結果之 t 檢定 (50 筆資料)

3 群			5 群		
	GCA	ACA		GCA	ACA
平均數	2.866513	2.731044	平均數	1.521355	1.374884
變異數	0.110211	0.071955	變異數	0.031071	0.019221
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	4.652613		t 統計	7.444536	
P(T<=t) 單尾	3.328E-05		P(T<=t) 單尾	1.665E-08	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

表 5.20 (續) ACA 與 GCA 分群結果之 t 檢定 (50 筆資料)

7 群			9 群		
	GCA	ACA		GCA	ACA
平均數	1.019087	0.890859	平均數	0.785677	0.623296
變異數	0.016734	0.011646	變異數	0.012795	0.005907
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	6.747133		t 統計	9.222164	
P(T<=t) 單尾	1.049E-07		P(T<=t) 單尾	2.004E-10	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

由表 5.20 可以看出，在 50 筆資料規模下，不同的分群規模 ACA 與 GCA 之 P 值皆小於 0.05 代表拒絕虛無假設，即在此條件下 ACA 有較 GCA 優良的分群結果。

經由前述分析可知，ACA 在資料規模為 50 筆資料時，與其他三種分群演算法比較求解變異並沒有一定的趨勢顯示是否會隨著問題複雜度提昇（即增加分群規模）而有較為顯著差異的差異性，但是在求解結果方面，經過驗證觀察各演算法的績效值（TWCV 值），可以發現隨著問題複雜度的增加越能顯現出 ACA 之求解分群結果，優於其他三種分群演算法。圖 5.18、5.19、5.20 與 5.21 為在 50 筆資料下，ACA 於不同分群規模的分群結果，圖 5.22 為不同分群規模各分群演算法於 50 筆資料規模時求解結果趨勢圖。

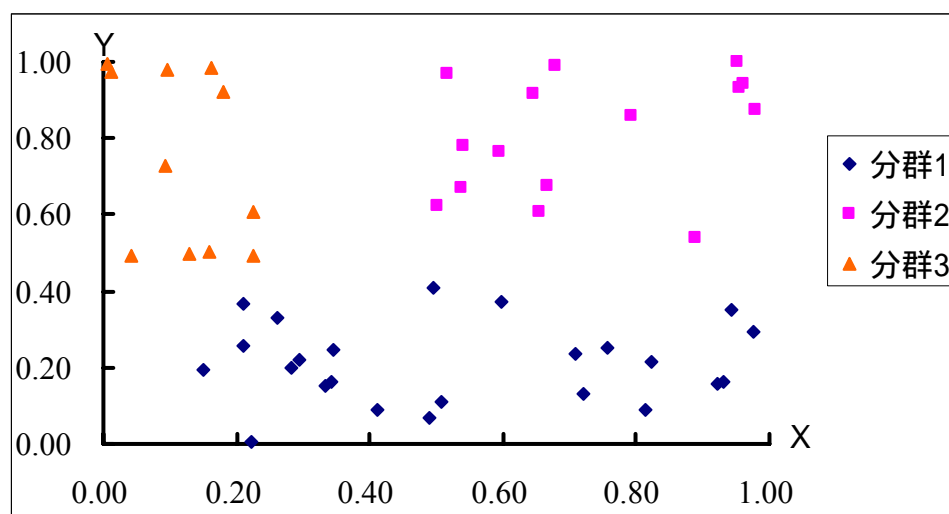


圖 5.18 50 筆資料下 ACA 分群結果 (3 群)

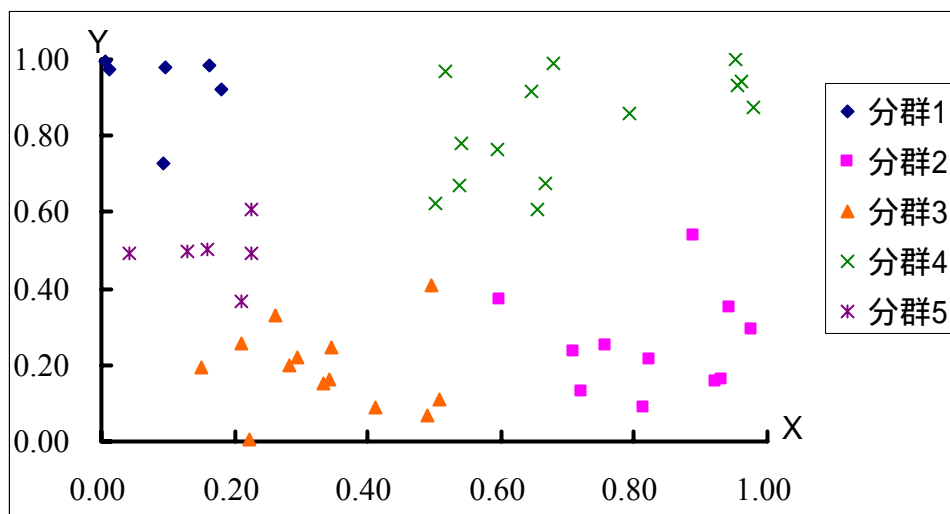


圖 5.19 50 筆資料下 ACA 分群結果 (5 群)

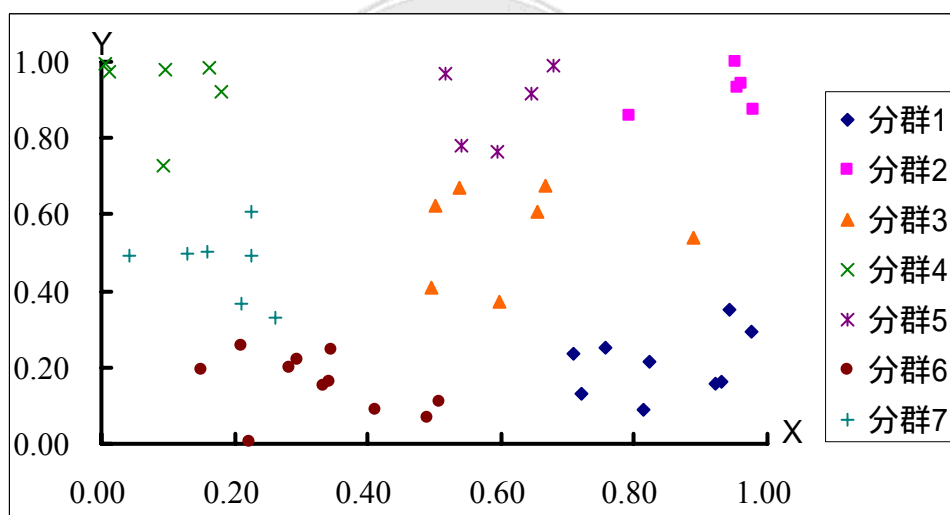


圖 5.20 50 筆資料下 ACA 分群結果 (7 群)

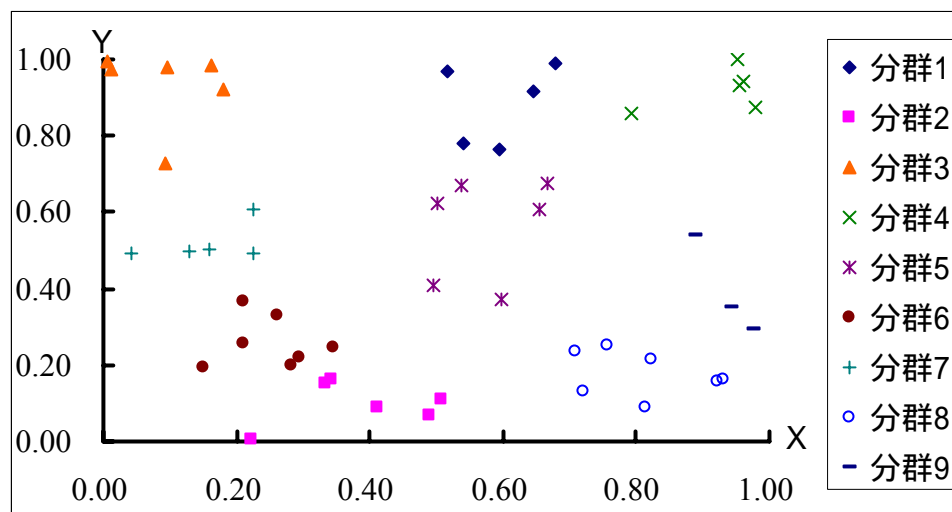


圖 5.21 50 筆資料下 ACA 分群結果 (9 群)

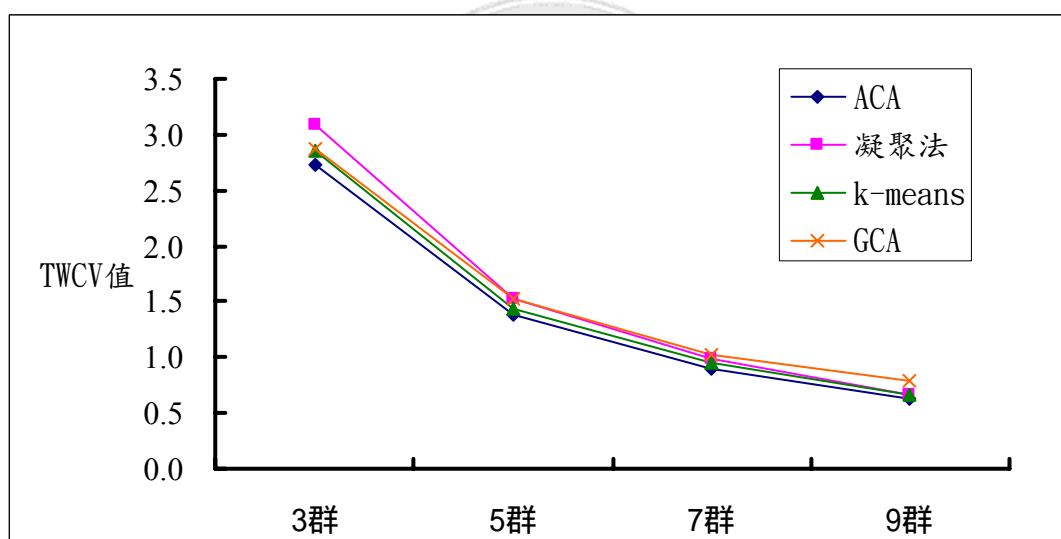


圖 5.22 不同分群規模各分群演算法求解結果趨勢圖 (50 筆資料)

5.3 大規模問題之比較

5.3.1 參數設定

在大規模問題中，相關參數設定如同其他問題規模一樣，採用隨機產生的一組資料樣本透過試誤法找到適當的參數設定，此狀況下固定 $\alpha = 1$ 在不同的 β 值變動找尋適當的參數設定，圖 5.23 為固定 α 值變動 β 值所得的 TWCV 曲線圖：

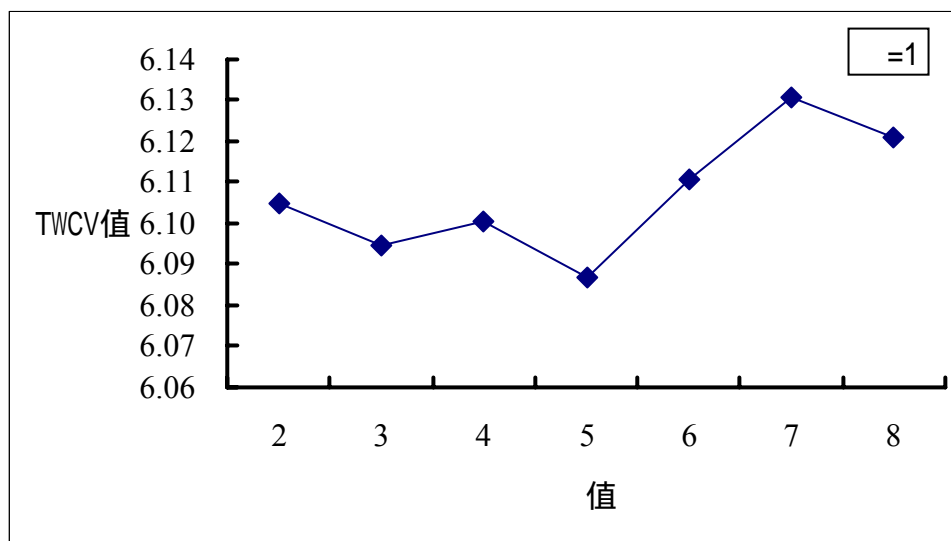


圖 5.23 決定 α 、 β 之曲線圖（大規模問題）

由圖 5.23 可知，在此情況下 α 、 β 值對於求解結果有較明顯的敏感度，而在 $\alpha = 1$ 、 $\beta = 5$ 時有最小的 TWCV 值，故在大規模問題時，各相關參數設定為 $\alpha = 1$ ， $\beta = 5$ ，費洛蒙衰退參數 $\rho = 0.4$ ，轉換規則參數 $q_0 = 0.8$ ，而螞蟻數即為分群群數，目標分群次數為 $\delta = 1000$ ，更新費洛蒙濃度之門檻次數 $\eta = 100$ 。

5.3.2 不同分群規模之比較

利用隨機方式產生 30 組資料規模為 100 筆資料的樣本，透過統計方法來驗證與其他分群演算法是否有較小的分群變異與較佳的分群結果；利用 F 檢定來針對演算法的變異數作一探討，並透過 t 檢定來針對演算法的分群結果與其他分群演算法作一比較分析。以下首先針對 ACA 與凝聚法在 100 筆資料規模不同分群規模進行分析比較：

表 5.21 ACA 與凝聚法分群結果之 F 檢定 (100 筆資料)

3 群			5 群		
	凝聚法	ACA		凝聚法	ACA
平均數	6.632101	5.921159	平均數	3.419648	3.111071
變異數	0.598256	0.152190	變異數	0.099658	0.032420
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	3.930988		F	3.073940	
P(F<=f) 單尾	0.000211		P(F<=f) 單尾	0.001724	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	
7 群			9 群		
	凝聚法	ACA		凝聚法	ACA
平均數	2.274569	2.070966	平均數	1.638305	1.515969
變異數	0.050550	0.018898	變異數	0.022439	0.007589
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	2.674868		F	2.956794	
P(F<=f) 單尾	0.004988		P(F<=f) 單尾	0.002342	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

由表 5.21 可知，在信賴水準 0.05 下，ACA 與凝聚法在不同分群規模下，ACA 與凝聚法之變異差異有顯著差異，表示 ACA 比凝聚法有較小的求解變異。

表 5.22 ACA 與凝聚法分群結果之 t 檢定 (100 筆資料)

3 群			5 群		
	凝聚法	ACA		凝聚法	ACA
平均數	6.632101	5.921159	平均數	3.419648	3.111071
變異數	0.598256	0.152190	變異數	0.099658	0.032420
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	6.773117		t 統計	8.644064	
P(T<=t) 單尾	9.789E-08		P(T<=t) 單尾	8.066E-10	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

表 5.22 (續) ACA 與凝聚法分群結果之 t 檢定 (100 筆資料)

7 群			9 群		
	凝聚法	ACA		凝聚法	ACA
平均數	2.274569	2.070966	平均數	1.638305	1.515969
變異數	0.050550	0.018898	變異數	0.022439	0.007589
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	6.901193		t 統計	5.585158	
P(T<=t) 單尾	6.956E-08		P(T<=t) 單尾	2.499E-06	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

由表 5.22 觀察，經過驗證 ACA 在此情況下，不同分群規模的分群結果變異與凝聚法比較下，皆有顯著的差異，即 ACA 有較小的求解變異。

表 5.23 ACA 與 k-means 分群結果之 F 檢定 (100 筆資料)

3 群			5 群		
	k-means	ACA		k-means	ACA
平均數	6.025375	5.921159	平均數	3.228114	3.111071
變異數	0.211024	0.152190	變異數	0.073609	0.032420
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.386586		F	2.270464	
P(F<=f) 單尾	0.191989		P(F<=f) 單尾	0.015401	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

7 群			9 群		
	k-means	ACA		k-means	ACA
平均數	2.138965	2.070966	平均數	1.537021	1.515969
變異數	0.025497	0.018898	變異數	0.009195	0.007589
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.349175		F	1.211624	
P(F<=f) 單尾	0.212430		P(F<=f) 單尾	0.304309	
臨界值：單尾	1.860811		臨界值：單尾	1.860811	

從表 5.23 可知，ACA 與 k-means 只有在分群規模為 5 群時，兩者間的求解變異差異顯著，其他情況下皆不顯著，即 ACA 與 k-means 在分群結果的變異上並無顯著差異。

表 5.24 ACA 與 k-means 分群結果之 t 檢定（100 筆資料）

3 群			5 群		
	k-means	ACA		k-means	ACA
平均數	6.025375	5.921159	平均數	3.228114	3.111071
變異數	0.211024	0.152190	變異數	0.073609	0.032420
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	2.379797		t 統計	4.645716	
P(T<=t) 單尾	0.012058		P(T<=t) 單尾	0.000034	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	
7 群			9 群		
	k-means	ACA		k-means	ACA
平均數	2.138965	2.070966	平均數	1.537021	1.515969
變異數	0.025497	0.018898	變異數	0.009195	0.007589
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	4.140089		t 統計	1.792226	
P(T<=t) 單尾	0.000137		P(T<=t) 單尾	0.041768	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

從表 5.24 可知，在資料規模為 100 筆的情況下，不同的分群規模 ACA 皆比 k-means 有較為優良的分群結果。接下來將針對 ACA 與 GCA 進行相同的統計分析。

表 5.25 ACA 與 GCA 分群結果之 F 檢定 (100 筆資料)

3 群			5 群		
	GCA	ACA		GCA	ACA
平均數	6.125255	5.921159	平均數	3.586631	3.111071
變異數	0.250539	0.152190	變異數	0.163945	0.032420
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	1.646231		F	5.056884	
P(F<=f) 單尾	0.092757		P(F<=f) 單尾	1.884E-05	
臨界值：單尾	1.860812		臨界值：單尾	1.860812	
7 群			9 群		
	GCA	ACA		GCA	ACA
平均數	2.918848	2.070966	平均數	3.254643	1.515969
變異數	0.158307	0.018898	變異數	0.571934	0.007589
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
F	8.376868		F	75.362314	
P(F<=f) 單尾	7.705E-08		P(F<=f) 單尾	1.667E-20	
臨界值：單尾	1.860812		臨界值：單尾	1.860812	

由表 5.25 可知，在分群規模為 3 群時，ACA 與 GCA 間的求解分群差異並不顯著，但在 5 群、7 群、9 群時則檢定值呈現顯著的狀態，即隨著分群規模的提升，ACA 有較 GCA 為小的求解變異。

表 5.26 ACA 與 GCA 分群結果之 t 檢定 (100 筆資料)

3 群			5 群		
	GCA	ACA		GCA	ACA
平均數	6.125255	5.921159	平均數	3.586631	3.111071
變異數	0.250539	0.152190	變異數	0.163945	0.032420
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	1.646231		t 統計	5.056884	
P(T<=t) 單尾	0.092757		P(T<=t) 單尾	1.884E-05	
臨界值：單尾	1.860812		臨界值：單尾	1.860812	

表 5.26 (續) ACA 與 GCA 分群結果之 t 檢定 (100 筆資料)

7 群			9 群		
	GCA	ACA		GCA	ACA
平均數	2.918848	2.070966	平均數	3.254643	1.515969
變異數	0.158307	0.018898	變異數	0.571934	0.007589
觀察值個數	30		觀察值個數	30	
自由度	29		自由度	29	
t 統計	11.810877		t 統計	12.982051	
P(T<=t) 單尾	6.657E-13		P(T<=t) 單尾	6.547E-14	
臨界值：單尾	1.699127		臨界值：單尾	1.699127	

從表 5.26 可知，在此資料規模下，不同的分群規模 ACA 與 GCA 的檢定值皆呈現顯著狀態，即 ACA 有較 GCA 優良的分群結果。

經由前述可知，在資料規模為 100 筆時，大部份情況下 ACA 皆有較其他三種分群演算法的求解結果較小的變異，且在求解分群結果方面，經過統計方法驗證，ACA 跟其他三種分群演算法比較明顯有較佳的分群結果。圖 5.24、5.25、5.26、5.27 為 100 筆資料規模時，ACA 分群演算法之分群結果，圖 5.28 為不同分群規模各分群演算法於 100 筆資料規模的求解結果趨勢圖。

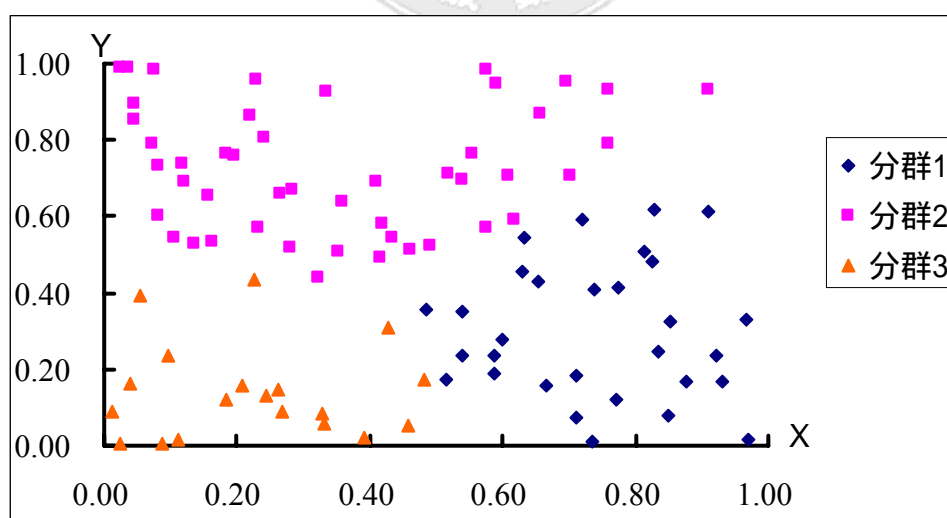


圖 5.24 100 筆資料下 ACA 分群結果 (3 群)

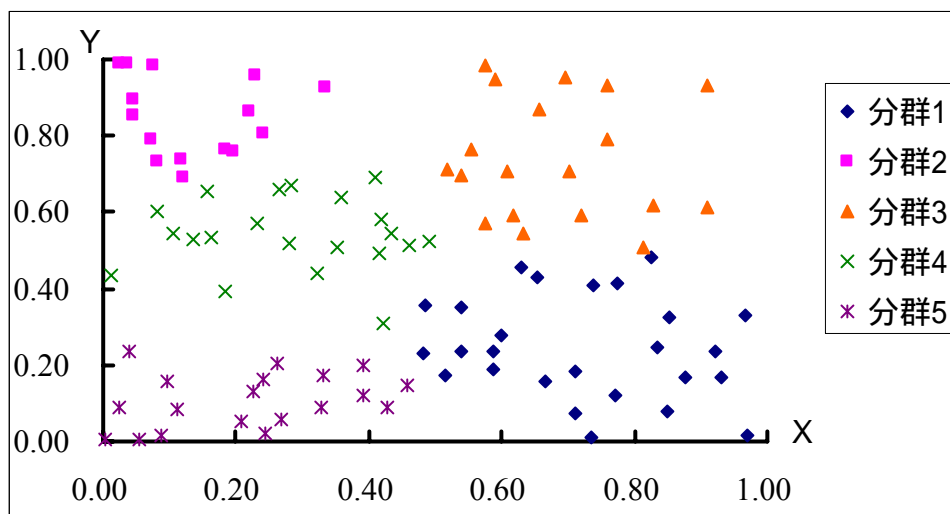


圖 5.25 100 筆資料下 ACA 分群結果 (5 群)

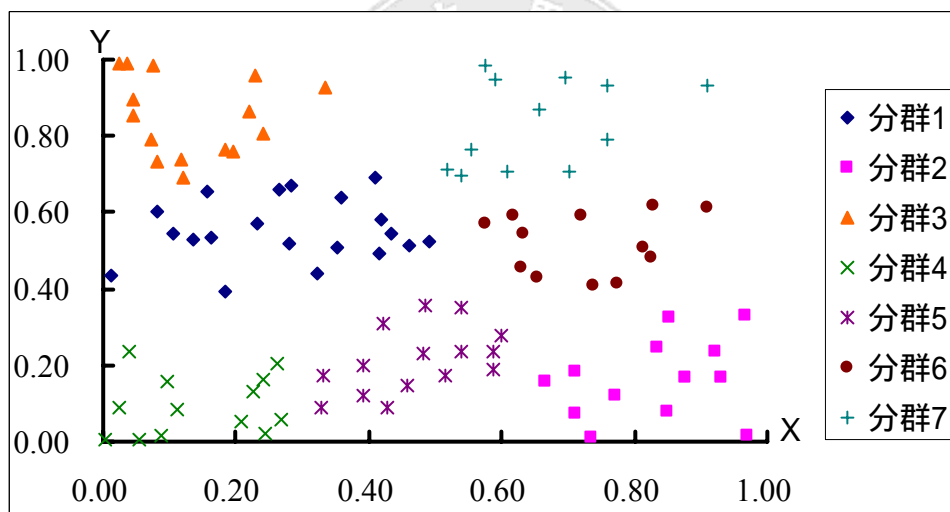


圖 5.26 100 筆資料下 ACA 分群結果 (7 群)

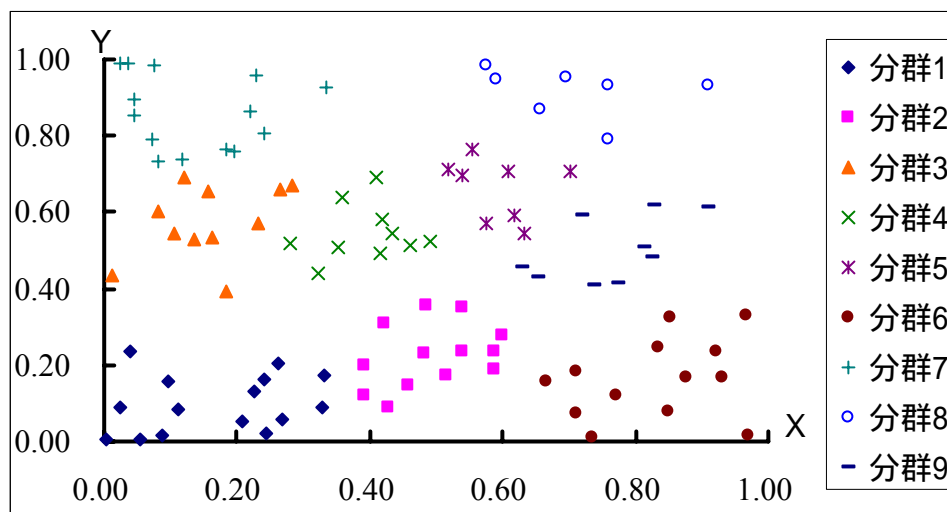


圖 5.27 100 筆資料下 ACA 分群結果 (9 群)

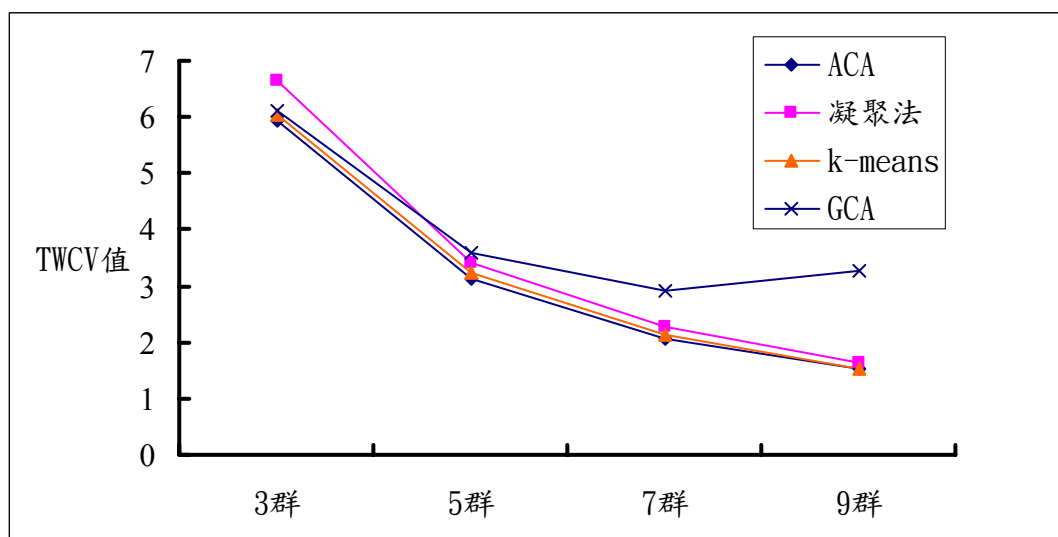


圖 5.28 不同分群規模各分群演算法求解結果趨勢圖 (100 筆資料)

根據本章節的分析比較之後，可以觀察到 ACA 與其他分群演算法在求解穩定度差異大部分情況下為沒有顯著差異，在求解的效度上，ACA 與其他分群演算法比較後，於大部分情況下，ACA 有較其他分群演算法較為明顯差異的分群結果，表 5.27 為 ACA 對於不同分群演算法求解結果的改善率：

表 5.27 ACA 對於其他分群演算法求解結果的改善率

	ACA 對凝聚法	ACA 對 k-means	ACA 對 GCA
小規模問題	1.67%~10.72%	3.67%~11.61%	1.05%~13.19%
中規模問題	7.21%~11.63%	4.40%~7.22%	4.73%~20.67%
大規模問題	7.47%~10.72%	1.37%~3.18%	3.33%~53.42%



第六章 實例應用分析

本研究利用自民國 89 年 3 月至民國 91 年 8 月之鑑定資料庫為基礎，從各地區鑑定委員會與台灣省覆議鑑定委員會鑑定結果一致的案例當中，隨機選取 100 筆案件資料作為本研究實例分析之資料，並利用此資料實際觀察 ACA 演算法在分群結果上是否有較為優良的分群表現。

6.1 敘述統計分析

本研究實例分析欲探討此肇事鑑定資料庫中，各鑑定變數對於鑑定責任的影響程度為依歸作為分群的基礎，此肇事鑑定資料庫說明及次數分配彙整如表 6.1 所示：

表 6.1 影響變數之代號與說明

變數名稱	變數內容及代號	次數	百分比
鑑定責任	1. 全部原因	24	24
	2. 主因	20	20
	3. 同為原因	12	12
	4. 次因	20	20
	5. 無因	24	24
區別	1. 北部地區（基隆、宜蘭、台北）	14	14
	2. 桃竹苗地區（桃園、新竹、苗栗）	18	18
	3. 中部地區（台中、彰化、南投）	24	24
	4. 南部地區（嘉義、台南、高雄、屏東、澎湖）	40	40
	5. 花東地區（花蓮、台東）	4	4
性別	1. 男	88	88
	2. 女	12	12
年齡	1. 20 歲以下	2	2
	2. 21 歲～35 歲	51	51
	3. 36 歲～50 歲	37	37
	4. 51 歲～65 歲	9	9
	5. 65 歲以上	1	1

表 6.1 (續) 影響變數之代號與說明

變數名稱	變數內容及代號	次數	百分比
車種	1. 小客車 (小客、營小客)	71	71
	2. 小貨車 (小客貨、小貨車)	19	19
	3. 大車 (大貨車、大客車)	10	10
超速	1. 嚴重超速 (超過速限 20km/hr)	9	9
	2. 超速 (超過速限 0~20km/hr)	13	13
	3. 無	78	78
飲酒	1. 有 (大於 0.55mg/l)	4	4
	2. 有 (0.25 mg/l~0.55 mg/l)	3	3
	3. 合格 (經酒測, 小於 0.25)	1	1
	4. 無	92	92
教育	1. 不識字	5	5
	2. 國中小	18	18
	3. 高中職	33	33
	4. 專科以上	12	12
	5. 不明	32	32
駕照	1. 有	92	92
	2. 無 (已達考照年齡)	2	2
	3. 無 (未達考照年齡)	0	0
	4. 不明	6	6
天色	1. 日間天然光線	56	56
	2. 夜間有照明	26	26
	3. 夜間無照明	12	12
	4. 不明	6	6
天候	1. 有雨	8	8
	2. 無雨	92	92
道路類別	1. 不明 (筆錄無紀錄)	14	14
	2. 省道	20	20
	3. 縣道	24	24
	4. 鄉道	8	8
	5. 市區道路	30	30
	6. 村里道路	4	4

表 6.1 (續) 影響變數之代號與說明

變數名稱	變數內容及代號	次數	百分比
道路型態	1. 路段 (直路)	40	40
	2. 行車管制號誌交叉路口	14	14
	3. 閃光號誌路口	19	19
	4. 無號誌路口	27	27
速限	1. 30km/hr	2	2
	2. 40km/hr	58	58
	3. 50 km/hr	4	4
	4. 60 km/hr	22	22
	5. 70 km/hr	6	6
	6. 不明	8	8
行向	1. 對向	36	36
	2. 同向	26	26
	3. 橫向 (左方車)	18	18
	4. 橫向 (右方車)	20	20
車損部位	1. 前、右側	46	46
	2. 右側、右後	16	16
	3. 後、左後	7	7
	4. 左側、左前	28	28
	5. 無或不明	3	3
筆錄速率	1. 30km/hr 以下	35	35
	2. 31km/hr~40km/hr	22	22
	3. 41km/hr~50km/hr	8	8
	4. 51km/hr~60km/hr	4	4
	5. 61 km/hr~70km/hr	3	3
	6. 70 km/hr 以上	8	8
	7. 不明	20	20
是否預見	1. 是 (有)	51	51
	2. 否 (無)	27	27
	3. 不明	22	22

表 6.1 (續) 影響變數之代號與說明

變數名稱	變數內容及代號	次數	百分比
採取措施	1. 無	46	46
	2. 閃(向右閃、向左閃)	8	8
	3. 減速(減速、停車)	4	4
	4. 其他(變換車道、倒車、繞、按喇叭、閃燈、超車)	6	6
	5. 不明	36	36
是否煞車	1. 煞後撞	23	23
	2. 撞後煞	0	0
	3. 不明	41	41
	4. 否	36	36
駕傷	1. 有	24	24
	2. 不明(無受傷之筆錄)	0	0
	3. 無	76	76
乘傷	1. 有	14	14
	2. 無乘客	84	84
	3. 不明(無受傷之筆錄)	0	0
	4. 無	2	2
駕亡	1. 有	6	6
	2. 不明(無筆錄資料)	0	0
	3. 無	94	94
乘亡	1. 有	3	3
	2. 無乘客	86	86
	3. 不明(無筆錄資料)	5	5
	4. 無	6	6
路權	1. 有道路優先權	50	50
	2. 無道路優先權	50	50

由於此肇事資料變數眾多，而本研究實例驗證目的在於針對與鑑定肇事責任有相關的變數進行分群，故檢定各變數與鑑定責任間的卡方值，於信賴水準 95%下，挑選出適當的變數在進行分群，此隨機挑選的 100 筆資料之各變數的卡方檢定結果如表 6.2 所示：

表 6.2 肇事責任與影響變數間之卡方檢定表

變數項目	卡方值	自由度	P 值	是否顯著
區別	27.441	16	0.037	*
性別	6.115	4	0.191	—
年齡	20.387	16	0.203	—
車種	13.214	8	0.105	—
超速	7.879	8	0.445	—
飲酒	23.831	12	0.021	*
教育	9.843	16	0.875	—
駕照	6.614	8	0.579	—
天色	22.101	12	0.036	*
天候	2.306	4	0.680	—
道路類別	23.583	20	0.261	—
道路型態	63.852	12	0.000	*
速限	20.792	20	0.409	—
行向	78.932	12	0.000	*
車損部位	21.495	16	0.160	—
筆錄速率	29.815	24	0.191	—
是否預見	6.080	8	0.638	—
採取措施	23.078	16	0.112	—
是否煞車	3.163	8	0.924	—
駕傷	9.162	4	0.057	—
乘傷	6.550	8	0.586	—
駕亡	4.170	4	0.383	—
乘亡	20.793	12	0.054	—
路權	106.113	4	0.000	*

經過檢定之後可知，24 個變數當中區別、飲酒、天色、道路型態、行向、路權六個變數，因其 P 值小於 0.05 為拒絕虛無假設所以判定為顯著變數，即此六個變數對於鑑定責任有顯著的差異，故在本研究中實例分析採用隨機挑選的 100 筆肇事鑑定資料，每筆資料包含六項特徵值，以此來進行群落分析。

6.2 分群結果之分析

經過前一小節的篩選資料變數之後，本節將利用此 100 筆資料，透過 ACA、k-means、凝聚法、GCA 四種演算法，來作一分群結果的比較，並對於各分群所代表的分群特徵作一敘述分析。實例分析當中，根據第五章的變數設定，在分群數設定為 3 群與 5 群藉此判斷 ACA 在實例當中是否能夠於不同的分群規模下，仍有比其他分群演算法較佳的分群結果，各分群演算法之分群結果如表 6.3 所示：

表 6.3 各分群演算法之求解結果

分群演算法		ACA	k-means	凝聚法	GCA
TWCV 值	3 群	295.8642	316.6462	328.9237	314.1897
	5 群	217.5459	221.9415	228.9284	247.8908

由表 6.3 可知，ACA 在 3 群及 5 群時，分群結果的績效值 TWCV 皆較其他分群演算法小，表示在此情況下，ACA 仍有較佳的分群結果；在改善率方面 ACA 於分群規模為 3 群時，對於其他分群演算法的改善率為 5.83%~11.17%，於分群規模為 5 群時，對於其他分群演算法的改善率為 1.98%~12.24%；接下來將對於 ACA 所求解出的各分群情況作一說明：

表 6.4 區別變數在每一分群之情況（3 群）

區別	分群		
	1	2	3
北部地區	0 0%	0 0%	14 53.85%
桃竹苗地區	0 0%	6 14.29%	12 46.15%
中部地區	12 37.5%	12 28.57%	0 0%
南部地區	18 56.25%	22 52.38%	0 0%
花東地區	2 6.25%	2 4.76%	0 0%

由表 6.4 可知在分群規模為 3 群時，分群 1 中南部地區資料所佔百分比最多（56.25%），分群 2 中南部地區資料所佔百分比最多（52.38%），分群 3 中以北部地區資料為最多（53.85%），總結上述，分群 1 與分群 2 大多為南部地區資料，分群 3 則多為北部地區資料。

表 6.5 區別變數在每一分群之情況 (5 群)

區別	分群				
	1	2	3	4	5
北部地區	0 0%	0 0%	0 0%	0 0%	14 63.64%
桃竹苗地區	0 0%	4 11.76%	4 28.57%	2 33.33%	8 36.36%
中部地區	9 37.5%	8 23.53%	4 28.57%	3 50%	0 0%
南部地區	13 54.17%	20 58.82%	6 42.86%	1 16.67%	0 0%
花東地區	2 8.33%	2 5.88%	0 0%	0 0%	0 0%

由表 6.5 可知在分群規模 5 群時，分群 1 以南部資料居多 (54.17%)，分群 2 以南部資料居多 (58.82%)，分群 3 以南部資料佔大多數 (42.86%)，分群 4 以中部地區資料居多 (50%)，分群 5 以北部資料居多 (63.64%)，綜合上述，分群 1、2、3 多為南部地區資料，分群 4 多為中部地區資料，分群 5 多為北部地區資料。

表 6.6 飲酒變數在每一分群之情況 (3 群)

飲酒	分群		
	1	2	3
有 (大於 0.55mg/l)	2 6.25%	0 0%	2 7.69%
有 (0.25 mg/l~0.55 mg/l)	2 6.25%	1 2.38%	0 0%
合格 (經酒測，小於 0.25)	1 3.125%	0 0%	0 0%
無	27 84.375%	41 97.62%	24 92.31%

由表 6.6 可知在分群規模為 3 群時，三個分群中皆無飲酒佔多數，分別為 84.375%、97.62%、92.31%，代表在此 100 筆資料中，飲酒肇事比率偏低。

表 6.7 飲酒變數在每一分群之情況 (5 群)

飲酒	分群				
	1	2	3	4	5
有(大於 0.55mg/l)	0 0%	0 0%	0 0%	4 66.67%	0 0%
有(0.25 mg/l~ 0.55 mg/l)	0 0%	0 0%	1 7.14%	2 33.33%	0 0%
合格(經酒測,小 於 0.25)	1 4.17%	0 0%	0 0%	0 0%	0 0%
無	23 95.83%	34 100%	13 92.86%	0 0%	22 100%

由表 6.7 可知在分群規模 5 群時, 分群 1、2、3、5 多為無飲酒狀況下肇事, 分群 4 多為酒後駕駛肇事 (66.67%)。

表 6.8 天色變數在每一分群之情況 (3 群)

天色	分群		
	1	2	3
日間自然光線	16 50%	28 66.67%	12 46.16%
夜間有照明	6 18.75%	10 23.81%	10 38.46%
夜間無照明	6 18.75%	4 9.52%	2 7.69%
不明	4 12.5%	0 0%	2 7.69%

由表 6.8 可知在分群規模為 3 群時, 分群 1 中日間自然光線比例最多 (50%), 分群 2 中日間自然光線比例最多 (66.67%), 分群 3 中日間自然光線比例最多 (46.16%), 綜合上述, 三個分群中日間自然光線所佔比例都是最多的。

表 6.9 天色變數在每一分群之情況 (5 群)

天色	分群				
	1	2	3	4	5
日間自然光線	12 50%	24 70.59%	8 57.14%	0 0%	12 54.55%
夜間有照明	6 25%	6 17.65%	4 28.57%	2 33.33%	8 36.36%
夜間無照明	3 12.5%	4 11.76%	2 14.29%	3 50%	0 0%
不明	3 12.5%	0 0%	0 0%	1 16.67%	2 9.10%

由表 6.9 可知在分群規模為 5 群時，分群 1 多為日間自然光線 (50%)，分群 2 以日間自然光線佔多數 (70.59%)，分群 3 以日間自然光線佔多數 (57.14%)，分群 4 則為夜間無照明為多數 (50%)，分群 5 以日間自然光線佔多數 (54.55%)。

表 6.10 道路型態變數在每一分群之情況 (3 群)

道路型態	分群		
	1	2	3
路段 (直路)	26 81.25%	0 0%	14 53.85%
行車號誌管制交叉路口	4 12.5%	2 4.76%	8 30.77%
閃光號誌路口	2 6.25%	15 35.71%	2 7.69%
無號誌路口	0 0%	25 59.52%	2 7.69%

由表 6.10 可知在分群規模為 3 群時，分群 1 中以路段 (直路) 所佔比例最多 (81.25%)，分群 2 則以無號誌路口佔多數 (59.52%)，分群 3 以路段 (直路) 所佔比例最多 (53.85%)，綜合上述，分群 1 與分群 3 大多為路段資料較多，而分群 2 則為無號誌路口資料較多。

表 6.11 道路型態變數在每一分群之情況 (5 群)

道路型態	分群				
	1	2	3	4	5
路段 (直路)	22 91.67%	0 0%	0 0%	5 83.33%	13 59.09%
行車號誌管制交叉路口	2 8.33%	2 5.88%	2 14.29%	1 16.67%	7 31.82%
閃光號誌路口	0 0%	11 32.35%	6 42.86%	0 0%	2 9.09%
無號誌路口	0 0%	21 70.59%	6 42.86%	0 0%	0 0%

由表 6.11 可知在分群規模為 5 群時，分群 1 (91.67%)、4 (83.33%)、5 (59.09%) 多為於直路肇事，分群 2 以無號誌路口肇事佔多數 (70.59%)，分群 3 則在閃光號誌路口與無號誌路口居多 (各為 42.86%)。

表 6.12 行向變數在每一分群之情況 (3 群)

行向	分群		
	1	2	3
對向行車	18 56.25%	2 4.76%	16 61.54%
同向行車	12 37.5%	6 14.29%	8 30.77%
橫向左方車	1 3.125%	16 38.10%	1 3.85%
橫向右方車	1 3.125%	18 42.86%	1 3.85%

由表 6.12 可知在分群規模為 3 群時，分群 1 中以對向行車所佔比例最多 (56.25%)，分群 2 則為橫向右方車所佔比例最多 (42.86%)，分群 3 為對向型車所佔比例最多 (61.54%)，綜合上述，分群 1 與分群 3 中以對向行車所佔比例較多，而分群 2 則以橫向右方車為主。

表 6.13 行向變數在每一分群之情況 (5 群)

行向	分群				
	1	2	3	4	5
對向行車	10 41.67%	0 0%	8 57.14%	6 100%	12 54.55%
同向行車	12 50%	0 0%	6 42.86%	0 0%	8 36.36%
橫向左方車	1 4.17%	16 47.06%	0 0%	0 0%	1 4.55%
橫向右方車	1 4.17%	18 52.94%	0 0%	0 0%	1 4.55%

由表 6.13 可知在分群規模為 5 群時，分群 1 多為同向行車 (50%)，分群 2 多為橫向右方車 (52.94%)，分群 3 多為對向行車 (57.14%)，分群 4 則皆為對向行車，分群 5 以對向行車佔多數 (54.55%)。

表 6.14 路權變數在每一分群之情況 (3 群)

路權	分群		
	1	2	3
有道路優先權	16 50%	21 50%	13 50%
無道路優先權	16 50%	21 50%	13 50%

由表 6.14 可知在分群規模為 3 群時，在三個分群當中有無道路優先權的資料所佔比例皆為各半。

表 6.15 路權變數在每一分群之情況 (5 群)

路權	分群				
	1	2	3	4	5
有道路優先權	13 54.17%	17 50%	7 50%	3 50%	10 45.45%
無道路優先權	11 45.83%	17 50%	7 50%	3 50%	12 54.55%

由表 6.15 可知在分群規模為 5 群時，分群 1 多為有道路優先權 (54.17%)，分群 5 則以無道路優先權居多 (54.55%)，分群 2、3、4 則有無道路優先

權的資料所佔比例皆為各半；在實例分析中，本研究依據肇事鑑定變數對於鑑定責任來作群落分析，接著將針對不同分群規模下每一分群中的鑑定結果中的變數逐一說明。

當分群規模為 3 群時：

分群 1：

此分群多為南部地區資料，肇事事故發生時，駕駛多在無飲酒狀況下肇事，而當時的天色多為日間自然光線，道路型態為直路型態，行車方向多為對向行車肇事，在有無道路優先權方面則各佔半數。

分群 2：

此分群多為南部地區資料，肇事事故發生時，駕駛多在無飲酒狀況下肇事，而當時的天色多為日間自然光線，發生地點的道路型態多為無號誌路口，行車方向多為橫向右方車，在有無道路優先權方面則各佔半數。

分群 3：

此分群多為北部地區資料，肇事事故發生時，駕駛多在無飲酒狀況下肇事，而當時的天色多為日間自然光線，發生地點的道路型態多為直路路段型態，行車方向多為對向行車肇事，而有無道路優先權方面，兩方案所佔比例是相同的。

表 6.16 各分群中變數分布之情形（3 群）

	地區	飲酒	天色	道路型態	行車方向	有無道路優先權
分群 1	南部	無飲酒	日間自然光色	直路	對向	各佔半數
分群 2	南部	無飲酒	日間自然光色	無號誌路口	橫向右方車	各佔半數
分群 3	北部	無飲酒	日間自然光色	直路	對向	各佔半數

當分群規模為 5 群時：

分群 1：

此分群多為南部地區資料，肇事事故發生時，駕駛多在無飲酒狀況下肇事，而當時的天色多為日間自然光線，道路型態為直路型態，行車方向多為同向行車肇事，在有無道路優先權方面則多為有道路優先權。

分群 2：

此分群多為南部地區資料，肇事事故發生時，駕駛多在無飲酒狀況下肇事，而當時的天色多為日間自然光線，道路型態為無號誌路口型態，行車方向多為橫向右方車肇事，在有無道路優先權方面則兩方案所佔比例是相同的。

分群 3：

此分群多為南部地區資料，肇事事故發生時，駕駛多在無飲酒狀況下肇事，而當時的天色多為日間自然光線，道路型態為閃光號誌路口與無號誌路口型態，行車方向多為對向行車肇事，在有無道路優先權方面則兩方案所佔比例是相同的。

分群 4：

此分群多為中部地區資料，肇事事故發生時，駕駛多酒後駕車肇事，而當時的天色多為夜間無照明，道路型態為直路型態，行車方向多為對向行車肇事，在有無道路優先權方面則兩方案所佔比例是相同的。

分群 5：

此分群多為北部地區資料，肇事事故發生時，駕駛多在無飲酒狀況下肇事，而當時的天色多為日間自然光線，道路型態為直路型態，行車方向多為橫向右方車肇事，在有無道路優先權方面則以無道路優先權佔多數。

表 6.17 各分群中變數分布之情形（5 群）

	地區	飲酒	天色	道路型態	行車方向	有無道路優先權
分群 1	南部	無飲酒	日間自然光色	直路	同向	有
分群 2	南部	無飲酒	日間自然光色	無號誌路口	橫向右方車	各佔半數
分群 3	南部	無飲酒	日間自然光色	閃光號誌路口 無號誌路口	對向	各佔半數
分群 4	中部	有飲酒	夜間無照明	直路	對向	各佔半數
分群 5	北部	無飲酒	日間自然光色	直路	對向	無

第七章 結論與建議

7.1 結論

1. 依據過去 Dorigo 等人所提出一系列螞蟻理論相關文獻，本研究對於其中的轉換規則與費洛蒙更新等觀念，構建出本研究的重心—螞蟻分群演算法（ACA）。
2. 本研究所提出的螞蟻分群演算法為依據螞蟻族群系統所構建出的群落分析方法，經過窮舉法找出資料規模為 10 筆資料時的最佳分群結果，再利用本研究中所提出的 ACA 求解，目的在於判斷 ACA 是否可以達到或是趨近於最佳解，其結果為與窮舉法所得的分群結果相同，由此可知 ACA 於此情況下能找到最佳解；小規模問題驗證其目的在於判斷不同的交換樣本方式對於求解分群結果是否有顯著的差異，隨著分群規模的增加，經過統計分析之後可以發現 ACA-2 較其他交換樣本方式有優良的求解穩定度與分群結果，因此在不同規模問題比較當中把 ACA-2 設定作為驗證的 ACA 分群演算法；在不同規模問題（20 筆、50 筆、100 筆）中，不同的分群規模（3 群、5 群、7 群、9 群）經過對於不同分群演算法的分群結果進行變異數與平均值檢定，其結果本研究所提出的 ACA 皆有優於 k-means、凝聚法與 GCA 的分群結果。
3. 針對不同分群演算法的求解結果變異探討，ACA 與其他三種比較的分群演算法在大部分的情況下，透過統計分析方法得知其檢定值並不顯著，即各演算法間的求解穩定度差異不大。
4. ACA 與其他分群演算法比較之後可知，在小規模問題時，於不同的分群規模下 ACA 能夠較其他分群演算法所求解出的分群結果改善 1.05%~13.19%，中規模問題於不同分群規模時，ACA 較其他分群演算法所求解出的分群結果改善 4.72%~20.67%，大規模問題於不同分群規模時，ACA 較其他分群演算法的求解結果改善 1.37%~53.42%；在實例驗證部份，ACA 與其他分群演算法的求解結果比較下，ACA 較其他分群演算法所求解出的結果改善 1.98%~12.241%。
5. 在實例驗證部分，本研究採用民國 89 年 3 月至民國 91 年 8 月之肇事鑑定資料庫的資料，從中隨機挑選 100 筆資料作為實例驗證的資料依據，在不同分群規模下（3 群與 5 群），比較 ACA、k-means、凝聚法與 GCA 四種演算法間的分群結果，在計算四種演算法的 TWCV 值，可以觀察到 ACA 分群演算法較其他三種分群演算法有較佳的分群結果，且 ACA 改善其他分群演算法求解結果達 1.98%~12.24%。

6. 在經過一連串的模式驗證與實例驗證之後，本研究所提出的 ACA 為一種可以採用的群落分析方法，在不同的分群規模（3 群、5 群、7 群、9 群）與不同的資料規模（20 筆、50 筆、100 筆）皆比傳統分群演算法（k-means 法與凝聚法）與 GCA 有較優良的分群結果。

7.2 建議

對於往後的相關研究，在此提出下列幾點建議：

1. 群落分析問題最困難之處在於如何設定每一個問題的分群數，本研究所提出的 ACA 演算法必須先設定分群數再進行分群，往後相關研究可以進行對於如何在不設定分群數的狀況下，即可進行分群而得到理想的分群結果方面來進行相關研究。
2. 本研究中所採用的 GCA 為透過套裝軟體所得到的結果，而其中的參數設定皆為預設值，並未對其中的相關參數如突變率進行敏感度分析，往後的相關文獻亦可對於 GCA 其中的參數設定進行分析，以了解是否參數設定的不同會造成求解結果的差異顯著。
3. 實例驗證部份本研究採用台灣區筆事鑑定資料庫資料，而交通運輸應用方面如旅次資料、客貨運排班等，其他運用方面如消費者消費行為分析、生產線排程等皆可運用 ACA 來做驗證。
4. 後續相關研究，可結合螞蟻理論與遺傳演算法利用螞蟻理論中費洛蒙更新觀念與遺傳演算法之選擇、交配、突變的特性，發展出混合（hybrid）分群演算法來求解群落分析問題。
5. 由於螞蟻理論為一種新近的理论，故其原始理論模型仍持續在改良當中，而本研究就是利用其中的重點觀念構建出 ACA 演算法，而其他運用方面也可依循此模式，利用螞蟻理論的特性透過現今電腦的強大運算能力，來求解複雜度較高組合最佳化問題。

參考文獻

1. 中文部分

- [1] 朱文正，民 92 年，「考量旅行時間可靠度之車輛途程問題—螞蟻族群演算法之應用」，國立交通大學交通運輸研究所碩士論文，台北市。
- [2] 周仕雄，民 92 年，「應用螞蟻系統於資料挖礦之集群分析」，國立台北科技大學生產系統工程與管理研究所碩士論文，台北市。
- [3] 彭國倫，民 90 年，「FORTRAN 95 程式設計」，基峰資訊股份有限公司。
- [4] 馮正民、邱裕鈞，民 93 年，「研究分析方法」，建都文化事業股份有限公司。
- [5] 黃俊英，民 88 年，「企業研究方法」，東華書局。
- [6] 楊翠微，民 90 年，「以重力理論為基礎的階層式資料分群演算法之研究」，國立台灣大學資訊工程學研究所碩士論文，台北市。
- [7] 葉承銓，民 91 年，「應用適應性基因演算法於資料分群的問題」，樹德科技大學資訊管理系碩士班碩士論文，高雄縣。
- [8] 詹金凌，民 92 年，「整合螞蟻理論與案例式推理於知識管理之應用」，國立台北科技大學生產系統工程與管理研究所碩士論文，台北市。
- [9] 蔡崇煒，民 91 年，「多重搜尋基因演算法：一個新的有效解決通訊網路及資料庫中複雜問題之方法」，屏東科技大學資訊管理系碩士班碩士論文，屏東縣。
- [10] 蕭宗勝，民 91 年，「螞蟻族群演算法應用在組合問題之研究」，私立銘傳大學資訊管理研究所碩士論文，台北市。
- [11] 藍坤銘，民 91 年，「應用資料挖掘之順序樣式法於螞蟻理論之改善」，國立台北科技大學生產系統工程與管理研究所碩士論文，台北市。
- [12] 蘇木春、張孝德，民 86 年，「機器學習：類神經網路、模糊系統以及基因演算法則」，全華科技圖書股份有限公司。

2. 英文部分

- [1] Bonabeau, E., Dorigo, M. and Theraulaz, G., 1999, "Swarm Intelligence," *Oxford University Press*.
- [2] Chiou, Y. C. and Lan, L. W., 2001, "Genetic clustering algorithms," *European Journal of Operational Research*, Vol. 135, Issue: 2, pp. 413-427.
- [3] Corne, D., Dorigo, M. and Glover, F., 1999, "New ideas in optimization," *McGraw-Hill*.
- [4] Cowgill, M. C., Harvey, R. J. and Watson, L. T., 1999, "A genetic algorithm approach to cluster analysis," *Computers & Mathematics with Applications*, Vol. 37, Issue: 7, pp. 99-108.
- [5] Dorigo, M., Maniezzo, V. and Colorni, A., 1996, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, Vol. 26, No. 1, pp. 29-41.
- [6] Dorigo, M. and Gambardella, L. M., 1997, "Ant colonies for the traveling salesman problem," *Biosystems*, Vol. 43, Issue: 2, pp. 73-81.
- [7] Dorigo, M. and Gambardella, L. M., 1997, "Ant colony system: a cooperative learning approach to the traveling salesman problem," *IEEE Transactions on Evolutionary Computation*, Vol. 1, No. 1, pp. 53-66.
- [8] Dorigo, M. and Di C., G., 1999, "Ant colony optimization: a new meta-heuristic," *Evolutionary Computation*, CEC 99. Proceedings of the 1999 Congress , Vol. 2 , pp. 1470-1477.
- [9] Dorigo, M., Bonabeau, E. and Theraulaz, G., 2000, "Ant algorithms and stigmergy," *Future Generation Computer Systems*, Vol. 16, Issue: 8, pp. 851-871.
- [10] Gambardella, L. M. and Dorigo, M., 1996, "Solving symmetric and asymmetric TSPs by ant colonies," *Evolutionary Computation*, Proceedings of IEEE International Conference, pp. 20-22.
- [11] Gámez, J. A. and Puerta, J. M., 2002, "Searching for the best elimination sequence in Bayesian networks by using ant colony optimization," *Pattern Recognition Letters*, Vol. 23, Issue: 1-3, pp. 261-277.
- [12] Jiang, J. H., Wang, J. H., Chu, X. and Yu, R. Q., 1997, "Clustering data using a modified integer genetic algorithm (IGA)," *Analytica Chimica Acta*, Vol. 354, Issue: 1-3, pp. 263-274.

- [13] Krishna, K. and Narasimha Murty, M., 1999, "Genetic K-means algorithm," *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, Vol. 29, Issue: 3, pp. 433-439.
- [14] Maniezzo, V., Cobrni, A. and Dorigo, M., 1994, "The ant system applied to the quadratic assignment problem," *Tech. Rep. IRIDIA/94-28*, Université Libre de Bruxelles.
- [15] Maulik, U. and Bandyopadhyay, S., 2000, "Genetic algorithm-based clustering technique," *Pattern Recognition*, Vol. 33, Issue: 9, pp. 1455-1465.
- [16] McMullen, P. R., 2001, "An ant colony optimization approach to addressing a JIT sequencing problem with multiple objectives," *Artificial Intelligence in Engineering*, Vol. 15, pp. 309-317.
- [17] Murthy, C. A. and Chowdhury, N., 1996, "In search of optimal clusters using genetic algorithms," *Pattern Recognition Letters*, Vol. 17, Issue: 8, pp. 825-832.
- [18] Randall, M. and Lewis, A., 2002, "A parallel implementation of ant colony optimization," *Journal of Parallel and Distributed Computing*, Vol. 62, Issue: 9, pp. 1421-1432.
- [19] Sarkar, M., Yegnanarayana, B. and Khemani, D., 1997, "A clustering algorithm using an evolutionary programming-based approach," *Pattern Recognition Letters*, Vol. 18, Issue: 10, pp. 975-986.
- [20] Yuan, Y. and Zhuang, H., 1996, "A genetic algorithm for generating fuzzy classification rules," *Fuzzy Sets and Systems*, Vol. 84, Issue: 1, pp. 1-19.