

第五章 結論與建議

自組織映射圖神經網路為採用競爭式學習最具代表性之非監督式類神經網路模式，不同於大部分類神經網路的應用是在已知叢集的前提下進行分派任務，該模式具備傳統分群法所欠缺的適應學習能力，並可按照資料內部產生的分派規則將觀察值分類，使得其分群效能更勝於傳統分群方法。然而隨著大型資料集合的普遍化、計算效能的大幅提升與商業快速回應的殷切需求，使得傳統分群技術必須改善已因應大量、特殊限制與即時調適等環境需求。因此，以人工智慧所發展出的軟性計算技術便相當適合用來改善原有的資料分析方法。以下分別將本研所得之結論與貢獻加以陳述，同時並展望 SOM 之未來發展趨勢，最後對於後續研究提出建議方向。

5.1 結論

SOM 的演算法與網路架構簡單且易於延伸應用，並同時具備資料降維的視覺化與拓樸保存特性，因此自 1982 年 Kohonen 提出迄今，已成功地應用在各種不同領域中，同時處理各學門的特定問題理論上持續受到探討與應用，並發表於類神經網路的相關國際期刊。另一方面，隨著時代變遷所造成上述主客觀環境不同，使得群集分析需要結合以人工智慧技術以發展更廣泛並具效能之演算法。常見的作法除了結合傳統分群法形成兩階段分群法之外，與模糊理論及基因演算法等軟性計算科學結合應用遂逐漸興起，成為群集分析、資料探勘與知識發掘等當前的主要課題。本研究有鑒於 SOM 整合性研究回顧之缺乏，因此首先設定「SOM 模式改善調查」與「SOM 分群綜合應用」兩大研究目標，接著參考 Cooper 的「五階段研究模型」，並結合「主題層面配對檢索策略」與本研究的「檢索策略矩陣」進行文獻檢索與資料蒐集，並將所得結果進一步篩選整理與分析歸納，最後呈現本研究結果如下

所述：

- A. 本研究檢索篩選的量化結果為兩大部分。在「SOM 模式改善之調查」方面，包括改善模式 31 篇相關研究文獻。而在「SOM 分群綜合應用」方面，包含經由資訊檢索得到 60 篇與文獻滾雪球法 72 篇，合計共 132 篇分群相關文獻，前者主要是以 SOM 為主，與其他方法結合或比較的分群應用；而後者主要是以傳統分群、模糊理論、遺傳演算法為主的分群研究。
- B. 在「SOM 模式改善調查」方面，本研究介紹 SOM 基本原理、演算法特性與優缺點，並搭配其發展沿革到近期應用，同時亦完成 SOM 相關期刊、國際研討會與著作等綜合性整理。此外本研究將篩選後的文獻進一步以四個主要特徵構面加以分析歸納，所得到的質化成果主要是建立一個 SOM 改善模式的綜合性參考目錄 (bibliography)，並得到以下結論：SOM 網路的改善架構應具動態適應性或發展最佳化機制，以突破傳統固定架構之缺點。搭配演化式的修剪機制，或建立群集相似性之評估準則，可使網路模式能根據現實世界資料的動態變化而即時修正，並且適用於今日大量高維且多元化的資料型態。而網路訓練參數的初始化、敏感性與最佳化，以非歐氏距離來衡量鄰近區域的相似性、非均勻密度的輸入資料型態與學習速率的退火機制等，皆為本階段研究回顧之結果。
- C. 在「SOM 分群綜合應用」方面，本研究介紹分群的基本定義、分群法分類與執行階段，以及評估分群效度等理論，並回顧以傳統分群、模糊理論與基因演算法等為三大主軸，同時探討與 SOM 結合應用之分群研究，並且進行分群法的評估比較。所得到的質化結果，同樣是建立一個 SOM 與上述方法結合的綜合性分群應用參考目錄。並得到以下結論：
 - (a) 就傳統分群法而言，階層式分群法的樹狀結構適合呈現資

料內部的結構性而有 GHSOM 之改善模式。而 k-means 快速簡易與即時修正特性，亦成功地與 SOM 結合為兩階段分群法，或者如 Fast SOM 輔以啟發式法重新定義 SOM 起始拓樸圖來加速傳統 SOM 學習速率。

- (b) 再就模糊理論觀之，FSOM 的結合分群不僅省去 Fuzzy C-Means 可能產生的龐大運算量，卻也同時保留對於輸入資料的模糊處理能力與最佳化機制，再加上原有 SOM 所具備的資料縮減、非線性的降維映射與鄰近區域內資料關係的保存特性，使其不僅在處理群組技術中工件辨識與分群問題能有效抗雜訊，並且能以多角度進行辨識，以提供多元化且高效能的視覺化資料分析策略。
- (c) 最後以基因演算法而論，先使 Kohonen 學習達到局部最佳化，再將 Kohonen 的訓練參數集合視為基因搜尋空間，以尋找最佳參數組合。如此在較複雜的資料集合中，成本函數最小化的機制將有效加速學習收斂。另外亦可利用基因演算法在二維網格中進行最佳化搜尋，使網路可依據適應度在演化個體之間發展組織化結構，以取代過去的參數化表現，使其具備行為模式基礎，並適用於時間序列資料的建模與即時修正，類似簡易的大腦記憶機制，可進一步應用於動態環境如錯誤診斷與評估之應用領域。

5.2 研究貢獻

- A. 完整提供 SOM 相關研究資訊，從基本原理、演算法特性與優缺點，到發展沿革及近期應用方向，同時還包括相關期刊、國際研討會與重點著作等綜合性整理介紹。
- B. 以「SOM」與「Learning Algorithm」為配對主題層面之檢索策略，

- 搭配「檢索矩陣」，針對五大著名資料庫進行文獻檢索與篩選，建立一個 SOM 改善模式的綜合性參考目錄，合計共 31 篇。
- C. 使用「向量量化」、「資料/集群視覺化」、「網路拓樸映射與保存特性」、「網路拓樸之動態網格模式」等四大構面綜合分析與探討上述 SOM 改善研究。
 - D. 基於上述 1~3 項研究成果，可提供後續研究者在 SOM 演算法改善之重要參考。
 - E. 介紹回顧分群定義、基本原理、分群法分類、分群階段與分群效度之評估方法等文獻探討。
 - F. 將「SOM」與「Clustering」配對為主題層面之檢索策略，搭配「檢索矩陣」，針對五大著名資料庫進行文獻檢索與篩選，建立一個與傳統分群、模糊理論與基因演算法等分群法結合的綜合性分群應用參考目錄。其中包含經由資訊檢索得到 60 篇與文獻滾雪球法 72 篇，合計共 132 篇分群相關文獻。而前者主要是以 SOM 為主，與其他方法結合或比較的分群應用；而後者主要是以傳統分群、模糊理論、遺傳演算法為主的分群研究。
 - G. 基於上述 4~6 項研究成果，可提供後續研究者在 SOM 與其他演算法結合應用之重要參考。

5.3 SOM 趨勢展望

1996 年 7 月 5 日複製羊桃麗(Dolly)複製成功，分子生物技術達到一個成功的高潮，複製羊變成活的製藥廠的夢想有可能成真了。1998 年 1 月美國芝加哥的科學家席德，提出開設複製人診所的想法，使得生物倫理學(bioethics)成為最近最熱門的討論課題。

美國國家衛生研究院(NIH)於 1989 年成立了人類基因組研究中心，邀請華森(James D. Watson)教授(DNA 雙螺旋結構的發現者之一)為第一任中心主任，目前已於西元 2003 年解讀完人類遺傳密碼的三十

億個 DNA 序列之任務。這當然必須依靠電腦的高速計算能力才有可能對這龐大的生物資料庫資料進行整理與分析。因此結合了分子生物學、生物物理學、統計數學、計算機科學...等跨領域學門便整合出一個新興的研究領域 - 生物資訊學 (Bioinformatics) 與計算生物學 (Biocomputing , Computational Biology)。狹義生物技術所衍生之生物產業，是以分子生物學為基礎。而分子生物學所研究主要對象即為三大生物巨分子:去氧核糖核酸 (DNA)、核糖核酸(RNA)、蛋白質(protein)。遺傳密碼的 DNA 是以一度空間的線性順序排列，而執行生化反應功能來維繫生命的是蛋白質，而蛋白質功能的正常與否，則是取決於三度空間的摺疊(folding)是否正確。因此計算生物學的主要課題便是基因序列與蛋白質的分析與預測。廣義而言，生物資訊學(Bioinformatics)意指所有應用電腦處理、分析的生物學研究過程，包含利用電腦、網路進行資料的收集與整合，利用影像處理設備將研究過程所得的圖像比對，以及多數分子生物學研究者比較熟悉的巨分子序列分析比對、結構預測等。原因是科學家將處理複雜的生物化學現象，轉化為只要研究 G, A, T, C 串列(string)【Olson, 1995】之間的關係即可【Heijne, 1987】。目前藉助電腦做 DNA 串列比對，已找出一些疾病的原因，也成功地應用於品種改良。如此將可加速遺傳工程進步，並獲得豐碩的突破性研究成果。

換句話說，將龐大的生物資料庫透過電腦的高速計算能力與 WWW 傳輸連結，並結合以人工智慧為基礎的資料處理技術，將徹底改變傳統的研究模式，降低研究成本及時間，以達到加速基因資訊的解讀、蛋白質立體結構與功能預測、生物品種改良與藥物設計等應用。而 SOM 今後主要研究方向除了前面第 3.2 節所列相關發展方向中，以 WEBSOM 為處理大型文件集合之組織結構，還包括生物資訊學中基因序列表現視覺化的分群研究將是 SOM 下一階段重要發展方向。

5.4 後續建議

- A. 有鑒於本研究屬於調查回顧性質，旨在探索 SOM 相關模式改善與分群綜合應用之研究脈絡，後續可以本研究為基礎，針對較著名的資料集合，如 IRIS、動物語意或工件外形等特定資料集合進行分群實證研究，以評估相關改善模式或結合性分群法在特定問題上之分群效能。
- B. 在 SOM 模式改善方面，為因應當前資料分析環境趨勢，建議網路模式需朝向能夠處理大量高維度的資料集合，或朝向能夠反應動態性輸入資料之機率分佈。因此，演算法必須簡單有效且具有適應性的網路架構，如此方能有助於複雜動態之資料處理。
- C. 根據上述建議，其實在 SOM 分群整合應用也是如此。與本研究所介紹各具特色的分群法相結合，相當有助於傳統 SOM 網路模式的最適化發展。因此後續研究可參考動態階層式分群法的樹狀結構，k-means 快速簡易與即時修正達到最適化之特性，Fuzzy C-Means 模糊處理能力與最佳化機制，基因最佳化搜尋法的動態適應性來進行網路拓樸的演化。
- D. 後續研究可接續第三章 SOM 構面分析，進一步探討各構面改善模式之演算法特性及其適合應用之問題範圍，如影像切割、市場區隔、工件辨識或機器學習等。