國立臺灣大學理學院地理環境資源所

博士論文

Department of Geography
College of Science
National Taiwan University

Doctoral Dissertation

地理視覺化輔助的空間知識探索—台北市交通流量的個案分析

Applying Geovisualization Techniques in Enhancing Knowledge Discovery Framework for Geographical Databases- The Case Study of Traffic Flow in Taipei City

陳致元

Chih-Yuan Chen

指導教授：朱子豪 博士

Advisor: Tzu-How Chu, Ph.D.

中華民國99年12月
December, 2010

國立臺灣大學理學院地理環境資源所

博士論文

Department of Geography
College of Science
National Taiwan University

Doctoral Dissertation

地理視覺化輔助的空間知識探索—

台北市交通流量的個案分析

Applying Geovisualization Techniques in Enhancing Knowledge Discovery Framework for Geographical Databases- The Case Study of Traffic Flow in Taipei City

陳致元

Chih-Yuan Chen

指導教授：朱子豪 博士

Advisor: Tzu-How Chu, Ph.D.

中華民國 99 年 12 月

December, 2010

國立臺灣大學博士學位論文
口試委員審定書

地理視覺化輔助的空間知識探索—

台北市交通流量的個案分析

Applying Geovisualization Techniques in
Enhancing Knowledge Discovery Framework
for Geographical Databases-
The Case Study of Traffic Flow in Taipei City

本論文係 陳致元 君（學號：D90228001）在國立臺灣大學地理環境資源學系、所完成之博士學位論文，於民國 99 年 12 月 23 日承下列考試委員審查通過及口試及格，特此證明。

口試委員：

朱子豪　　博士　　（簽名）

（指導教授）

孫志鴻　　博士

蔡博文　　博士

李瑞陽　　博士

周學政　　博士

中文摘要

地理視覺化探索分析是一種利用地理空間的互動展示來探索多維度時空資料的方法，自組織映射圖網路的視覺化技術是目前最有效的方法，但仍存在部分問題，如資料的分類方法以及類神經網路的大小需依賴研究者主觀判斷。本研究提出「先分解再重新聚類」的方法，整合自組織映射圖及空間自組織映射圖兩種類神經網路演算法以及群聚演算法，利用逐步合併神經元的群聚資料，達到動態調整分類數量的方式，再配合其他視覺化及空間地圖的展示，來探索資料隱含的關係與空間特徵。本研究使用兩個案例進行視覺化探索，第一案例使用台北市早晚尖峰時間車流量資料進行分析，發現無法找出資料關連的原始資料，再經由地理視覺化探索分析後，可在尖峰時間的東西向汽車車流量資料當中找到兩個主要的集合。此兩集合不但在相關係數及判定係數上均有顯著的改善，同時在空間上也可區分為兩個不同的分區。第二個案例使用機車車流量進行分析，在早晨尖峰時間的資料當中，經過地理視覺化探索分析後，可以發現六個主要的空間群聚以及線型的空間特徵。顯示地理視覺化探索分析對於發現地理資料的空間特性，以及辨識不同區域資料的差異性具有不錯的效果。研究結果顯示本研究提出的方法容易觀測到空間或是資料的關連特徵，未來可應用本研究的方法分析其他的時空資料。

關鍵字：地理視覺化、自組織映射圖、空間自組織映射圖、交通流量、群聚分析

Abstract

Geovisualization is a method of exploring spatial knowledge hidden in multidimensional geographic and temporal data via interaction with map and graph. The visualization of Self-Organizing Map (SOM) is one of the most effective methods but still has problems of what size the network should be. This research proposed a novel method called "Divide and Regroup", to integrate clustering analysis and two SOM algorithms (SOM and Geo-SOM) interactively and dynamically for finding hidden data relations and spatial patterns. Two different rush-hour traffic flow data of Taipei City were selected and two cases were done to demonstrate the effectiveness of this novel method. In the first case, the correlation coefficient and coefficient of determination of the unclassified data were low. Two major groups of traffic flow data were recognized using the Geovisualization approach. The correlation coefficients and coefficients of determination of the classified data have improved significantly. In the second case, six major spatial clusters and liner feature groups were recognized. Furthermore, these groups showed different data patterns indicate that the Geovisualization approach is useful for identifying spatial and data characteristics hidden in geographic data. The results demonstrated the effectiveness of the novel method of Geovisualization.

Keywords: Geovisualization, SOM, traffic flow, clustering analysis

目錄

國立臺灣大學博士學位論文口試委員審定書 iii

中文摘要 iv

Abstract v

第一章 緒論 1

第一節 研究動機 1

第二節 研究目的 4

第二章 文獻回顧 6

第一節 空間資料庫知識探索與資料探勘所面臨的挑戰 6

一、資料庫知識探索與資料探勘基本架構 6

二、知識探索與資料探勘所面臨的問題 9

三、時空地理知識探索的新需求 11

第二節 地理視覺化方法的發展歷程及理論基礎 13

一、地理視覺化技術的興起 14

二、地理視覺化研究的理論基礎 16

三、地理視覺化工具的分類 23

第三節 地理視覺化與知識探索的整合研究 36

一、地理視覺化與知識探索的架構整合 36

二、利用地理視覺化探索的實用性原則以及工具整合 38

三、交通流量資料研究作為一整合案例的可行性評估 40

第四節 小結 42

一、傳統知識探索架構及資料探勘架構的問題 42

二、視覺化工具的研究具有其理論基礎以及成功的要素 42

三、整合性的地理視覺化知識探索架構需要被提出 43

四、利用空間角度來思考交通流量資料知識探索的可能性 43

第三章 研究方法 44

一、以交通流量為對象的案例探索 46

二、利用探索原則改善 SOM 以及相關演算法 49

三、視覺化工具的選用以及探索環境的整合 53

第四章 研究結果與討論 55

第一節 傳統 SOM 演算法與群聚演算法的結合 55

一、階層式群聚演算法的先期評估 55

二、原始車流量的相關分析與簡單回歸分析 58

三、 四組車流量的地理視覺化探索分析 58

四、集合一的相關分析與簡單迴歸分析 62

五、集合二的相關分析與簡單迴歸分析 64

六、不同年度的比較 66

七、小結 70

第二節 Geo-SOM 演算法與群聚演算法的結合 71

一、Geo-SOM 網路大小的選定 71

二、機車流量資料的 Geo-SOM 分類結果 71
三、階層式群聚演算法的資料再聚類 73
四、經過群聚演算及探索後的結果 77
五、小結 79

第五章 結論與建議 81
一、地理視覺化知識探索架構之操作型方法 81
二、本研究所達到的貢獻 83
三、未來可努力的研究方向 83

參考文獻 86

圖目錄

圖 1：知識探索過程 (Fayyad, 1997) 7
圖 2：資料探勘過程(Koperski et al., 1996) 8
圖 3：地理動態性與空間屬性架構 (Yuan, 2009) 12
圖 4：時空知識探索的內部及外部結構 (Yuan, 2009) 13
圖 5：視覺化分析的涵蓋範疇 (MacEachren and Kraak, 2001) 15
圖 6：推論鏈結概念架構 (Gahegan, 2009) 18
圖 7：視覺化技術在探索式科學架構中所扮演的角色 (Gahegan, 2009) 19
圖 8：雙變數面量圖(擷取自 GeoViz Toolkit 軟體) 26
圖 9：不同視覺化方法介面展示 (擷取自 GeoViz Toolkit 軟體) 27
圖 10：雙變數面量圖與散步圖展示矩陣 (擷取自 GeoViz Toolkit 軟體) 28
圖 11：SOM 六角形拓樸映射圖視覺化範例 (擷取自本研究開發軟體) 30
圖 12：SOMVIS 自組織映射圖網路範例 30
圖 13：階層式自組織映射圖網路範例(Bação et al., 2005) 31
圖 14：空間自組織映射圖網路範例 (Bação et al., 2005) 32
圖 15：星狀圖展示介面圖 (擷取自 GeoViz Toolkit 軟體) 33
圖 16：星狀圖面量地圖展示 (擷取自 GeoViz Toolkit 軟體) 33
圖 17：Space-filling 視覺展示(Gahegan, 2009) 34
圖 18：Map cube 交通資料展示(Sekhar et al., 2002) 34
圖 19：Codex 系統畫面(Pike and Gahegan, 2007) 35
圖 20：地理視覺化技術與視覺化資料探索方式 (Maceachren et al., 1999) 37
圖 21：視覺化與知識探索之概念結合架構 (Gahegan, 2009) 37
圖 22：本研究之地理視覺化整合的研究架構 45
圖 23：研究方法流程 45
圖 24：路口調查點分佈示意圖 47
圖 25：路口轉向調查示意圖 47
圖 26：路口轉向車流原始資料 48
圖 27：四種不同的群聚演算法的分類結果 57
圖 28：89 年原始資料的簡單線性迴歸分析 59
圖 29：89 年原始資料資料展示及視覺化處理後之結果 60
圖 30：89 年經過地理視覺化重新聚類後之 SOM 視覺化及空間分布 61
圖 31：89 年資料經地理視覺化分析後的集合一簡單線性迴歸分析 63
圖 32：89 年資料經地理視覺化分析後的集合二簡單線性迴歸分析 65
圖 33：90 年資料經地理視覺化分析後的集合一簡單線性迴歸分析 67
圖 34：90 年資料經地理視覺化分析後的集合二簡單線性迴歸分析 68
圖 35：90 年資料經地理視覺化分析後的 SOM 視覺化及空間分布 69
圖 36：經過 Geo-SOM 演算後的資料空間分布 72
圖 37：經過 Geo-SOM 演算後的原始資料及平均值平行座標圖 73
圖 38： n=15 時，四種群聚演算法合併後視覺化地圖展示 74
圖 39： n=10 時，四種群聚演算法合併後視覺化地圖展示 75
圖 40： n=5 時，四種群聚演算法合併後視覺化地圖展示 76
圖 41： n=6 時， Ward’s 演算法合併後之 SOM 視覺化展示 78

圖 42： n=6 時， Ward's 演算法合併後之資料空間分布.....78
圖 43：合併後之原始資料平行座標圖及分組平均值平行座標圖.....79

表目錄

表 1：地理視覺化工具的分類 (本研究整理).....25
表 2：89 年原始資料相關係數及判定係數.....59
表 3：89 年資料分群後資料相關係數及判定係數 (集合一).....62
表 4：89 年資料分群後資料相關係數及判定係數 (集合二).....64
表 5：90 年原始資料相關係數及判定係數.....66
表 6：90 年資料分群後資料相關係數及判定係數 (集合一).....66
表 7：90 年資料分群後資料相關係數及判定係數 (集合二).....67

第一章 緒論

第一節 研究動機

近年來由於電腦資訊科學的快速進步，地理資料的數量與其複雜度也不斷地增加，用來儲存與記錄的地理資料倉儲 (data warehouse) 等技術也持續的發展當中。對於資料分析能力的需求也隨著資料的膨脹而日益增加。此外，時間維度分析的需求，也引領了時空理論與發展研究工具的研究議題。地理資料的特徵同時具有空間性與時間性，在資料的分析上涉及了趨勢分析及變遷偵測等複雜的過程。因此，地理資訊科學目前研究的重點是如何將時空地理資料做整合性的分析探索，同時發展出有效的研究技術與工具。

空間知識來自於對空間資料的知識探索 (KDD, knowledge discovery in database) ，在資料庫持續成長的過程中逐漸成為一個重要的研究手段。諸多技術例如資料挖掘 (data mining)、人工智慧演算法等研究陸續的被提出，並且廣泛的應用到地理資訊科學的領域當中。在過去的研究當中，雖然已經有許多成功的案例運用在決策或是預測的分析之中，但仍有許多學者提出了許多有待解決的瓶頸。 許多學者提出在他們的研究當中，有許多隱藏以及有意義的資料型態仍未被發掘 (Jin and Reynolds, 2000; Yang and Srinivasan, 2006) 。 當前地理空間資訊科學研究的重點主要如大規模資料高度複雜運算、平行處理與時空資料擷取、資料視覺化、時空現象的分類、模擬，與預測、時空尺度，以及時空知識探索與資料挖掘等等。而若要改進這樣的問題，必須善用空間認知，發展更適合人類思考與推理的知識整合方法，並應用在未來的研究(林祥偉，2003)。

在這個時空背景當中，地理資料視覺化（Geovisualization）的技術逐漸成為一新興的研究分枝而受到地理學家的重視。地理視覺化是利用地理空間的互動展示方法探索資料的空間特性，從探索的過程或是結果裡產生新的假設、新的解決方案，或是建構新知識的方法 (Kraak, 2003)。

從九○年代起便有一群地理學家開始針對地理視覺化進行不斷的探索。Maceachren et al. (1999) 針對地理視覺化與知識探索的整合進行了一個初探研究，指出視覺化技術可有助於 KDD 探索的過程，提供研究者更多不同的角度來觀察資料。幫助研究者在複雜多維的資料庫當中，找出可能的空間關連以及特徵型態，產生對於研究主題的新認知或是產生新的基本假設，進而發現新知識。Andrienko and Andrienko (1999) 針對互動地圖進行了初步的研究，發現探索式資料分析需要高度的互動性及動態性的展現，而這種互動的特性近年來也被認知科學領域的研究所提出，這類研究發現視覺化資料展示其設計及互動技術會影響使用者對不同現象特徵的認知 (Masters and Edsall, 2000)。因此面對著不斷成長的地理資料，地理視覺化技術可以提供一個更佳有效率的方式，在複雜多維的資料庫當中找出可能的空間關連以及特徵型態。在過去的幾年當中，已經有許多研究提出地理視覺化整合知識探索的架構，以及不同的視覺化技術應用在地理知識的探索上 (Maceachren and Kraak, 1997; Andrienko and Andrienko, 1999; Wachowicz, 2000; Gahegan et al., 2001; Takatsuka and Gahegan, 2002; Edsall, 2003; Koua and Kraak, 2004; Dykes et al., 2005; Koua et al., 2006; Kraak, 2007; Guo, 2009)。但是目前地理視覺化輔助知識探索的相關研究，仍然存在著許多問題，例如 Brachman and Anand (1996) 提出知識探索本身不容易自動化，它需要一個像人類一樣的高級智能體在其中心才能夠運作。因此要進行地理視覺化的研究，必須要對於人類的空間認知進行進一步研究，才能找出其主觀視覺認定背後所隱含的視覺化知識，進而利用這樣的知識輔助空間知識的探索與思考。Miller and Han (2001) 則提出視覺化是整合人類智慧和知識探索程序的一種強大的手段，它可以應用在知識探索過程裡的各個階段，協助研究者進行線上分析、查詢、技術選擇以及結果闡釋。因此，地理視覺化與知識探索的研究必須在技術發展，人類認知，以及工具使用三方面進行探討。

哲學家Michael Polanyi 在其 1967 年所著「個人知識 (Personal

Knowledge)」當中，首先提出了隱性知識 (tacit knowledge) 的概念。他針對人類對於知識的探索提到：

「我們所知道的比我們能說出來的還多。」

由於隱性知識較不容易被發現以及傳播，因此在知識探索的研究當中，如何去挖掘並且正式化這樣的知識型態，一直是學者努力的目標。這樣的哲學基礎，也啟發了視覺化研究的理論探討。目前人類視覺感知的能力，甚至遠遠超過目前電腦所能處理的範圍，因此也引發許多學者的思考這樣的事實：人類視覺對於事物的認知能力遠超過人類對於視覺感知能力本身的認知。

Arnheim 在1997年所寫的 Visual Thinking一書中，提到了視覺化的分析基礎，他說視覺化分析是：

「用科學的角度去了解一個事件或是事情，意味著去找到一某一種型態的支配力，能夠有效的解釋在研究當中系統內相關連的特徵。」

過去數十年間傳統的空間分析方法，不但資料不易取得，電腦的計算速度也較慢，因此當資料的數量和複雜度（如變動性）增加的時候，許多已發展的分析方法往往無法負擔。另外這些傳統方法要求研究者使用時必須先有較為嚴謹的假設。因此，傳統空間方析方法受限於現有已知知識的框架之下，比較不容易發現新的或未預期的特徵。從過去對於視覺化的理論或是哲學的探討裡，可以找到一些解決這些問題的線索。

在交通流量資料的相關研究當中，過去的研究從巨觀 (macro view) 的質性研究到目前微觀 (micro view) 的數值模式推估，相較過去不論在資料上或是研究深度上都有長足的進步，過去學者也發展出許多預測模式，來推估當時可能之交通狀況。此類模式在預測能力以及預測精度上已有不錯成果。但在車流量預測相關研究裡，以空間分布為切入的觀點較少被討論及應用，交通行為彼此之間在空間上互相影響且相當複雜，當區域不同的時候，其影響交通行為的主要因子也有所差異，因此要將空

間納入交通流量模式裡加以考量有其困難性。交通運輸是一種空間運動，受到空間特性的限制與引導，而交通資料本身也具有連續性、週期性、區域性等特徵，非常適合於時空資料庫知識探索之研究。

因此，本研究認為地理視覺化研究在知識探索領域上具有相當的潛力，若可以整合各別領域的研究成果，藉由一個視覺化知識累積的平台，整理出相關的視覺隱性知識，如此應該可以增加後端知識探索的正確性或是幫助發掘資料當中隱含的新知識，在資料不斷膨脹的現況當中加入更有效率的探索模式。但由過去的文獻當中可以發現，此一領域除了尚在發展的階段之外，還牽涉到人類認知與參與的部份，此外視覺化工具在使用操作上的知識在許多研究領域還是一個未探索的面相，導入這些知識以後所帶來的更多新知識發現也非常值得期待，因此本研究認為地理視覺化技術不論是在研究的新穎性以及重要性上皆有其價值存在，並希望站在前述的基礎上，對於視覺化以及知識探索的結合進行研究。

第二節 研究目的

本文將整理空間資料庫知識探索架構以及地理視覺化技術的相關理論與文獻，提出一個地理視覺化技術整合空間資料庫知識探索架構，設定視覺化工具在各個知識探索過程當中所扮演的角色，以及歸納視覺化工具使用的操作行知識及工具選擇的知識指導，再透過台北市交通流量資料進行雛形驗證，說明視覺化技術導入後對知識探索本身所產生的結果和影響，以提供後續相關研究參考。在方法上本研究先就過去文獻進行回顧，再根據分析出過去整合研究所存在的可改進之空間，提出一新的地理視覺化探索研究架構，再整合視覺化工具及群聚分析、導入實際交通資料進行雛形的模式架構驗證。如同過去知識探索領域的研究所述，如何發現隱性知識，對於知識探索有關鍵性的影響。因此本研究認為，發展可以幫助前知識產生的地理視覺化分析架構將可以有助於地理學個

領域的研究專家發展新的知識探索方法。由於探討人類視覺的分析運作模式涉及到人類對於空間視覺之處理能力、對於研究課題資料所引發之前置知識和假設，以及人類大腦引發創意的過程，屬於大腦生理學的範疇，超出地理研究者能力之範疇，因此本研究將不討論此部分，而專注於歸納人類可辨識之特徵關連及規則形態，以及視覺化工具組合與新前置知識類別之對應關係。在這樣的研究認知下，本研究著重在於如何利用地理視覺化的研究面向來發現可能可以導向新知識產生的新的空間型態 (spatial pattern) 或是資料型態(data pattern)。並把這樣的過程加以正式化，以期未來能夠將此架構更加的完備。

在這樣的研究基礎下，本研究所要達到的目的如下：

1.在知識探索架構中導入地理視覺化工具，利用其發現隱藏在資料中的新空間特徵及資料特徵。

2.發現不同的空間特徵及資料特徵，以及歸納各種視覺化工具的使用特性。

3.分析隱性知識發掘的過程，並嘗試把過程加以它正式化。

4.提出新的視覺化時空知識探索架構，並在交通流量的資料導出過去尚未歸納的新知識。

第二章 文獻回顧

本研究之文獻回顧主要可分為四大部分，首先先針對過去知識探索以及資料探勘技術的研究架構與方法進行分析，並提出其架構上所存在的不足以及目前對於新的時空研究分析需求上所遇到的瓶頸，其次針對地理視覺化整合的知識探索研究進行統合回顧分析，整理出地理視覺化在知識探索工作上所應該扮演的角色與整合的方法，然後再提出地理視覺化目前所面對到的問題以及可能的解決方法，在研究議題方面，本研究以交通資料為例，探討空間知識面相加入的新觀點以及地理視覺化所能帶來的改進，最後再針對文獻回顧進行一個總結。

第一節 空間資料庫知識探索與資料探勘所面臨的挑戰

一、資料庫知識探索與資料探勘基本架構

1929 年分類法專家布利斯 (Evelyn Bliss) 在其著作 < The Organization of Knowledge and the System of the Science> 當中最早提出了知識組織的概念， 布利斯主要從事分類學研究，同時強調知識有機結構的重要性，因此布利斯採用的主要方法是將知識的單元區分出來，再依照彼此之間的關係來建立連結，因此早期的知識探索的主要工作是在做知識分類的研究。Mark (1993) 也提到分類方法是人類科學的一個基礎工具，他利用空間資訊的本質將空間知識分為三個範疇，分別為宣告性知識、程序性知識和結構性知識等。 而空間知識可依存在形式的外顯式主題、形式（表現方式）、精鍊程度等來分類。若依知識的表現方式分類，則可以分為法則式、語意網、程序式，以及數值模式等。因此不論是過去的文字描述性知識，或是進入數位時代的數值模式型知識，過去的研究當中知識探索的主要目標，就是要找出這些分類或是結構當中所存在的有機連結。

電腦的發明以及數位資料的儲存技術對於人類的知識探索產生了革命性的影響。Fayyad (1997) 提出資訊時代空間知識的來源為對於空間資料的知識探索（knowledge discovery，KD）過程，KD 是一種從資料庫中粹取出有用的資訊並將知識正式化的有效技術，其過程包含： (1).資料倉儲（data warehousing）(2).目標資料選取（target data selection）(3).清理（cleaning）(4).前處理（preprocessing）(5).轉換及歸納（transformation and reduction）(6).資料探勘（data mining）(7).模式的選取或連結（model selection (or combination) (8).評估解釋（evaluation and interpretation）(9).知識擷取（finally consolidation and use of the extracted knowledge）等。

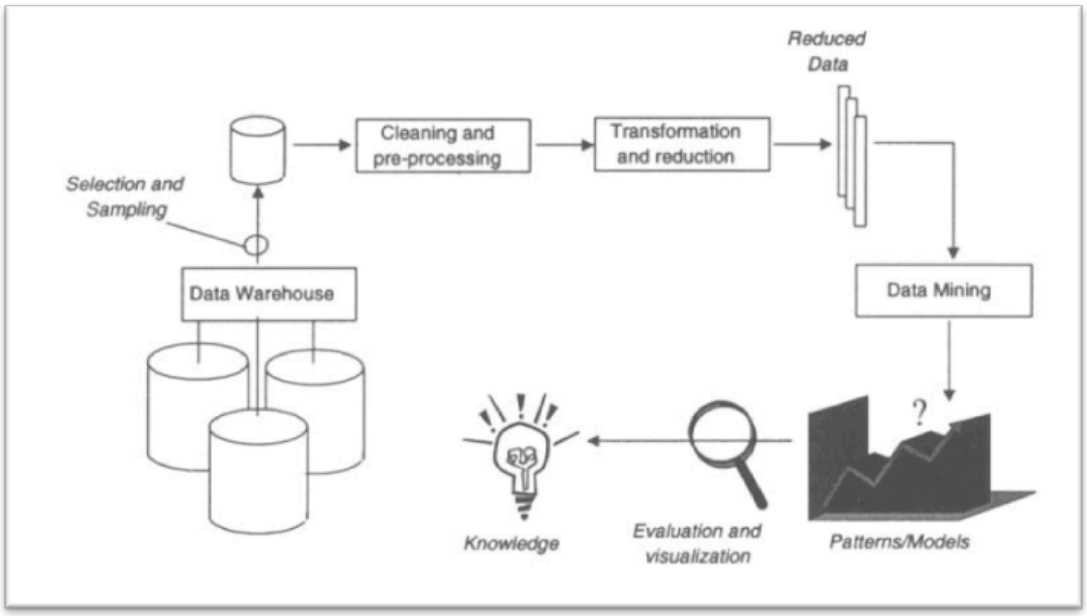


圖 1：知識探索過程 (Fayyad, 1997)

過去在資料倉儲以及資料探勘的研究較為蓬勃發展。在知識探索的過程當中，扮演基礎建設角色的就是資料倉儲。資料倉儲的興起是由於資料儲存和處理成本的下降、以及資訊本身價值的提升。在資料倉儲裡面資料通常只是一個唯讀的資料錄，主要功能是提昇對龐大的資料處理分析能力，而又不會去破壞到原始資料的完整性 (Jarke et al., 2003)。

資料探勘也是資料庫知識探索的過程中重要的一部分。資料探勘在最近幾年來已經成為資料庫系統中相當重要的議題，其方法是利用適當的自動化或半自動化方法從大量的資料中探索和分析，去找出有意義的特徵型態 (patterns) 或是條件規則 (rules) 的過程 (Koperski et al., 1996)。

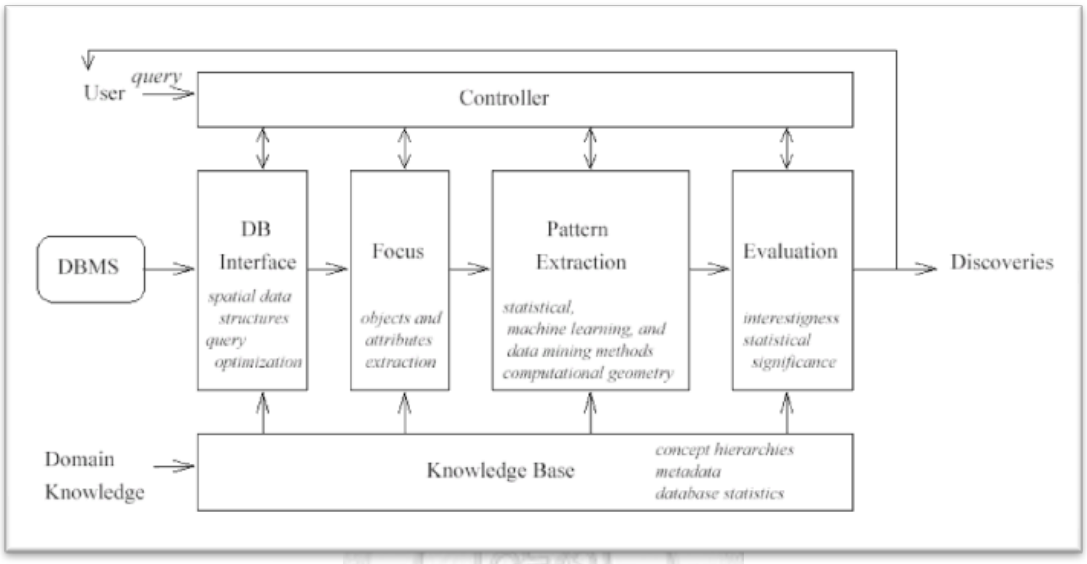


圖 2：資料探勘過程(Koperski et al., 1996)

在地理資料的探索上，過去關於空間資料探勘的分類研究頗多，有許多學者針對在資料庫當中進行空間探索的方法整裡 (Matheus et al., 1993; Koperski et al., 1996; Miller and Han, 2009)，其主要類別可分為：(1). 資料分群與分類 (clustering and classification)。(2). 資料與空間關聯 (spatial and non-spatial data association)。(3). 偏差值偵測 (deviation detection)。(4). 趨勢與迴歸分析 (trends and regression analysis)。(5). 歸納以及特徵化 (generalization and characterization)。 空間資料探勘可以用在瀏覽空間資料庫、理解空間資料、探索空間關係及空間資料與非空間資料之間的關係，或是重新組織空間資料庫，建構空間知識架構，或是空間查詢最佳化等。 由資料探勘而得來的空間知識可以包括許多種的形式，例如特徵性規則 (chatacteristic rules)、區別性規則 (discriminant rules)、關連式規則 (association rules)、誤差與演化規則 (deviation and evolution

rules)、或是描述顯著結構或是群聚的規則等。 關於知識探索與資料探勘的研究相當豐富，過去也得到許多相當成功的結果，但是目前仍然存在許多有待解決的問題。

二、知識探索與資料探勘所面臨的問題

Miller & Han (2009) 對於資料庫知識探索的基本概念做了一個總合的整理，他們認為在大量的資料庫裡，許多有用的資訊是隱藏在一些有趣的資料特徵當中的。而這些非隨機發生的現象或關係是具有根據且非特例的（valid）、具有創新的 (novel) ，具有價值的 (useful) ，且要容易被人所闡釋接受的 (ultimately understandable)。要發掘這些隱藏的知識或是資訊，則必須要發展一套更有效的方法及架構。Han and Kamber (2006) 指出面對未來龐大的資料庫，資料探索必須要能夠適應不同的資料尺度，另外必須要發展以使用者為中心 (user centric) 的互動式探索方法，才能適應不同的探索需求。另外若是要增加探索大量複雜資料庫的效率，視覺化是一個有效的手段。

Mannila (1997) 提出資料探勘必須要考慮如何善用背景知識，同時發展工具將探索到的新知識整理或是視覺化。 Yang et al. (2006) 訪問資料探勘研究的專家，指出資料探勘的技術發展過於特定化，缺少一個統一的理論或是方法來指導。另外與資料品質相關的前處理技術，目前缺少相關領域的知識來輔助。同時面對高維度以及時間序列資料的處理能力仍不足，必須要發展在圖形當中探勘新資訊的方法。最後則是資料探勘的程序不容易自動化，必須要有一個明確的方法論來指導使用者避免在探索的過程當中發生錯誤。 Guo (2009) 也提出近年來發展的運算模式可以處理大型的資料庫，但無法自動給予問題解釋或是意義。統計式比較嚴謹且可驗證的，但是統計式的需要一個先驗假設或是關連形式，且統計對於少量的特殊資料常僅以偏差值的方式對待處理。

Fayyad (1997) 所提出的KDD過程當中，在資料探勘過程之前，必須經過資料選取、資料前處理、以及資料轉換等步驟。Kriegel et al. (2007) 在其針對未來資料探勘的趨勢討論當中，曾提出真正被用在進行知識探索的時間，只佔了整體過程當中的10％。而未來隨著資料分析的複雜度增加，以及時空資料分析的需求，資料探勘所產生的特徵型態也會更加複雜。因此在資料前處理（在導入資料探勘技術之前的步驟）的所扮演的角色重要性將會有顯著的增加。Hall et al. (2009) 在其所發展的WEKA系統軟體當中，便包含了數十種的資料前處理功能函數，包含資料表操作、過濾、轉換、切割、分類等等功能。而過去的研究也顯示運用在資料前處理的技術會大大的影響資料挖掘的成果。因此資料探勘的目標將從過去的「從資料當中找到特徵」逐漸演變成為「找到特徵以及其自身演化的過程」。Famili et al. (2010) 在其研究當中提出了真實世界資料所存在的三種主要的問題，並且整理了有關資料前處理的三大類方法。 在其研究中提出了資料前處理的意義在於解決資料的問題以便進行資料探索、了解資料的特性以便套用合理的分析，以及從原始資料當中抽取更多有意義的知識。

由於目前資料探索當中存在著這些問題，因此許多學者在他們的研究當中指出仍然有許多隱藏以及有意義的資料型態仍未被發掘出來，而人類對於這些資料的解釋及應用的發展仍遠不及資料增加的速度 (Jin and Reynolds, 2000; Yang and Srinivasan, 2006; Li and Wang, 2007) 。 因此一個更有效以及更快速的資料挖掘技術是有必要在進一步探索的。 林祥偉 (2003) 提出應結合地理學的認知研究，發展更適合人類思考與推理的知識整合方法。Yang and Srinivasan (2006) 也在他們的研究結論當中提出了三項非常重要的研究面向：如何捕捉 (capturing) 其他更多型態之空間關係例如拓樸及方向性的關連；如何探索 (discovering) 稀少但是重要的特徵型態；如何分析 (analyzing) 以物件為基礎的軌跡。 Bedard et al. (2009)則強調知識探索的過程對資料探勘的結果有重要的影響，只有良好的探索

過程才能確保資料探勘能獲得有意義的結果。

三、時空地理知識探索的新需求

過去受限於資料模式，地理空間的知識只能從位置和幾何當中找到型態和關連。過去有相當多的文獻著重在探討關於時間與空間的本質性討論(Reichenbach et al., 1958; Schrodinger, 1985)。但近年來時空資料的不斷生產帶來了地理動態性 (geographic dynamics) 的課題，各種不同的資訊平台和技術如行動通訊裝置、3D/4D展示、網路等等都大大的改變了我們看待時間與空間的角度 (Maguire, 2007)。Yuan (2009) 整理提出地理動態性是一個重要概念，動態性是一個普遍用來描述各種影響力去驅使系統中的整體或是系統內部個體的結果。而地理動態性同時也是一個複雜的概念，要找到一個可以適合各種面相的時空結構是很困難的。各種行為 (activity)、事件 (event)、或是過程 (process) 都是構成地理動態性的單元 (圖3)。行為本身需要主體去執行，而事件需要地點去發生。事件可能是靜態或是動態的、間歇性或是連續性的。行為會留下足跡 (footprint)，一連串的足跡形成了一系列順序 (sequence)，而觀察這些順序我們可以了解變遷 (change) 或是運動 (movement) 發生的過程。因此所謂發掘地理動態性當中所存在的知識就是指去統合這些行為、事件、或是過程的資訊，並在這樣的過程當中找出關連或是規則。

在時空地理資訊探索裡面，變遷 (change) 是最重要的一個部分，Andrienko and Andrienko (2005)將時空變遷分為存在性改變、空間特性改變，以及屬性改變三種，Yuan (2009)則提出變遷包括無改變 (no change)、出現 (appearance)、消失 (disappearance)、擴張 (expansion)、收縮 (contraction)、扭曲 (deformation)、分割 (split)、合併 (merge)、移位 (displacement)、旋轉 (rotation)、收斂 (convergence) 等等。但是要分辨這樣的變遷並非簡單的工作。因此 Yuan 主張要發展這些自動的知識探索演

算法需要先探索足跡，了解造成這些足跡背後的地理動態性，並且合理的安排他們的順序，建立出一條地理動態性時空的軌跡。而以架構來看的話時空資料庫知識探索必須要分為特性、存在、和結構的角度來分析 (圖4)，這一部分來自於內部的分析，同時也有來自於外部的觀察。 前半部的工作著重於利用領域知識來分割定義問題和現象，後面的工作著重於連結，最重要的事是要如何抽取出有意義的足跡集合。

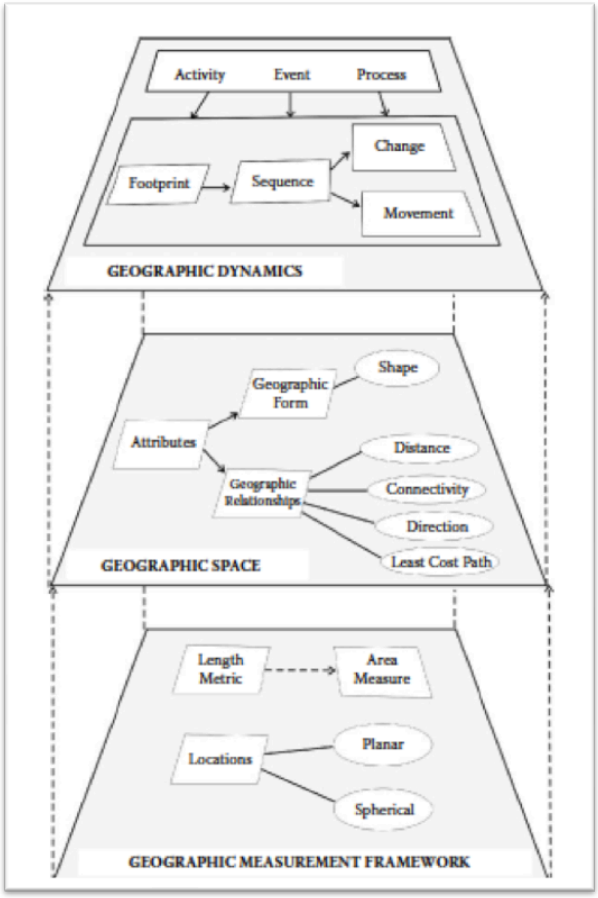


圖 3：地理動態性與空間屬性架構 (Yuan, 2009)

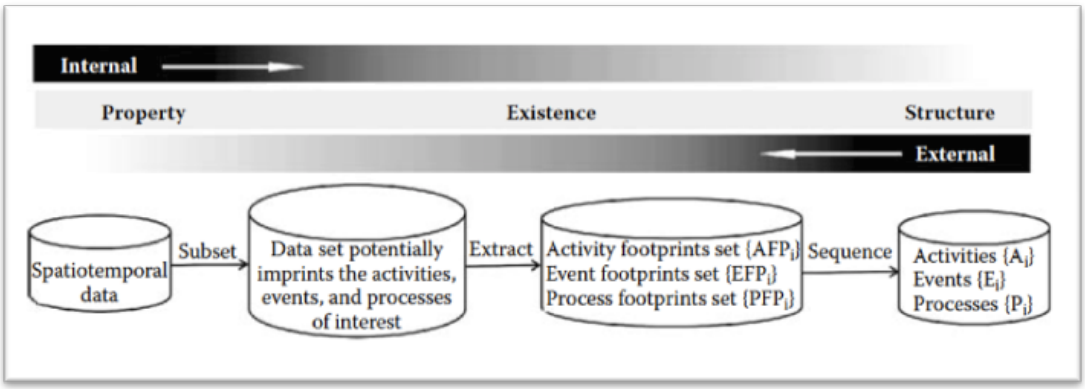


圖 4：時空知識探索的內部及外部結構 (Yuan, 2009)

由於資料的數量或是複雜度增加，要直接從資料當中抽取出有用的規則也變得更加的困難，因此在時空探索的研究中便有學者提到詮釋規則探索 (meta-mining) 的概念 (Abraham and Roddick, 1997; Roddick and Spiliopoulou, 2002)，說明可先從資料的摘要 (summary) 或是已被抽取出來的結果當中進行規則探索的工作，先抽取出資料當中的詮釋規則 (meta-rules)，之後再以這些詮釋規則為基礎進行更深入的探索。由以上回顧可以了解，地理現象為動態性，其可能產生變化的部分包括空間、時間、或是屬性特質，在時空資料庫當中要能夠做知識探索必須要分為特性、存在、和結構的角度來分析，這一部分來自於內部的分析，同時也有來自於外部的觀察，其各別皆需要不同的資料儲存或是展現型態，因此一個新的整合性時空資料模式必須要能夠針對地理現象的特質，針對不同的使用需求，整合現有的地理資訊系統環境運作來動態完成處理以及分析的工作(Yuan, 1997; Yuan, 2009)。

縱上所述，可知目前知識探索架構存在的問題主要為缺乏一有效的探索架構來處理及展現大量複雜的資料、缺乏相關領域知識來引導資料的前處理與轉換、缺乏對於時空資料的處理能力，以及缺乏以使用者為中心的互動式探索的方法論等。過去的研究裡有學者提出圖形或是視覺化可以增加探索的效率，但是在前人所提出的知識探索架構裡，視覺化僅扮演著資料或是知識展現的角色。人類視覺感知的能力遠超過目前電腦所能達到的範圍，因此近年來地理視覺化 (geovisualization) 在這樣的研究趨勢下逐漸受到重視。因此本研究第二節將先就地理視覺化的發展歷程及理論基礎進行回顧。

第二節 地理視覺化方法的發展歷程及理論基礎

一、地理視覺化技術的興起

視覺化技術在90年代以來逐漸受到電腦科學領域的重視，當大量及多維資料不斷膨脹時，電腦計算的複雜度也成指數的增加，地理學有地圖學長久的傳統，視覺化技術對於人類的地圖資訊展現、空間認知，以及時空知識探索等面相，皆有其潛在研究的價值，在這樣的背景之下一項新興的研究方式—地理視覺化 (geovisualizaiton) 的探索方式，開始受到地理學者的注意。

地理視覺化是利用地理空間的互動展示方法探索資料的空間特性，從探索的過程或是結果裡產生新的假設、新的解決方案，或是建構新知識的方法 (Kraak, 2003)。地理視覺化與一般資訊視覺化最大的不同在於，它必須要使用三維 (three dimension, 3D) 展示或是平面地圖展示方法來表示實體空間，它處理的是空間問題，而不只是抽象的屬性資料 (Andrienko and Andrienko, 1999) 。Card et al. (1999) 認為地理視覺化與資訊視覺化 (information visualization) 的不同之處為：「地理視覺化專注於探索與地理現象有關的時間或是空間資料」。因此， 地理視覺化可以幫助研究者尋找隱藏於原始時間與空間資料中的資料關連性、空間分布或是趨勢等資料特徵。

Dykes et al. (2005) 認為地理視覺化是關於人、地圖、處理，以及資訊和知識的獲取。Robinson et al. (2005) 則提出地理視覺化在本質上是高互動性和動態性。相對之下 GIS 的傳統研究方式是高結構性的、主要目的是為了建構靜態的空間資料描述。

Kraak (2007) 也提出視覺化應該要跟分析結合在一起來談論。他認為視覺分析是一種利用互動式視覺介面進行分析推理的一種科學。簡而言之就是“偵測到所預期的結果以及發現不能預期的現象”。

目前所發展的時空資料視覺化軟體工具，一方面是拜電腦科學所賜，另一方面則是沿襲了傳統地圖學的研究方式 (Andrienko et al., 2003)。由

於視覺化工具有互動性及動態性，因此使用者可以選擇多樣化的資料表現方式，有助於資料探索的進行，同時也幫助辨識從資料當中找出資料的型態，以及資料的結構與意義 (Edsall, 2001)。因此地理視覺化整合了資訊視覺化科技、探索式地圖學、影像分析、資料探索分析、以及GIS等等 (圖5)，並提供了視覺探索、地理資料的整合與表現的一套理論、方法以及工具 (MacEachren and Kraak, 2001)。

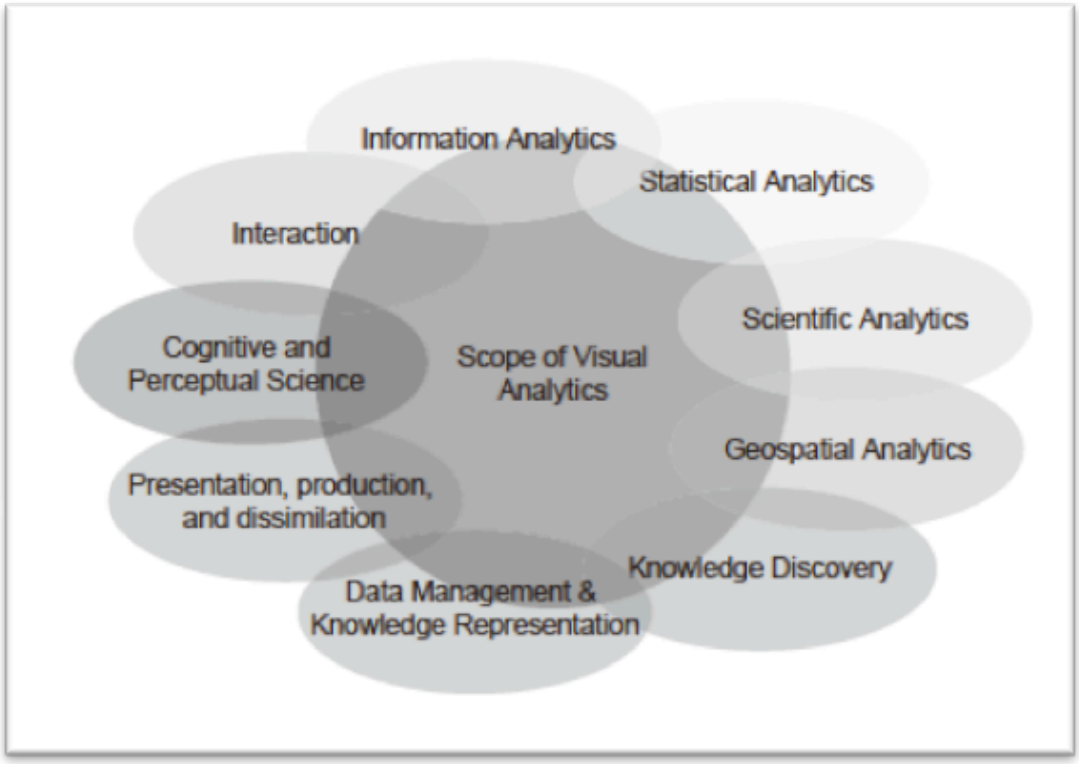


圖 5：視覺化分析的涵蓋範疇 (MacEachren and Kraak, 2001)

Ferreira de Oliveira and Levkowitz (2003) 在他們的研究當中指出未來視覺化技術會逐漸朝向視覺化資料探勘技術 (visual data mining technique) 的方向前進，而要完成這樣的目標必須要仰賴資訊視覺化以及資料探勘研究者的共通合作。

從 2000 年開始有越來越多的學者針對地理視覺化和空間知識探索

的整合進行討論，國際製圖學會 (The International Cartographic Association, ICA) 在其視覺化研究的議程當中，便提出以高速運算地理學、地理視覺化、和資料庫知識探索等三大技術作為解決地理研究問題的主要手段。 Gahegan et al. (2001) 針對 Koperski et al. (1996)關於空間資料挖掘架構中的Controller的部份進行了進一步的解釋，認為視覺化技術可以代替這一部份的功能，並且可以加強使用者對於整體過程的理解和闡釋。而也有學者提出利用人類的視覺系統觀察是有利的，在許多趕鮒]如遙測、醫學）的工作上目前尚未有更優秀的感知系統可以比人類視覺在特徵與結構的辨識上表現的更優秀 (Miller and Han, 2001)。

由以上回顧可得知地理視覺化的定義做一簡單的結論，亦即若是處理的資料是與空間或時間有關的地理現象，同時將這樣的資料進行視覺化分析必須要配合地圖及其他圖表的使用展示，另外探索這些資料必須要仰賴使用者的地理知識或是空間認知等能力的情況下，則這樣的縱合研究面相可以稱為地理視覺化，而相關的技術及分析可以稱為地理視覺化技術及分析。

地理視覺化研究已經成為一個新興的研究趨勢，過去的知識探索及資料探勘並未善用人類視覺感知的優異能力，因此無法發掘隱藏在複雜資料當中的知識，若是可以將地理視覺化與資料庫知識探索技術相結合，應可以改進過去方法所面臨的許多問題。因此本研究針過去對地理視覺化的相關理論基礎進行回顧，並將過去相關研究進行統整分析，提出視覺化技術在知識探索當中所應當扮演的角色。

二、地理視覺化研究的理論基礎

查爾斯·桑德斯·皮爾士 (Charles Sanders Santiago Peirce) 在其 1891 年的著作 The Architecture of Theories 裡面提出了人類發現新知識的三種模式，分別為演繹法 (deduction)、歸納法 (induction)，以及溯因法

(abduction)。演繹法是指在一個既定的前提之下，依照邏輯法則所推推導出來的結論，這也是合理性思考的核心。而歸納法則是依照實際發生的事件，所推導出來的整體一般性主題或法則。一般演繹法並不會產生新的類別或是關係，而歸納法則是必須仰賴統計或是機率來其斷定式的邏輯盲點。溯因法與兩者不同的地方在於它是由推理到最佳解釋的過程，溯因法利用事實的集合來推導出它們的最合適的解釋。這意謂必須生成假設來解釋所觀察的事實或推導結論，溯因法比較像是演繹法的反向操作，演繹法允許推導 B 作為 A 的結論，溯因則是允許推導 A 作為 B 的解釋，簡單來說，溯因是解釋已知事物的過程，也就是由案例來生成理論，因此常被稱為「發現的邏輯」，例如觀察到一件使人感興趣的 B 事實（例如：飼養雞不斷的死去），若是A（例如：禽流感流行）為真，則 B 便成為理所當然的事實，進而產生認為A理由影響 B 事實為真。不過這樣的方法雖然可以提出一完整說明的假設，卻同時缺乏演繹法的邏輯嚴謹度以及歸納法的統計基礎。

這三種邏輯推論方法各有其優缺點，但單獨的方法皆無法滿足目前資料庫知識探索的所有需求，因此有學者提出推論鏈結 (inferential chains) 的概念（圖 6），強調人類在推論的過程當中是同時在這三種方式當中進行轉換，歸納法所衍生的類別範疇 (categories) 以及溯因法所產生的解釋 (explanations) 除了可以用來建構出演繹法所需要的規則 (rules) 之外，當規則的例外以及類別範疇的特例發生時，可以將結果回饋到假設的產生、或是概念的修正(Gahegan, 2009; Gahegan and Brodaric, 2002; Sowa and Majumdar, 2003)。

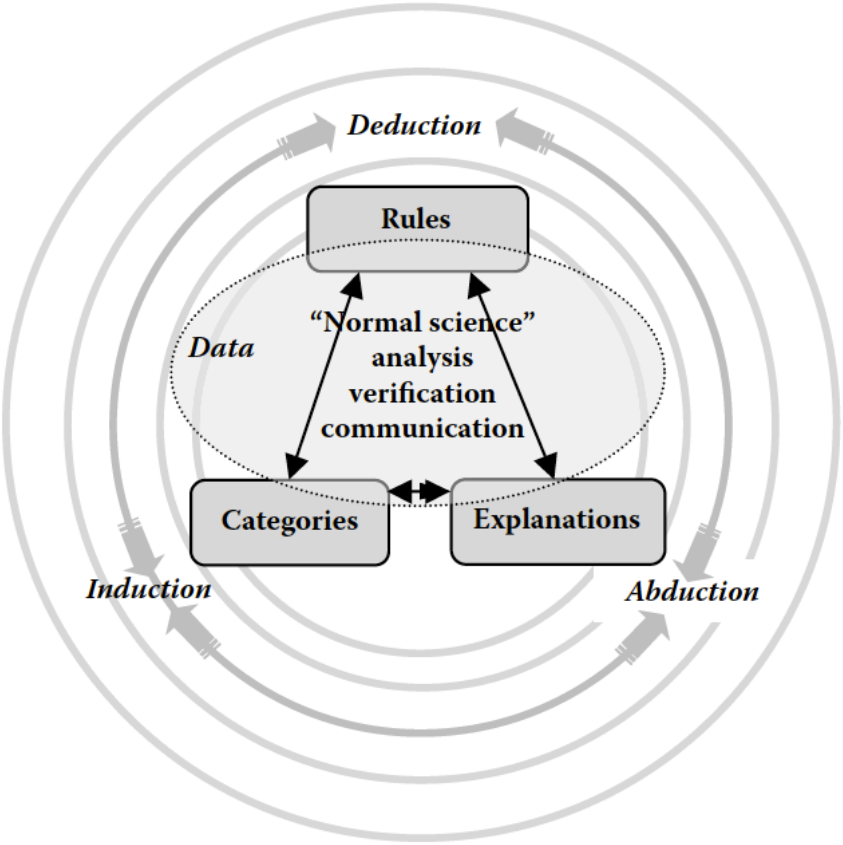


圖 6：推論鏈結概念架構 (Gahegan, 2009)

在視覺化工具整合的環境當中，這些探索工作是由視覺化工具的創造者和使用者共同來合作產生出來的，探索和歸納以及形成假設等工作是同時一起在進行的，視覺化的部分可以提供一些較為寬鬆假設的描繪，但是使用者可以利用它去產生一些更為細緻的內部理論或是結構。Gahegan and Brodaric (2002) 則整理了視覺化技術在探索式科學架構中所扮演的角色 (圖 7)。說明了視覺化技術不論是在以資料為主的探索及溯因，或是到模式為主的演繹及驗證，視覺化技術皆可輔助建構下一步推論工作或是幫助發現問題而進行修正。在探索的階段知識探索的過程著重在於資料的檢視，視覺化在此扮演的角色是透過各種不同的手段，讓使用者能夠更容易發現任何特殊的資料型態或是偏差值，這些特殊的資料可以當做是被過濾的對象，讓下一階段的類別或是模式建構更加的嚴謹和正確，同時這些資料也有可能是打破原有假設或是形成新假設或理論

的證據，必須要被獨立出來加以檢證，這部份視覺化輔助的是溯因的推論過程。在類別範疇建構的階段視覺化扮演的角色是利用視覺化的展示來呈現類別或是架構的概念，透過與對於類別的檢視以及互動調整使用者可以即時得到反饋，另外視覺化的介面如心智圖，階層架構圖等也有助於使用者對於資料結構的理解，這部份視覺化輔助的是歸納的推論過程。在模式建構與分析的階段，透過視覺化的模式追蹤或是模式操縱，使用者可以比較不同的模式選用及參數設定之差異，對分析結果所產生的影響，這部份視覺化輔助的是演繹的推論過程。而在結果評估的階段視覺化主要的角色則是如何去闡釋資料及資料當中仍然存在的不確定性，這部份視覺化輔助的對於模式的改進或是引發新的假設或是概念。

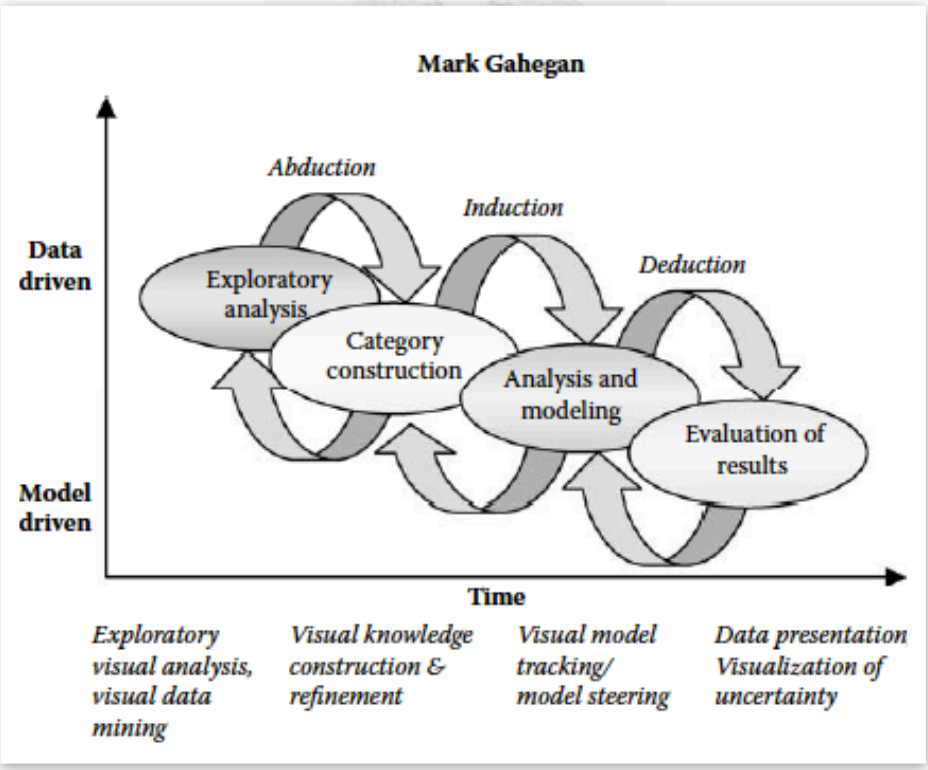


圖 7：視覺化技術在探索式科學架構中所扮演的角色 (Gahegan, 2009)

Andrienko and Andrienko (1999) 在其針對互動式地圖的研究中，指出地圖式的視覺化典範包括三個主要的要素：個別的地圖使用 (使用者

可依其需要產生不同的地圖)、展現未知的資料而非展示已知資料，以及高度的人類與地圖的互動。因此基於以上三種要素以及演繹、歸納，以及溯因法的結合，可知視覺化研究是一附加在探索式科學研究架構之上的一們新興領域，它以人類為中心，透過視覺感知系統以及推論鏈結的整合，透過高度的互動性及動態性的展現，讓知識探索的架構更為完整。

Fabrikant (2001) 提出任務本體論 (task-ontologies) 的概念，強調有效的視覺化工具應該要提供使用者必要的視覺化協助，來進行問題解決或是決策支援。因此藉由任務本體論的探討，將有助於地理視覺化互動式工具的介面設計。

Preece et al. (2002) 以使用者為中心 (user centric) 的程式設計觀點，說明一個可用並且實用的資訊系統，最基本的要求就是對於使用者和其使用系統的了解。這項要求也是一個程序發展的哲學，在每一個設計階段，都必須針對資訊系統的使用者的需求、要求，還有限制等等進行考慮。

Aditya and Kraak (2005) 提出情境式問題分析以及人機互動 (HCI) 的研究取向，強調可用性的研究應該在工具發展的初期就被導入，而工具的設計重點在於如何了解使用者並且設計可以真正協助他們的使用介面。而要達成這些目的必須要先具體化使用者的需求、以及加強工具的使用介面，以及要不斷的改善系統的設計。

根據這樣的理論基礎，過去有許多研究者近一步去探討該如何設計有效的地理視覺化架構，以及該如何驗證這樣的設計架構的有效性以及可用性。例如 Gahegan (2009) 提出如果要了解如何才能輔助人類的視覺探索，必須要充分理解下列兩個問題：

1.人是如何去連結不同的思考過程來輔助他的研究方向？我們如何去在不完整的概念架構中辨識出有用的？

2.哪些視覺化分析研究和技術是可用的？而且他們可以被用到哪種程度來支援人類的探索。

Gahegan (2009) 也提出根據人類視覺感官的本質，視覺化研究可以支援知識探索的成功要件為：

1.透過視覺感知將非容易直接觀察的資料展現給人。

2.提供方法手段去展現不同的概念架構如顏色，形狀、大小、位置、移動等。

3.透過即時動態的介面工具讓使用者可立即得到回饋與修正概念。

Slocum et al. (2001) 主張視覺化的新研究方法如果不經過認知理論的架構之下且利用可用性工程原則 (usability engineering principles) 反覆測試的話，無法被採用為有效的研究方法。該研究並提出多個重要的研究議題，其中一個就是地理視覺化方法的有效性評估。而過去針對這些有效性的評估，曾經有不同的研究結果出現， Koussoulakou and Kraak (1992) 以及 Patton and Cammack (1996) 發現動態地圖能對其研究成果有所助益, 但是Slocum and Egbert (1993) 以及 Johnson and Nelson (1998) 的研究成果則指出與靜態地圖差異不大。Morrison et al. (2000) 的研究則指出動畫一般來說在教育目的上效果比較不如靜態地圖。 Harrower et al. (2000) 發現brushing和focusing方法對於學生來說不太有效用。不過 MacEachren et al. (1998) 的研究則顯示流行病學家利用這些工具找到了重要的空間型態。

在 Uhlenkuken et al. (2000) 的研究中，則提出了視覺化工具軟體目前仍存在整合性的問題，大部分的軟體僅供特定分析用途，無法大量運用到不同目的之分析需求之上。不過近年來仍然有許多成功的研究案例，例如 Andrienko et al. (2003)、Dykes and Mountain (2003) 以及Zhang (2006) 的研究成功的應用了視覺化方法於大量的時空資料分析之上。

Jones et al. (2009) 在其文章中強調地理視覺化研究必須要回到以個

使用者導向的觀點。作者認為過於複雜的使用者介面設計是過去幾年來GIS 和地理視覺化工具的使用著沒有顯著的增加的原因之一。本研究認為一個較為精簡的使用者介面可以幫助 GIS 專家以及非專業人士建立新的知識了解。因此採用了少即是多 (less-is-more) 的觀點，藉由可用性矩陣分析系統效率、有效性、以及可學習性三方面來探討。該文章也認同Philbrick 在1953年所發展出的的地圖學準則，這些準則如今仍然適用於GIS 以及地理視覺化。而過去的gis發展已經逐漸忽略了簡化這個概念在使用者介面設計當中所扮演的重要的角色。 Fuhrmann等人(2005)也指出地理視覺化缺乏有效性的原因是未以使用者為中心的方式來設計系統。

Guo (2009) 則提出地理視覺化的研究方式可以強化探索和對複雜型態的了解，但是光靠視覺化的方式比較不能夠處理較大型的資料，並且使用者可能有自我的偏好及預期心理造成偏差。因此一些整合使用者及視覺化模式的研究 (MacEachren et al., 2003; Swayne et al., 2003; Swienty et al., 2008) 陸續被提出。

在Pennsylvania State University的 GeoVISTA Center，是近年來致力於地理視覺化探索研究最多的研究中心，該中心所研發的 GeoVISTA Studio系統軟體，是以JAVA語言撰寫而成，採用開放性資源 (Open Source) 及免撰寫程式碼 (codeless) 的元件導向式設計 (component-oriented)，提供使用者開放的系統架構環境，GeoVISTA Studio最主要的目標便是藉由提供多方面的視覺化展現及運算工具的環境，來提升地理資訊科學資料的分析 (Takatsuka and Gahegan, 2002)。在該系統的可用性評估方面，過去已經有一些相關研究証明對流行病學的資料分析有正面幫助 (Edsall, 2003a; Robinson, 2005b; Robinson, 2005a; Robinson et al., 2005)。Demsar (2007) 則是觀察人類如何進行地理空間資料的視覺探索，利用放聲思考法 (thinking aloud method)，採用人機互動 (HCI) 的方式收集受試者操作GeoVISTA Studio的類型。由此可以發現，過去地理視覺化的研究集中在技術的觀點上，使用者並不佔有一個重要的角色，但是近年來以使

用者為中心的研究取向漸漸受到重視。

在實際的空間知識探索應用上，Robinson et al. (2005) 的研究中便利用GeoVISTA Studio軟體進行社會經濟時空統計資料的探索，來探討這些統計因子與流行病學的時空關係以及空間分佈特性，而在他的研究當中利用互動式的視覺化探索技術，也成功的找出了阿拉帕契山區的高肺癌致死率與其他社會經濟指標的關連性。另外近年來發展的工具則有以Frank Hardisty主要開發者的 GeoViz Toolkit等 (Hardisty, 2005)。在這些工具裡面所含有的互動式視覺化分析工具，其最大的特點在於所有的分析工具和地圖顯示都是互動且即時的，另外一個特點則是其各種工具支援使用者的自行調整、分類、部分選取、動畫展示等。有了這些工具的發展，讓地理學家得以利用互動式視覺介面進行分析推理，對於時間及空間型態能夠更有效率的認知。

由以上回顧，可以發現地理視覺化的研究取向已經漸漸朝向地理視覺分析的統合研究前進，而視覺化研究的發展歷程以及理論基礎的探討也漸漸聚焦到以人為中心的研究出發點，而過去知識探索以及資料探勘的研究架構雖然也有提到使用者的角色，但是在過去並未能有效發揮其所扮演的角色與功能。而視覺化在過去的系統架構上也僅被定為成展示知識成果的工具與手段，而視覺化工具並非是要取代人類的角色，而是要在人類探索新知識的模式上，輔助人類去發現更多隱藏以及有意義的資料特徵、資料關係、資料趨勢以及極端值等。因此地理視覺化可以說是輔助工具的發展，同時又是研究人類空間認知與視覺分析架構的統合研究。依照這樣的角色定位，過去發展出了許多視覺化工具，因此在建立本研究的方法與架構之前，先針對過去文獻進行統合的整理並加以分類，以作為研究方法階段工具選用的參考。

三、地理視覺化工具的分類

Hinneburg et al. (1999) 在其針對高維度資料的視覺化資料挖掘研究當中，歸納出四種視覺化資料挖掘的主要分類如幾何投影轉換技術、圖像化技術、像元展示技術，以及階層展現技術等。Gahegan (2009) 針對地理視覺化工具的功能性，提出視覺化工具可分為視覺編碼或展示、時間序列行為、資料整體特性，以及互動性支援等四種大類。資料個體、整體，或是時間行為等若是透過不同的展示方式或角度，觀察的節果以及後續的推論也可能會有不同，因此互動性的支援提供了是用者操作的彈性，利用人類的視覺系統以及大腦歸納推理邏輯能力，希望能夠補足視覺化展示的不足以及刺激人類研究者產生新發現。但此研究也提出隨著時空資料動態分析的需求增加，未來這些工具的功能性也會趨向於整合。因此整合此類文章可以得知，視覺化工具其發展與應用的主要目的是要幫助找出資料整體或是個體的特別特徵，或是找到容易發現這些特徵的新方法或是觀察角度。以這樣的出發點，近年來許多研究者開發出了許多地理視覺化工具，本研究依其特性整理如下表1。

（一）地圖式展示技術

地圖和一般圖表不同，他可同時展現空間概念以及屬性資料特性，地圖不但是展示的介面同時也是搜尋空間上的特殊型態、空間分區，以及空間群聚的探索空間，地理視覺化的地圖展示著重在於地圖與屬性資料的即時互動性，同時多維資料的展示方法也常和地圖整合，例如雙變數面量圖 (bivariate choropleth) 便是一種可將多維屬性資料對應到空間地圖上的方式 (Andrienko and Andrienko, 1999)。雙變量的展示方式與過去的地圖學中的面量圖相似，但不同的地方在於利用了雙色顏色漸層套疊，如此可以在顏色的表現上多出一個維度的變化，例如利用綠色深淺代表廢氣工廠數量的高低，紫色的深淺代表肺癌的死亡率高低，便可以在地圖上尋找出工廠數量多以及死亡率高的深色地帶。

表 1： 地理視覺化工具的分類 (本研究整理)

展示方式	產生來源	展示方法	搜尋目標	相關研究
地圖式展示技術	使用者	Choropleth map,	空間上的特殊型態、空間分區，空間群聚	(Goodchild and Dubuc, 1987; Andrienko and Andrienko, 1999)
圖畫式展示技術	使用者	Scatterplot Scatterplot matrix PCP	資料中的特殊型態、資料分類，資料群聚，資料偏差值	(Cook et al., 1997; Ankerst et al., 1998; Koua et al., 2006)
資料投射或轉換	演算法	Principal component analysis Multidimensional scaling RadViz Self-organizing map (SOM)	資料群聚、偏差值，資料分界	(Flexer, 2001; Penn, 2005; Muller et al., 2006; Jing et al., 2007)
圖像或符號化轉換	使用者	Star Plot	符號特殊型態、規律性	(Klippel et al. , 2009)
網格或像素展示方法	演算法	Space-filling Pixel-based Treemap Map Cube	群聚、熱點	(Johnson and Shneiderman, 1991; Shneiderman , 1992; Keim, 2000; Stolte et al., 2003)
階層或概念網路展示圖	詮釋資料	Treemap Concept Map	節點及連結特殊型態	(Canas et al., 2005)
動畫或是聲音	使用者	Sound Map Animation	規律性、變化趨勢、偏差值	(Krycier, 1994; Evans, 1997; Brauen, 2006)

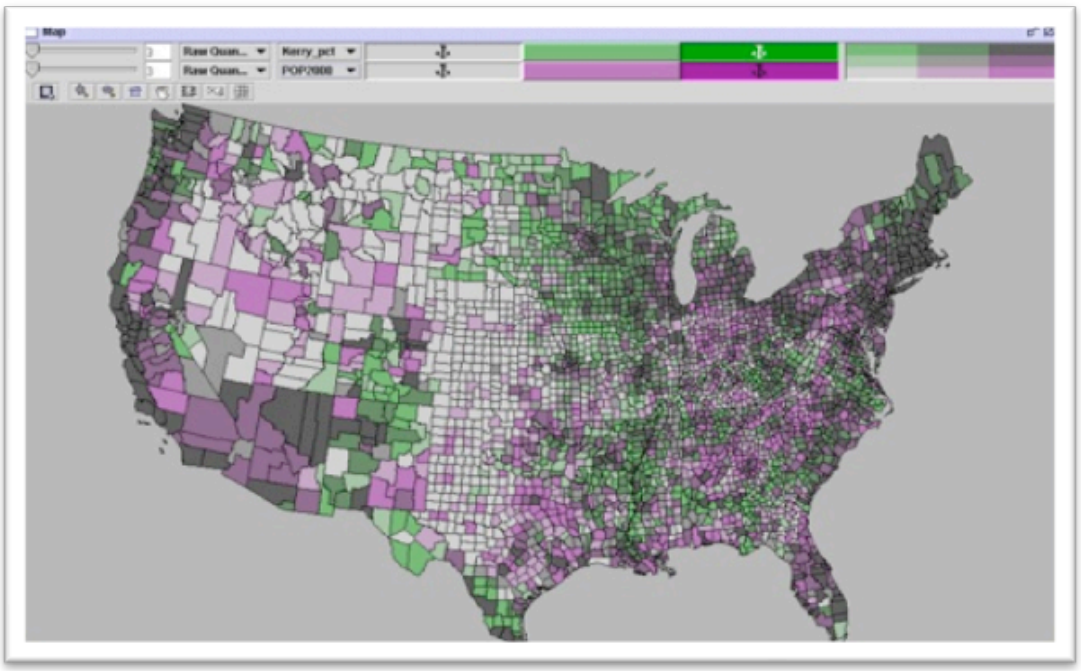
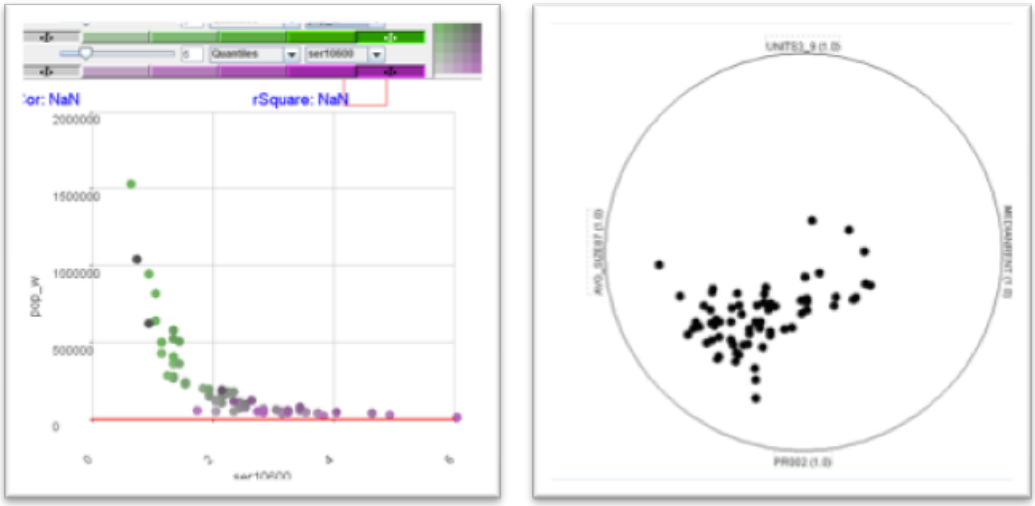


圖 8：雙變數面量圖(擷取自 GeoViz Toolkit 軟體)

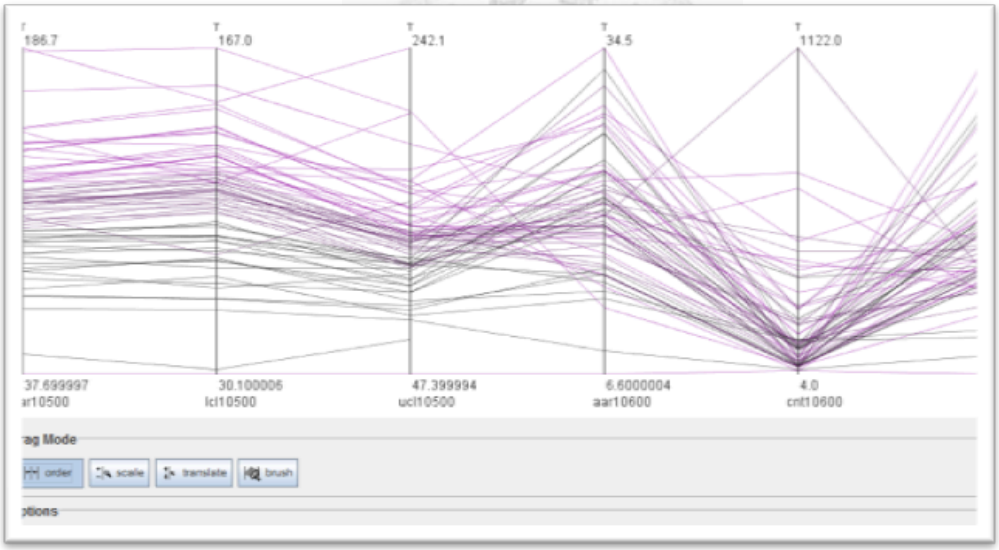
（二）圖表式展示技術

圖表式展示技術發展較早，過去有許多技術早被用於統計資料的展示，而在地理視覺化當中，此類方法主要的搜尋對象為資料中的特殊型態、資料分類，資料群聚，資料偏差值等等，當我們可以過濾出這些特殊的形態以後，我們可以在把它相對應的空間部分進行進一不的分析比對，以了解其空間之特性。這類的展示方式如散布圖 (scatterplot)、平行座標圖 (parallel coordinates plot, PCP)、放射圖 (radviz)，以及利用多維變數所建立出來的矩陣 (例如scatterplot matrix)等。散佈圖由於展示之基本限制，因此對於複雜多維資料較不容易呈現，但近年來也筑i許多新的方式例如利用展示陣列或是不同顏色和大小等，來增加其展示資料的維度。平行坐標圖的概念最早由Inselberg所提出(Edsall et al., 2000)，其展示方法是將一筆資料的各欄位描繪成相互平行的軸線，再將資料的數值依照其位置連接成各自不同的折線 在Edsall (2003b) 的研究中指出平行座標圖對於促進對多維度資料庫的瞭解具有有效的幫助。但是在Koua et

al. (2006) 針對各種視覺化工具有效性的分析研究中則指出PCP在使用者正確性的表現上不如傳統地圖以及自組織映射地圖網路。



(a) 散布圖　　(b) 放射圖



(c) 平行座標圖

圖 9：不同視覺化方法介面展示 (擷取自 GeoViz Toolkit 軟體)

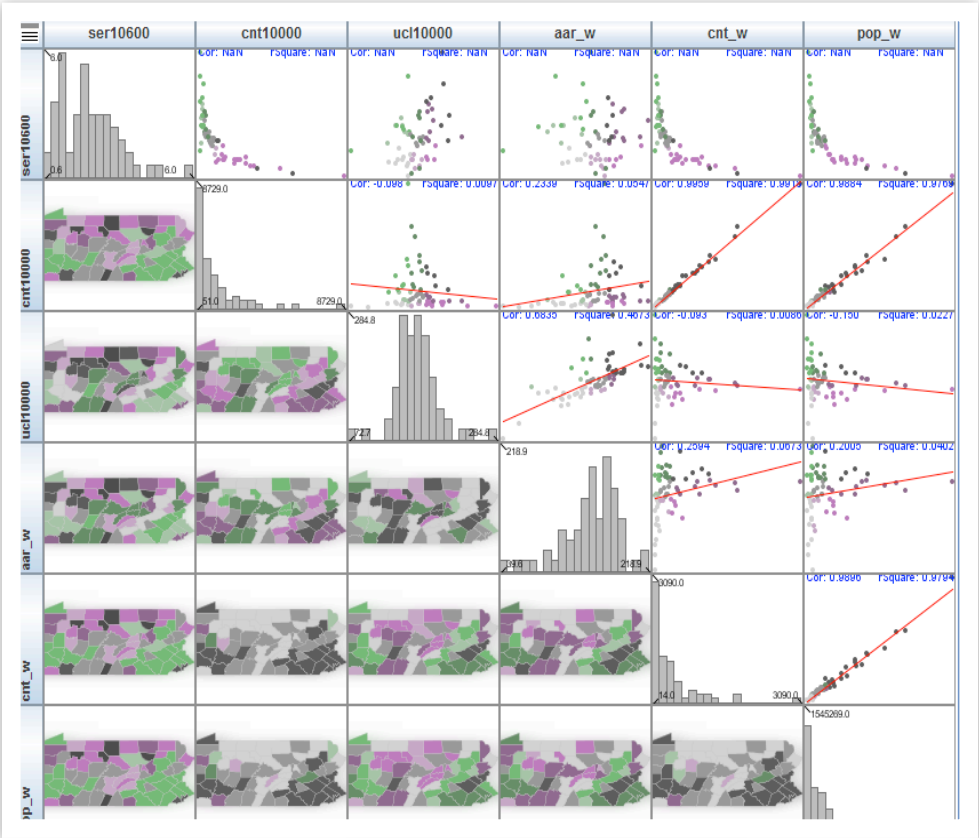


圖 10：雙變數面量圖與散步圖展示矩陣 (擷取自 GeoViz Toolkit 軟體)

（三）資料投射或轉換

提出視覺化的最大挑戰是在於資料維度的問題，以及資料的數量的膨脹。還有對於複雜資料型態的了解或感知。因此過去的許多研究方法都著重在於如何減少資料的維度，但同時又可以保留最多有用的資訊，如多元尺度法 (Multidimensional Scaling) ，主成分分析法 (principle component analysis, PCA)，放射圖視覺分析 (radviz spring visualization)，自組織映射圖網 (self organizing map, SOM) 等(Gahegan, 2009)。其中近年來受到最多地理學家注目的便是自組織映射地圖網。

自組織映射圖網於1980由T.Kohonen提出，它是屬於無監督式學習網路模式的一種。它的基本原理是模仿人類大腦中，具有相似功能的腦細胞會聚集在一起的特性，所發展出的類神經網路。因此，只要在學習

範例中有內在的聚類規則，那麼就適用於自組織映射圖網路。SOM在網路拓樸的選擇上多採用二維座標系，其中以又以六角形拓樸最為常用，因其展示較為直覺易瞭解，其拓樸關係是利用Unified Distance Matrix (U-matrix) 的展示方式，網格顏色的深淺來代表神經元之間的距離差異，另外漸層色的顏色差異代表不同的神經元，以及圓形的大小代表配置到神經元中資料的數量。在最近的地理視覺化之研究中發現 SOM 在發掘資料當中可能存在的群聚上有不錯的成效(Koua et al., 2006)。SOM 的地理視覺化探索工具又以GeoVISTA Studio的主要程式發展者Guo後來所獨立發展的SOMVIS程式介面較為友善及容易使用。近年來最受到地理視覺化學者重視和使用的便是自組織映射圖網路 (self-organizing map, SOM) 。SOM 是一種非線性、具有組織的且平滑的演算法，可將高維度資料進行壓縮，映射到較低維度的神經元矩陣裡面 (Kohonen, 2001)。因此地理學者開始注意到將 SOM 整合到地理視覺化的探索方法 (Koua and Kraak, 2005; Thill et al., 2008) 。

由於 SOM 的網路結構複雜，研究者常使用成份平面展示 (component plane presentation, CPP) 與統一距離矩陣 (U-matrix) 的著色方法展示其拓樸映射圖。CPP是將映射的網路利用規則的幾何排列圖形展示，U-matrix則是利用灰階顏色的深淺來表示神經元彼此之間在拓樸映射平面上的歐幾里德距離，顏色越深代表神經元的距離越遠，資料的差異較大。神經元所在之位置的顏色則是計算其與鄰近神經元的距離平均值而來(Skupin and Agarwal, 2007) 。 Koua and Kraak (2004) 在其研究中利用 SOM 視覺化展示來進行世界人口衛生資料的探索，其研究結果也顯示 SOM 配合地圖圖及其他視覺化，透過互動與選取的方式，可以發現資料當中潛在的關係或是關連性。Guo (2005) 提出一個3D結構的配色方法，利用色彩矩陣以及亮度的變化，讓靠近的神經元的配色較為接近，但同時又可以在圖面上區分出不同特性的資料，可以更有效的幫助研究者探索新的地理資料特徵。

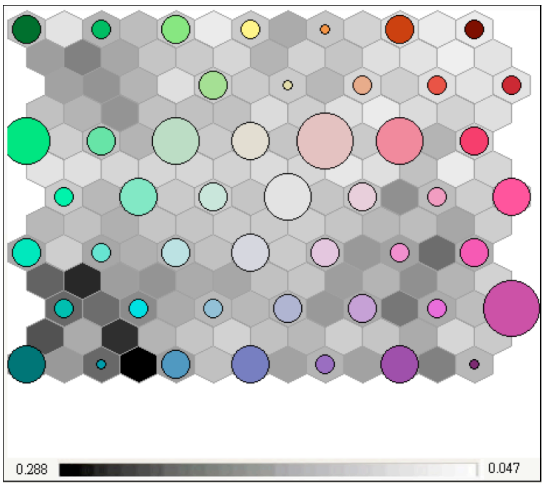


圖 11： SOM 六角形拓樸映射圖視覺化範例 (擷取自本研究開發軟體)

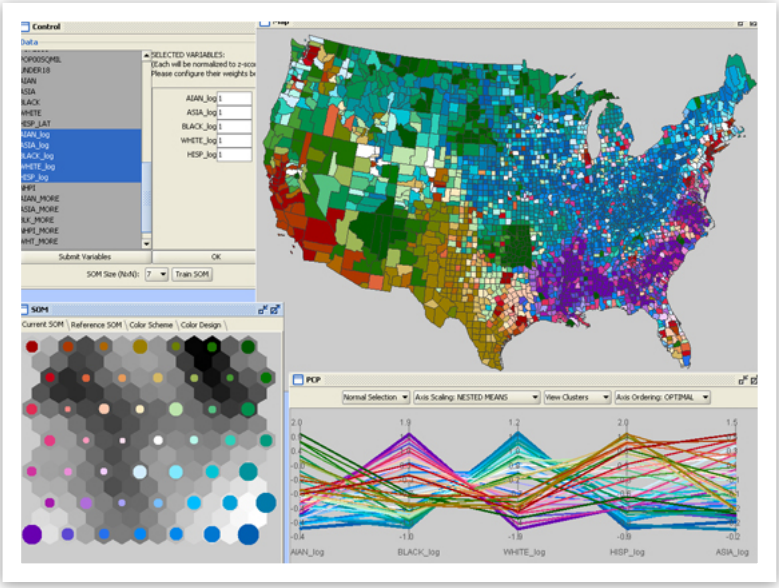


圖 12：SOMVIS 自組織映射圖網路範例
(取自 http://www.spatialdatamining.org/software.htm)

由於SOM的原始演算法僅針對資料空間進行自組織，過去雖然有許多改進的演算法被提出，但是大多並未考慮應用在地理資訊科學的應用上 ，因此近年來有學者針對原始演算法進行改進，加入了空間特性的考慮，其主要的改進方法有三種 (Bação et al., 2005)：

1. 座標強化自組織映射圖網路 (Geo-enforced SOM)

這類方法主要分為兩種，一種直接使用空間座標作為 SOM 的

輸入向量，但會依照不同的使用需要調整座標所佔的權重值，此法當權重值調整時，其分析結果也會改變。另一種則是利用參考之地理中心點，將座標值與地理中心之距離計算後作為輸入向量進行計算，此方法也可以將多個地理中心納入計算。

1. 階層式自組織映射圖網路 (Hierarchical SOMs)

此法由 Ichiki et al. (1991) 所提出，他利用三層式的網路來進行符號及語意圖的分析，將 SOM 分為個個較低階的特殊化 SOM 地圖，再將個別地圖組合成較高階的 SOM 地圖，在地理空間應用上，此法可以觀察不同屬性的輸入向量的低階 SOM 地圖，是否在空間上具有相同或是特殊的分佈型態。

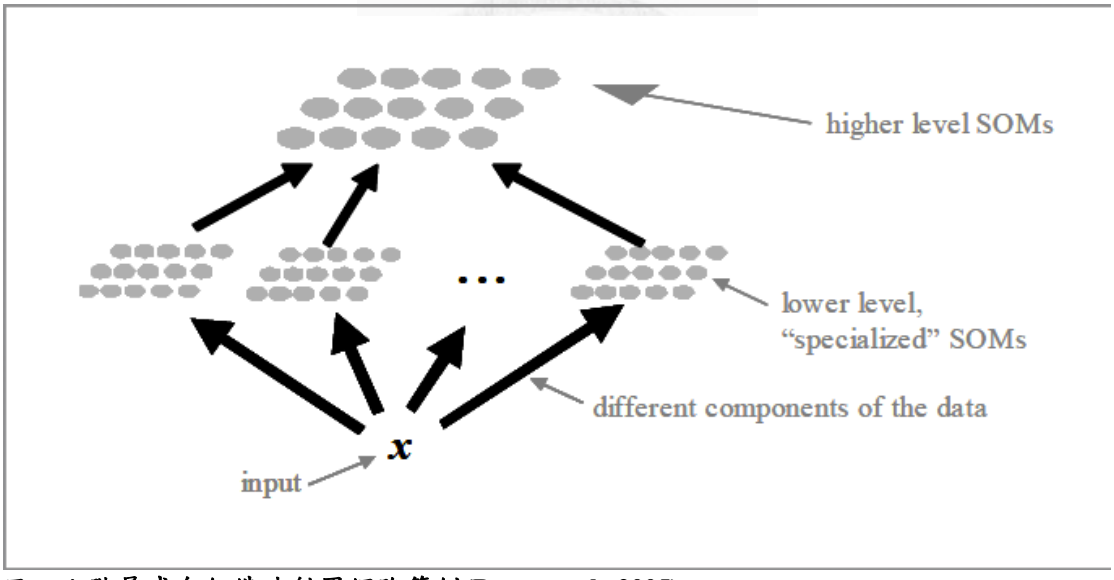


圖 13：階層式自組織映射圖網路範例(Bação et al., 2005)

1. 空間自組織映射圖網路 (Geo-SOM)

此法由 Bação et al. (2005) 所提出，其演算方法與傳統 SOM 類似，但是在網路的競爭演算的過程中，僅在空間位置較為接近的神經元當中選出資料最相近的做為優勝神經元 (best matching unit, BMU)，因此 Geo-SOM 的演算結果會將地理位置相近的資料進行結合，

Bação et al.在其研究中利用模擬資料針對傳統 SOM 以及 Geo-SOM 進行比較，發現 Geo-SOM 的分類結果較能反應空間分布型態，但是傳統 SOM 在資料的聚類上表現較佳。

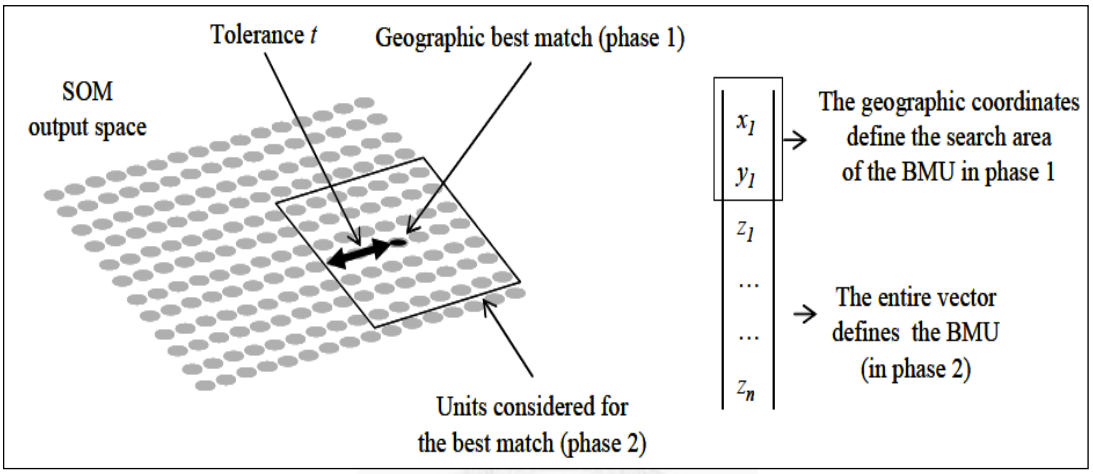


圖 14：空間自組織映射圖網路範例 (Bação et al., 2005)

（四）圖像或符號化轉換

過去地圖學的傳統多應用圖像及符號在靜態的展示上，並未將其應用在互動式的探索上面，在2007年GeoVISTA Center的Frank Hardisty以GeoVISTA Studio的元件為核心，開發了一套整合更多視覺化工具的軟體GeoViz，主要用於醫療及流行病學的研究計畫方面。其中包括星狀地圖的展示，它在雙變量展示地圖上面在多程載了一層符號化的星狀圖設計，透過人類視覺對符號的辨識與認知，將可以更靈活的運用在多維度資料的展示方面。

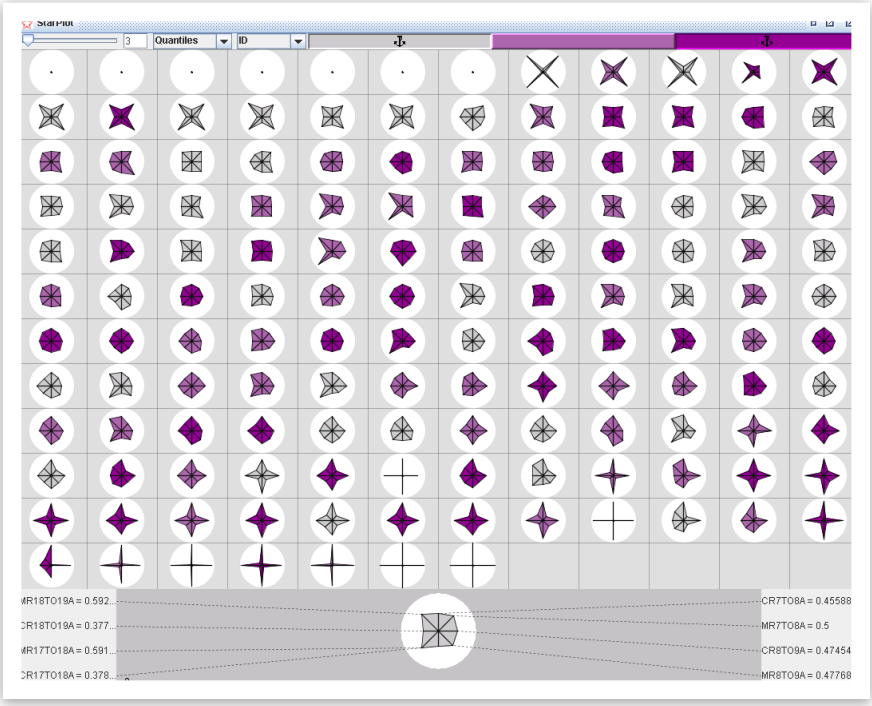


圖 15：星狀圖展示介面圖 (擷取自 GeoViz Toolkit 軟體)

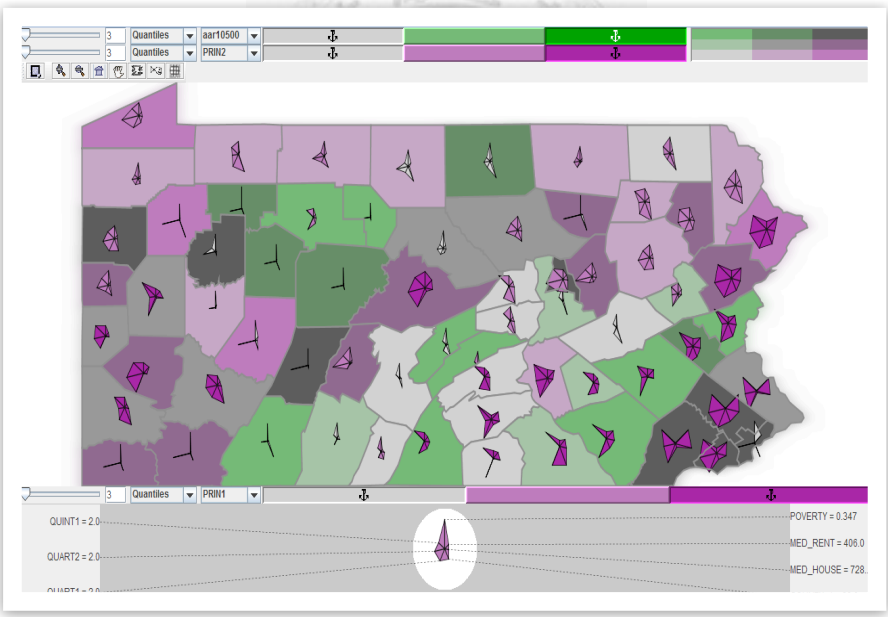


圖 16：星狀圖面量地圖展示 (擷取自 GeoViz Toolkit 軟體)

（五）網格或像素展示方法

網格式及像素展示的方法主要是利用顏色的分類及資料的排序，將大量的資料展示在二維的介面當中，希望能夠觀察出資料當中的群聚、熱點， Space-filling展示方法可用各種不同的方式來排序料，Map Cube 提

供了一組空間資料集合運算的方法，並且可幫助研究者快速且有效的去分析資料型態以及跨維度的資料比較。交通資料便包括了這些時空特性的特質。非常適合用 Map cube 來分析。

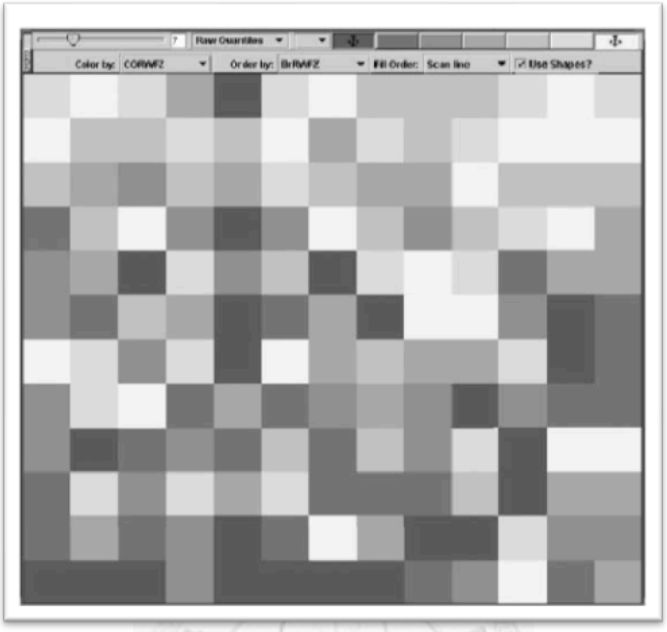


圖 17：Space-filling 視覺展示(Gahegan, 2009)

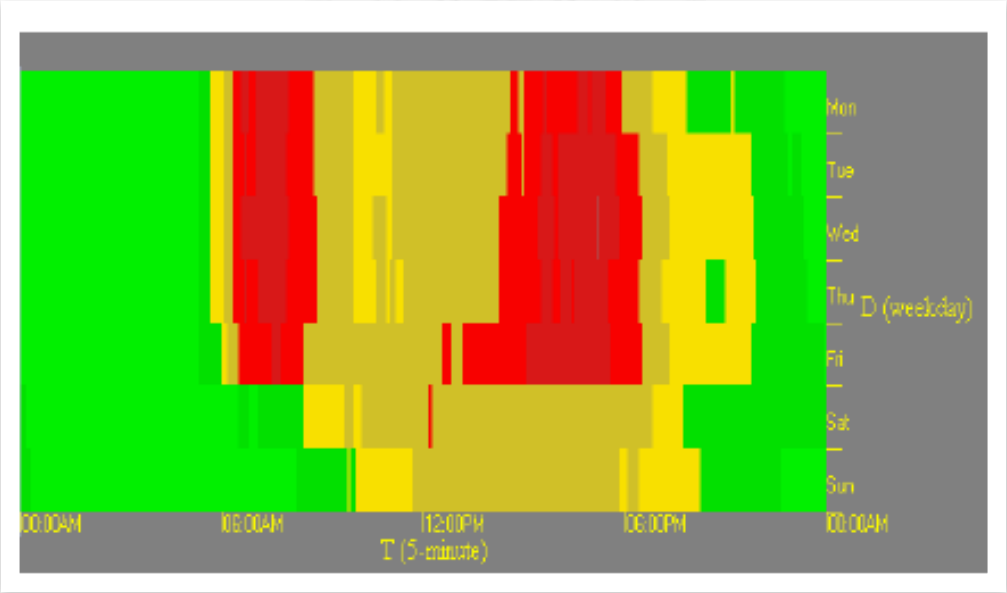


圖 18： Map cube 交通資料展示(Sekhar et al., 2002)

（六）概念階層或網路展示圖

概念階層或是網路架構的展示模式通常是要發掘出節點或是連結的特殊型態，這一類方法必須要先依據資料決定某種階層分類方法或是概

念結構，近年來針對本體論 (ontology)、社群網路、或是消費關連分析等的研究逐漸增加，此類視覺化的方法不但是一個實用的展示介面，同時也是套用這些研究成果去發掘新關連或是架構產生的工具。近年來較著名的軟體如由美國國家科學基金會所資助的研究工具 Codex 以及所獨立發展出來的 ConceptVISTA 軟體等。

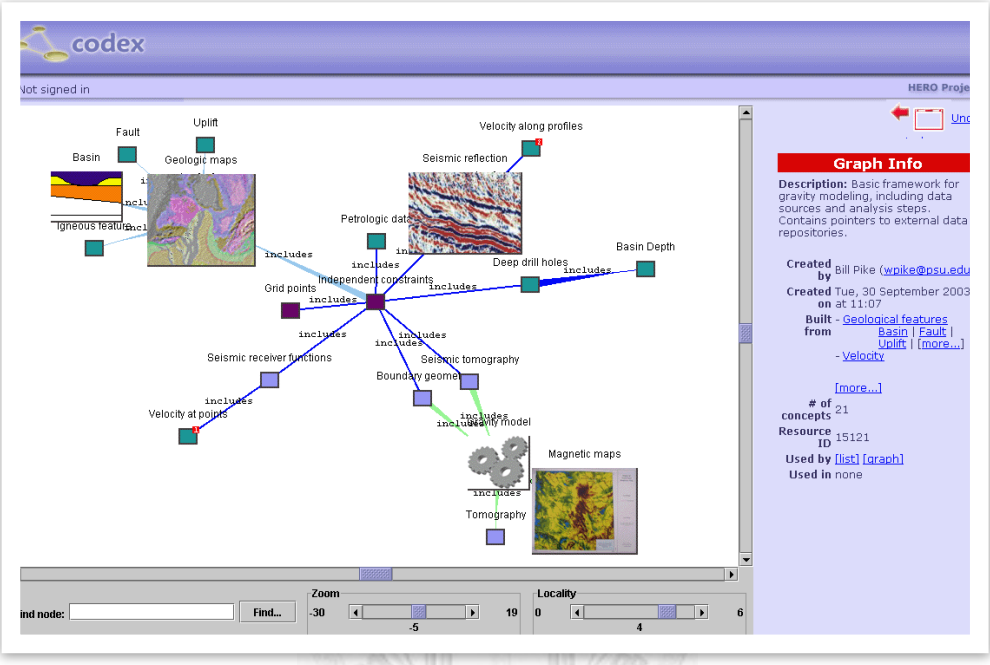


圖 19： Codex 系統畫面(Pike and Gahegan, 2007)

（七）動畫或是聲音

Kyrcier (1994) 最早提出聲音結合地理視覺化的概念，他提出雖然聲音本身雖然並非視覺化，但是研究的證據指出聲音可以作為視覺化的一個有用的輔助，他並提出了視覺化工具的設計應該包括3D、動畫 (時間維度)、即時互動功能及聲音。 Evans在其研究結果當中則發現動畫可以有效的提昇使用者在描繪主題以及對資料的識別能力。 Brauen (2006) 針對互動性的聲音地圖進行研究，提出聲音在地理視覺化當中有諸多潛力及不同的角色可以扮演，但其結果也包含了正反兩面的成果，說明聲音有其優點卻也有其缺失，且近年來關於此領域的研究仍然闕如，其研究最後指出聲音還是人類生活當中重要的一種感官資訊，未來其應用仍存

在著諸多的可能性。 Harrower (2007) 在其針對人類認知與動態地圖的研究中，提出應該分散一部分的視覺資訊轉換為聲音感知，同時動態的資料更需要有效的在空間上或洫他〕W進行整理或是排序。

由以上回顧，本研究發現地理視覺化技術平台提供了一個較傳統地理資訊系統 (GIS) 軟體互動性更高的探索平台。藉由不同視覺化的展示即顏色的安排，輔助研究者更為容易的去發現可能隱藏在資料裡面的關連性以及特徵型態。在空間資料探索的過程當中，因此若是能善加利用資料視覺化技術，應能有助於在後續資料挖掘程序的效率以及改善。因此第三節將針對地理視覺化整合知識探索架構的研究，進行一統合的分析與回顧，同時評估應用地理視覺化技術探索交通資料的適用性。

第三節 地理視覺化與知識探索的整合研究

現行針對地理視覺化以及知識探索的整合，並沒有一完整的理論被提出，過去的討論大多是屬於概念性架構的提出，本研究將針對過去相關的整合概念架構進行回顧，分析過去學者針對視覺化工具與技術所提之概念與實用性的原則，接著針對交通流量資料過去的相關研究進行回顧，最後說明為何地理視覺化探索架構在交通流量資料的探索上具有其進一步研究的潛力。

一、地理視覺化與知識探索的架構整合

Maceachren et al. (1999) 在其研究中針對了地理視覺化技術與KDD的整合進行了初步的探討，並提出了一個概念性的整合架構，將地理視覺化的三個功能要素 (特徵辨認、比較、及闡釋) 與KDD的三個主要任務 (概念架構抽取、類型抽取與分類，以及現象抽取) 加以組合，得到了九個改良的視覺化資料探索方式 (圖20)。研究認為這些視覺化方法都有相同的目標和方法，就是希望能夠找到、關聯、並解釋有趣且有意義或是

非預期的特徵 (物體或是型態)。

GVis Operations \ KDD Operations	concept hierarchy and structure extraction	categories extraction and classification	phenomenon extraction
feature identification	can help an analyst notice levels at which features/ patterns appear, thus the key levels to be modeled in the concept hierarchies	helps to find category bounds, explore their fuzziness, identifies specific data individuals that are prototypical of a category and those that are anomolous	can be used to find exemplars of real world phenomena that match criteria derived in the category extraction operation
feature comparison	can help to relate concept hierarchies for different attributes a step toward creating an appropriate overall multidimensional structure for relating them	facilitates comparison of categories to determine the key attribute dimensions along which they relate or differ, determine which categories are similar	can be used to compare phenomena to derive rules that describe relationships (perhaps causal ones) between them
feature interpretation	facilitates linking an internal version of concept hierarchy and/or concept structure to domain knowledge as part of the process of refining each to fit a particular theme	facilitates the explanation of categories that are extracted in relation to physical parameters of the real world	can be used to help understand real world phenomena by developing abstract characterization of those phenomena and modeling their behaviors

圖 20：地理視覺化技術與視覺化資料探索方式 (Maceachren et al., 1999)

Maceachren et al. (1999) 也指出視覺化技術在 KDD 探索過程中的前、中、後期皆可扮演不同的角色。Gahegan (2009) 則指出人類對於視覺資料的檢視和推論是一個不斷反覆的過程，人類不斷利用這樣的過程中產生一些新的發現，並用這些發現作為繼續誘發新探索的工具 (圖21)。而這樣的能力對於知識的不斷發現及更新具有正面的效果，例如在探索前期對於極端值的反應可以改善中期資料探勘的正確性等。

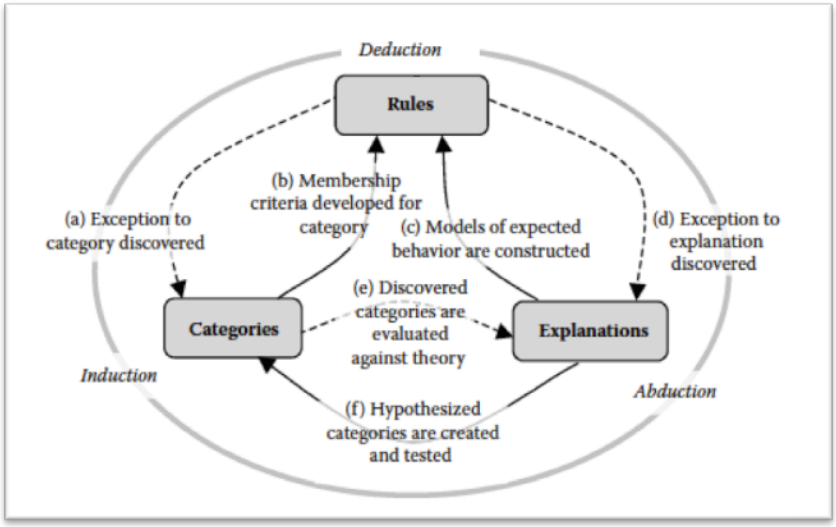


圖 21：視覺化與知識探索之概念結合架構 (Gahegan, 2009)

過去的視覺化研究主要聚焦在資料視覺特性的實體化。例如Keim and Kriegel (1994) 在其研究中利用視覺化工具輔助多資料庫查詢及線上分析。Lee and Ong (1996) 發展 WinViz 軟體，結合多維視覺展示、監督式與非監督性分類方法來進行線上分析以及結果的闡釋。近年來認知科學領域的研究，也提出視覺化資料展示的設計以及互動技術會影響使用者對不同現象特徵的認知(Masters and Edsall, 2000)。

在時空資料探勘裡面另外一個重要的問題就是多維度資料 (Multidimensional Data) 的處理和分析議題。Keim and Kriegel (1994) 利用 Space-filling 技術加上三色套疊的展示處理五維以及更高維的複雜資料集，並以模擬資料說明如何在視覺化介面中進行探索。 Shneiderman (1992) 在他的研究當中利用樹狀地圖 (treemap) 來展示具有階層概念架構的多維資料，說明樹狀圖可以展示階層以及數量之間的相對關係。Nelson (1997) 則利用不同符號代表不同地區的犯罪率及不同的案件類別，並對使用者的反應時間進行了測試和研究。

二、利用地理視覺化探索的實用性原則以及工具整合

目前雖然沒有關於地理視覺化的知識探索的正式理論被提出，但是已經有少數學者開始進行初步理論的架構，Ben Shneiderman 在 1996 年提出的資訊視覺化的操作型定義如下：

(1)綜觀 (overview)。

(2)變焦(zoom)。

(3)過濾(filter)。

(4)取得更細緻資訊(detail-on-demand)。

(5)關連 (relate)。

(6)記錄 (history)。

(7)抽取 (extract)。

Shneiderman所提的操作型定義，實質上已經包含了工具操作以及資訊處理兩個層面。 Lee and Ong (1996) 則是在他們的研究當中進一步提出了互動性 (interaction) 的概念，強調視覺化工具主要的優勢在於即時互動式的視覺搜尋 (visual query) 以及展示 (display) 功能，能幫助領域專家及研究人員對於資料進行橫向或是縱向的篩選 (sift)、聚焦 (focus)、再聚焦 (refocus)、區隔 (segment)等工作。North and Shneiderman (2000) 則提出了同步更新展示(snap-together)的概念，並對統計學家，以及程式設計師進行測試。以上研究大部分著重在於工具的使用與組合上，但是對於知識探索本身，這樣的操作型原則並沒有說明該如何跟知識探索本身建立連結。

Andrienko and Andrienko (2006) 在其所著的專書裡面首次嘗試以一個經驗研究出發的角度，來整理視覺化技術的發展對於知識探索所帶來的影響。作者並整理出不同種類的視覺化工具可以對哪些知識探索任務帶來一些幫助。雖然這些整理來自於作者本身對於過去所做研究的經驗結論，但是這些經驗研究當中所用到的探索任務和需求，已經囊括許多知識探索的一般性需求。作者在其著述當中整理出知識探索裡面重要的探索任務 (task) ，以及整理出視覺化工具 (tools) 的不同類別。但是在書中作者提出他的經驗研究裡面並無法將所有任務以及工具利用相互參照 (cross-refererce) 的方式羅列，其原因是因為關於知識探索的任務具有其特定性，因要列舉出所有方法有技術上的困難，另外一個原因則是作者在其書中對於任務的分類範疇屬於一般性的分類，因此也無法對應到是用目的較為明確的工具上。基於上述的原因，作者並未提出相互參照的工具使用說明，而是歸納十個視覺化知識探索的基本原則如下：

(1)觀察全體 (See the whole)。

(2)簡化以及摘要(Simplify an abstract)。

(3)分解與重組(Divide and Group)。

(4)觀察關係 (See in relation)。

(5)尋找可辨識的 (Look for recognizable)。

(6)變焦與聚焦 (Zoom and focus)。

(7)注意特殊性 (Attend to particulars)。

(8)建立連結 (Establish linkages)。

(9)建立架構 (Establish structure)。

(10)加入領域知識 (Involve domain knowledge)。

這些實用性原則是知識探索當中使用者會使用的任務，但是同時這些原則也是視覺化工具使用的引導(guide line)，而根據不同研究的目的，以及資料的特性，研究者可以將這些原則進行操作的組合，而這些實用性原則會引導視覺化的探索去發掘新的知識。

三、交通流量資料研究作為一整合案例的可行性評估

在交通研究領域方面，交通流量的預測與推估算是近年來學者們努力的面向之一，過去已發展出許多不同的交通流量預測模式，用以推估未來可能的交通流量與交通狀況等 (Yin et al., 2002; Vlahogianni et al., 2005) 。近年來交通運輸的相關研究著重在於車流量的預測，希望藉由運算模式的發展，達到即時的車流量推估或是車速或旅行時間的估算。過去學者發展了許多不同種類的模式，例如歷史資料統計式方法、時間序列迴歸模式、類神經網路和細胞自動機等等(Demetsky, 1997; Schadschneider, 2002; Yin et al., 2002; Clark, 2003; Vlahogianni et al., 2005) 。此類研究的目標在於增加車流量的預測能力及準確度，希望達到線上即時預測的效果。近年來人工智慧於交通運輸研究日漸增多，如類神經網路（Neural Network），案例式推理（Case Base Reasoning）、模糊理論（Fuzzy Theory）等，可見交通問題具有對高度複雜以及非線性的系統特性。

過去的這些研究已有不錯成果，但是過去的研究忽略一個主要的因子，就是交通行為也是一種空間行為，但在車流量預測相關研究裡，空間分析相關理論較少被討論及應用，主要原因在於交通行為彼此之間互相影響且相當複雜，加上不同區域的主要影響因子也有所差異，因此要將空間納入交通流量模式裡加以考量有其困難性。 Hai and Fengxiang (1998) 利用類神經網路對於高速公路車流進行分析分級，但是其分類依據主要是資料的特性，而並沒有加入空間的考量。Lingras (2000) 則是利用不同道路的使用分類，將資料分別導入類神經網路裡面去進行短期交通預測。 Turochy (2006) 將鄰近道路路口流量納入時間序列模式來計算，但沒有考量空間特性，而是假設越鄰近的道路彼此之間的影響程度越高。

視覺化技術與地理資訊系統近年來也開始被用在交通研究上，Claramunt et al. (2000) 提出了一個整合時空資料庫、運輸地理資訊系統以及視覺化工具的概念架構，利用互動的介面來幫助使用者在交通資料的視覺化，辨識出不同的交通行為類別。Shekhar et al. (2000) 利用群集分析以及資料視覺化方法，研究雙城市 (twin-cities)的高速公路交通型態，研究結果發現視覺化可以有效幫助研究者了解研究區的資料型態，並且十倔化技術對於極端值的偵測具有發展的潛力。Zhang et al. (2004) 提出全域觀點 (global) 的交通資料觀察角度是有其必要的，在其研究當中利用內插法展現交通流量的三維暈渲圖以及等值線圖，並在圖面上指出了交通壅塞點 (hot spot) 的位置，但是其研究中也說明內差的交通資料表現方法並不能有效的反映實際在路網上移動的交通行為。Yudong et al. (2006) 的研究當中利用了自組織映射圖網路 (SOM) 來分類研究區的短期車流量資料，並且利用視覺化的展示來分析其時間序列上的趨勢以及規律。

由以上回顧，可以發現傳統交通流量模式裡，較少對於整體空間現象的觀察與描述，大部分研究聚焦在路口的流量預測分析。少數考慮整體資料以及空間因子的研究，也沒有提出具有空間分布、區域性等空間

特徵的相關研究結果。本研究認為在這個面相上，地理視覺化的知識探索有其潛在的研究價值，交通流量資料本身也具有連續性、週期性等時間特徵，因此如果能夠發現與交通流量相關的時間與空間知識，將可以用來強化原有交通流量理論的假設，補強目前交通流量理論在空間知識的不足之處。

第四節 小結

根據以上文獻回顧，本研究整理目前地理視覺化的知識探索研究，得到以下四點結論；

一、傳統知識探索架構及資料探勘架構的問題

本研究整理相關知識探索架構與資料探勘方法的文獻，可知傳統知識探索架構已經漸漸無法處理複雜及大量的資訊，尤其是在資料的前處理流程，目前比過去更需要領域知識的指導，去幫助研究者過濾以及轉換資料。而過去在這些面相當中，較少以使用者為中心的研究觀點，去定義或是指導研究者及其領域知識在這些過程當中所應當扮演的角色。而視覺感知能力是人類最具有優勢的探索工具，視覺影像它提供研究者更直接的連結到腦中的概念架構。過去視覺化在資料探勘的工作及任務裡面，較少被討論以及應用。

二、視覺化工具的研究具有其理論基礎以及成功的要素

視覺化的理論基礎探討顯示它是輔助人類知識探索最重要的工具之一，過去傳統地圖學建立了許多典範，許多是承襲自我們過去所建立的對地球的概念與關係的描述。但在這些典範常常必須建立在資料是完整以及經過適當處理的基礎上。但雖然這是一個架構良好的典範，但就如同 Gahegan (2009) 所提出的觀點，GIS 並不是唯一可以使用的技術，我們

還有另外的角度可以來觀察世界 。溯因法是根據觀察而產生由推理到最佳解釋的過程，若是如同本研究所假設，研究者看到的比其所知道的還多，則提供多樣的角度來觀察資料將有助於推理或是假設的產生。再者，地理視覺化分析的潛力的原因在於，它並非是單純的展示工具或是電腦模式，它真正依賴的是專家的領域知識。如同Andrienko and Andrienko (2006) 在其研究當中所提，在資料探索裡面最具支配性的工具是人腦，視覺化只是一種輔助，但是做為一個輔助工具，視覺化似乎是裡面最為重要的方法之一。

三、整合性的地理視覺化知識探索架構需要被提出

過去針對視覺化與知識探索的整合研究多屬於概念性以及實用性的原則整理，過去研究專注於某些特定主題的探索，較沒有辦法以一般性的原則來建構視覺化探索的標準過程。另外針對視覺化探索後所抽取出之新的空間型態 (spatial pattern) 或是資料型態(data pattern)，過去也較少討論該如何利用這些特徵來對研究資料進行進一步的處理或是闡釋。關於空間及資料特徵如何被外顯化，進而改變資料分析結果的過程仍然缺少相關的研究提出，因此一個整合性的地理視覺化知識探索架構及流程有必要在近一步的研究及提出，以期未來能夠將此架構更加的完備。

四、利用空間角度來思考交通流量資料知識探索的可能性

在過去交通車流模式研究當中，較為缺乏對於整體空間現象的觀察與描述，但是交通行為也是一種時空行為，其資料特徵非常適合用來進行地理視覺化探索架構的建立與操作。因此本研究希望整合上述研究，提出一個新的整合地理視覺化之知識探索架構，以交通流量資料作為案例討論對象，希望透過方法的整合與改進，探索資料當中隱含的特徵，將來可以形成新的空間理論假設，提供資料在空間的差異性描述，將有助於交通資料在空間應用面相的加值。

第三章 研究方法

本研究整理文獻回顧，認為一個地理視覺化整合的操作型方法，必須針對過去傳統知識探索架構的缺失加以改善。因此採用以使用者為中心的視覺化探索概念，並將過去研究當中所發展的視覺化及知識探索原則加以整合，開發出一個探索隱性知識的地理視覺化環境，用以改進過去傳統方法當中，對於資料的前處理以及領域知識的套用不足的問題，同時被探索出來的隱性知識，本研究也利用地理研究者對於空間資料處理的需求加以整理歸類。這些歸納整理的結果，將可以用來把原始資料進行進一步精鍊或是分類，再套用資料挖掘模式的分析，建構出一個地理視覺化整合的知識探索架構。如此不但對於資料本身的理解以及分類會更為細緻，所得到資料探勘結果也將會比未經過視覺化處理的資料優異。視覺化研究方法要達到這樣的目標，首先必須要能夠在原始資料當中將隱含的資料或是空間特徵外顯化，此部份也是本研究的核心所在。本研究所提之地理視覺化整合架構如圖 22。

基於以上目標，本研究套用 Andrienko and Andrienko (2006) 所提出之視覺化知識探索的基本原則，並將其中可用以改善過去視覺化方法或是流程的原則加以重新組合，並以視覺化當中目前較受到研究者重視的 SOM 視覺化分析演算，以及其根據空間資料所改進的空間自組織演算法，作為本研究視覺探索的主要演算法。本研究並將視覺化知識探索的基本原則當中的「分解與重組 (divide and groupe) 」原則套用，利用本研究也將階層式群聚演算法 (clustering analysis) 與 SOM 相關演算法整合，提出「先分解再重新聚類」的演算策略， 再配合其他相關視覺化圖表展示，增加地理視覺化探索分析成功的機會，其研究方法流程如圖 23。

在研究對象上，本研究認為車流量的分析研究，過去缺少地理空間知識的探索面向，若能引入地理視覺化探索分析，應可以改善過去交通流量研究僅聚焦在道路路口的微觀角度，提供一個比較巨觀的、空間觀

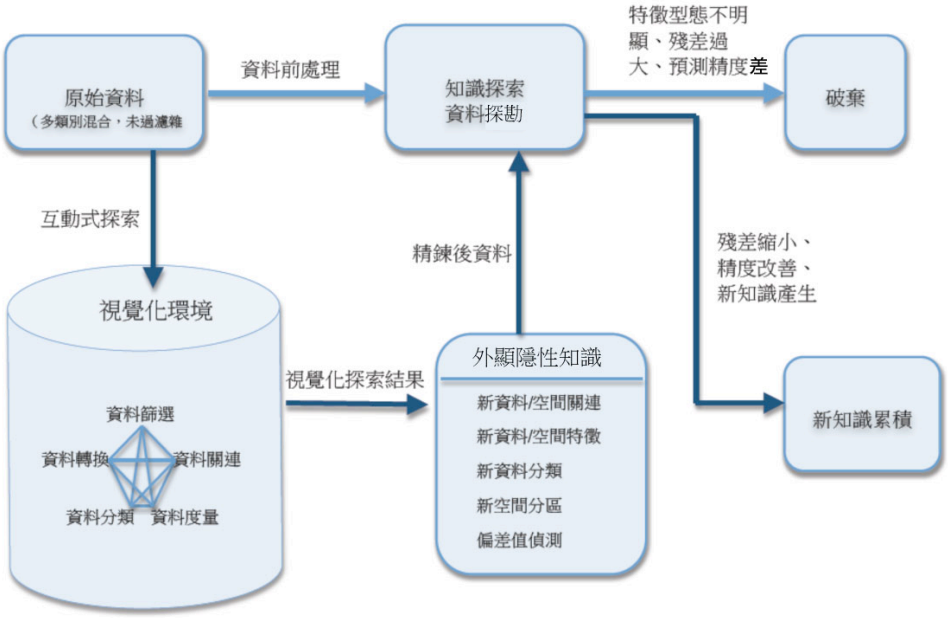


圖 22： 本研究之地理視覺化整合的研究架構

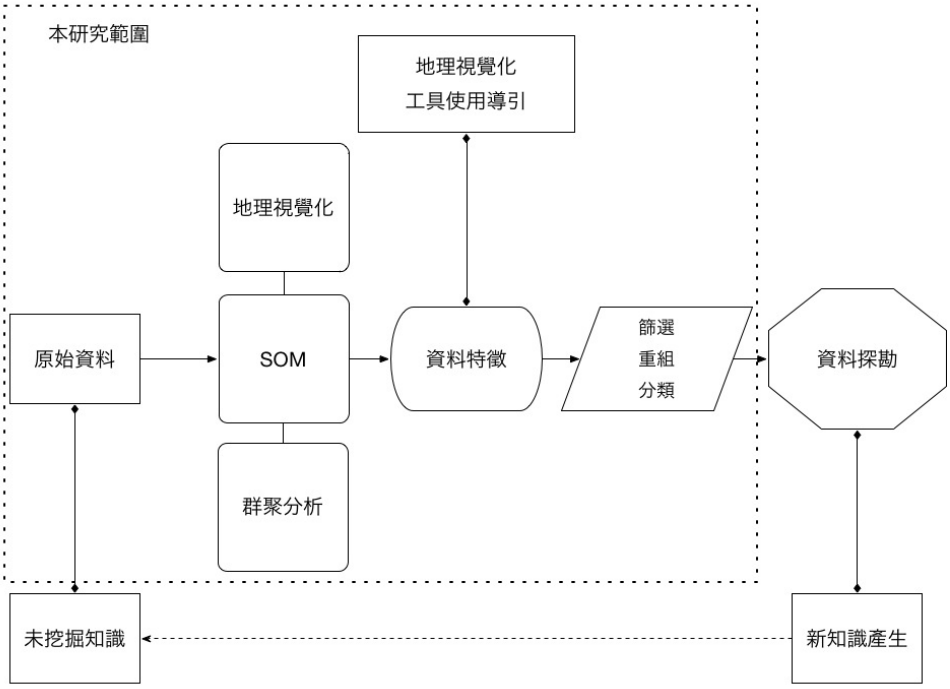


圖 23： 研究方法流程

察的交通流量特性描述，進而改進現有模式的分析效果。因此，本研究嘗試發展車流量資料的地理視覺化探索分析，並使用台北市的交通流量資料進行探討，一方面可以充實地理視覺化探索分析的相關研究案例，另一方面可補強傳統交通流量研究在空間分析上的不足。但本研究的主要目的並非是提出一新的交通流量空間理論，而是透過地理視覺化技術，找出交通資料當中隱含的資料關連與空間特徵，透過這些新的發現說明地理視覺化的有效性以及交通流量的空間分析面向有其必要，而這些新的發現也可以作為未來改善相關理論的基礎。

一、以交通流量為對象的案例探索

本研究擬使用交通車流量資料作為知識探索及模式建構驗證之主要對象，其原因有以下幾點；

(1).車流量資料為具空間分佈之時間序列資料，不論是在時間面向或是空間面向皆適合以不同的尺度去進行知識探索，且其數值的資料特性適合以視覺化的方法呈現。

(2).關於車流量的理論與研究在過去皆有大量的文獻及研究成果，提供了時空知識探索相當數量的研究基礎。

(3).在過去的車流量研究理論當中，缺少了空間關係引入之研究面向，空間知識技術的導入有對交通資料加值的潛力。

本研究已取得台北市交通管制工程處委託民間業者調查之路口交通量資料，包括各路口單位時間內之各種車輛之15分鐘為單位之車流數量及轉向資料，車輛種類主要包含大客車、小客車、和機車三大類，路口行為分為直行、左轉及右轉，同時記錄了道路車道數、快慢車道、日期、天氣及號誌時相等。另外由交通部運研所所出版之新世紀道路路網圖當中可取得相關路口之座標位置。

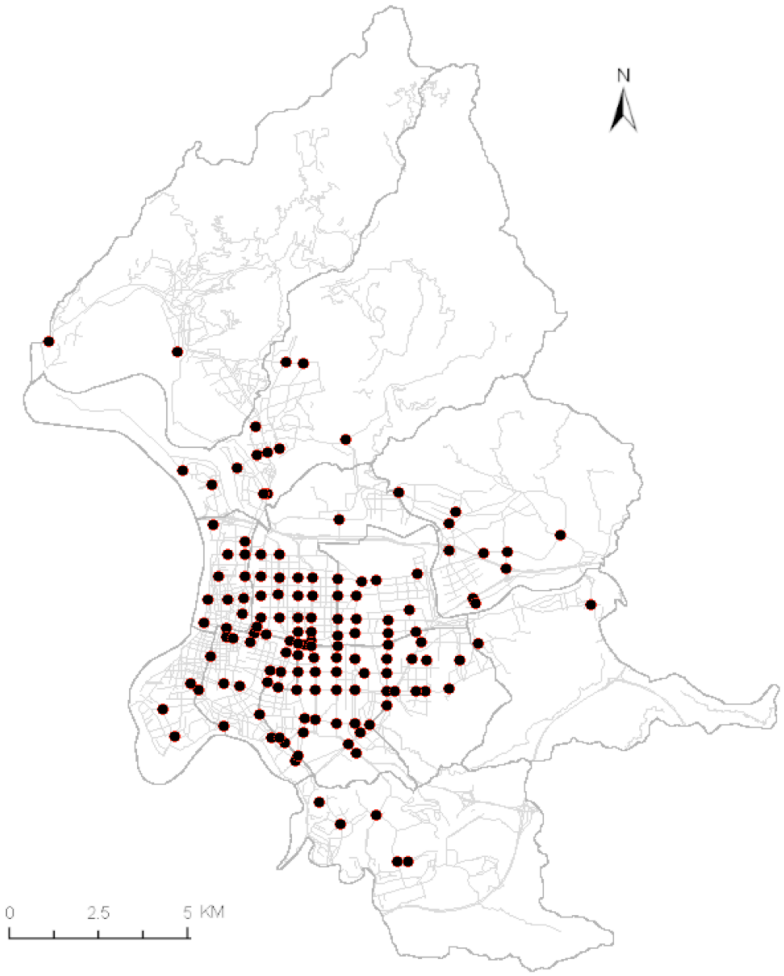


圖 24：路口調查點分佈示意圖

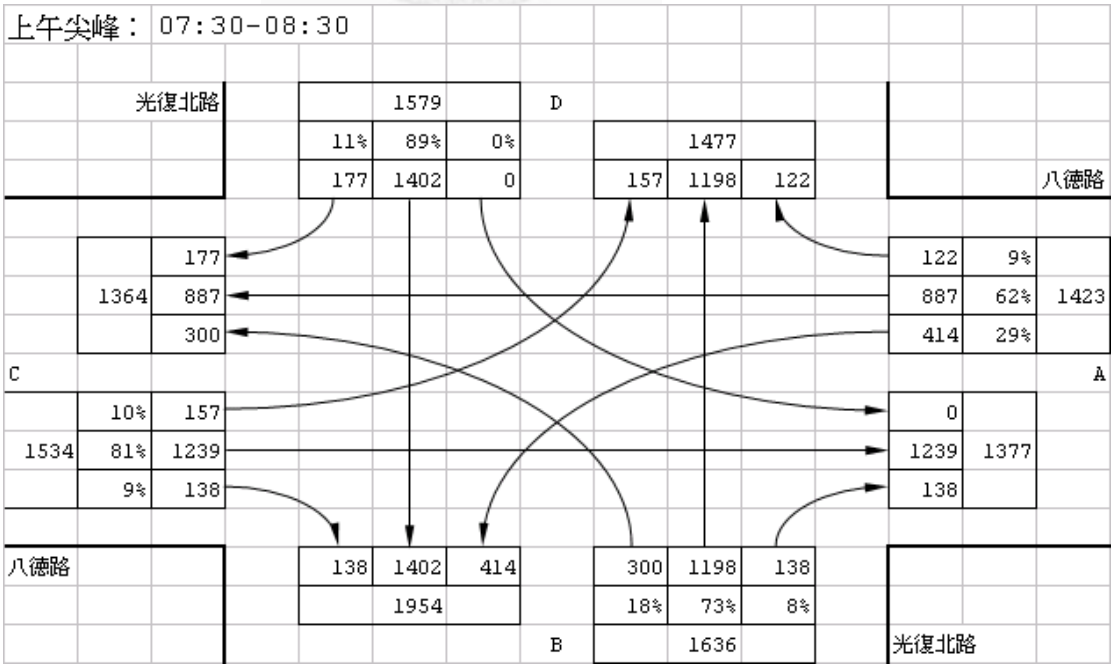


圖 25：路口轉向調查示意圖

Four90資料.xlsx - Microsoft Excel

	G	H	I	J	K	L	M	N	O	P	Q
2	TD2	Weather	Weather2	SC0715A	SC0730A	SC0745A	SC0800A	SC0815A	SC0830A	SC0845A	SC0900A
3	2	1	1	113	192	190	213	313	230	321	245
4	2	2	2	4	5	8	4	8	11	18	7
5	3	2	2	431	133	127	117	203	168	78	142
6	2	2	2	40	78	121	112	82	102	87	82
7	2	1	1	85	177	146	201	144	114	73	78
8	3	2	2	115	163	278	261	414	254	320	313
9	2	1	1	23	32	34	62	50	92	41	91
10	2	1	1	167	178	153	173	212	154	205	135
11	2	1	1	146	196	274	337	247	338	301	294
12	2	1	1	103	219	222	404	441	448	283	285
13	2	1	1	87	59	106	92	154	130	150	93
14	2	1	1	102	160	182	255	246	232	274	253
15	2	1	1	200	374	148	428	292	219	122	194
16	2	1	1	294	295	296	383	496	442	458	476
17	2	1	1	170	292	293	256	389	413	383	409
18	3	1	1	66	149	204	190	240	276	163	322
19	2	1	1	64	71	118	145	142	184	154	185
20	4	1	1	23	45	69	60	69	80	51	51
21	2	1	1	94	107	221	196	296	289	340	254
22	2	1	1	39	36	42	80	96	90	99	97
23	2	1	1	1	3	1	3	2	2	4	1
24	2	1	1	47	91	73	124	111	168	132	204
25	2	1	1	105	114	126	157	247	192	198	187
26	2	1	1	98	90	102	188	263	319	225	213
27	2	1	1	273	289	314	331	352	377	358	327
28	3	2	2	45	56	53	57	85	108	135	92
29	3	2	2	0	0	0	0	0	0	0	0
30	2	1	1	101	162	213	242	256	276	271	277
31	2	1	1	239	397	358	491	418	331	411	346
32	2	1	1	192	208	300	313	368	569	423	417

原始資料 / 方向別車種別轉向分析 / 轉向前後車流資料 / 方向別車種比率 / 方向別轉向分析PCU / 方向別轉向量

圖 26：路口轉向車流原始資料

由於此資料在時間尺度上為每15分鐘的車流量加總資料，本研究探索的目標在於資料的整體特性，因此根據 Andrienko and Andrienko (2006) 所提出簡化以及摘要的原則，將資料整合加總為每小時，而在轉向資料方面本研究也將轉向後的車流進行統計。

本研究嘗試以空間分析的觀點進行交通流量的地理視覺化探索分析，由於尖峰時間車流量大多以通勤為主，因此通勤行為應具有空間的相關性以及差異性，期待能藉由地理視覺化探索分析找出交通流量在尖峰時間的空間特性。

本研究使用台北市交通管制工程處委託民間業者調查的路口交通流量資料進行地理視覺化探索分析，由於每年的調查路口數量及位置皆不盡相同，因此本研究採用路口調查數量及重複調查路口較多的2000年到2001年的調查統計資料，每年約調查一百五十至兩百點不等。調查的主要時間區段為上下班的尖峰時間 (7:00-9:00、17:00-19:00) ，每年取非假日當中之一日作為調查日。由於方向性是觀察的重要資料特徵之一，因

此本研究剃除掉十字路口以外的資料，僅取四個方向皆有車流的路口作為研究對象 (圖24) ，調查點的座標位置由交通部運研所出版的新世紀道路路網圖取得。

雖然此交通量數據的時間區段記錄較少，但是在方向的紀錄上屬於較為完整的人工調查資料，十字路口的四個方向皆有資料紀錄。上午尖峰時間由東向西的交通流量與晚上尖峰時間的西向東對象車流量進行比較與分析，如此早晚兩個尖峰時段可以產生四組的對應關係，因此本研究將針對此四組關係去進行地理視覺化探索分析。此交通流量資料包括各路口單位時間內各種車輛以15分鐘為單位的車流數量及轉向資料，車輛種類包含： (1) 大客車； (2) 小客車； (3) 機車等三大類，但各種車輛種類的空間特性也不盡相同，因此本研究僅針對「小客車」及「機車」車流流量資料進行地理視覺化探索分析。將15分鐘的資料進行前置處理，彙整為以小時為單位，並將車道數目納入求出平均車道流量，以進行地理視覺化探索分析。

二、利用探索原則改善 SOM 以及相關演算法

本研究在地理視覺化探索分析的方法採用近年來受到地理視覺化探索分析重視與採用的 SOM 以及利用空間演算改進的 Geo-SOM 演算為核心演算法。 針對 SOM 相關演算法過去研究中較少被提到的分類數量的問題，本研究也將階層式群聚演算法 (clustering analysis) 與 SOM 相關演算法整合，提出「先分解再重新聚類」的演算策略，配合視覺化探索，增加地理視覺化探索分析成功的機會。在一般知識探索的研究當中，當資料已經進行了初步的前處理以及篩選後、接著就是導入資料挖掘的技術進行料分析，本研究採用自組織映射圖網路相關演算以及群聚分析作為此交通流量視覺化分析之主要工具，其主要原因如下：

1. 自組織映射圖網路可將多維的資料壓縮投射至二維或是三維的平面

上，較容易利用視覺化工具展示，對於複雜的屬性資料可以有效壓縮其資訊量。

2. 自組織映射圖網路的本質可以說是一種非監督性分類方法，因此非常適合用來觀察資料的聚集關係，透過與地圖動態的展示連結可以用來探索特殊的空間型態。
3. 自組織映射圖網路是利用競爭學習法，每一個神經元並不會被強迫分配到資料，因此具有相同特性的資料可以有效的形成聚類。
4. 自組織映射圖網路可有效的將資料聚類，過去應用在其他領域例如商業資料的分類上有不錯的成果。

SOM 採用競爭式的學習法則，其完整詳細的數學公式可以參考其原始演算法 (Kohonen, 1998) ，本文僅針對演算法的流程步驟進行說明：

1. 初始化：首先將神經元的鏈結值向量以隨機方式設定，若有 N 個神經元個數便要設定N個鏈結值向量之初始值。
2. 輸入資料向量：根據資料的排序或是隨機順序，依序輸入網路之中。
3. 尋找優勝神經元：將輸入向量與神經元特徵向量當中，找出歐幾里德距離最小者，作為優勝神經元 (best matching unit, BMU) 。
4. 調整神經元的鏈結值： SOM 演算法找出優勝神經元後，會配合學習率參數調整優勝神經元鄰近區域的神經元鏈結值向量，此學習函數和鄰近區域都是隨時間遞減的函數，會隨著演算法執行的過程中逐漸遞減以達到微調的目的。
5. 重複演算：將資料向量不斷的重複輸入至網路中進行調整，直到達到預設的演算循環 (本研究預設值為 4,000 次) 或是設定的收斂值後完成網路。

Geo-SOM 演算法 (Bação et al., 2005) ，其演算原則與傳統 SOM 演算類似，其最大的不同在於尋找優勝神經元的方法，在尋找的過程當中

只有地理空間分布較為靠近的神經元會被納入進行特徵向量的比較，其演算法的流程步驟如下：

1. 初始化：首先將神經元的鏈結值向量以隨機方式設定，若有 N 個神經元個數便要設定N個鏈結值向量之初始值，同時每個神經元還有另一組空間向量，記錄每個神經元的空間位置。
2. 輸入資料特徵向量：根據資料的排序或是隨機順序，依序輸入網路之中。
1. 尋找地理優勝神經元：在輸入向量當中，找出空間向量距離在容許值 t 的範圍之內的神經元，作為地理優勝神經元 (geographic best matching units) 。
1. 在地理優勝神經元當中，找出輸入向量與神經元向量歐幾里德距離最小者，作為最後的優勝神經元 (best matching unit, BMU) 。
2. 調整神經元的鏈結值： 找出優勝神經元後，配合學習率參數調整優勝神經元以及鄰近區域的地理優勝神經元的空間及資料向量鏈結值，此學習函數和鄰近區域都是隨時間遞減的函數，會隨著演算法執行的過程中逐漸遞減以達到微調的目的。
3. 重複演算：將資料向量不斷的重複輸入至網路中進行調整，直到達到預設的演算循環 (本研究預設值為 5,000 次) 或是設定的收斂值後完成網路。

SOM 相關演算應用在地理視覺化的研究面向中，目前所存在最大的問題就是網路的拓樸類型以及大小問題，二維排列的 SOM 演算法一般採用對稱的四角形網路及六角形網路兩種，目前以六角形拓樸網路較常被採用。 SOM 網路大小的採用上可以分為三種類型 (Bação et al., 2008)：大型網路、中型網路與小型網路。當分群目標較為確定或已知的時候，研究者則可以參考需求選擇網路的大小，例如 Thill et al. (2008) 利用小

型的 SOM 來做美國南方語言分類的研究。若是研究對象分類未知或是必須要探索時，如何設定網路大小則是需進一步探討的議題。由於究竟應採用多少類別的分類並非絕對，因此目前相關研究關於 SOM 網路大小的訂定多未著墨 (Koua and Kraak, 2004; Koua et al., 2006; Guo, 2009) 。

因此本研究提出「先分解再重新聚類」的方式，先利用資料聚類能力較為優異的 SOM 以及 Geo-SOM 演算法將資料以中型網路分為較小的群體，再導入運算成本較小的階層式群聚演算法進行資料的再聚類，利用逐步縮小聚類數量的方式，並藉由地圖或是其他視覺化圖表探索空間資料的關連性、空間分布或是趨勢等。對於資料的分類數量，本方法也不預設特定數量，而是依據Koua et al. (2006) 採用的互動式探索模式，讓使用者進行動態的探索來決定對其探索目標最有意義的分類方式。

本研究基於以上因素自行撰寫SOM類神經網路演算法及階層式群聚演算法，利用不同的SOM網路大小，進行視覺化的互動式探索。而由於過去缺少相關演算法結合之研究，因此本研究將先行就主要的階層式群聚演算法進行一適用性的評估後，再選定其中一種最適合的演算法進行地理視覺化的分析探索工作。

過去常用的階層式群聚演算法有中心法(Centroid Method)、單一連結法(single-linkage clustering method, SLINK) 以及完整連結法(complete-linkage clustering method, CLINK) 及華德法(Ward's clustering method) 等四種主要群聚分析演算法，其詳細之演算法公式說明如下：

1.中心法 (centroid method)，也稱平均法 (average-linkage clustering method) ，其定義為群聚C_i與Cj的距離為兩集合之中點與點之間距離總和的平均，其公式如下：

$$D_{\text{ave}}(C_{\text{i}},C_{\text{j}}) = \frac{1}{|C_{\text{i}}||C_{\text{j}}|}\sum_{a\in C_{\text{i}},b\in C_{\text{j}}} d(a,b) \qquad \text{(式1)}$$

2.單一連結法 (single-linkage clustering method, SLINK)，其定義為

群聚C_i與Cj的距離為兩集合之中最近兩點之間的距離，其公式如下：

$$D_{\min}(C_i,C_j)=\min_{a\in C_i,b\in C_j} d(a,b) \quad \text{……(式2)}$$

3.完整連結法 (complete-linkage clustering method, CLINK)，其定義為群聚C_i與Cj的距離為兩集合之中最遠兩點之間的距離，其公式如下：

$$D_{\max}(C_i,C_j)=\max_{a\in C_i,b\in C_j} d(a,b) \quad \text{……(式3)}$$

4.華德法 (Ward's clustering method) ，其定義為計算群聚C_i與C_j合併後資料a的總變異數，來作為兩集合之間的距離，此方法需要先計算其兩群聚合併後的算數平均中心 μ，其公式如下：

$$D_{\text{ward}}(C_i,C_j)=\sum_{a\in C_i\cup C_j}\left\|a-\mu_{i,j}\right\|^2 \quad \text{……(式4)}$$

三、視覺化工具的選用以及探索環境的整合

本研究針對過去發展之工具進行回顧後發現並沒有一單一軟體可以完整支援本研究所提之研究架構之需求，但透過工具整合的方式可以達到本研究所要表達之研究目標，因此本研究自行利用Visual Basic程式語言開發地理視覺化探索工具。將散布圖、自組織映射地圖網路、空間地圖等工具整合至一地理視覺化分析系統，並加入群聚分析模組，利用其演算速度較快的特性，配合動態展示達到較有效率的半自動化視覺探索，來搜尋可能的空間特徵或是資料特徵。

由於過去尚未發展地理視覺化探索的相關確定理論，因此視覺化工具使用的知識累積尚不充足，因此在初期階段採用人機互動的模式，須等到知識的累積到達一定的程度才能加以自動化以及人工確認輔助的功能。但必須要說明的是，本研究的視覺化探索，並非是以盲目無目的的搜尋的方式進行，而是希望利用視覺化的輔助，幫助研究者定位出具有空間關連性的資料位置，以及利用資料特徵外顯化的技術，幫助研究者

找出特定資料在空間上所隱含的特徵關係或是意涵。因此針對不同的主題，利用本研究架構來協助其發現隱含在資料中的空間特徵關係，以及幫助研究者解釋出在資料之間的關聯性。

本研究採用的 SOM 以及 Geo-SOM 演算法，兩者最大的差別在於後者再聚類的過程中將空間的接近性視為判定的條件，因此以「先分解再重新聚類」的方法將其再度組合，前者的重點將著重在經過重組的神經元是否可以在資料面上顯示出有意義的特徵型態，而後者的重點則是在代表資料特徵以及空間位置皆相近的神經元，在經過重組之後，可否在空間上顯示出有意義的空間特徵型態。因此本研究將根據此兩種不同的探索策略，針對所收集到的交通資料進行探索，同時針對探索前後資料的特徵分類或是空間型態進行比較與分析。本研究將設計兩個案例來進行視覺化的探索及分析，首先先利用傳統 SOM 演算法與群聚演算法結合，針對尖峰時間的對向車道的汽車流量，進行相關性的視覺化探索，同時比較視覺化技術導入之後，在相關係數以及判定係數等資料特徵描述上的改進。接著第二個案例探索研究區內的機車車流量，是否在空間上具有其特殊的分佈或是區塊特性，同時利用平行座標圖 (PCP) 的展示方式，來對多維度資料進行視覺化。而探索的主要目的是分析視覺化技術導入之後，是否在空間的分區以及資料分布的辨識上，可以達到輔助的效果。

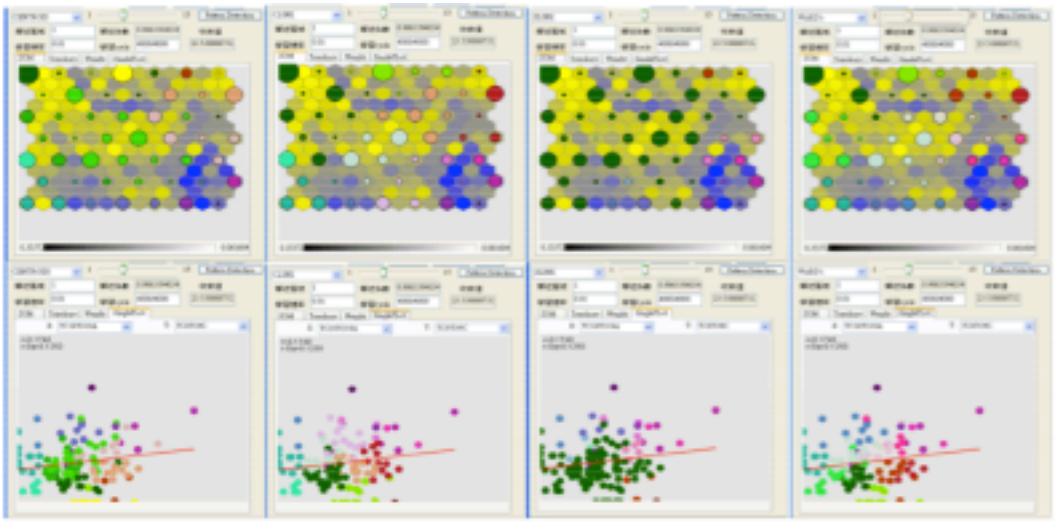
第四章 研究結果與討論

第一節　傳統 SOM 演算法與群聚演算法的結合

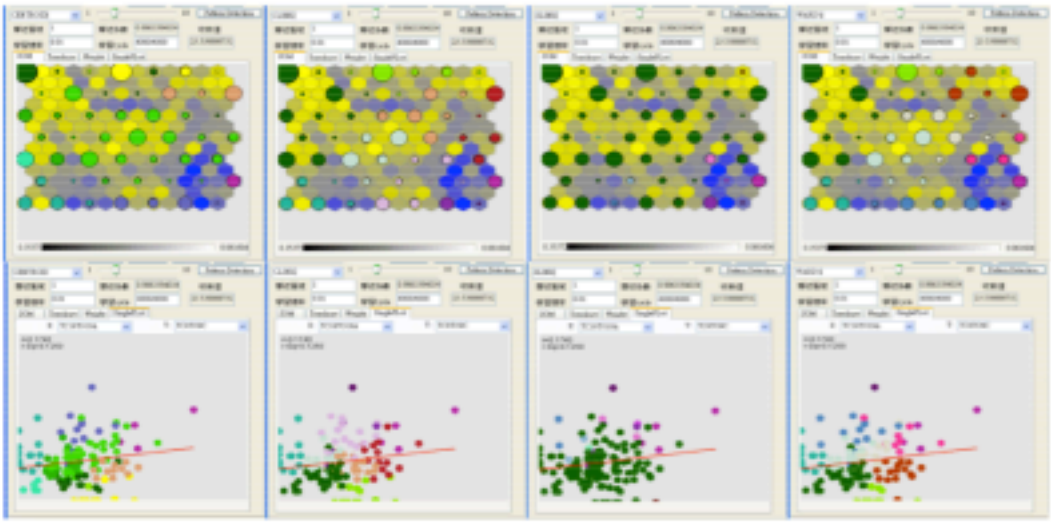
一、階層式群聚演算法的先期評估

本研究針對民國 89 與 90 年所調查早晨尖峰東西向以及晚間西向東各兩個時段的車流量資料，作為 SOM 的輸入向量。 針對研究區資料，本研究初步先針對四種不同的群聚演算法進行視覺化的評估，首先利用7 × 7的中型 SOM 網路，學習函數初始值為 0.1（不小於0.01），預設鄰近半徑為 4 ，採用高斯函數，經計算後將資料細分為具有類似特性的49個群聚，再利用四種演算法進行重組聚類，本研究檢視合併後的群駐數量 n 在 15 、 10 、以及 5 的時候的分類結果，發現除了 Ward's 法以外，其他的三種演算法在合併的過程當中，會產生群組成員過度集中的問題（圖 27），而 Ward's 法在這三種分類情況下，其分配較其他三種演算法來的平均。

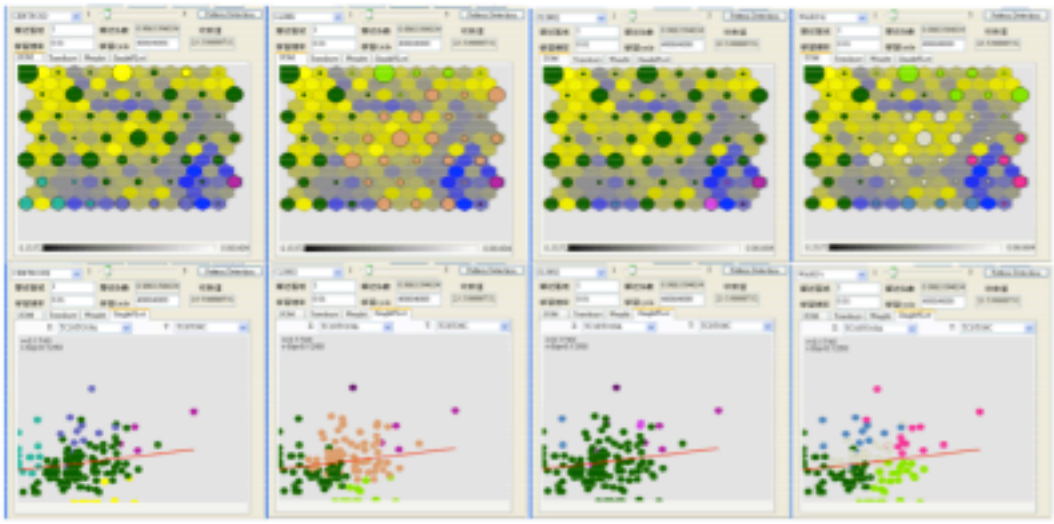
在四種演算法當中，可以發現 CLINK 方法的分類結果在資料的分配上最不平均，此結果雖然不利於「先分解再重新聚類的演算策略，但是在極端值的辨識上卻具有較好的效果，其完整連結的特性可以在群集合併的初期便找到具有較大變異性的個體或是小群集。因此若是分析的目標在於找到極端值或是特殊型態的資料，則應可採用此法，但由於本研究之「先分解再重新聚類」的演算策略，其一部分工作必須依賴探索者本身的視覺觀察，在空間展示或是資料展示上面去辨識有意義的特徵，在本研究所實驗的案例當中，Ward's 法演算所提供的結果，在視覺探索上較為其他三種演算法保留了較大的資料分析以及重組的彈性。因此 SOM 網路的合併分析將以 Ward's 法作為主要的合併演算法。



(a) n=15時的U-Matrix與散布圖（由左至右依序為Centroid、Slink、Clink及Ward's群集演算法）



(b) n=10時的U-Matrix與散布圖（由左至右依序為Centroid、Clink、Slink及Ward's群集演算法）



(c) n= 5時的U-Matrix與散布圖（由左至右依序為Centroid、Slink、Clink及Ward's群集演算法）

圖 27：四種不同的群聚演算法的分類結果

二、原始車流量的相關分析與簡單回歸分析

以 89 年的資料為例，初步觀察此四組原始資料的散布圖可發現資料分布較為分散，缺乏明顯的趨勢或是資料群聚的特性 (圖 28) 。以四組原始資料進行相關分析與簡單迴歸分析 (表 2) ，發現四組的相關係數都偏低，每一組的相關係數都在0.2以下，顯示這四組資料彼此之間幾乎沒有存在相關性。此外，簡單迴歸的配適程度也不佳，每一組的判定係數都在0.04以下，顯示原始資料進行迴歸分析，會幾乎沒有解釋能力。

由這樣的分析結果顯示，無法支持進一步的推論或是假設，來說明台北市的早晚尖峰時間的車流量具有潛在的相關特性，同時也無法在這樣的原始資料裡找到空間上的差異或是明顯的群聚。

三、四組車流量的地理視覺化探索分析

前述數據與分析結果為傳統統計分析所得出的結果，由於傳統方法並沒有考慮空間差異的因素，因此在分析後得到了無法解釋的結果。本研究提出「先分解再重新聚類」的方法是要利用地理視覺化探索分析，解決過去傳統方法無法考慮資料的空間差異的問題。

先將四個車流量的屬性資料，利用 7 × 7 的中型 SOM 網路，將資料細分為具有類似特性的49個群聚，再利用 Ward’s 法進行逐步的重組聚類，將分到同一群組的神經元用相同的顏色來表示，並在資料散布圖以及地圖展示上進行互動式的探索。利用 Ward’s 法將分組數逐漸減少為13組的情況下，在散布圖上觀察到三種主要的資料特徵類型 (圖 30) ，以及其餘不具明顯特徵的少數資料。

在這三類當中有一種類別屬於單向道路，其空間位置對應為限定東向行駛的仁愛路路段以及限定西向行駛的信義路路段 (圖30 a, b) 。在 SOM 資料歸類初期便可以發現被歸類在鄰近的神經元當中，代表其資料特徵相當一致，並且與其他資料具有較大的區隔性，同時在空間上又具

表 2：89 年原始資料相關係數及判定係數		
尖峰時段	TC17to18C (西向東)	TC18to19C (西向東)
TC7to8A (東向西)	r = 0.1682 R^2 = 0.0283	r = 0.1874 R^2 = 0.0351
TC8to9A (東向西)	r = 0.1855 R^2 = 0.0344	r = 0.1941 R^2 = 0.0377

註：TC 代表平均車道流量，7to8 代表 7-8 時車流 (其餘可類推)，A 代表東向西車流，C 代表西向東車流，r 代表相關係數，R^2 代表迴歸判定係數。

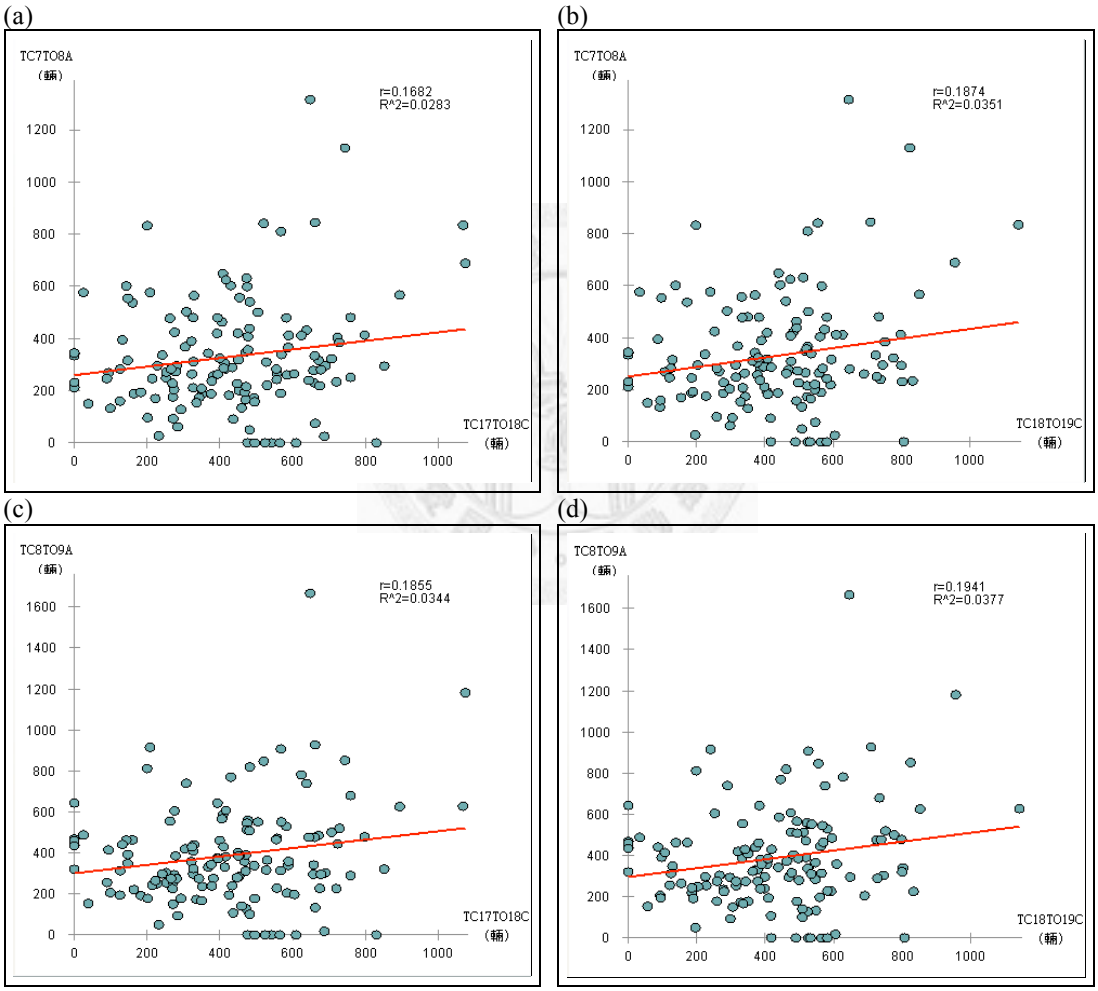


圖 28：89 年原始資料的簡單線性迴歸分析：圖中紅線為原始資料的簡單線性迴歸線，r 為相關係數、R^2為判定係數

有比較明顯的線性型態，因此在視覺化的探索當中，屬於較容易觀察到的資料特徵。此部份由於缺乏對象路段車流，因此對向車流不具有相對關係，若是要考慮簡單迴歸分析則此部份資料不應該納入分析。單行道的資料在本研究的案例裡應屬於被剔除的極端值資料，但是卻不容易由原本的表單資料裡剔除。本研究發現 SOM 演算法對於極端值的偵測有不錯的成效，若是採用中型或是大型的網路拓樸，此類具有空間特徵的極端值資料可有效歸類在單一或是鄰近神經元，可在地理視覺化探索分析的過程裡偵測得出。

將此兩集合分開顯示在 SOM 視覺化展示與地圖上，可以更容易觀察到車流量資料的分類與其空間特徵。觀察集合一的 SOM 視覺化展示 (圖30 c, d) ，發現集合內的U-matrix彼此之間的顏色較深，代表差異較大，且被分配到單一神經元的資料群較多，說明集合一雖然有較高的相關係數和判定係數，但其神經元與神經元之間的差異性較高。觀察集合二的 SOM 視覺化展示 (圖30 e, f) ，發現集合內的U-matrix的顏色較淺，代表資料差異性較小。

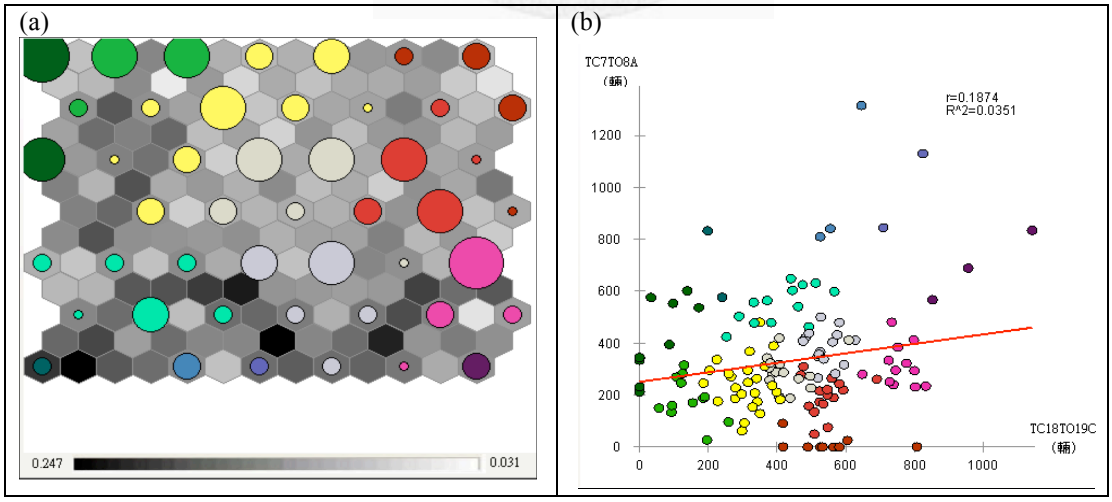


圖 29 : 89 年原始資料資料展示及視覺化處理後之結果：(a)為 SOM 視覺化及 U-Matrix 之計算結果；(b)為經地理視覺化處理過後的三種分類數據散布圖

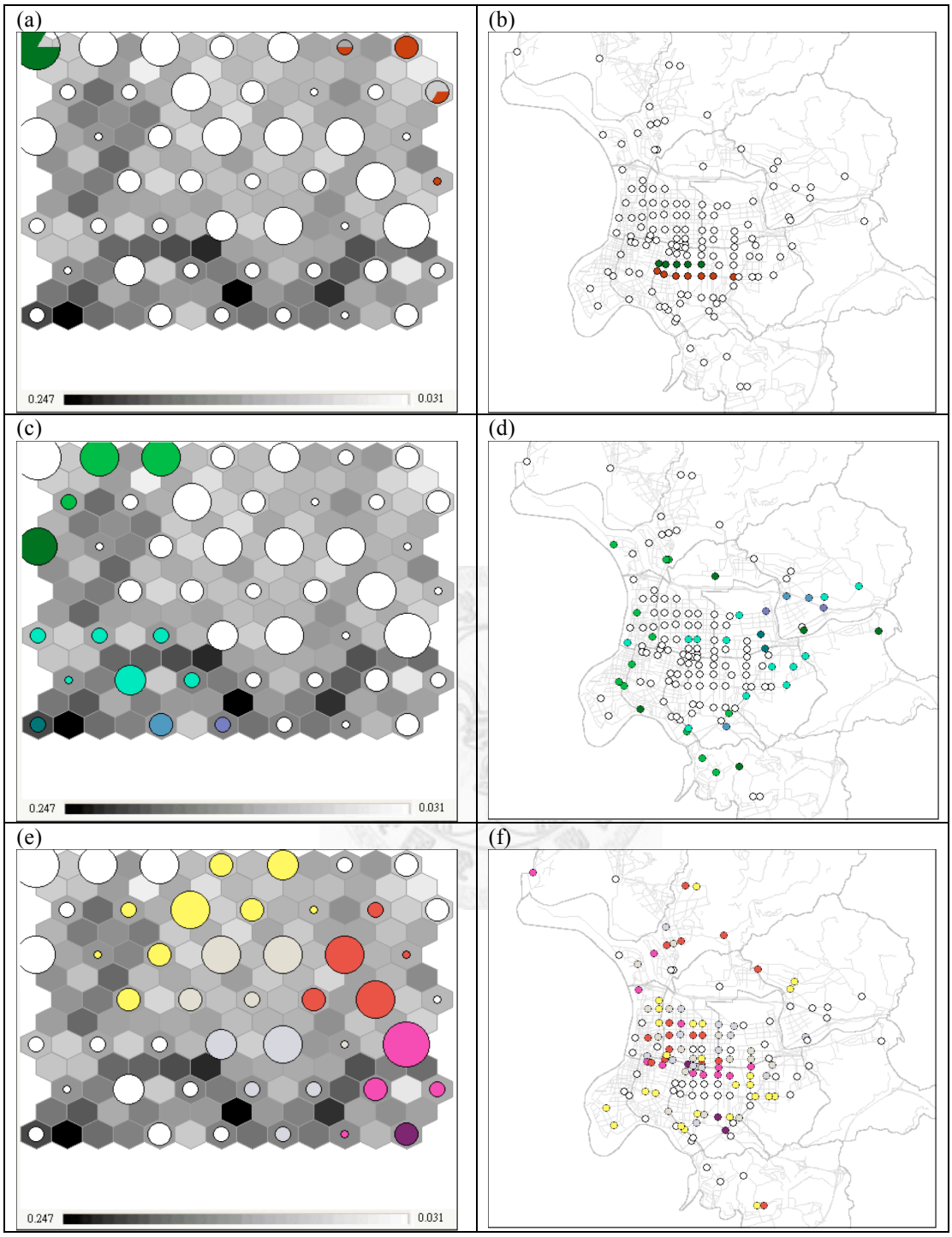


圖 30：89 年經過地理視覺化重新聚類後之 SOM 視覺化及空間分布：圖中(a)、(b)為兩條單向路段；(c)、(d)為集合一；(e)、(f)為集合二

四、集合一的相關分析與簡單迴歸分析

兩集合分別以早晨 (7-8時) 與傍晚尖峰 (17-18與18-19時) 車流量資料進行分析，集合一的相關係數與判定係數與原始資料相較有非常明顯的提昇 (表 3、圖 31) ，最高者為0.75，最低者也有0.68，代表此集合內的早晚尖峰流量具有中至高相關性。此外，大部分早上尖峰時間的車流量大於晚上尖峰時間的對向車流，且呈現正相關的趨勢。簡單迴歸的配適程度與原始資料相較也有非常顯著的提昇，判定係數最高者為0.56，最低者也有0.47。

觀察集合一的空間分布發現，其分布大多位於市中心周邊的道路路口，其車流量分布符合早上入城及傍晚出城的流量較高的特性，車流量有從東向西逐漸遞減的趨勢 (圖30 c, d) 。

表 3：89 年資料分群後資料相關係數及判定係數 (集合一)

尖峰時段	TC17to18C (西向東)	TC18to19C (西向東)
TC7to8A (東向西)	r = 0.7500 R^2 = 0.5626	r = 0.7398 R^2 = 0.5473
TC8to9A (東向西)	r = 0.7057 R^2 = 0.4980	r = 0.6875 R^2 = 0.4727

註：TC 代表平均車道流量，7to8 代表 7-8 時車流 (其餘可類推)，A 代表東向西車流，C 代表西向東車流，r 代表相關係數，R^2 代表迴歸判定係數。

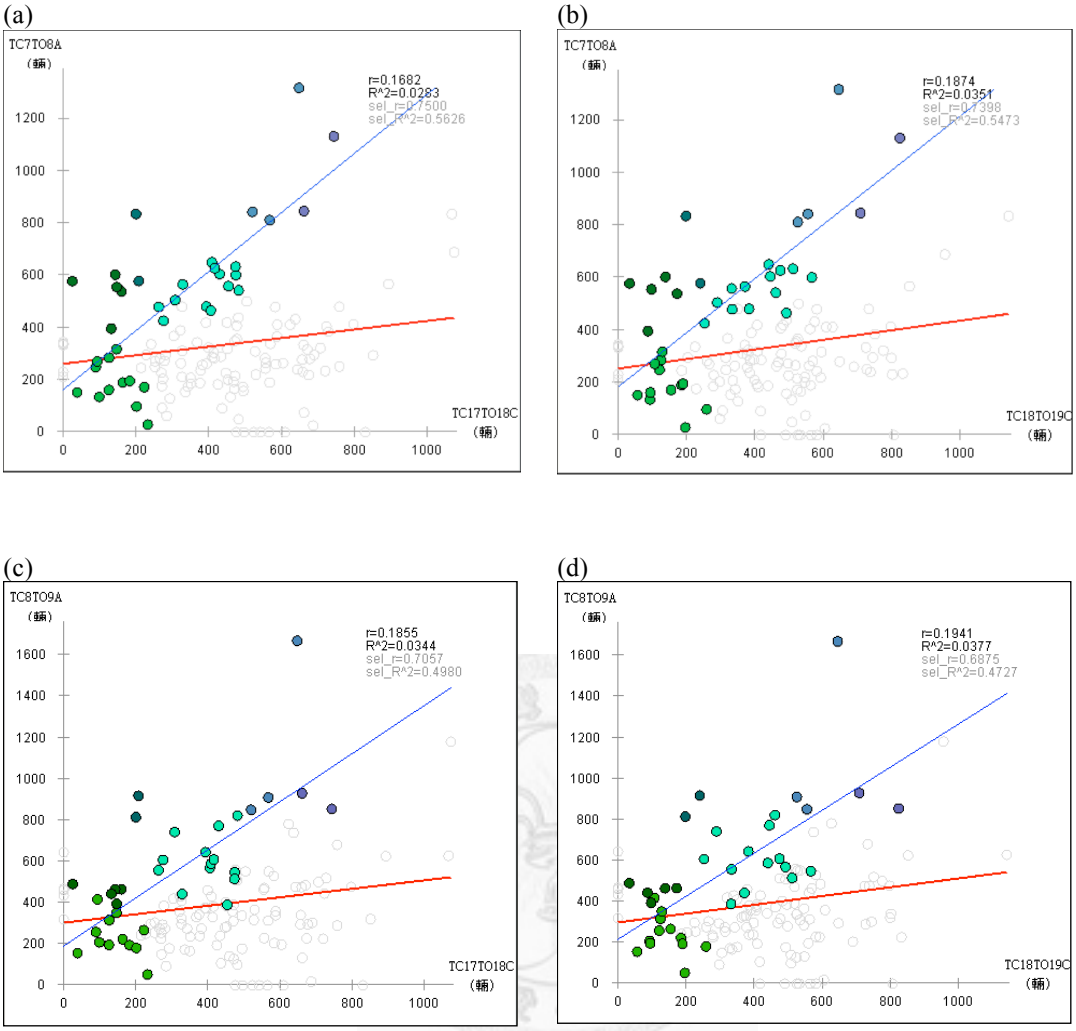


圖 31 : 89 年資料經地理視覺化探索分析重新聚類的集合一簡單線性迴歸分析：a, b 為早晨 (7-8 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 c, d 為早晨 (8-9 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 圖中紅線為原始資料的簡單線性迴歸線、r 為其相關係數、R^2為其判定係數。藍線為集合一的線性回歸線，sel_r 為其相關係數、sel_ R^2為其判定係數

五、集合二的相關分析與簡單迴歸分析

集合二的相關係數與判定係數與原始資料相較有非常明顯的提昇 (表4、圖 32) ，最高者為0.51，最低者也有0.48，代表此集合內的早晚尖峰流量具有中相關性且呈現正相關的趨勢。簡單迴歸的配適程度與原始資料相較也有非常顯著的提昇，判定係數最高者為0.29，最低者也有0.23。

觀察集合二的空間分布後可以發現，大多位於市中心棋盤式的道路路口，但也有少部分資料位於市中心外圍部分，其車流量分布則與集合一不同，有越往市中心早晚流量越高的趨勢，車流量從外圍逐漸向中心遞增，其中以忠孝東路的路口資料可以觀察到相同顏色的線形特徵 (圖30 e, f) 。

表 4：89 年資料分群後資料相關係數及判定係數 (集合二)

尖峰時段	TC17to18C (西向東)	TC18to19C (西向東)
TC7to8A (東向西)	r = 0.5107 R^2 = 0.2608	r = 0.4843 R^2 = 0.2345
TC8to9A (東向西)	r = 0.5385 R^2 = 0.2900	r = 0.5009 R^2 = 0.2509

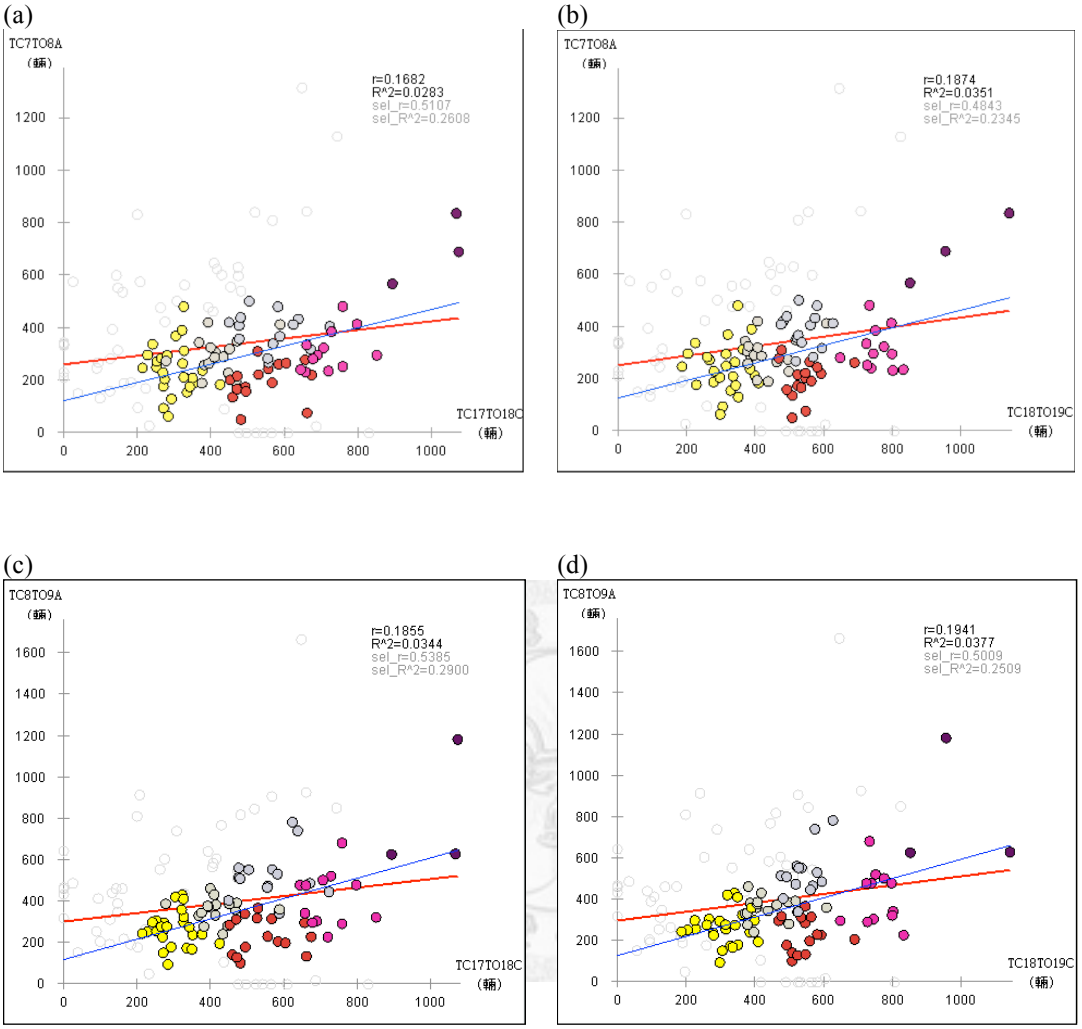


圖 32：89 年資料經地理視覺化探索分析重新聚類的集合二簡單線性迴歸分析：a, b 為早晨 (7-8 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 c, d 為早晨 (8-9 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 圖中紅線為原始資料的簡單線性迴歸線、r 為其相關係數、R^2為其判定係數。藍線為集合一的線性回歸線，sel_r 為其相關係數、sel_ R^2為其判定係數

六、不同年度的比較

將 89 年的資料資料進行初步分析之後，再利用同樣的 7×7 中型網路，針對 90 年的相同欄位資料進行分析，也可以在資料當中觀察到類似的現象。90 年的資料每一組的相關係數都在 0.21 以下，顯示這四組資料彼此之間幾乎沒有存在相關性。此外，簡單迴歸的配適程度也不佳，每一組的判定係數都在 0.045 以下，顯示原始資料進行迴歸分析，並無法解釋資料之間的相關關係。

表 5： 90 年原始資料相關係數及判定係數

尖峰時段	TC17to18C (西向東)	TC18to19C (西向東)
TC7to8A (東向西)	r = 0.1838 R^2 = 0.0338	r = 0.1824 R^2 = 0.0332
TC8to9A (東向西)	r = 0.1977 R^2 = 0.0391	r = 0.2122 R^2 = 0.0450

註：TC 代表平均車道流量，7to8 代表 7-8 時車流 (其餘可類推)，A 代表東向西車流，C 代表西向東車流，r 代表相關係數，R^2 代表迴歸判定係數。

而將 90 年的原始資料經過分類以及重組之後，也得到了類似 89 年資料的三種主要集合，同時集合的相關係數與判定係數的改善結果也類似。排除單項道路後資料仍然可以區分為兩大集合，集合一的相關係數與判定係數與原始資料相較有明顯的提昇 (表 6、圖 33) ，最高者為0.66，最低者也有0.62。 簡單迴歸的配適程度與原始資料相較也有顯著的提昇，判定係數最高者為0.44，最低者也有0.39。

表 6：90 年資料分群後資料相關係數及判定係數 (集合一)

尖峰時段	TC17to18C (西向東)	TC18to19C (西向東)
TC7to8A (東向西)	r = 0.6685 R^2 = 0.4469	r = 0.6298 R^2 = 0.3966
TC8to9A (東向西)	r = 0.6578 R^2 = 0.4328	r = 0.6451 R^2 = 0.4162

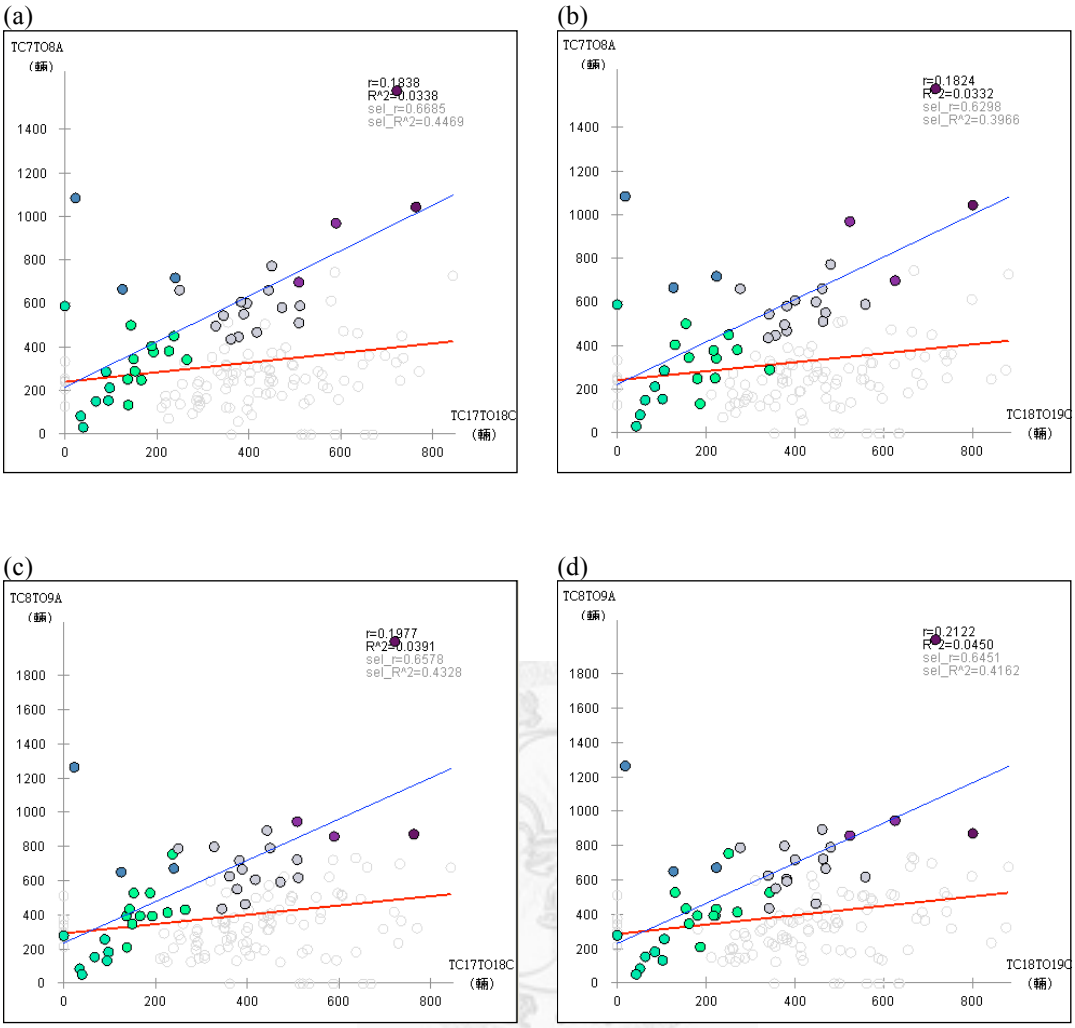


圖 33：90 年資料經地理視覺化探索分析重新聚類的集合一簡單線性迴歸分析：a, b 為早晨 (7-8 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 c, d 為早晨 (8-9 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 圖中紅線為原始資料的簡單線性迴歸線、r 為其相關係數、R^2為其判定係數。藍線為集合一的線性回歸線，sel_r 為其相關係數、sel_ R^2為其判定係數

集合二的相關係數與判定係數與原始資料相較也有明顯的提昇 (表 7、圖 34) ，最高者為0.54，最低者也有0.52。 簡單迴歸的配適程度與原始資料相較也有顯著的提昇，判定係數最高者為0.30，最低者也有0.27。

表 7：90 年資料分群後資料相關係數及判定係數 (集合二)

尖峰時段	TC17to18C (西向東)	TC18to19C (西向東)
TC7to8A (東向西)	r = 0.5256 R^2 = 0.2763	r = 0.5393 R^2 = 0.2909

TC8to9A (東向西)	r = 0.5218 R^2 = 0.2723	r = 0.5496 R^2 = 0.3020

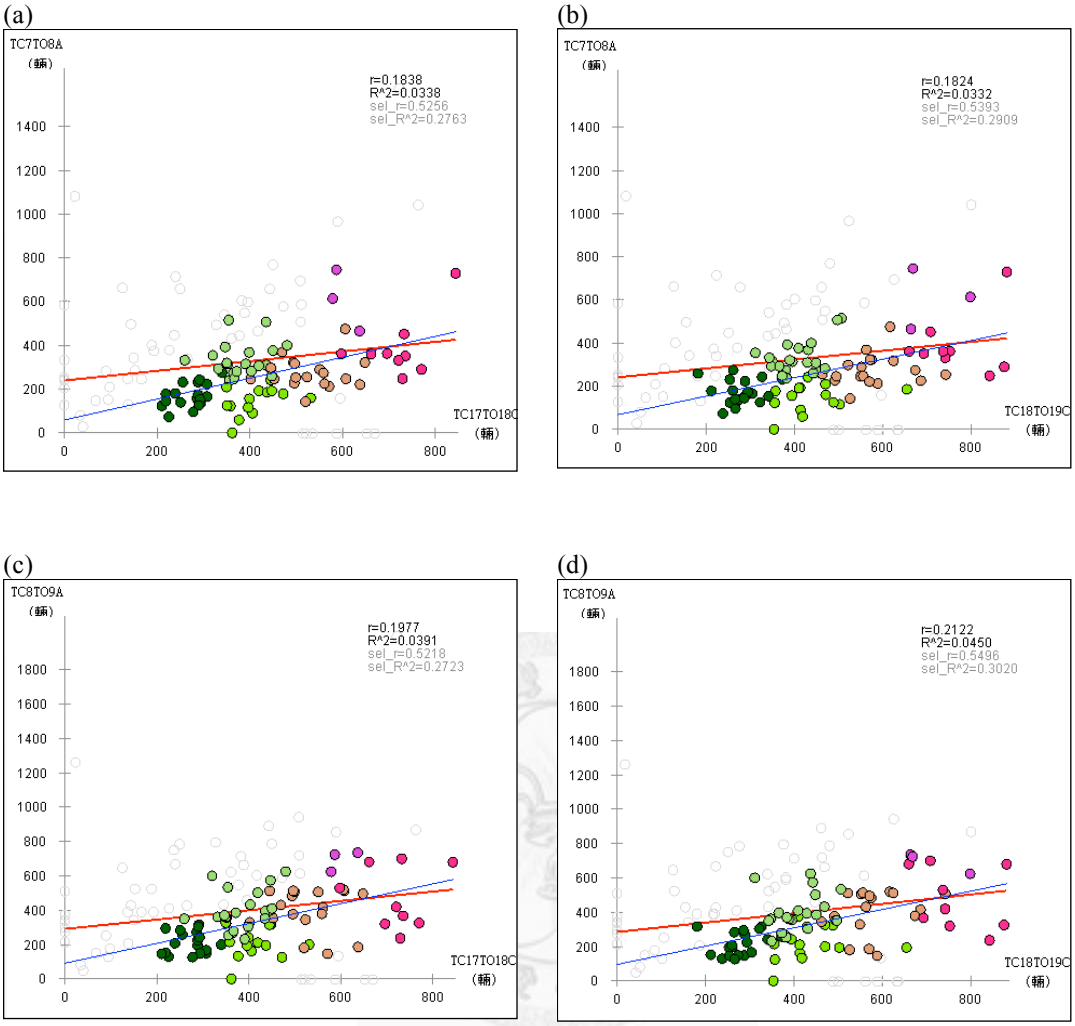


圖 34：90 年資料經地理視覺化探索分析重新聚類的集合二簡單線性迴歸分析：a, b 為早晨 (7-8 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 c, d 為早晨 (8-9 時) 與傍晚尖峰 (17-18 與 18-19 時) 車流量對應。 圖中紅線為原始資料的簡單線性迴歸線、r 為其相關係數、R^2為其判定係數。藍線為集合二的線性回歸線，sel_r 為其相關係數、sel_ R^2為其判定係數

同樣將此兩集合分開顯示在 SOM 視覺化展示與地圖上，可以發現其不論在 SOM 的資料拓樸或是地理空間上與 89 年資料均相當類似。觀察集合一的 SOM 視覺化展示 (圖35 c, d) ，發現集合內的 U-matrix 彼此之間的顏色較深， 且被分配到單一神經元的資料群較多， 代表各神經元所代表的資料群集差異較大。觀察集合二的 SOM 視覺化展示 (圖35 e, f) ，發現集合內的U-matrix的顏色較淺，代表整體資料差異性較小。

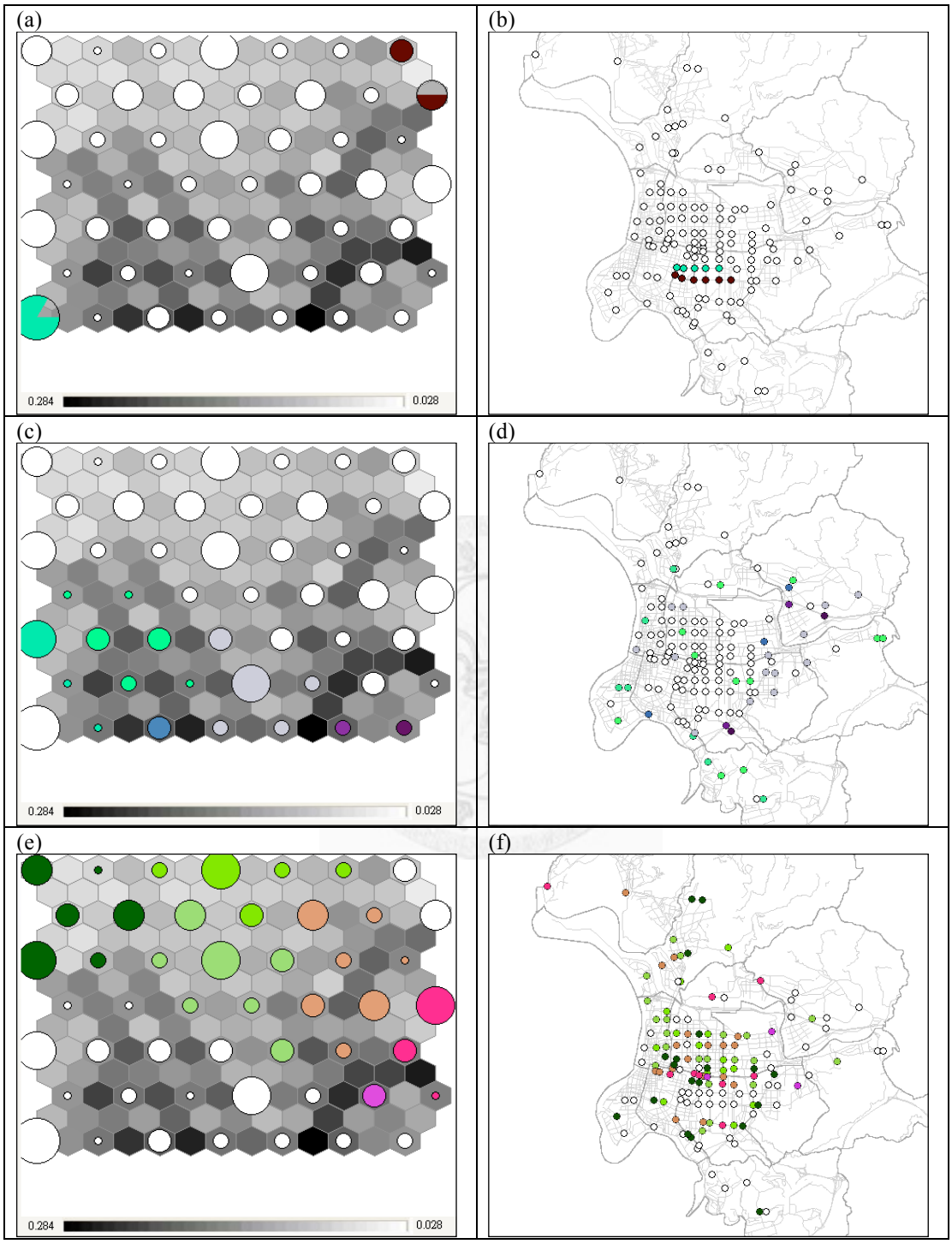


圖 35：90 年資料經過地理視覺化重新聚類後之 SOM 視覺化及空間分布：圖中(a)、(b)為兩條單向路段；(c)、(d)為集合一；(e)、(f)為集合二

在空間分布方面，集合一大多位於市中心周邊的道路路口，其車流量分布符合早上入城及傍晚出城的流量較高的特性，車流量有從東向西逐漸遞減的趨勢。集合二的空間分布大多位於市中心棋盤式的道路路口，車流量分布有越往市中心早晚流量越高的趨勢，從外圍逐漸向中心遞增，在集合二當中可以觀察到東西向的相鄰路口其資料接近，在地圖上比較容易觀察到東西向的線形特徵。

七、小結

傳統的交通流量研究並未考慮資料本身具有的空間差異，因此當使用傳統統計分析時，台北市交通流量的時空資料並未呈現相關性。但是當使用地理視覺化探索分析方法時，可發現台北市交通流量的時空資料可分成三大類型，顯示交通流量在市中心棋盤式道路以及周邊區域有不同的空間分布，以及此空間分布在其內部流量資料上特性的差異。

利用經地理視覺化探索分析得到的新集合進行相關分析與簡單迴歸分析後發現，比原始資料的相關係數與判定係數大幅提昇。這意謂著當經過地理視覺化探索分析後，資料經過分類，呈現新的空間特徵，也在統計分析上顯現新的關係。因此，在晨間和傍晚的尖峰時段車流量的相關性上，若是用此空間型態或是資料型態來區分資料，其解釋度會比原始資料未經不分類為佳。

此結果也可以解釋為，台北市晨間和傍晚的尖峰時段車流量具有正相關的特徵，但應可以區分為市中心棋盤式路網與市中心周邊道路兩種不同的空間分區的角度來討論，此兩區具有不同的資料內部結構以及空間的分佈特徵，若是將資料以這樣的觀察結果分類，套用到相關車流量的分析預測模式上，其結果應會比未分類的資料的結果為佳。

再者，簡單迴歸模式若是能夠經由地理視覺化探索分析的輔助，在

空間上找出具有意義的資料分區，則其模式的配適能力也可以大幅的增加。唯本研究之重點在於利用地理視覺化找出有潛在意義的資料及空間特徵，因此在本研究的案例當中，相關統計假設與檢定並非本研究所關注的焦點，僅用來證明地理視覺化探索分析確實能在找出具有意義的空間特徵時，對於找出分類後數據間的關係有顯著的提昇。

第二節　Geo-SOM 演算法與群聚演算法的結合

一、Geo-SOM 網路大小的選定

由於 Geo-SOM 同時考慮了資料以及空間特性，其目標是在空間上尋找有意義的特徵型態或是分區，依據本研究「先分解再重新聚類」的演算方式，在 SOM 的網路大小上希望先能夠將資料切割成較小的空間單元，之後再利用資料的群聚性將小的空間單元組合，並在地圖的展示上進行觀察。因此本研究採用 11 × 11 的 Geo-SOM網路，先將資料細分為具有相同資料特性以及空間位置較為接近的小群集，然後再利用不同的群聚演算法進行第二步的重新聚類，若是再聚類的過程當中，在地圖展示上觀察到特殊的空間型態或是區域分布，再深入觀察其資料分布的特性，以了解整體資料的特性以及內部結構。

二、機車流量資料的 Geo-SOM 分類結果

將 90 年度的上午尖峰兩個時段 (7:00-8:00、8:00-9:00) 的十字路口機車車流量，作為 Geo-SOM 的輸入向量，同時利用兩個不同的平行座標圖 (PCP) 來展示多維度原始資料的原始資料，以及經過群組過後的資料的平均值分布，而群組平均值的平行座標圖，也利用不同粗細的折線來表示其組內所含的資料數目多寡。

四個方向的尖峰時段機車車流，經 Geo-SOM 演算後得到結果如圖

36，觀察地圖分布可發現其空間展示上較容易觀察到空間群聚的現象，除了信義路的單向道路可以觀察到明顯的線性特徵以外，在市中心棋盤式道路的靠北、東、南的周邊路口，皆可以觀察到明顯鄰近的路口其顏色也較為相近，而觀察 Geo-SOM 的 U-matrix 也發現這些周邊路口的彼此之間的顏色較淺，代表其資料的差異較小，的確有空間及資料的群聚的現象存在。

而觀察經過 Geo-SOM 演算後的平行座標圖 (圖 37)，可以發現不論是原始資料或是經過平均值計算過後的平行座標圖 (圖 37)，其分類或是資料的特徵展示皆顯得雜亂且不容易觀察，由於本案例採用 11 × 11 的 Geo-SOM網路，因此在重新聚類前的資料分組便至少有121組，因此在這樣的情況下，較不容易對於研究區內的空間分區或是資料分布進行觀察。因此針對原始的 Geo-SOM 演算結果，本研究再將群聚演算法導入，一邊將資料進行動態的合併，一邊進行資料的空間分布呈現以及資料分布的觀察。

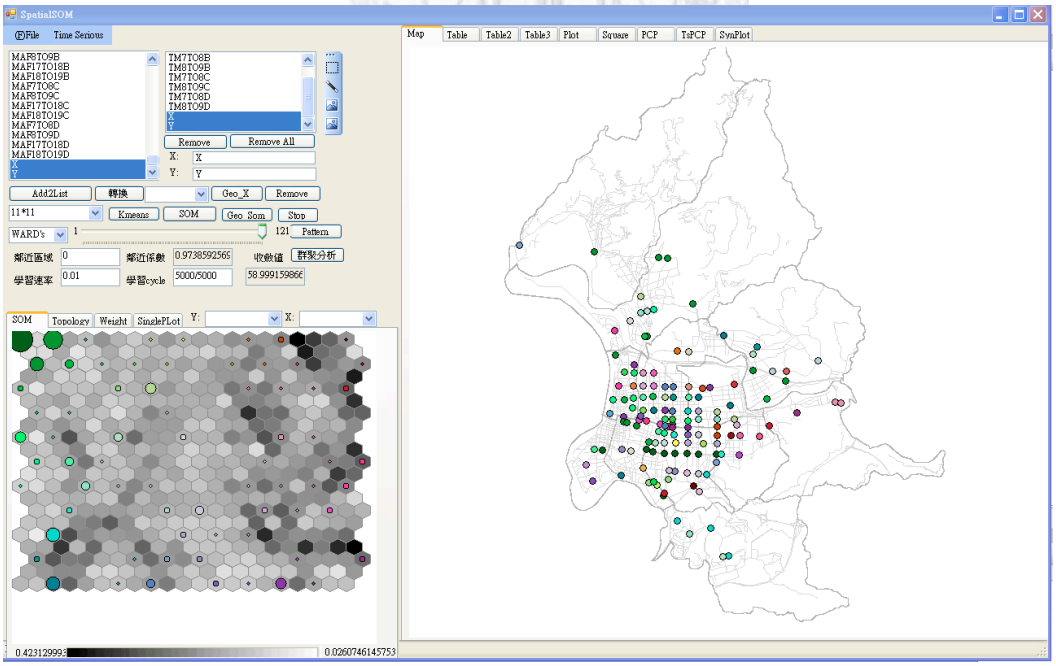


圖 36：經過 Geo-SOM 演算後的資料空間分布

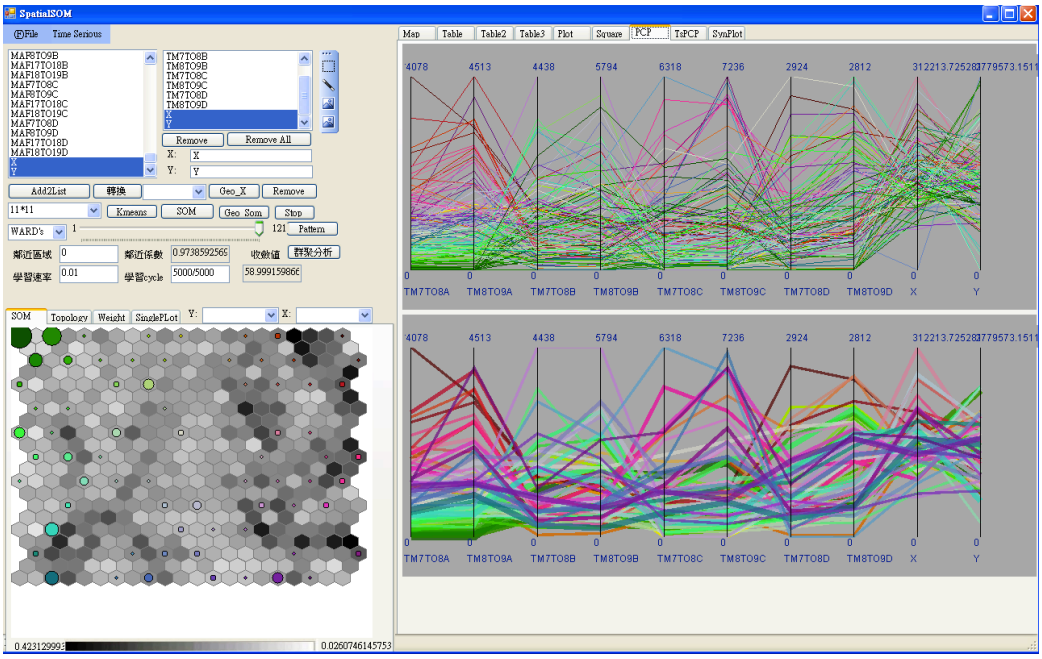


圖 37： 經過 Geo-SOM 演算後的原始資料及平均值平行座標圖

三、階層式群聚演算法的資料再聚類

與 SOM 演算的先期評估法類似，本研究針對民國 90 年所調查早晨尖峰兩個時段四個方向的機車車流量資料，作為 Geo-SOM 的輸入向量。針對四種不同的群聚演算法進行視覺化的評估，但跟傳統 SOM 演算的目標不同的地方在於，對於群聚演算法的評估主要是觀察其在空間分布上的變化。因此本研究首先利用 11 × 11 的中型 SOM 網路，將資料細分為具有類似特性的49個群聚，再利用四種演算法進行重組聚類，檢視合併後四種演算法的群駐數量 n 在 15 、 10 、以及 5 的時候的分類結果在空間上所展現的分布。 經過合併演算後，在 n = 15 的情況下，其空間分布結果如圖 38。觀察其空間分布發現，除了Slink 演算法有資料過於集中在同一群聚，空間分布不明顯的情形之外，其他三種演算法皆可觀察到與原始資料分布類似的周邊道路的資料型態。同時除了Slink 演算法之外，其他三種演算法也可以觀察到市民大道以及信義路等部分路段的空間資料呈現線性的資料群聚特徵。但是其餘的資料在空間上多屬於破碎的小集合，較難觀察到明顯的群聚或是空間分區的特徵。

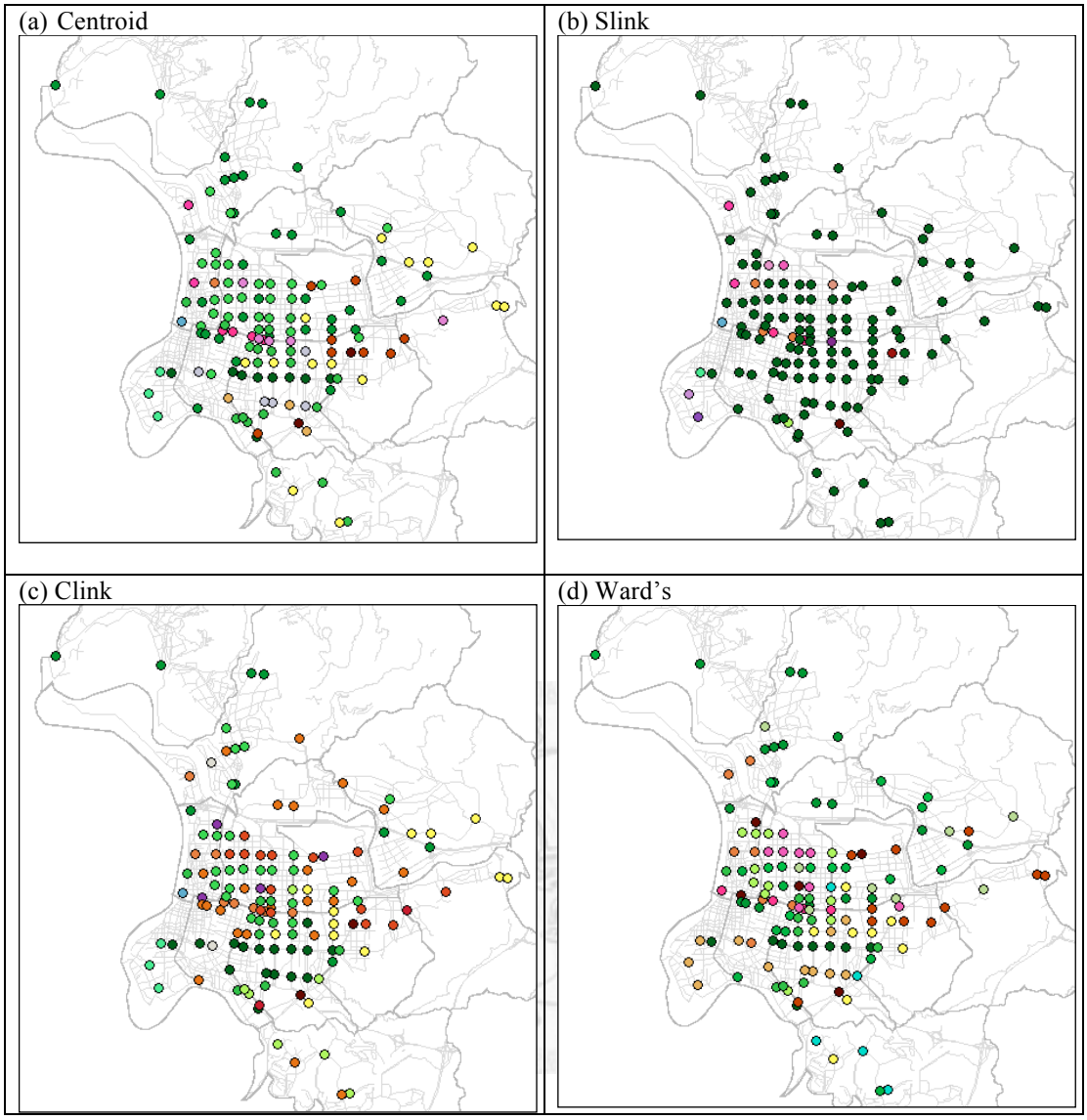


圖 38：n=15 時，四種群聚演算法合併後視覺化地圖展示。(a) Centroid 法，(b) Slink 法，(c) Clink 法，(d) Ward's 法

而在 n = 10 的情況下（圖 39），Slink 演算法的資料較 n = 15 的聚類結果更為集中， 只剩下多個各含有一個資料的群聚以及包含所有其他資料的大群聚，路段的線型群聚特徵已經無法被觀察到。在 Centroid 法以及 Clink 法的聚類結果當中開始出現資料數量較多的群聚，但是在空間的群聚表現上以 Centroid 法的結果較為明顯，而 Ward's 法的空間分布結果相對其他的演算法來說變化少，但是聚類當中的資料數量較為平均。

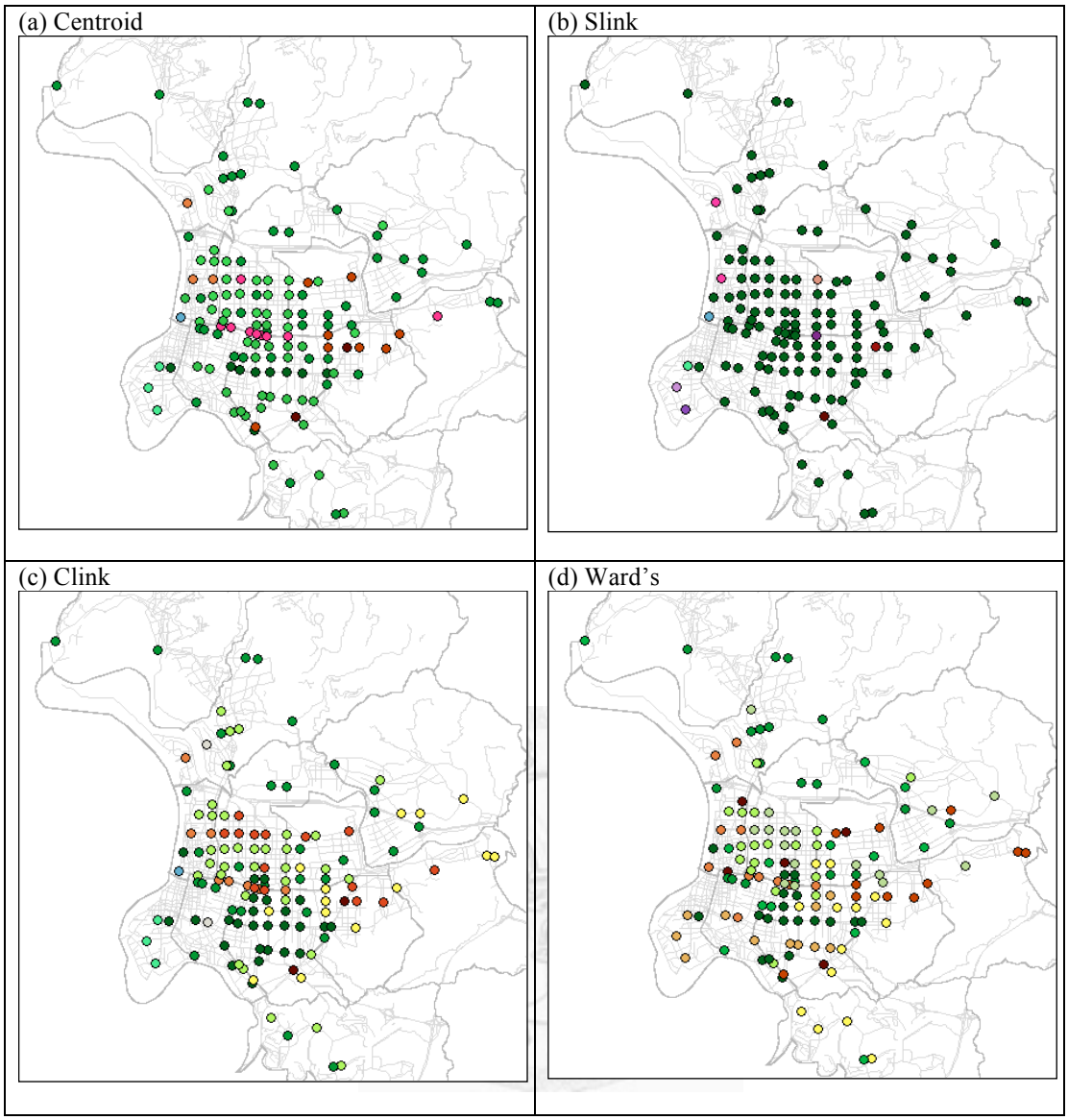


圖 39： n=10 時，四種群聚演算法合併後視覺化地圖展示。(a) Centroid 法，(b) Slink 法，(c) Clink 法，(d) Ward's 法

在 n = 5 的情況下（圖 40），Slink 演算法的資料只剩下四個各含有一個資料的群聚以及包含所有其他資料的大群聚，路段的線型群聚特徵已經無法被觀察到。在 Centroid 法的聚類結果開始出現大群聚，只剩下市民大道的線性特徵。另外 Clink 法及 Ward's 法的聚類結果保留較多的大群聚。而在所有的空間分布結果當中，以 Ward's 法在棋盤式道路區可以觀察到較多的空間群聚以及線性特徵。

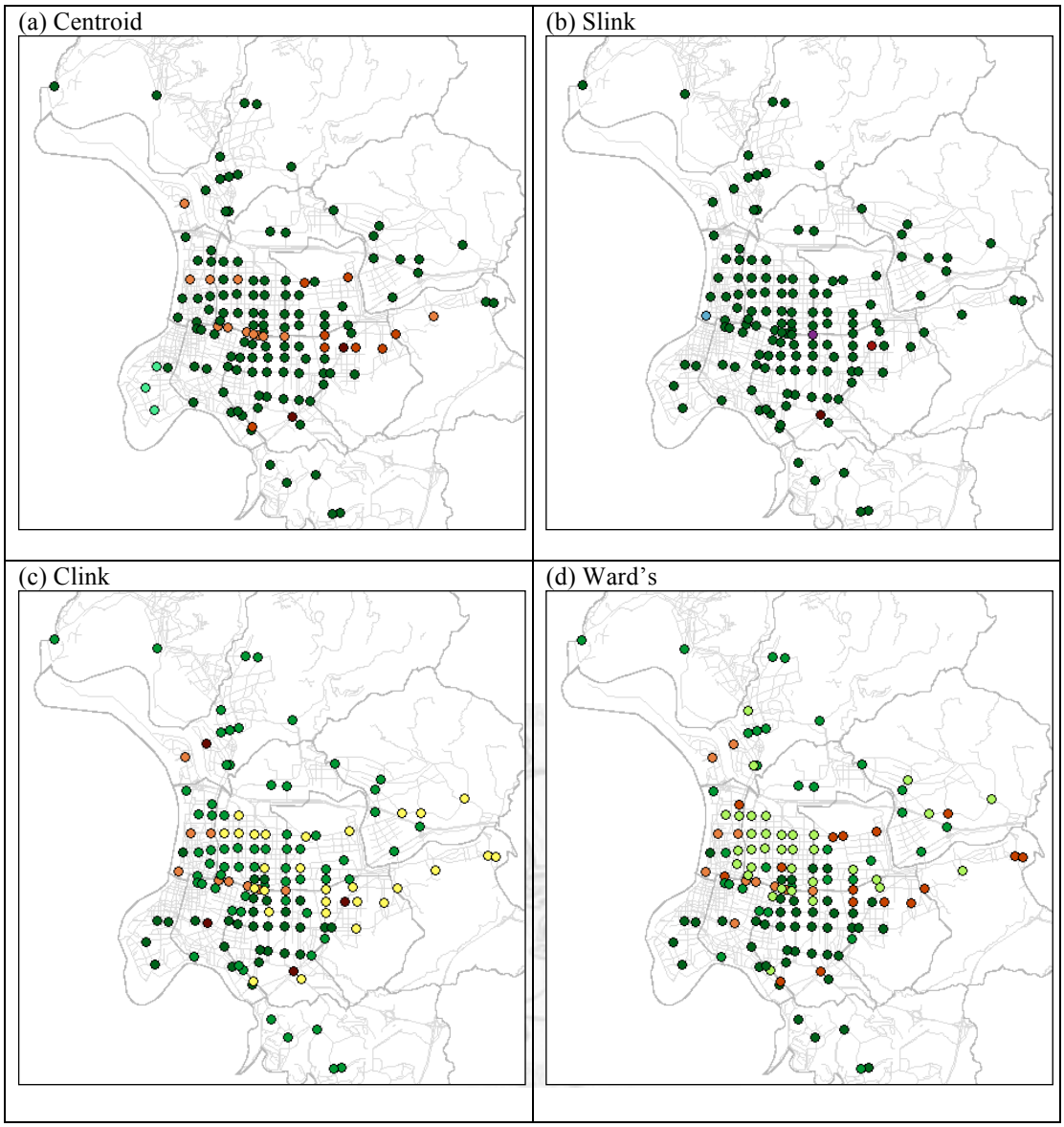


圖 40：n=5 時，四種群聚演算法合併後視覺化地圖展示。(a) Centroid 法，(b) Slink 法，(c) Clink 法，(d) Ward's 法

由以上的分析結果，可發現 Slink 法在聚類的過程當中有過度集中的問題，因此較不適合用於資料的再聚類演算上， Centroid 演算法也容易形成較大的聚類，但是其空間區域的分布較 Slink 法具有群聚性。Clink 和 Ward's 法群聚大小較為均勻，但是在合併的過程當中， Clink 在空間的展示上較為破碎，而 Ward's 法比較容易觀察到空間區塊的特徵。由於本案例採用 Geo-SOM 方法有考慮到資料的空間位置，因此其「先分解再重新聚類」探索重點在於找出在空間上具有特殊型態或是區域分布

的特徵。在不同數目的合併結果中，以 Ward's 法演算所提供的結果較為其他演算法在空間上的群聚特徵較為明顯。因此 Geo-SOM 網路的合併分析將以 Ward's 法作為主要的合併演算法。

四、經過群聚演算及探索後的結果

本研究先利用 Ward's 法將分組數重 121 組逐漸減少為 6 組的情況下，在空間地圖上可觀察到空間群聚以及道路的線性特徵 (圖 42)，例如市民大道和信義路路段在地圖上的展示可觀察到線性特徵(分組 2、分組 6)，因此對照平均值平行座標圖觀察 (圖 43)，可以發現視覺化所觀察到的空間群聚在資料的展示上也具有容易辨識的特徵。市民大道的資料特性是在早晨尖峰時間的東向西及西向東的機車車流量的平均值較高，而信義路路段則是在早晨尖峰時間僅有南往北的車流量較高。

在市中心棋盤式道路的部份 (分組 4)，可觀察到市民大道以北的路口資料有明顯的空間群聚，對照平均值平行座標圖觀察，發現其資料特徵為西向東的流量較高。在市中心以東的部分周邊道路 (分組 1)，其空間群聚雖然較不明顯，但是資料特徵也非常容易辨識，在機車車流量的部份僅有東向西車流量較高。

在市中心南部的周邊道路(分組 3)，可以在觀察到部分資料群集，和平東路和仁愛路部分路段有觀察到線型特徵，對照其資料分組以及平行座標圖，可知道其資料特徵與分組 6 的信義路路段類似，但是資料特性不同的地方在於信義路為西向東單向道路，但是分組 3 和分組 6 在南向北的機車流量皆較台北市其他地方為高。

在地圖上還可以觀察到有一資料數量不少的分組 (分組5)，其空間分布較為分散，主要在研究區靠近周邊的位置，而資料特徵是東西南北向的車流量皆較低。

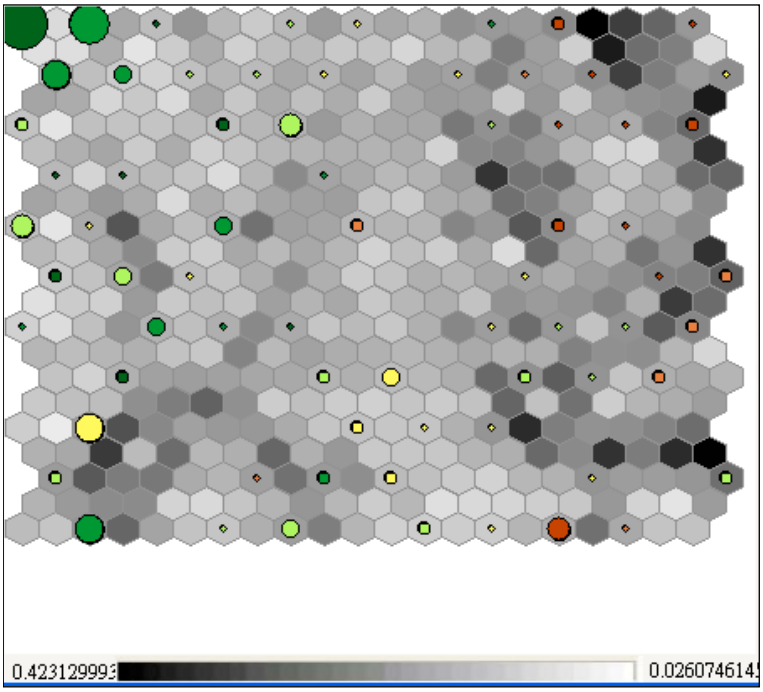


圖 41： n=6 時， Ward's 演算法合併後之 SOM 視覺化展示

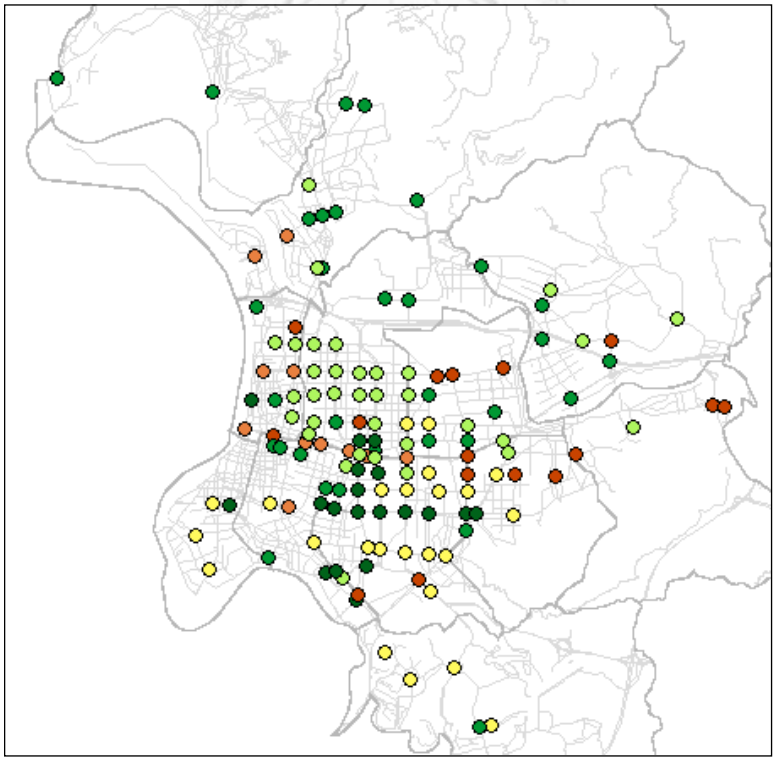


圖 42： n=6 時， Ward's 演算法合併後之資料空間分布

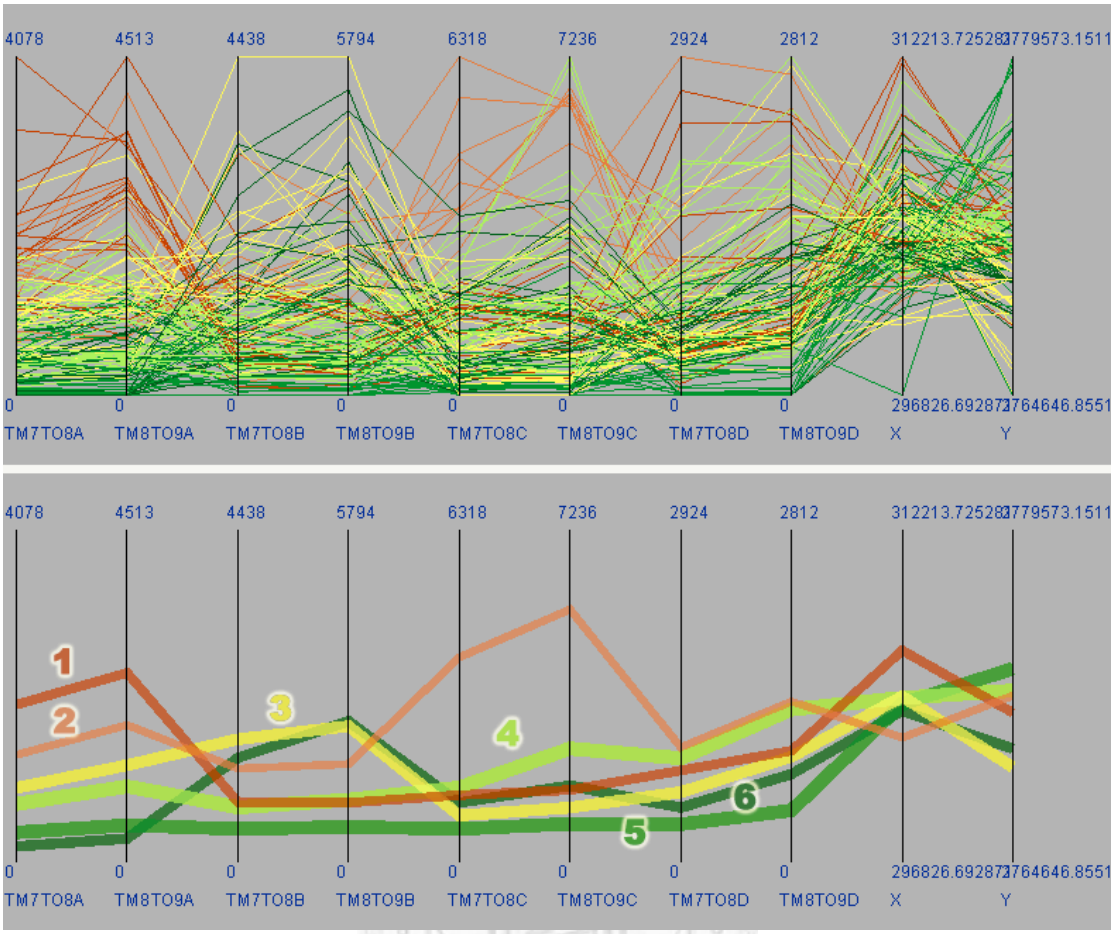


圖 43：合併後之原始資料平行座標圖及分組平均值平行座標圖（ TM 代表機車流量，7TO8 代表 7-8 時車流 (其餘可類推)，A 代表東向西車流，B 代表南往北車流，C 代表西向東車流，D 代表北往南車流，圖中數字代表分組代號。）

由以上重組結果可發現，利用視覺化探索所得到的分組，可以在空間地圖上發現具有明顯的群聚或是空間特徵，同時對照平均值平行座標圖在資料面上也可以分辨出群聚所代表的資料特徵。

五、小結

當地理資料的維度變的越複雜，要在資料當中找到空間特徵也變得更加困難。若是利用傳統的資料觀察或是分類的方法，其分類或是資料的特徵皆不容易觀察及分析。本研究利用 Geo-SOM 以及群聚演算法，先將資料細分為空間位置及資料特徵皆類似的小群集，再將這些小群集重

新組合，經過測試發現的確可以觀察到空間上的群聚或是特殊的空間特徵。顯示機車的流量分布具有空間特性，而這些不同空間的資料可以經由視覺化的方法找到其個別的差異。

觀察地理視覺化探索分析得到的資料分組，可以發現平均值平行座標圖比原始資料的平行座標圖更容易了解資料的特性，同時也比較容易進行不同分組之間資料差異的比較。因此原始資料經過視覺化的分析之後，可以得到新的空間分布以及資料分組方式，同時透過平均值平行座標圖的比較，可以在原始資料當中辨識出不同的分組或是不同的空間區域，同時視覺化分析也可以提供區分其流量資料特性的方法。

在 Geo-SOM 的分析結果中，可以發現在不同的屬性資料當中，不同分組的流量資料值域會有重疊的現象，例如分組 3 和分組 6 僅有東向西道路的流量型態不同，但是由於 Geo-SOM 考慮資料的空間位置，因此在分組結果當中可以將兩者的空間分布區分出來。此分組結果顯示本研究的地理視覺化方法可以協助發現空間特徵以及了解資料的區域差異。

第五章 結論與建議

一、地理視覺化知識探索架構之操作型方法

本研究根據對於過去文獻的回顧與整理，提出了一個新的視覺化工具整合的知識探索架構，認為空間視覺化探索的過程在於對於資料本身的特徵抽取以及資料的空間關連性抽取，因此配合視覺化工具的整合，經由案例的驗證，本研究發現視覺化技術導入後的確對知識探索本身產生了不同的結果和影響。因此本研究歸納出以下幾種地理視覺化可用之探索操作型方法，分別敘述如下；

1. 資料群聚對應空間群聚

對於變數較多的複雜資料，要直接觀察出資料空間的群聚現象較為困難，因此便非常適合利用可將維度縮減的資料轉換模式，SOM映射網路對於資料的處理類似於遙感探測領域的非監督性分類，會將原始資料輸入後產生之權重值神經元排列為二維矩陣，由於越鄰近的神經元代表其資料的內在聚類規則越類似，因此非常適合用來尋找資料的群聚所對應的空間群聚。在空間上，對地理資料有意義且可以辨認之空間特徵，如道路的線狀外形、資料的群聚現象等。在屬性資料的關係，則是使用散布圖、相關分析的相關係數或簡單迴歸分析的判定係數作為判斷依據。在 SOM 的成份平面展示則是要觀察U-matrix的顏色深淺，用以找出極端值群集或是與其他群集差異較大的資料。研究者在使用散布圖及 SOM 映射圖探索時，也要配合地圖展示同時觀察空間上的變化。

2. 資料群聚/特殊形態對應特殊空間形態

與尋找空間群聚的方法類似，SOM的內在聚類規則也可以用來觀

察是否具有特定的空間形態如線型、環狀、或是與特殊地物形狀位置的相對應關連性等。被抽取出來的空間型態也可以導入新的空間資料，用來找出與其它空間地物之間未知的關係，或是利用不同的資料群聚用以分割為不同的空間區域。 由於 SOM 已經將資料做較細緻的分類，但是此細緻分類在視覺化展示或是地理空間展示上可能過於破碎，較不容易探索有意義的資料特徵，因此將 SOM 的神經元進行合併，利用視覺化探索找到潛在的空間關係及資料特徵。

3. 視覺化資料分級及分類造成空間分區

此方式不需要將原始資料進行轉換，可利用較為關鍵的影響因子進行資料值域的分類探索，例如利用SOM配合散布圖的互動方式探索資料的值域也可以定位出特定的空間位置，但資料維度過多也會造成視覺化資料分級的困難度。

4. 視覺化空間分區造成資料分類或分級

這部份可利用如 Geo-SOM 等具有空間聚類能力的演算法。藉由先決定空間分區，再對應回資料空間進行資料的分類或是分級。此方式可以解決少數資料位在資料分級邊緣所造成的分類差異問題，可利用資料的空間特性找出差異性。但此方法必須要注意空間上或是資料面上可能會產生極端值，以及單純利用空間來分類資料可能造成的資料分區範圍重疊的問題。

5. 先處理及偵測偏差或極端值

由於資料處理階段的資料補遺部分不在本研究的研究範圍內，且視覺化工具也無法改善本身資料缺乏的問題，因此視覺化工具在資料前處理的階段主要是伴演資料當中的偏差或極端值偵測的角色。資料挖掘的結果取決於輸入資料的品質與正確性，因此如何利用視覺化工具檢測出

偏差值或是異常資料形態，也是地理視覺化相當重要的一環。此部分SOM 有不錯的成果，未來可以繼續研究PCP或是星狀圖等視覺化工具。

二、本研究所達到的貢獻

本研究提出一個新的地理視覺化的探索研究方法，此方法在交通流量分析上面得到初步不錯的成果，將來可以修改套用至其他面向的時空分析研究當中，因此具有再研究的價值及潛力。

在視覺化的研究理論原則與工具的操作指導上，本研究將「先分解再重新聚類」的知識探索原則導入 SOM 以及 Geo-SOM 演算，配合執行效率較佳的群聚演算法，逐步合併群集。此方法不但運算時間快速，同時在分群的數量也比傳統的 SOM 演算與視覺化展示更具彈性。而針對台北市的車流量資料進行傳統統計方法分析與地理視覺化探索分析，發現後者的確能有效地找出資料潛藏的空間特徵，並且在具有空間特徵的資料分類後，可有效分析得出資料間的相關性。而導入 Geo-SOM 後，則發現的確有助於了解台北市機車流量的空間分布以及不同區域車流量的資料特性差異。

本研究的貢獻，應該從地理視覺化在資料庫知識探索所扮演的角色，以及地理視覺化本身技術發展現況兩個層次來說明，在前者，本研究提出了一個以人類視覺分析為中心的研究架構，說明以視覺化處理、課題假設，以及前知識的輔助之下，可以在此探索環境下得到一些新的發現或是新的認知。在後者，本研究整合上述概念，提出了一個可以輔助前知識、假設、人腦空間認知等概念的視覺化探索方法。研究結果顯示，不論是在資料的空間細緻度或是資料的相關性或是可用的分類標準上，都比未經過視覺化探索之前來的更加外顯。

三、未來可努力的研究方向

雖然在本研究的案例探索當中，得到了比過去改善的結果，但是本研究在探索的過程當中，仍然發現了許多研究資料本身的限制以及研究方法不足的面相，未來還需要進一步的去分析及研究，茲將本研究所發現的問題分述如下：

1. 本研究限於時間與研究者各自專精的領域知識，並不以大量不同領域的資料進行個別的案例探討，而是希望以有限的案例，說明在空間資料的視覺化探索當中，該如何去發掘可能潛在的視覺化知識以及將其利用在知識探索的改進之上，未來希望各種不同領域的研究也能導入此整合式的探索架構，發掘各自不同的專業領域的視覺化知識，用以改進各領域的研究成果。
2. 本研究雖然是以地理視覺化工具為核心，但是在屬性資料上由於研究本身的限制，目前只能以數字資料（numeral data）作為主要的對象，視覺化提供了對數字資料多元化的觀察角度，本研究相信對於空間資料或是數字資料來說，資料的觀察角度會決定研究所能做出的結論與看法，因此在這樣可能有無限多解的課題當中，導入人類視覺分析的感知能力，應該可以彌補部分過去傳統研究方法可能產生的盲點或是不足。
3. 在不同群聚演算法的使用上，本研究發現若是群數量較少的情況之下，除 Ward’s 法以外，其他演算法容易發生資料集中在某些群聚之內的情況，因此比較難觀察出空間的特殊型態，此部分又以SLINK演算法較為明顯。因此本研究建議在不希望資料過度集中的狀態下，群聚演算法以 Ward’s 法較為容易進行視覺化的探索觀察。
4. 關於資料群聚的數量，由於目前的研究並沒有一個明確的指導，因此本研究採用的方式為可以動態調整群及數量的互動性方式。利用SOM及 Geo-SOM 演算法自我聚類的特性以及群聚演算法本身運算時間短的優點，來達到即時展示探索的視覺化環境。在本研究中所

展示的不同數量的分析結果，實際上為探索多種可能之後的結論。

5. 在集合的空間分布上，目前的分類方法雖可以大致區分資料的空間分布，但是仍然可以觀察到某些群集資料的空間分布具有不規則性，由於本研究採用的是點資料，並且研究去的路口資料分布並不均勻，由於這樣的資料特性，利用 Geo-SOM 演算法比較容易產生零散的群集，要達到比較明顯的空間區域劃分，還需要未來研究的努力。

本研究以地理視覺化探索分析應用在台北市交通流量資料，發現傳統方法與地理視覺化探索分析之間的差異，分析結果顯示地理視覺化探索分析可有效找出交通流量資料的空間特徵，且依據地理視覺化探索分析分類後的資料也找出其間的關係，因此若是能夠探索出交通資料的隱含空間知識，提供資料在空間的差異性描述，將有助於交通資料在空間應用面相的加值。本研究雖就交通資料進行了視覺化工具的初步評估，未來若要推廣到不同的資料上，則必須要針對各種不同的資料進行測試，建立視覺化工具使用與探索任務之間適用性的參照。Andrienko et al. (2006) 認為「任何的思考都必須要有先前的認知或是感知」，綜合本研究結果，顯示利用地理視覺化探索分析的方法未來有潛力在時空資料探索的領域，扮演輔助人類思考的重要角色。因此，本研究認為地理視覺化探索分析未來可應用至分析其他時空資料，具有學術研究的價值。

參考文獻

Abraham, T. & Roddick, J. (1997) Discovering meta-rules in mining temporal and spatio-temporal data. *Data mining, data warehousing & client/server databases: proceedings of the 8th International Database Workshop.* Hong Kong: Springer Verlag, 29-31.

Abraham, T. & Roddick, J. (1999) Survey of spatio-temporal databases. *GeoInformatica,* **3** (1), 61-99.

Aditya, T. & Kraak M. (2005) Reengineering the Geoportal: Applying HCI and Geovisualization Disciplines. *Proceedings of 11th EC-GI & GIS Workshop, ESDI: Setting the Framework*, Alghero.

Aggarwal, C. & Yu, P. (2001) Outlier detection for high dimensional data. *ACM SIGMOD Record,* 30 (2), 37-46.

Andrienko, G. & Andrienko, N. (2005) Visual exploration of the spatial distribution of temporal behaviors. *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, 799-806.

Andrienko, G. & Andrienko, N. (1999) Interactive maps for visual data exploration. *International Journal of Geographical Information Science,* **13** (4), 355-374

Andrienko, N., Andrienko, G. & Gatalsky, P. (2003) Exploratory spatio-temporal visualization: An analytical review. *Journal of Visual Languages & Computing,* **14** (6), 503-541

Andrienko, N. & Andrienko, G. (2006) *Exploratory analysis of spatial and temporal data: A systematic approach*: Springer Verlag.

Ankerst, M., Berchtold, S. & Keim, D.A. (1998) Similarity clustering of dimensions for an enhanced visualization of multidimensional data. *Information Visualization, 1998. Proceedings. IEEE Symposium on*, 52-60, 153.

Armstrong, M. (1988) Temporality in spatial databases. *Proceedings of GIS/LIS '88.* Falls Church, VA: American Congress on Surveying and Mapping, 880-889.

Bação, F., Lobo, V. & Painho, M. (2005) The self-organizing map, the geo-som, and relevant variants for geosciences. *Computers & Geosciences*, **31** (2), 155-163.

Bedard, Y., Merrett, T. & Han, J. (2009) 3 fundamentals of spatial data warehousing for geographic knowledge discovery. *Geographic data mining and knowledge discovery*, **45**.

Brachman, R. & Anand, T. (1996) The process of knowledge discovery in databases: A human-centered approach. *In* Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. & Ulthurusamy, R. eds. *Advances*

in knowledge discovery and data mining. Cambridge, MA: MIT Press, 37-57.
Brauen, G. (2006) Designing interactive sound maps using scalable vector graphics. *Cartographica: The International Journal for Geographic Information and Geovisualization,* **41** (1), 59-72.
Breiman, L. (1984) *Classification and regression trees*: Chapman & Hall/CRC.
Breunig, M., Kriegel, H., Ng, R. & Sander, J. (2000) Lof: Identifying density-based local outliers. *ACM SIGMOD Record,* **29** (2), 104.
Brimicombe, A. (2003) A variable resolution approach to cluster discovery in spatial data mining. *Lecture Notes in Computer Science*, 1-11.
Canas, A., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T., Lott, J. & Carvajal, R. (2005) Concept maps: Integrating knowledge and information visualization. *Lecture Notes in Computer Science,* 3426, 205.
Cheeseman, P. & Stutz, J. (1996) Bayesian classification (autoclass): Theory and results. *Advances in knowledge discovery and data mining,* 180.
Chen, L., Chang, E., Garcia-Molina, H. & Wiederhold, G. (2002) Clustering for approximate similarity search in high-dimensional spaces. *Knowledge and Data Engineering, IEEE Transactions on,* **14** (4), 792-808.
Claramunt, C., Jiang, B. & Bargiela, A. (2000) A new framework for the integration, analysis and visualisation of urban traffic data within geographic information systems. *Transportation Research Part C-Emerging Technologies,* **8** (1-6), 167-184.
Clementini, E. & Di Felice, P. (2001) A spatial model for complex objects with a broad boundary supporting queries on uncertain data. *Data & Knowledge Engineering,* **37** (3), 285-305.
Clementini, E., Di Felice, P. & Koperski, K. (2000) Mining multiple-level spatial association rules for objects with a broad boundary. *Data & Knowledge Engineering,* **34** (3), 251-270
Cook, D., Symanzik, J., Majure, J.J. & Cressie, N. (1997) Dynamic graphics in a gis: More examples using linked software. *Computers & Geosciences,* **23** (4), 371-385
Demsar, U. (2007) Investigating visual exploration of geospatial data: An exploratory usability experiment for visual data mining. *Computers, Environment and Urban Systems,* **31** (5), 551-571.
Dykes, J.A. & Mountain, D.M. (2003) Seeking structure in records of spatio-temporal behaviour: Visualization issues, efforts and applications. *Computational Statistics & Data Analysis,* **43** (4), 581-603.

Dykes, J., Maceachren, A. & Kraak, M. (2005) *Exploring geovisualization*: Pergamon.
Edsall, R.M. (2001) *Interacting with space and time: Designing dynamic geovisualization environments*. Penn State University.
Edsall, R.M. (2003a) Design and usability of an enhanced geographic information system for exploration of multivariate health statistics *Professional Geographer,* **55** (2), 146-160.
Edsall, R.M. (2003b) The parallel coordinate plot in action: Design and use for geographic visualization *Computational Statistics and Data Analysis,* **43** (4), 605-619.
Edsall, R.M., Harrower, M. & Mennis, J.L. (2000) Tools for visualizing properties of spatial and temporal periodicity in geographic data. *Computers & Geosciences,* **26** (1), 109-118.
Ester, M., Kriegel, H. & Sander, J. (1997) Spatial data mining: A database approach. *Lecture Notes in Computer Science,* 1262, 47-68.
Evans, B.J. (1997) Dynamic display of spatial data-reliability: Does it benefit the map user? *Computers & Geosciences,* **23** (4), 409-422
Fabrikant, S. (2001) Building task ontologies for geovisualization. *Proceedings of ICA Commission on Visualization and Virtual Environments Pre-Conference Workshop on Geovisualization on the Web*, Beijing.
Famili, F., W. Shen, R. Weber & Simoudis E. (2010) Data Pre-processing and Intelligent Data Analysis. *International Journal on Intelligent Data Analysis*, 1.
Fayyad, U. (1997) Knowledge discovery in databases: An overview. *Inductive logic programming.* 1-16.
Ferreira De Oliveira, M.C. & Levkowitz, H. (2003) From visual data exploration to visual data mining: A survey. *Visualization and Computer Graphics, IEEE Transactions on,* **9** (3), 378-394.
Flexer, A. (2001) On the use of self-organizing maps for clustering and visualization. *Intelligent Data Analysis,* **5** (5), 373-384.
Gahegan, M. (2009) Visual exploration and explanation in geography analysis with light. *In* Miller, H.J. & Han, J. eds. *Geographic data mining and knowledge discovery.* Second ed. London: Taylor & Francis, 291-324.
Gahegan, M. & B. Brodaric. (2002) Computational and visual support for geographical knowledge construction: filling in the gaps between exploration and explanation. *Proceedings of the 10th International Symposium on Spatial Data Handling,* 11. *Springer Verlag*.
Gahegan, M., Wachowicz, M., Harrower, M. & Rhyne, T.-M. (2001) The integration of geographic visualization with knowledge discovery in databases and geocomputation. *Cartography and Geographic*

Information Science, **28**, 29-44.
Goodchild, M. & Dubuc, O. (1987) A model of error for choropleth maps, with applications to geographic information systems. *Eighth International Symposium on Computer Assisted Cartography: Proceedings.* Baltimore, Maryland: American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, 165.
Guo, D. (2009) Multivariate spatial clustering and geovisualization . *In* Miller, H.J. & Han, J. eds. *Geographic data mining and knowledge discovery.* Second ed. London: Taylor & Francis, 321-345.
Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann & Witten I. (2009) The WEKA data mining software: An update. *ACM SIGKDD Explorations Newsletter*, 11, 10-18.
Han, J. & Kamber, M. (2006) *Data mining: Concepts and techniques*: Morgan Kaufmann.
Hardisty, F. (2005) The geoviz toolkit. *Auto- Carto.* Las Vegas, NV.
Harrower, M. (2007) The cognitive limits of animated maps. *Cartographica,* **42** (4), 349-357 Available from: 10.3138/carto.42.4.349
Harrower, M., Maceachren, A. & Griffin, A. (2000) Developing a geographic visualization tool to support earth science learning. *Cartography and Geographic Information Science,* **27** (4), 279-293.
He, Z., Deng, S. & Xu, X. (2005) An optimization model for outlier detection in categorical data. *Advances in intelligent computing.* 400-409.
Hinneburg, A., Keim, D.A. & Wawryniuk, M. (1999) Hd-eye: Visual mining of high-dimensional data. *Computer Graphics and Applications, IEEE,* **19** (5), 22-31.
Jarke, M., Lenzerini, M., Vassiliou, Y. & Vassiliadis, P. (2003) *Fundamentals of data warehouses*: Springer Verlag.
Jin, X. & Reynolds, R.G. (2000) Mining knowledge in large scale databases using cultural algorithms with constraint handling mechanisms. *Proceedings of the 2000 Congress on Evolutionary Computation*, La Jolla, CA, 1498-1506.
Ichiki, H., Hagiwara, M. & Nakagawa, M. (1991) Self-organizing multilayer semantic maps. *Proceedings of IJCNN IEEE International Joint Conference on Neural Networks,.* Seattle.
Jing, Y., Daniel, H., Matthew, O.W., Elke, A.R. & William, R. (2007) Value and relation display: Interactive visual exploration of large data sets with hundreds of dimensions. *Visualization and Computer Graphics, IEEE Transactions on,* **13** (3), 494-507.

Johnson, B. & Shneiderman, B. (1991) Tree-maps: A space-filling approach to the visualization of hierarchical information structures. IEEE Computer Society Press Los Alamitos, CA, USA, 284-291.
Johnson, H. & Nelson, E. (1998) Using flow maps to visualize time-series data: Comparing the effectiveness of a paper map series, a computer map series, and animation. *Cartographic Perspectives,* **30**, 47-64.
Johnson-Laird, P.N. (1988) *The computer and the mind*: Harvard University Press.
Jones, C., Haklay, M., Griffiths, S. & Vaughan, L. (2009) A less-is-more approach to geovisualization-enhancing knowledge construction across multidisciplinary teams. *International Journal of Geographical Information Science,* **23** (8), 1077-1093.
Keim, D. (2000) Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics,* **6** (1), 59-78.
Keim, D.A. & Kriegel, H.P. (1994) Visdb: Database exploration using multidimensional visualization. *Computer Graphics and Applications, IEEE,* **14** (5), 40-49.
Klippel, A., Hardisty, F. & Weaver, C. (2009) Star plots: How shape characteristics influence classification tasks. *Cartography and Geographic Information Science,* **36**, 149-163.
Knorr, E. & Ng, R. (1996) Finding aggregate proximity relationships and commonalities in spatial data mining. *IEEE transactions on knowledge and data engineering,* **8** (6), 884-897.
Koperski, K., Adhikary, J. & Han, J. (1996) Spatial data mining: Progress and challenges. *ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, Montreal, Canada.
Koperski, K. & Han, J. (1995) Discovery of spatial association rules in geographic information databases. *Lecture Notes in Computer Science,* 951, 47-66.
Koua, E. & Kraak, M. (2004) Evaluating self-organizing maps for geovisualization. *Exploring Geovisualization*.
Koua, E.L., Maceachren, A. & Kraak, M.J. (2006) Evaluating the usability of visualization methods in an exploratory geovisualization environment. *International Journal of Geographical Information Science,* **20** (4), 425-448.
Koussoulakou, A. & Kraak, M. (1992) Spatia-temporal maps and cartographic communication. *Cartographic Journal,* **29** (2), 101-108.
Kraak, M.-J. (2003) Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing,* **57** (5-6), 390-399.
Kraak, M.-J. (2007) Geovisualization and visual analytics.

Cartographica, **42** (2), 115-116.
Krycier, J. (1994) Sound and geographic visualization. *Visualization in modern cartography*, 149.
Ladner, R., Petry, F. & Cobb, M. (2003) Fuzzy set approaches to spatial data mining of association rules. *Transactions in GIS,* **7** (1), 123-138.
Langran, G. & Chrisman, N. (1988) A framework for temporal geographic information. *Cartographica: The International Journal for Geographic Information and Geovisualization,* **25** (3), 1-14.
Lee, H. & Ong, K. (1996) Visualization support for data mining. *IEEE Expert,* **11** (5), 69-75.
Maceachren, A., Boscoe, F., Haug, D. & Pickle, L. (1998) Geographic visualization: Designing manipulable maps for exploring temporally varying georeferenced statistics. *Proceedings, Information Visualization,* 98, 19-20.
Maceachren, A.M. & Kraak, M.-J. (1997) Exploratory cartographic visualization: Advancing the agenda. *Computers & Geosciences,* **23** (4), 335-343.
Maceachren, A., M., Wachowicz, M., Edsall, R., Haug, D. & Masters, R. (1999) Constructing knowledge from multivariate spatiotemporal data: Integrating geographical visualization with knowledge discovery in database methods. *International Journal of Geographical Information Science,* **13** (4), 311-334.
Maceachren, A.M., Brewer, I., Cai, G. & Chen, J. (2003) Visually-enabled geocollaboration to support data exploration and decision-making. *21st International Cartographic Conference.* Durban, South Africa.
Maceachren, A.M. & Kraak, M.-J. (2001) Research challenges in geovisualization. *Cartography and Geographic Information Science,* **28** (1), 1-11.
Maguire, D. (2007) The changing technology of space and time. *In* Drummond, J., Billen, R., JoãO, E. & Forrest, D. eds. *Dynamic and mobile gis: Investigating changes in space and time.* Boca Raton, FL: CRC Press.
Mannila, H. (1997) Methods and problems in data mining. *Database Theory XICDT'97*, 41-55.
Mark, D.M. (1993) Toward a theoretical framework for geographic entity types. *COSIT '93 Proceedings: Spatial Information Theory*, Italy: Springer-Verlag, 270-283.
Masters, R. & Edsall, R. (2000) Interaction tools to support knowledge discovery:a case study using data explorer and tcl/tk. *Visualization Development Environments 2000 Proceedings*.
Matheus, C.J., Chan, P.K. & Piatetsky-Shapiro, G. (1993) Systems for

knowledge discovery in databases. *Knowledge and Data Engineering, IEEE Transactions on,* **5** (6), 903-913.
Miller, H. & Han, J. (2001) *Geographic data mining and knowledge discovery*: CRC Press.
Miller, H.J. & Han, J. (2009) Geographic data mining and knowledge discovery: An overview. *In* Miller, H.J. & Han, J. eds. *Geographic data mining and knowledge discovery.* Second ed. London: Taylor & Francis, 1-26.
Morrison, J., Tversky, B. & Betrancourt, M., (2000) Animation: Does it facilitate learning, 53-59.
Muller, W., Nocke, T. & Schumann, H. (2006) Enhancing the visualization process with principal component analysis to support the exploration of trends. Australian Computer Society, Inc., 130.
Nelson, E.S., Dow, D., Lukinbeal, C. & Farley, R. (1997) Visual search processes and the multivariate point symbol. *Cartographica,* **34** (4), 19.
Ng, R. & Han, J. (2002) Clarans: A method for clustering objects for spatial data mining. *IEEE transactions on knowledge and data engineering*, 1003-1016.
Patton, D. & Cammack, R. (1996) An examination of the effects of task type and map complexity on sequenced and static choropleth maps. *Cartographic design: Theoretical and practical perspectives*, 237-252.
Peirce, C. (1891) The architecture of theories. *The Monist,* 1 (2).
Penn, B.S. (2005) Using self-organizing maps to visualize high-dimensional data. *Computers & Geosciences,* **31** (5), 531-544
Pike, W. & Gahegan, M. (2007) Beyond ontologies: Toward situated representations of scientific knowledge. *International Journal of Human-Computer Studies,* **65** (7), 674-688.
Quinlan, J.R. (1993) *C4.5* San Mateo, Calif. :: Morgan Kaufmann Publishers.
Reichenbach, H., Reichenbach, M., Freund, J. & Carnap, R. (1958) *The philosophy of space & time*: Dover Pubns.
Robinson, A.C. (2005a) *Assessing geovisualization in epidemiology: A design framework for an exploratory toolkit*. The Pennsylvania State University.
Robinson, A.C. (2005b) Geovisualization and epidemiology: A general design framework. *Proceedings of the International Cartographic Association.* La Coruna, Spain.
Robinson, A.C., Chen, J., Lengerich, E.J., Meyer, H.G. & Maceachren, A.M. (2005) Combining usability techniques to design geovisualization tools for epidemiology. *Cartography and Geographic Information Science,* **32**, 243-255

Roddick, J., Hornsby, K. & Spiliopoulou, M. (2001) An updated bibliography of temporal, spatial, and spatio-temporal data mining research. *Lecture Notes in Computer Science,* 2007, 147-164.

Roddick, J. & Spiliopoulou, M. (1999) A bibliography of temporal, spatial and spatio-temporal data mining research. *ACM SIGKDD Explorations Newsletter,* **1** (1), 34-38.

Roddick, J.F. & Lees, B.G. (2009) Spatio-temporal data mining paradigms and methodologies. *In* Miller, H.J. & Han, J. eds. *Geographic data mining and knowledge discovery.* Second ed. London: Taylor & Francis, 27-44.

Roddick, J.F. & Spiliopoulou, M. (2002) A survey of temporal knowledge discovery paradigms and methods. *Knowledge and Data Engineering, IEEE Transactions on,* **14** (4), 750-767.

Schrodinger, E. (1985) *Space-time structure*: Cambridge University Print.

Sekhar, S., Chang-Tien, L., Pusheng, Z. & Rulin, L. (2002) Data mining for selective visualization of large spatial datasets, 41-48.

Shekhar, S., Lu, C., Chawla, S. & Zhang, P. (2000) Data mining and visualization of twin-cities traffic data. *Dept. of Computer Science Technical Report TR 00-015, U. of Minnesota*.

Shneiderman, B. (1992) Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG),* **11** (1), 92-99.

Slocum, T. & Egbert, S. (1993) Knowledge acquisition from choropleth maps. *Cartography and Geographic Information Science,* 20 (2), 83-95.

Slocum, T.A., Blok, C., Jiang, B., Koussoulakou, A., Montello, D.R., Fuhrmann, S. & Hedley, N.R. (2001) Cognitive and usability issues in geovisualization. *Cartography and Geographic Information Science,* **28**, 61-75.

Sowa, J. & Majumdar, A. (2003) Analogical reasoning. *Lecture Notes in Computer Science*, 16-36.

Stolte, C., Tang, D. & Hanrahan, P. (2003) Multiscale visualization using data cubes. *IEEE Transactions on Visualization and Computer Graphics*, 176-187.

Swayne, D., Lang, D., Buja, A. & Cook, D. (2003) Ggobi: Evolving from xgobi into an extensible framework for interactive data visualization. *Computational Statistics and Data Analysis,* **43** (4), 423-444.

Swienty, O., Reichenbacher, T., Reppermund, S. & Zihl, J. (2008) The role of relevance and cognition in attention-guiding geovisualisation. *Cartographic Journal,* **45**, 227-238

Takatsuka, M. & Gahegan, M. (2002) Geovista studio: A codeless visual programming environment for geoscientific data analysis and

visualization *Computers and Geosciences,* **28** (10), 1131-1144.
Tang, J., Chen, Z., Fu, A. & Cheung, D. (2002) Enhancing effectiveness of outlier detections for low density patterns. *Advances in knowledge discovery and data mining.* 535-548.
Tang, J., Chen, Z., Fu, A. & Cheung, D. (2007) Capabilities of outlier detection schemes in large datasets, framework and methodologies. *Knowledge and Information Systems,* **11** (1), 45-8.
Uhlenkuken, C., Schmidt, B. & Streit, U. (2000) Visual exploration of high-dimensional spatial data: Requirements and deficits. *Computers & Geosciences,* **26** (1), 77-85
Wachowicz, M. (2000) The role of geographic visualization and knowledge discovery in spatio-temporal modeling. *Publications on Geodesy,* **47**, 27 - 35.
Wan, Y. & Bian, F. (2008) Cell-based outlier detection algorithm: A fast outlier detection algorithm for large datasets. *Advances in knowledge discovery and data mining.* 1042-1048.
Wang, L., Xie, K., Chen, T. & Ma, X. (2005) Efficient discovery of multilevel spatial association rules using partitions. *Information and Software Technology,* **47** (13), 829-840.
Worboys, M., (1992) A model for spatio-temporal information. *In*: Bresnahan, P., Corwin, E. & Cowen, D., ed. *Proceedings of the 5th International Symposium on Spatial Data Handling*, Charleston, SC: IGU Commission of GIS, 602-611.
Yang, H. & Srinivasan, P., (2006) Mining spatial and spatio-temporal patterns in scientific data. *IEEE Conference on Data Engineering Workshops*, Atalanta, 146.
Yuan, M. (1996) Temporal GIS and Spatiotemporal Modeling. *Integrating GIS and Environmental Modeling* edited by M. Goodchild (CD-ROM).
Yuan, M. (1997) Knowledge acquisition for building wildfire representation in geographic information systems. *The International Journal of Geographic Information Systems,* **11** (8), 723-745.
Yuan, M. (2009) Toward knowledge discovery about geographic dynamics in spatiotemporal databases. *In* Miller, H.J. & Han, J. eds. *Geographic data mining and knowledge discovery.* Second ed. London: Taylor & Francis, 347-365.
Yang, Q., Wu, X., Han, J., Heckerman, D., Keim, D., Liu, J., Madigan, D., Piatetsky-Shapiro, G., Raghavan, V. & Rastogi, R. (2006) 10 challenging problems in data mining research. *International Journal of Information Technology and Decision Making,* **5** (4), 597-604.
Yudong, C., Yi, Z., Jianming, H. & Danya, Y. (2006) Pattern discovering

of regional traffic status with self-organizing maps. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 647-652.

Zhang, H.S., Zhang, Y., Li, Z.H. & Hu, D.C. (2004) Spatial-temporal traffic data analysis based on global data management using mas. *Ieee Transactions on Intelligent Transportation Systems,* **5** (4), 267-275.

Zhang, J., Gao, Q., Wang, H., Liu, Q. & Xu, K. (2009) Detecting projected outliers in high-dimensional data streams. *Database and expert systems applications.* 629-644.