

逢甲大學

土地管理學系碩士班

碩士論文

應用 DBSCAN 群聚演算系統於空間資料之分析

Spatial Data Analysis Base on DBSCAN
Clustering Algorithm

指導教授：周天穎

研究生：蔡宗欣

中華民國九十七年六月

逢 甲 大 學

土地管理學系碩士學位論文

應用 DBSCAN 群聚演算系統於空間資料之分析
Spatial Data Analysis Base on DBSCAN
Clustering Algorithm

研究生：蔡宗欣

經碩士學位考試合格特此證明

評審委員

楊龍士

劉永禧

周宜強

指導教授

周天領

系主任

洪本善

考試日期：97年6月5日

謝 誌

二年的時間說長不長，能夠如期完成碩士學業以及論文的撰寫真的要感謝很多人，此刻的心情有點複雜，二年的學習開啟了對研究的熱誠但也發現自己的不足。從不曾想過自己可以念到研究所，其實我也曾經是逢甲人，只是年少無知的我未能完成，心中總是有些遺憾，1993年進入臺中市政府後就跟 GIS 結緣，2001年心裏萌生了想再進修的念頭，有幸能夠再回到逢甲，心中早就下定決心兩年一定要完成。

二年的時光過得很快，如今要感謝周天穎老師的指導，引導我的思考邏輯讓我得到許多的啟發。感謝口試委員蕭家旗副市長、楊龍士副校長、邱景升老師，提供許多寶貴的建議。在學期間，感謝雷祖強老師、楊龍士老師、洪本善老師、謝靜琪老師、楊文燦老師的教導，激勵大家更多的思考及發揮，了解自己並表現自己，戰場上孤軍奮戰是很寂寞，還好有卓惠玲同學並肩作戰，時時叮嚀。感謝市府給予本人進修的機會，主管及其他同仁的體諒，因為有你們的幫忙與鼓勵，讓我能在工作及學業上兼顧，最後還要感謝是我最愛的老婆及二個寶貝，有他們的支持才能如期完成。

每學期的 Seminar 真是讓人又愛又怕，我們一起討論，一起被檢討，一起看到大家的成長及茁壯，從茫茫然到知道自己的方向。感謝我的爸媽除了精神上的鼓勵外還提供兩年學費獎勵，讓我沒有後顧之憂，我覺得自己很幸福。最後，謹以此論文，獻給曾經支持與協助過的我親友，使我能順利的完成碩士學業，在致上最誠摯的敬意與謝意。

蔡宗欣 謹誌 2008 年 7 月

摘要

本研究以 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)群聚演算法為例，使用台中市政府工商資料庫 96 年 12 月 31 日止 2,061 家合法登記生產中之工廠資料為測試範圍。再配合台中市 GIS(Geographic Information System)地理資訊倉儲中千分之一數值航照地形圖資料庫所提供的地址門牌點位坐標檔 494,028 筆資料，使用自行開發之系統，提供比對程式，自動完成地址欄位比對及坐標匯入。DBSCAN 演算程式部份，採用資料庫重複搜尋方式來開發執行，透過輸入變數半徑(Eps)及門檻值(Pict)之設定，可直接產生不同的群聚資料成果。將成果資料直接匯入一般 GIS 地理資訊軟體(本研究使用 MAPINFO 7.0)搭配原有 GIS 資料倉儲中提供的圖資，如都市計劃圖、行政區圖、道路街廓圖等資料，套圖分析展現空間群聚成果。

本研究是利用資料庫的搜尋功能來協助達成空間群聚演算法，因為採用資料庫重複搜尋的方式取代原來矩陣搜尋，因此減少矩陣搜尋及矩陣化簡運算。同時解決陣列數目過大記憶體不足之現象。此對於執行運算之電腦相關設備等級需求不用太高，只要使用一般等級的電腦及資料庫就能分析處理龐大 GIS 空間資料。且資料庫採用目前個人電腦最普遍支援的 xBASE 資料庫(檔案格式為.dbf)，是相當經濟且方便的分析工具。

關鍵字：DBSCAN、空間群聚、地理資訊

Summary

This research performs the example of clustering algorithms with DBSCAN (Density-Based Spatial Clustering of Applications with Noise), uses Taichung government industrial and commercial database, which contains 2061 legally registered factories which were at work by the data of 12/31/2007, as the test domain. And then to coordinate with 494,028 data records which contain address doorplate and position coordinates in the database of 1/1000 aerial photography terrain data, which is located in Taichung City Government GIS data warehouse, uses self-developed system to provide comparison of program and accomplishment of automatically addresses comparison and coordinates import. The part of DBSCAN algorithm uses the iterative query and search ability of database system to perform in this research. Through the setting of radius (Eps) and threshold (Pict), the approach can directly produce different clustering results. The result can be imported into some general GIS systems (MAPIFO 7.0 in this research) as well as cooperating with terrain data in GIS data warehouse, such as urban project map, administrative area map, road and street profile map,...etc, to present the analyzed results of spatial clustering.

This research makes use of query and searching ability of DBMS to help to achieve spatial clustering algorithm. As introducing the iterative searching approach in a database system to substitute original matrix searching, the matrix operations such as searching in matrix will be reduced in our method. The approach in this dissertation also solves the main memory insufficiency problem while adopting large dimension of matrix, it also helps to process large scale GIS spatial data with less requirement for computer capabilities such as running speed of CPU, memory size,...etc. The approach also adopts xBASE database (.dbf as the file format), which was general supported by current PC database system, provided as a economical and convenient analytic tool.

Keywords: DBSCAN, spatial clustering, GIS

目 錄

第一章、緒 論	1
第一節、問題陳述與分析.....	1
第二節、研究目的.....	4
第三節、研究範圍及限制.....	6
壹、研究範圍.....	6
貳、研究限制.....	6
第四節、研究內容與流程.....	7
壹、研究內容.....	7
貳、研究流程.....	8
第二章、文獻回顧	9
第一節、群聚演算法之比較.....	9
第二節、DBSCAN演算法.....	16
第三章、研究設計	18
第一節、資料庫選擇.....	18
第二節、DBSCAN群聚演算法.....	21
第三節、系統開發工具.....	24
第四章、實證分析	26
第一節、資料庫轉換.....	26
第二節、DBSCAN群聚分析系統.....	27
壹、地址標準化.....	27
貳、地址比對定位.....	27
參、DBSCAN群聚分析.....	28
第三節、結果分析.....	30

第五章、結論與建議	46
第一節、研究結論.....	46
第二節、研究建議.....	47
第六章、參考文獻	48
附錄一、程式執行原始碼.....	53
附錄二、DBSCAN演算系統模組結構表.....	57
附錄三、歐基里德距離對稱矩陣.....	58
附錄四、DBSCAN空間群聚演算系統.....	59
附錄五、台中市工商資料庫(Microsoft SQL SERVER 2000).....	62
附錄六、台中市門牌點位資料(MAPINFO 7.0).....	63
附錄七、開發程式工具主畫面(Dbtools for windows 32bit)...	65
附錄八、台中市政府GIS資料倉儲系統(數位地圖館).....	66



圖 目 錄

圖 1 研究流程圖.....	8
圖 2 群聚技術發展概況圖.....	14
圖 3 群聚算法比較表.....	15
圖 4 DBSCAN搜尋方法圖示.....	16
圖 5 DBSCAN演算流程圖.....	23
圖 6 DBTOOLS系統示意圖.....	25
圖 7 資料庫轉換流程圖.....	26
圖 8 DBSCAN群聚分析系統.....	28
圖 9 DBSCAN群聚分析系統.....	29
圖 10 工廠分佈圖(工廠數量=2061 家).....	30
圖 11 台中市都市計畫圖.....	31
圖 12 台中工業區工廠分佈圖.....	31
圖 13 DBSCAN群聚演算成果比較表(1).....	33
圖 14 Eps=200M Minpts=50.....	34
圖 15 Eps=300M Minpts=50.....	35
圖 16 Eps=400M Minpts=50.....	36
圖 17 Eps=500M Minpts=50.....	37

圖 18	Eps=600M Minpts=50	38
圖 19	群聚數量與Eps值折線圖	39
圖 20	DBSCAN群聚演算成果比較表(2).....	40
圖 21	Eps=500M Minpts=200	41
圖 22	Eps=500M Minpts=150	42
圖 23	Eps=500M Minpts=100	43
圖 24	Eps=500M Minpts=50	44
圖 25	群聚數量與MinPts值折線圖	45



第一章、緒論

第一節、問題陳述與分析

資料挖掘(Data Mining)又稱為資料採礦，一般也有稱為知識發現，主要是指從資料庫中快速簡易的找出未知或隱含其中有用的資訊，提供該資訊給予決策與管理之應用。群聚分析就是其中的一種方法，然而群聚分析演算的方式直至目前為止還是有很多種方法，至於針對 GIS (Geographic Information System, 簡稱 GIS) 空間資料來做群聚分析的工具卻不是很多，一般 GIS 系統如 ArcGIS、Mapinfo 等都只提供簡易的分析功能如環域分析與搜尋、最短路徑等，其他功能大部份都要透過客製化的方式來開發系統，因此要完成本研究之 DBSCAN 演算功能同樣也需要自己來開發系統，在參考許多文獻資料中大多是使用 GIS 系統本身所提供之開發介面來達成，極少有使用其他程式語言來撰寫，在考慮 DBSCAN 演算法本身對於空間分析的需求只有使用到環域搜尋及距離計算，因此本研究大膽採用一般資料庫開發工具來取代，利用資料庫本身所提供之基本工具(如搜尋、索引、迴圈等)來完成群聚演算及分群工作。

在平面空間群聚分佈來說，不考慮地形高低變化及河流屏障限制之下 DBSCAN(Density-Based Spatial Clustering of Applications with Noise)演算方法是一種較為有效率的演算方法，能快速完成不規則形狀之群聚，但是對於半徑及門檻值卻是相當敏感，因此這兩個參數之選定需要透過實際運算後才能判別出來，因此演算程式的便利性是最大的關鍵，傳統的演算程

式是求出各點之間的距離，然後是利用矩陣陣列方式展開成一個對稱陣列(如附件三)，當資料庫為N筆資料時展開矩陣就需有 N^2 個元素，因為是對稱矩陣因此可再簡化，但還是需要 $N^2/2$ 個元素，對於一般個人電腦分析演算GIS空間資料來說是一個很大的負擔，如果使用較高階的程式語言在硬體資源不足的情況下，則是大量使用虛擬記憶體(virtual memory)方式來解決(虛擬記憶體就是利用硬碟的記憶空間來模擬主記憶體)，另外一種方法是再透過陣列的化簡運算來減少陣列元素的量來降低佔用記憶體使用空間，但是同樣的陣列簡化運算還是會耗用不少系統資源及時間，如果再加上虛擬記憶體的I/O時間(input/output)，整體的運算時間常常不是一般人所能忍受的，也因此當遇到GIS空間資料時，往往動輒幾十萬筆資料，那一般使用者的容忍程度確實是一大考驗，最後還是只能提升軟硬體設備方式來改善，為解決此問題本研究嘗試使用資料庫搜尋方式來取代矩陣及行列式運算，整個演算過程中沒有使用二維陣列函數，因此並未產生對稱陣列，陣列變數及行列式運算由資料庫搜尋方式來達成，原本由GIS系統中運算的工作交由資料庫完成，也分擔GIS系統中主程式的運算壓力，減少使用虛擬記憶體(virtual memory)，提升整體效率，加速達成演算結果。

台中市政府於民國 78 年積極推動行政資訊化，成立資訊單位(計畫室資訊課)推動資訊相關業務，民國 83 年開使建置「台中市政府工商資料庫」，此業務當時為經濟局工商課主辦，其主要業務針對全台中市公司行號及工廠設立登記與核發業務，單就工廠設立登記而言資料庫累積至今就有 12,820 筆登記資料(其中包含生產中、歇業、歇業-遷廠、歇業-關廠、公告註銷、設立未登記等)，實際有生產的工廠登記也有 3,311 家，登記資料庫中除工廠名稱、統一編號、設立時間、負責人、工廠電話及住址等相

關基本資料外，比較與空間方位相關資料有關的，僅有住址中行政區及鄰里欄位，更無空間坐標資訊，因此以往對於這些工廠的分佈及群聚情形也只能透過區里統計來概略分析，如果能將這些資料空間化並透過圖像式的表現出來，不但能觀察工廠位置分佈情形，配合群聚分析演算分群，利用不同的顏色來表現差異性，可大大提升資料利用價值並從中發掘出更多有用的資訊，為達此功能必需增加空間屬性登記欄位，最簡單的方式就是坐標欄位，對於新登記之資料沒有甚麼問題，對於舊資料部份是一大問題，全面普查測量是一種方式，但是金錢與時間都是一大花費，台中市政府於民國 82 年起推動 GIS 地理資訊系統，先後完成都市計畫圖、街廓圖，民國 85 更獲得內政部補助台中市千分之一數值航照地形圖建置，完成建物門牌點位資料，並且該資料每年定期更新，如此資料能有效應用就能解決前述之問題，本研究嘗試開發此系統，將工商管理系統上的原始資料與建物門牌點位資料做地址比對，比對成功就將門牌點位坐標匯入工商資料庫中，比對不成功再用人工比對，如此可節省大量經費與時間，而台中市千分之一數值航照地形圖建置全案同時依照內政部的規劃區分建置建物、道路、水系、門牌點位等近 200 個圖層資料，為有效管理更新此一龐大圖資，於民國 95 年建置完成台中市 GIS 資料倉儲系統，不但提供本身業務相關需求，也提供一般市民線上查詢及申購，經過空間對位的工商資料就可匯入該資料倉儲中供市府應用，因此在如此豐富的資源下有效搭配利用這些資料必能分析出更多有價值的資訊。

第二節、研究動機

台中市總面積為 163.4 平方公里，東西長 14.2 公里，南北寬 11.3 公里，盆地地形，在不考慮道路路網與地點區位且又沒有河流湖泊或高山之天然屏障，選擇以 DBSCAN 演算法來分析空間分佈之群聚狀況是較為適合，但是該演算法有一個缺點，就是輸入參數”半徑”(Eps)及”門檻值”(Pict)，這二個參數是相當敏感，對於運算成果有直接的關係，必需透過實際運算後看出品質優劣，往往必需多次重複執行運算才能找出較好成果，如何能方便且快速的完成結果是很重要，快速便利的找出最適合之半徑(Eps)與門檻值(Pict)。

目前一般市售 GIS 系統並沒有提供 DBSCAN 演算工具，如果要開發就必需使用該系統所提供之開發工具，雖然大部份開發工具都支援外掛函數，但是遇到一些特殊需求，用外掛函數是無法達成，最後還是只能用它所提供開發介面，各系統間的開發介面也不同，因此就算在甲系統開發之演算程式，是無法拿到乙系統來執行，為有將資料直接演算處理完成後，任何 GIS 系統只要支援此格式檔案就可直接使用，為實現此一目標，本研究將自行開發獨立演算之系統程式，利用一般個人電腦及最普遍的 xBASE 規格之資料庫，其結構和 dBASE III、CLIPPER 都是相同，開發工具 DBTOOLS 之 INT 模組，語法也相容於 CLIPPER 程式語言之語法，以此開發 DBSCAN 群聚演算系統，達成演算成果。

配合國土資訊系統政策之推動，台中市政府多年來已陸續建置多項地理資訊服務系統及資料庫，而為滿足各項業務之需求，亦逐年增加不同的

圖資來源，藉以提昇資料的多元化與完整性。然而，傳統的圖資資料均使用檔案方式進行維護管理，隨著日漸龐大的圖資資料已使得原始的管理方式開始無法負荷。再者，各系統的資料來源、時間並不完全一致，使得系統之間的資料無法進行整合，若要進行圖資更新，各系統還需要協調廠商方能完成。上述這些問題都將隨著系統、圖資資料的成長而成正比，若未能提出有效之解決對策，將成為推動國土資訊系統之落實資料流通共享政策中，最大的阻礙。為解決如此刻不容緩的問題，建立完善機制將圖資資料進行有效率的維護管理，建置「台中市政府地理資訊倉儲管理中心」，整合所有圖資資料、詮釋資料，以提供完整且單一的地理資訊資料服務，實現台中市政府建置地理資訊資料之最大效益。本研究將可提供工廠點位分佈圖，配合資料倉儲系統中提供之道路、區里等資料做更深入之套圖分析，提供市政規劃之重要參考依據為本研究的主要目的。

第三節、研究目的

壹、開發住址比對程式

將千分之一航照資料庫中建物門牌點位檔與工商資料庫中工廠登記資料檔，做地址比對。

貳、開發 DBSCAN 群聚演算系統

利用資料庫搜尋方式達成 DBSCAN 群聚演算，完成工廠群聚分群。

第四節、研究範圍及限制

壹、研究範圍

本研究將針對 GIS 空間資料庫於 DBSCAN 群聚演算法中常會遇到之問題，利用資料庫搜尋法輔助 DBSCAN 演算法來實際驗證，測試資料以台中市政府「工商登記資料公示查詢系統」之資料庫(資料庫格式 Microsoft SQL SERVER 2000)為主，去除歇業、歇業-遷廠、歇業-關廠、公告註銷、設立未登記等之資料後，最後得到合法登記”生產中”之工廠資料，以此作為資料範圍。

台中市在 GIS 基本地形圖資料建置相當完備，為有效管理此一龐大資料庫，台中市政府建置「台中市地理資訊資料倉儲系統」，該系統除了有效率管理 GIS 圖資外，並提供不同年份的版次管理，可依不同時間不同版次來分析應用，本研究除將利用其中之台中市千分之一數值航照地形圖資料及門牌地址點位資料來定位外，還可利用其他圖資來套圖分析，最後利用地理資訊系統(MAPINFO7.0)軟體產出成果圖。

貳、研究限制

本研究將不使用 GIS 系統所提供之開發工具，而是使用資料庫中之搜尋功能取代 GIS 系統中之環域搜尋，針對空間平面坐標相臨距離及密度來分析群聚情形，資料部分將以台中市工商資料庫中狀態欄位為”生產中”之資料為主要分析內容，不考慮資料本身相關屬性資料如產業別、資本額、設立時間等，空間部分僅考慮坐標資料，不考慮道路、區位、土地使用分區、天然屏障(如高山、河流等)、人工屏障(如建築物、橋樑、鐵路等)。

第五節、研究內容與流程

壹、研究內容

本研究內容主要使用台中市千分之一數值航照地形圖門牌檔地址欄位與工商資料庫地址欄位比對得到工廠坐標點位後，將對位成功之門牌點位坐標轉入工商資料庫中 X、Y 坐標欄位，再使用 DBSCAN 演算系統執行群聚分析，分析演算方法是採用一般資料庫之基本搜尋加迴圈功能，重複搜尋檢驗條件參數，條件符合者存入暫存區，當暫存區數量符合第二參數值，則對暫存區各元素再次搜尋檢驗，直至暫存區內所有元素檢驗完成，重複上述方式完成所有資料搜尋，最後完成的結果是可直接使用 GIS 地理資訊系統軟體匯入應用，主要內容說明如下：

1. 建立台中市千分之一數值航照地形圖地址檔欄位及工商資料庫地址欄位標準化及格式轉換。
2. 建立以台中市千分之一數值航照地形圖之地址檔、門牌點位檔與工商資料庫地址欄位比對坐標對位程式。
3. 建立以資料庫搜尋方式之 DBSCAN 群聚演算系統，提升加強 GIS 空間資料演算之效率及便利性。
4. 產製成果檔可直接由 GIS 地理資訊系統軟體匯入，並可搭配台中市現有 GIS 資料倉儲中所有資料，可以使用疊圖與分析等多項功能，還可將分析資料回饋給資料倉儲再利用。

貳、研究流程

本研究流程分為資料收集、資料篩選、資料比對定位、資料群聚分析運算、套圖比對分析、成果產出等五大步驟：

步驟一、資料收集，千分之一數值航照地形圖之地址檔、門牌點位檔、工商資料庫。

步驟二、資料篩選，主要是使用工商資料庫中狀態欄位為”生產中”之資料為本研究之主要資料。

步驟三、資料比對定位，千分之一數值航照地形圖之地址欄位與工商資料庫地址欄位，比對定位。

步驟四、資料群聚分析運算，執行 DBSCAN 群聚演算系統，產生成果檔。

步驟五、成果檔產出，匯入 GIS 系統套圖比對分析。



圖 1 研究流程圖

第二章、文獻回顧

第一節、群聚演算法之比較

壹、切割式群聚演算法

最早期發展的群聚技術是切割式群聚演算法，這類演算法使用者通常是必需先決定所要分割群聚的數目，再利用重心點基礎 (Centroid-based) 或中心點基礎 (Medoid-based) 的方式進行分群演算。該演算法是利用距離為評估依據，評估的方法有：曼哈頓距離 (Manhattan distance) 與歐幾里得距離 (Euclidean distance)。K-means 是最典型以重心基礎的切割式群聚演算法 [Mac67]，它以群聚的重心作為群聚的代表點 (Representative object)，因為代表點不一定要是群聚中的一點，所以可以找到最佳的群聚。而，此方法所得的群聚的品質很容易受到雜訊 (Noise) 或是離群值 (Outlier) 所影響。另一種方法是以中心點作為代表點 (如 PAM 演算法 [KR90])，這些群聚技術對於小型的資料集合 (Data sets) 有著不錯的處理能力，但是隨著資料集合的增加，處理的效率也越來越差，所以，通常處理大型資料庫是採用取樣的方式來解決 (如 CLARA 演算法 [KR90])。然而取樣的演算法會受到樣本的數量以及取樣方法所影響，倘若樣本數量太少，則群聚的結果不足以代表整個資料庫資料分佈的狀況及意義；若取樣的方法不佳，則會影響到群聚的品質。CLARANS [NH94] 是架構於 PAM 與 CLARA 上的中心點基礎的切割式演算法，它是第一個針對於空間資料庫所設計的切割式群聚演算法，但只能發掘簡易的資料點分佈形狀 (Object shapes)，對於呈凸多邊形 (Convex shapes) 或是巢狀 (Nested) 分佈的資料處理效

果不佳，也無法有效率了的針對高維度資料（High-dimensionality data）進行群聚分析。

貳、階層式群聚演算法

該演算法呈現一種樹狀的架構，可分成凝聚法（Agglomerative）及分裂法（Divisive）。凝聚法是由下而上（Bottom-up）一開始我們將資料庫內每一筆資料均表示成一個群聚，接著我們依資料屬性的相似度開始做合併，每一次合併兩個相似度最高的群聚，直到我們所設定的終止群聚數目為止。而分裂法與凝聚法相反，它是為由上而下（Top-down）一開始將資料庫內所有的資料視為同一個群聚，然後將資料相似度較低的分裂呈現出不同群聚表現，一直到群聚的數目到達設定之終止條件為止。絕大部分階層式演算法是由下而上凝聚式的方式，而目前的演算法有 AMOEBA [EL00a]、AUTOCLUST [EL00b]、BIRCH [ZRL96]、CURE [GRS98]、CHAMELEON [KHK99]、ROCK [GRS99]等六種。

CURE 演算法是以群聚之間的相似（Closeness）程度做為合併依據，它不同於以往中心點或重心點的群聚演算法而是以一個代表點來取代一個群聚，而 MST（Minimum spanning tree）方法，是全部的點皆是代表點，CURE 演算法固定選擇 C 個點（ $2 \leq C \leq N$ ， N 為全部資料集合）代表群聚，接著將全部的代表點向重心收縮參數 α （ $0 \leq \alpha \leq 1$ ），以此將相似的群聚合併，直到群聚數為止。CURE 演算法適用於大型的資料集合，可以辨識任何形狀的群聚，對離群值處理較完整，但它不處理屬性（Categorical）資料。ROCK 演算法強調不同群聚間的資料鏈結（inter-connectivity）關係，它對於類別屬性（Categorical）有較好的處理能力。

ROCK 演算法著重於群聚間互相鏈結關係，忽略了群聚相似程度，而 CURE 演算法著重於群聚相似程度，忽略不同群聚間鏈結的程度，在 CHAMELE 演算法中針對 CURE 演算法與 ROCK 演算法的缺點，提出修正。它是以相對互相鏈結 (Relative Interconnectivity, RI) 與相對相似 (Relative Closeness, RC) 為基礎的階層式群聚演算法，它先以最近鄰近法 (K-nearest neighbor) 將所有的資料聚集為一群聚，接著將群聚切割為若干子群聚 (Sub-clusters)，最後以 RI 與 RC 為評估方法，將子群聚合併直到所訂定的群聚數目。CHAMELEON 對於群聚的形狀有較好的辨識能力。

階層式群聚演算法也被使用於空間資料庫上，例如 AMOEBA、AUTOCLUST 演算法等。這兩種演算法皆屬於分裂式演算法，AMOEBA 是利用 Delaunay 三角形 (Delaunay triangles) 計算出所有資料的最近鄰居並連接成 Delaunay 圖形 (Delaunay diagram)，也就是將資料集合聚集為一個群聚，接著計算圖形上所有連結邊 (Edges) 總合平均數 (Means)，剔除所有大於平均數的邊，將群聚分裂成若干個子群聚，在第二層分析中針對每一個子群聚計算其群聚計算所有邊總合平均數，剔除該群聚大於平均數的邊，以漸層式的分析，直到理想的群聚數目為止。AMOEBA 使用漸層式的分析，在每一層裡，子群聚個別計算其區域性之參數，這樣的演算法適用於地理資訊系統上，而 AUTOCLUST 演算法與 AMOEBA 類似，不同的是 AUTOCLUST 去除邊的標準差以外特別長或短的邊，這兩種演算法的時間複雜度只需要 $O(n \log n)$ ，且不需輸入任何的參數，也可發現任意的群聚形狀。

參、密度基礎群聚演算法

於一資料集合內，假設有某些資料點分佈相當密集，則我們將這些資料點視為一個群聚，也就是在群聚內資料分佈的密度應該大於群聚外資料分佈的密度，而密度基礎群聚演算法，是基於以上的觀念所發展出來的群聚方法，較重要的方法有 DBSCAN [EKSX96]、OPTICS [ABK99]。

DBSCAN 演算法是較早使用密度觀念處理群聚問題的演算法，使用者只需設定鄰近區域的半徑 (Eps) 及鄰近區域的資料點數 (MinPts) 最小值兩個參數，只要在半徑 Eps 鄰近區域內的資料點數大於門檻值 MinPts，則成群聚，接著繼續開始向外擴展群聚範圍，利用使用者所訂定參數所形成的鄰近區域，密度到達 (Directly density-reachable) 或是藉由其他資料點，密度間接到達 (Density-reachable) 包含到的資料點，分配到此一群聚，以此方法，一直到所有的資料點完成分配為止。利用密度方法之特性，DBSCAN 可以有效分辨出並控制離群值，但是對於任意形狀群聚的辨識效果較差，特別是參數設定，常會嚴重影響到群聚品質與群聚結果，這些參數必需要根據經驗或是觀察成果來判斷，使用者很難決定此參數。為解決此問題，OPTICS 演算法是以計算一個密度基礎之分群順序方式，它擴充 DBSCAN 演算法，根據此順序自動去處理參數，OPTICS 與 DBSCAN 演算法的結構是一樣的，其時間複雜度皆為 $O(n \log n)$ 。而 DBCLASD [XEKS98] 是另一種不需要輸入參數的演算法，它是用分配基礎群聚演算法 (Distribution based clustering algorithms)，作者觀察群聚內資料點距離最近鄰居的距離會小於到群聚外的距離，換句話說，在同一群聚內資料點到最近鄰居距離短的出現頻率應該高於距離長的頻率，如此會呈現機率分配，雖然 DBCLASD 不需輸入參數，但其計算執行時間大於 DBSCAN。

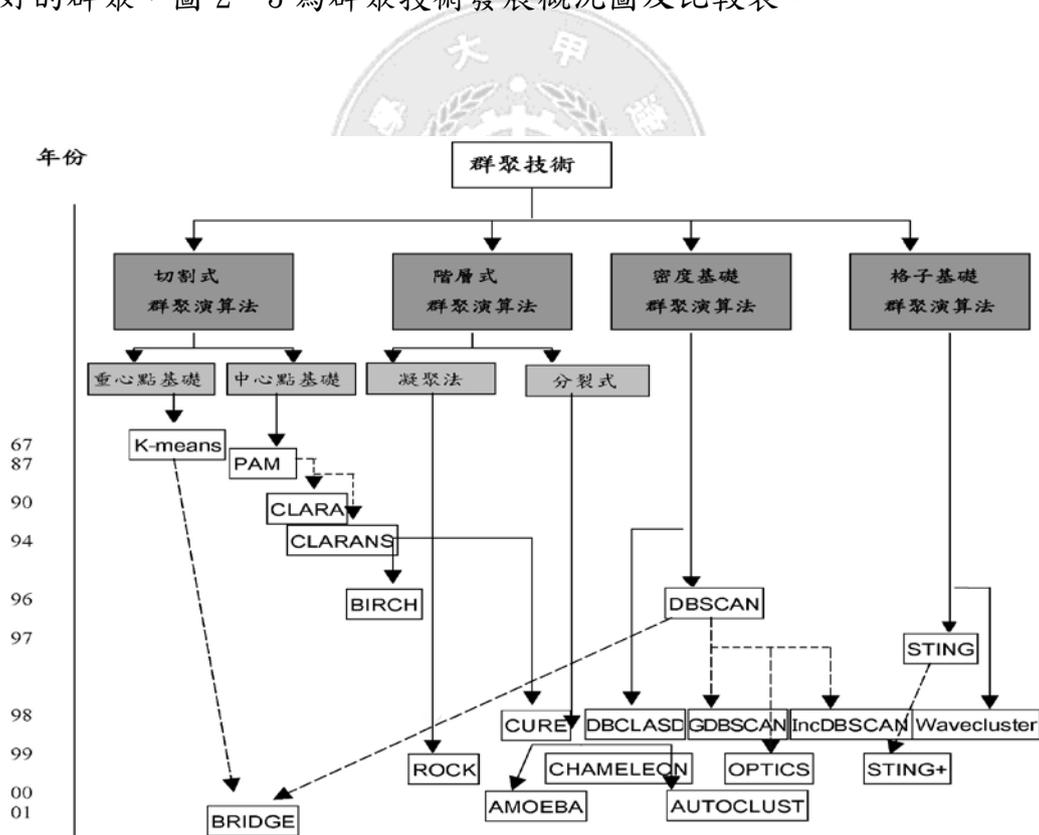
以往的群聚演算法大多只討論靜態資料庫的群聚，而動態資料庫群聚處理更為複雜，IncDBSCAN [EKSWX98] 便是針對在資料倉儲環境下處理群聚問題的密度基礎群聚演算法，因為所要處理的資料集合不是固定，所以需考慮到有資料加入時，對原來群聚的影響。同樣的，在刪除資料情況中，我們也要考慮群聚否會因為刪除資料點而分裂成兩個或多個群聚，或是使原本的群聚變成雜訊。

肆、格子基礎群聚演算法

格子基礎群聚演算法將資料空間量化成許多格子，此方式，大量的減少群聚的時間。在這類演算法中，較具代表性的 STING [WYM97]，WaveCluster [SCZ98]。STING 演算法由上而下的，利用廣度的搜尋，將格子內的群聚作合併。STING 探索存在格子的統計資訊，然後群聚，其缺點是分群形狀其邊不是水平就是垂直，儘管有快速處理時間，但會有損品質及正確率。此外，STING 是呈現階層式架構，較高階層會儲存較低階層資訊總合，所以查詢的速度非常快速，其時間複雜度為 $O(k)$ (k 為最底層格子的數)。

Agrawal 等學者於 1998 年提出 CLIQUE 演算法 [AGR98]，其使用投影 (Projection) 方式，將分佈高維度空間的資料點投影至較低維度的空間上，並且利用群聚分佈密度稠密的特性，若是資料點在高維度的空間中形成群聚，將之投影至低維度空間，必定可以在某個較低維度找到高的投影次數而形成群聚，以此方式找出部分維度的群聚，這樣的方法稱為部分維度群聚演算法 (Subspaces clustering algorithms)。CLIQUE 演算法採用 Apriori

的方式，首先將資料庫中所有的資料投影至一維的屬性空間上，計算投影次數（稱為 Coverage），若是 Coverage 高於次數門檻值者，稱為稠密單位（Dense units），接著合併兩個一維屬性稠密單位，形成二維屬性空間候選單位（Candidate units），再檢查二維空間候選單位是否稠密單位，以此計算到最高維度止，最後使用連通（Connected component）演算法，將相鄰的稠密單元合併，而形成存在於部分維度的群聚。ENCLUS 以密度（Density）[CFZ99]、投影次數（Coverage）與維度的相關性(Correlation of dimensions)來評估群聚性質，再以數學證明方式將三種評估方式轉成熵值（Entropy）的計算，應用於群聚分析的過程，並藉由熵值找出部分維度中較好的群聚，圖 2、3 為群聚技術發展概況圖及比較表。



群聚技術發展概況(實線代表新方法，虛線代表由上游方法所衍生之演算法)

(資料來源：第三屆網際網路應用與發展研討會 2002 陳榮昌、林育臣)

圖 2 群聚技術發展概況圖

演算法類型	演算法名稱	效能	適用資料庫類型	是否需要參數
切割式群聚演算法	k-means	$O(nkt)$, n 代表總數量, k 為分群數, t 為程式迴圈數	數值型資料	分群數 k
	PAM	$O(k(n-k)^2)$	數值型資料	分群數 k
	CLARA	$O(k(10-k)^2+k(n-k))$	數值型資料	分群數 k
	CLARANS	$\theta(kn)^2$	空間資料庫	最大鄰居數(Maxneighbor)
				局部最小值 (Numlocal)
階層式群聚演算法	CURE	低維度: $O(n^3)$, 高維度: $O(n^2 \log n)$	數值型資料	收縮參數 α
	ROCK	$O(n^2+nm_m m_a+n^2 \log n)$, m_m 為最大鄰居數, m_a 為平均鄰居數	類別型資料	需要參數
密度基礎群聚演算法	DBSCAN	$O(n \log n)$	空間資料庫	鄰居區域半徑 (Eps) 資料點數門檻值 (MinPts)
格子基礎群聚演算法	CLIQUE	$O(c^d+nd)$ d 為維度數 c 為常數	空間資料庫	ξ 代表切割格子的寬度 τ 表示投影至該格子之數量門檻值
	STING	$O(k)$, k 為最底層格子的數),	空間資料庫	需要參數

圖 3 群聚算法比較表

(資料來源：第三屆網際網路應用與發展研討會 2002 陳榮昌、林育臣)



第二節、DBSCAN 演算法

DBSCAN 演算法即是基於密度的一種群聚演算方法，由 Ester Martion 等人於 1996 年提出。該演算方式將具有高密度的區域劃分為一群，並可以在帶有“雜訊”(Noise)的空間資料中發現任意形狀的群聚。此演算法是利用群聚的密度連通性可以快速發現任意形狀的群，主要運算方法是對於指定範圍中的每個對象，再給其指定之半徑的領域中包含的對象不能少於某一給定的最小數目，為了發現一個群，DBSCAN 先從對象集合 A 中找到任意一對象 P，並搜尋 A 中關於半徑 Eps 和最小對象數 Minpts 的從 A 密度可達的所有對象，如果 P 是核心物件，即半徑為 Eps 的 P 的鄰域 (neighborhood) 中包含的物件不少於 Minpts，則根據演算法，可以找到一個關於參數 Eps 和 Minpts 的群，如果 P 是一個邊界點，則半徑為 Eps 的 P 鄰域包含的對象少於 Minpts，P 被暫時標註為雜訊點，然後，處理 A 中的下一個物件，透過不斷執行區域查詢來完成該群(如圖 4 圖示)。

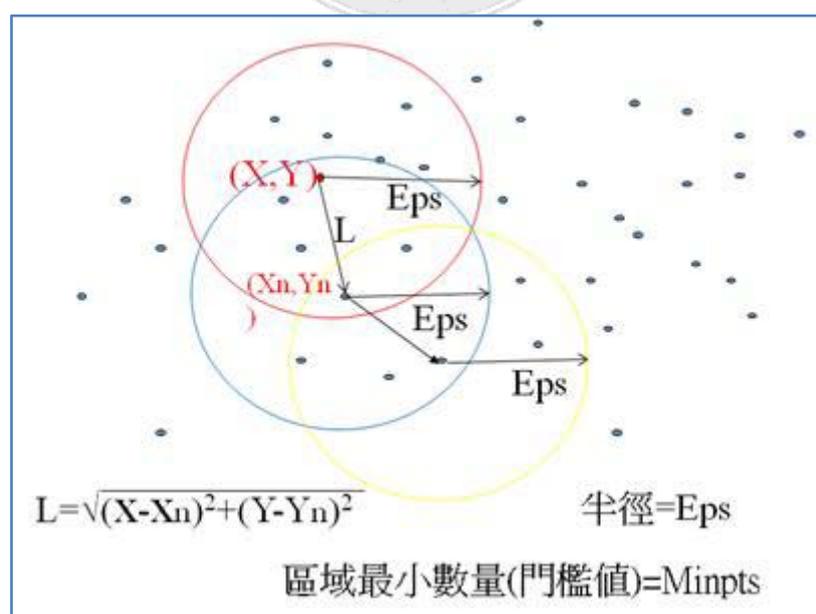


圖 4 DBSCAN 搜尋方法圖示

為了有效的執行區域查詢，DBSCAN 演算法使用了空間查詢 R-樹結構，另外，DBSCAN 要求用戶指定一個全域參數 Eps(為了減少計算量，需預先確定參數 Minpts)，為了確定該數值，DBSCAN 演算過程中首先，利用歐基里德距離公式(距離= $\sqrt{(X - X_n)^2 + (Y - Y_n)^2}$)，計算出任意對象與它的第 k 個最臨近的對象間距離，然後，根據求得的距離由小到大排序，並繪出排序後的圖，稱做 k-dist 圖，k-dist 圖中的橫坐標表示數據對象與它的第 k 個最近的對象間的距離；縱坐標為對應於某一 k-dist 距離值的數據對象的個數，R-樹的建立和 k-dist 圖的繪製是非常消耗系統資源及時間，此外，為了得到較好的聚類結果，用戶必需根據 k-dist 圖，通過試探選定一個比較合適的 Eps 值，DBSCAN 演算法不進行任何的預處理而直接對整個資料庫進行群聚操作，當數據量非常大時，就必需有大記憶體支援，I/O 消耗也非常大，其時間複雜度為 $O(n \log n)$ (n 為數據量)，群聚過程中的大部分時間用在區域查詢操作上，DBSCAN 演算法對參數 Eps 及 Minpts 非常敏感，且這兩個參數很難確定，常常需要透過實際試算成果來判斷，此為最大缺點。

第三章、研究設計

第一節、資料庫選擇

本研究之資料來源為台中市工商資料庫，資料庫為Microsoft SQL SERVER 2000 之格式，因開發工具所使用的資料庫為DBASE¹之.dbf格式資料檔，該資料庫有提供.dbf格式檔案匯出之功能，而且在個人電腦及桌上型地理資訊系統中是相當普遍在使用的資料庫，而大部份的軟體程式開發工具都有支援此格式，資料庫最大筆數可到達 10 億筆，適用於本研究的資料範圍故選擇此格式為本系統之資料庫。表 1～3，為本研究開發系統所使用的兩個主要資料檔，門牌點位資料檔欄位定義表及索引標識定義表、工廠資料檔欄位定義表，所有欄位均依原始資料格式轉入，僅增加坐標及分群欄位，欄位長度與欄位型態(Type)也保持原始檔格式，本系統只對使用到的欄位抬頭(Title)才定義，其他欄位均設空白。

台中市 GIS 資料倉儲(Data Warehouse)系統於 2005 年 12 月完成上線，目前共有千分之一地形圖資料庫、都市計畫資料庫、地政資料庫、公共管線資料庫、其他專業資料庫、交通資料庫、環保資料庫、綠化資料庫、重大建設資料庫、社政資料庫等十大資料庫，提供 Mapinfo、ESRI ArcGIS Shape file、Autodesk .dxf 等三種資料格式，Mapinfo 資料格式提供.DBF、.ID、.MAP、.TAB 等四個檔案，由於千分之一地形圖資料庫中門牌點位資料之 DBF 檔案中有完整地址及坐標資料，因此以該資料檔為地址比對定位依據。

1. dBASE是第一個在個人電腦上被廣泛使用的資料庫管理系統 (DBMS)，由Ashton-Tate在CP/M系統上發行，之後又發行Apple II、Apple Macintosh、UNIX、VMS和IBM PC的DOS平臺版本，而在DOS平臺上成為幾年中最暢銷的軟體。但是dBASE並沒有能成功轉換到Microsoft Windows平臺漸漸被Paradox、Clipper、FoxPro和Microsoft Access等新產品所替代。dBASE在1991年被賣給Borland，1999年所有的產品線權利轉賣給新成立的dBASE公司(資料來源：<http://wiki.ccw.com.cn/DBASE>)。

[DBF] 欄位定義								
模組名稱:DBF.94DOOR67				日期:97.05.09		頁次: 1		
Field	Type	Len	Dec	*Title(C)	*Picture(C)	Mode	Inq	*InqCmb
ID	N	11	0	'流水號'	"@X11"			
A	C	8	0	'圖層代碼'	"@X8"			
B	C	8	0		"@X8"			
C	C	4	0		"@X4"			
D	C	4	0		"@X4"			
E	C	5	0		"@X5"			
F	C	6	0	'區'	"@X6"			
G	C	8	0	'里'	"@X8"			
H	C	5	0	'鄰'	"@X5"			
I	C	20	0	'路段'	"@X20"			
J	C	12	0		"@X12"			
K	C	20	0	'巷'	"@X20"			
L	C	16	0	'弄'	"@X16"			
M	C	40	0	'號樓之號'	"@X40"			
N	N	19	9	'經度'	"@X19"			
O	N	19	9	'緯度'	"@X19"			
P	N	11	3	'X坐標'	"@X11"			
Q	N	11	3	'Y坐標'	"@X11"			
CL	N	5	0	'群'	"@X5"			
		232						

表 1 門牌點位檔欄位定義表

[DBF] 索引標識定義				
模組名稱:DBF.94DOOR67		日期:97.05.09		頁次: 1
索引標識	索引欄位運算式	昇/降	唯一鍵值	中文名稱
DOORID	ID	0	0	
DOORF	F	0	0	
DOORG	G	0	0	
DOORI	I	0	0	
DOORFGI	F+G+I+K+M	0	0	

表 2 門牌點位檔索引標識定義表

[DBF] 欄位定義									
模組名稱:DBF.HCDB3001					日期:97.05.09		頁次: 1		
Field	Type	Len	Dec	*Title(C)	*Picture(C)	Mode	Inq	*InqCmb	
A	C	33	0	'工廠名稱'	"@X33"				
B	C	8	0		"@X8"				
C	C	9	0	'區'	"@X9"				
D	C	7	0		"@X7"				
E	C	33	0		"@X33"				
F	C	64	0	'地址'	"@X64"				
G	C	9	0	'負責人'	"@X9"				
H	C	9	0	'生份證字 號'	"@X9"				
I	C	9	0		"@X9"				
J	C	9	0	J	"@X9"				
K	C	14	0		"@X14"				
L	C	9	0	L	"@X9"				
M	C	9	0	'狀態'	"@X9"				
N	C	9	0	N	"@X9"				
O	C	9	0	O	"@X9"				
P	C	9	0	P	"@X9"				
Q	C	36	0	'業別'	"@X36"				
R	C	9	0	R	"@X9"				
S	C	9	0	S	"@X9"				
X	N	11	3	"X坐標"	"999999.999"				
Y	N	11	3	"Y坐標"	"999999.999"				
CL	N	5	0	"群"	"99999"				
		330							

表 3 工廠資料檔欄位定義表

第二節、DBSCAN 群聚演算法

本演算方式是利用一般資料庫本身所提供之搜尋、排序、索引等功能，取代一般程式語言開發中常用之陣列方式，在本研究之演算過程中，也沒有 R-樹的建立和 k-dist 圖的繪製，而是透過一再重複搜尋的方式完成該演算(如圖 6 DBSCAN 演算流程圖)，主要目的是要降低使用系統資源，提升資料演算容量。

演算步驟如下程如：

步驟一、輸入 Eps 及 Minpts 的值，到系統中設定兩個全域變數中。

步驟二、開啟資料庫從第一筆資料坐標對其他各點坐標計算距離，公

式為 $\sqrt{(X_1 - X_n)^2 + (Y_1 - Y_n)^2}$ ，其中第一筆資料坐標 $= (X_1, Y_1)$ ，各點坐標 $= (X_n, Y_n)$ ， $n =$ 第 N 筆資料 (Record Number)。

步驟三、將步驟二計算出的距離對 Eps 值進行比較，當距離小於等於 Eps 值時 $C = C + 1$ (C 值累加)， n 值存入 $D(n)$ 暫存工作陣列，距離大於 Eps 值則往下一筆資料。

步驟四、完成第一輪迴圈後檢查 C 值是否大於等於 Minpts 值，如果大於等於 Minpts 值則形成第一群， C 值是小於 Minpts 值，則 $D(n)$ 歸零，跳步驟二，資料指向下一筆，再次計算檢查。

步驟五、步驟四完成第一輪迴圈檢查 C 值大於等於 $Minpts$ 之值後，對 $D(n)$ 暫存工作陣列中符合之資料再對全資料庫計算檢查，符條件值者檢查 $D(n)$ 陣列，有無重複，無重複者繼續存入 $D(n+1)$ 陣列中，直到 $D(n)$ 陣列最後一筆，始完成第一群。

步驟六、將該群結果寫回資料庫 cl 分群欄位中。

步驟七、跳回步驟二，重複演算前，需檢查 cl 欄位內容，為空值時，才能往下繼續執行運算。

步驟八、完成全部迴圈 DBSCAN 群聚演算亦完成。



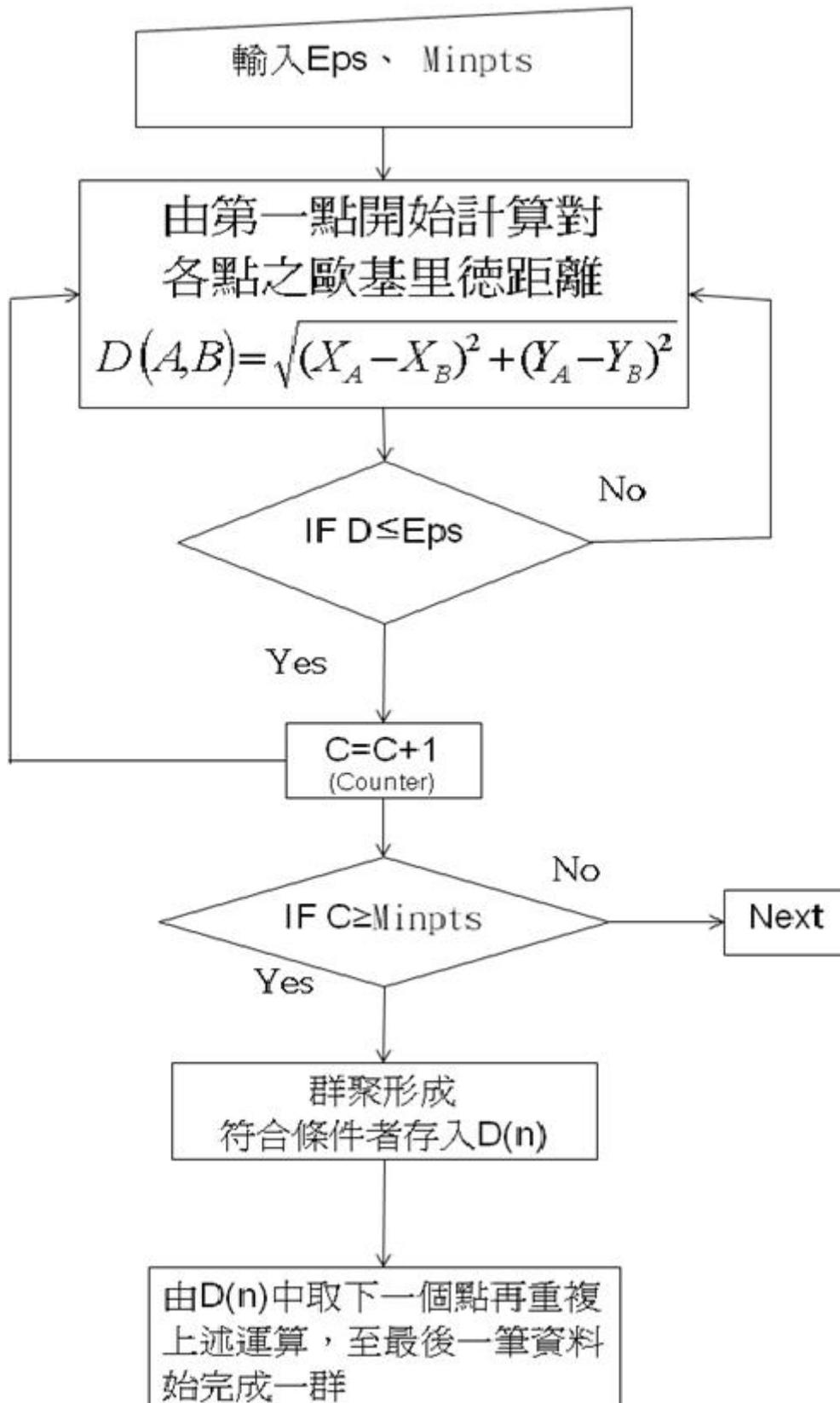


圖 5 DBSCAN 演算流程圖

第三節、系統開發工具

本研究 DBSCAN 演算系統，採用 DBTOOLS 開發工具，全名為「DBTOOLS 快速軟體產生器」(簡稱：DBTOOLS)(出版公司：訊光科技有限公司 Infolight System Corp.)，為國人自行開發之軟體，初版於 1988 年 4 月發行，DBTOOLS V1.0-DBTOOLS V3.2(1992.02)主要為 DOS 版，所發展的程式語言為 ASSEMBLY(NETWARE API、INTERPRETER、KERNEL)、CLIPPER 88 SUMMER FOR DBASE 3+、BLINKER V2.01；更新版之後 DBTOOLS V5.X 則以 BORLAND C++為主，後續發行 DBTOOLS FOR WINDOWS 版、DBTOOLS FOR WINDOWS 32bit、Visual DBTOOLS 等，本研究是以 DBTOOLS FOR WINDOWS 32bit 之版本為系統開發工具。

該程式開發工具同樣支援.DBF檔案格式，也就是相容於xBASE規格之資料庫，開發工具本身也提供改良後專屬.DBX及.MTX(MULTI-INDEX)檔案格式，.MTX就是單一索引檔案可同時包含多重鍵值，改良.DBT(備註欄)為.DBX提升資料庫搜尋效率特別是大型資料特別顯著，使用可依自己需求選擇，這一些資料庫都可相互轉檔使用，程式撰寫部分提供了一個快速開發介面，不用編譯及連結(COMPILE&LINK)程式，直接執行及除錯(debugging)。

程式開發分為前台(操作者介面)及後台(設計師介面)(如圖 7 系統示意圖)，開發人員可直接利用功能鍵切換前台與後台，邊開發邊執行，隨時可以觀看成果，特別是一般常用的程式，都已使用表格化輸入參數直接產生，如輸入畫面、報表產生、轉檔、過帳等一般資料庫用的到的模組都有，而對於特別演算則提供 INT 模組，本研究於附錄一中提供原始程式碼供參考，INT 模組的語法、指令及函數，有百分之九十以上相容於 CLIPPER，因此只要熟悉 DBASE、CLIPPER、FOXPRO 等此類語言者都可輕鬆上手，唯一缺點就是畫面較呆板，因此目前大部份使用於大陸產業界 ERP 系統上，

也因為是全中文化很快就進入中國市場，近年來該公司繼續開發 Visual DBTOOLS 系統，全面改版升級，不只提供原有之資料格式，更開發 ODBC 介面，可以直接連結大型資料庫如 ORACLE、SYBASE 等，是很有前瞻性的開發工具。

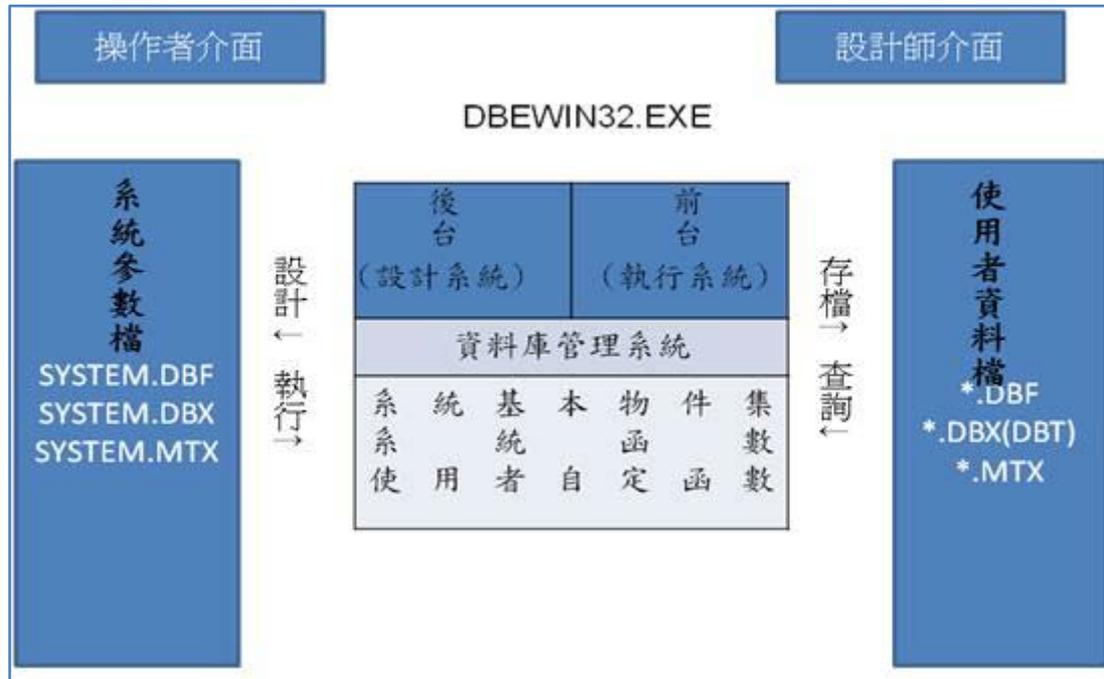


圖 6 DBTOOLS系統示意圖

第四章、實證分析

第一節、資料庫轉換

本研究測試資料主要使用台中市政府工商資料庫之工廠登記資料，因為該資料庫中住址欄位共分為四個欄位：縣市別、行政區別、里別、路街巷弄號，資料庫中並沒有空間坐標欄位，因此於原資料庫中新增三個欄位(X、Y、CL)，作為住址比對成功後，由門牌點位檔坐標欄位轉入工廠登記資料檔坐標欄位(X、Y)，DBSCAN 群聚分析演算後分群結果寫入欄位(CL)中，此欄位為群聚演算完成後，分群編號記錄之主要欄位，工商資料庫本身使用 Microsoft SQL Server 2000 資料庫，其中還包括商業登記等其他資料庫，透過該資料庫所提供之轉檔工具，將所需要的資料轉成 xBASE(.dbf) 格式資料檔，匯入本研究開發之系統，開使執行 DBSCAN 群聚分析演算，資料庫轉換流程如圖 7 資料庫轉換流程圖。

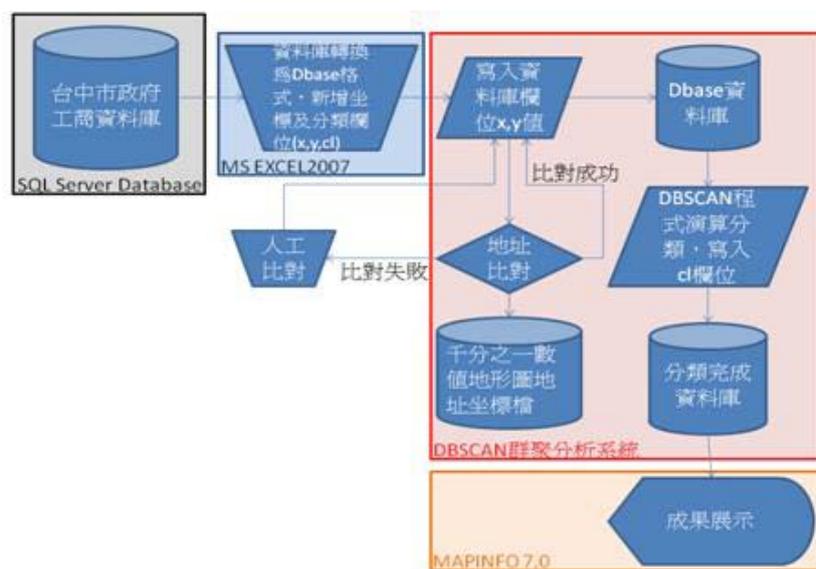


圖 7 資料庫轉換流程圖

第二節、DBSCAN 群聚分析系統

壹、地址標準化

比對之前有一項很重要的工作，就是資料標準化，比對成功率的高低與標準化關係是非常密切，本研究所用的台中市工廠資料檔中，地址欄位共分成四個欄位(縣市別、行政區別、里別、路街巷弄號)，而門牌點位資料檔是分區、里、鄰、路(段)、巷、弄、號(之號、樓)等八個欄位，工商聯合櫃台收件後，由輸入人員依照工廠登記申請資料，逐項欄位鍵入，因此，在比對之前必需先將地址欄位格式標準化，而人工輸入人員不同，輸入習慣也有差異如：“台中市”與“臺中市”、“10”跟“10”，前後空白與中間空白都要考慮，而實做過程中，還需考慮資料中繁體字與簡體字的問題，實務上最後採合併欄位方式，加上逐字檢查，去除前後空白及轉換字形為全形字，包含所有數字、英文、標點符號，最後新增比對欄位為標準化後的資料欄位，作為比對欄位。

貳、地址比對定位

台中市工廠資料檔中，並沒有空間坐標資料欄位，因此，必需透過系統來比對地址欄位及定位，從主系統畫面(圖 8 為 DBSCAN 群聚分析系統主畫面)中選擇「地址對位」項目直接對資料庫執行比對，比對坐標定位資料來源，是利用台中市一千分之一數值航照地形圖，門牌點位資料檔為比對依據，透過標準化後的地址欄位，由系統將兩個資料庫地址欄位自動比對，比對方式採完全比對，比對成功後，將門牌點位檔中(X、Y)坐標欄位的內容，轉入工廠資檔(X、Y)坐標欄位，雖然資料已完成標準化，實際執

行第一次成果，3,311 筆資料比對成功 2,061 筆，成功率為 62%，以完全比對方式而言，比對成功率超過 50% 以上算是很高，如果搭配分區或模糊比對，可提高到 90% 以上。

檢視比對未成功資料，絕大部份為住址不完整，因此，在未修正比對方式前，只能利用索引鍵排序，以人工辨識方式來完成定位，本研究只針對 DBSCAN 演算系統，對於地址比對未成功之部份，暫不處理，本研究僅對比對成功之資料執行後續演算分析。



圖 8 DBSCAN 群聚分析系統

參、DBSCAN 群聚分析

系統主畫面(如圖 8)，選擇 DBSCAN_ALL 項目，會開啟輸入視窗(如圖 9)，視窗中除 Eps 及 Minpts 兩參數外，還有一個「群組編號」，該值為群聚產生後，編號的起始值，一般使用預設值"1"就可以，完成參數輸入後

按”確定”，開始DBSCAN演算工作，從畫面中剛才的輸入視窗會保持不變，「群組編號」會隨群聚形成而增加，演算完成後會關閉輸入視窗，回復原主畫面，演算完成的成果檔案為(HCDB_0.DBF)，檢查成果檔案，如果為空資料庫表示此參數無群聚產生。將成果檔複製更名，再輸入其他參數繼續演算，Eps 值不能超出區域中之最大距離，最小大於零，Minpts 值最大不超過總數量，最小同樣大於零。將最後結果檔案匯入 GIS 系統中分層使用，透過 GIS 系統中之篩選功能，將不同群以不同顏色表示，如此，就很容易分別出各群，搭配街廓圖、都市計畫使用分區圖、行政區圖等，套疊分析，可分析出許多有價值的不同資訊，這些資訊將可提供都市計畫及城市規劃之參考資料。



圖 9 DBSCAN群聚分析系統

第三節、結果分析

由於 DBSCAN 演算法的兩個輸入參數 Eps、Minpts 與最後結果相當敏感，因此，需要透過實際重複運算後，所得到的結果再由人工觀察來判視，因此本研究分別以第一項 100M/50、200M/50、300M/50、400M/50、500M/50、600M/50(Eps/ Minpts)共六組參數，及第二項 500M/100、550M/100 共二組參數，透過實際執行演算分析來觀察比較結果。第一項特別將門檻值 (Minpts)固定不變，透過距離(Eps)的變化，觀察演算過後的成果比較，第二項則用不同的門檻值(Minpts)，距離(Eps)固定不變，從這兩項成果來分析。

圖 10 是未執行演算前之工廠分佈情形，圖中右方因屬大坑風景區，沒有工廠設立，為使圖面解析度提高將其切除未顯示，藍色線條為道路街廓，工廠點位資料，黑色點位為工廠位置，共有 2,061 筆。

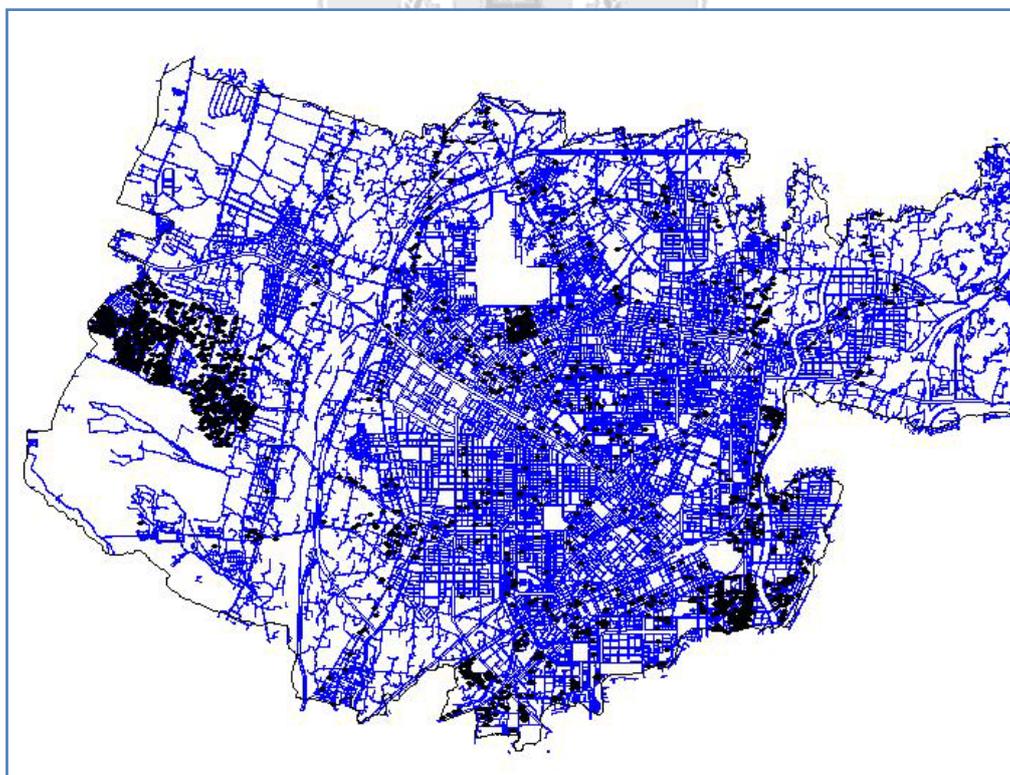


圖 10 工廠分佈圖(工廠數量=2061 家)

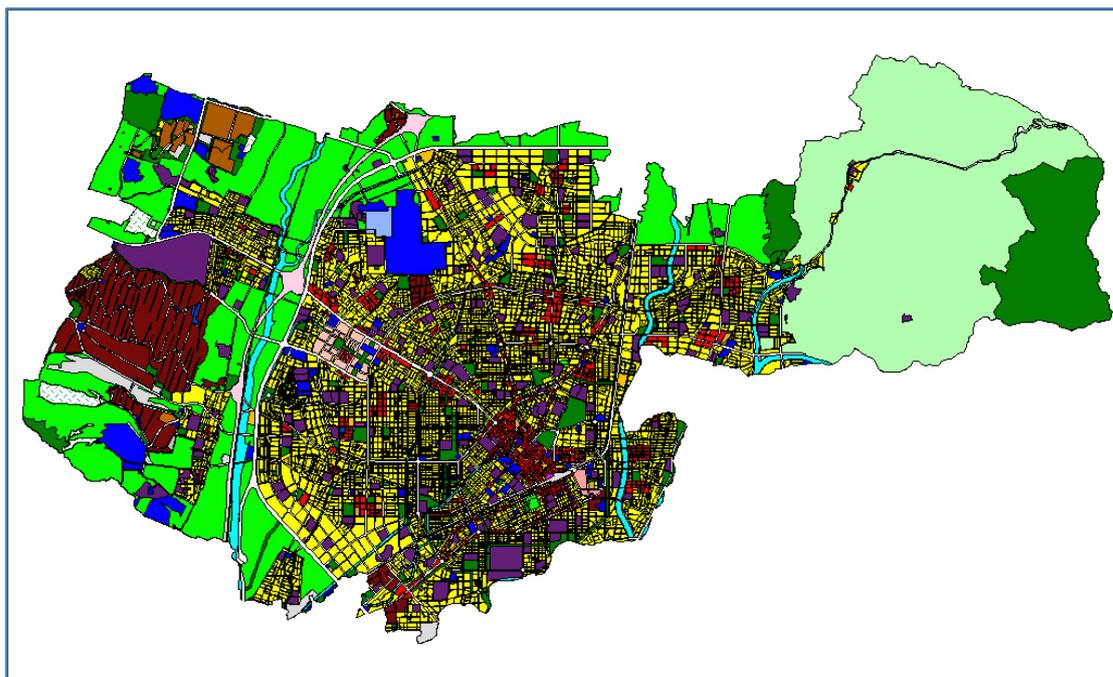


圖 11 台中市都市計畫圖

圖 11 為台中市都市計畫圖，從該圖與圖 10 比較，咖啡色部份屬工廠用地，而圖 10 中最大一塊黑點區域，是在左邊縣市界地方，依規定工廠應該設立於工廠用地上，我們只需將這兩個圖層套疊，如圖 12 台中工業區工廠分佈圖，就可以很容易的看出，工廠位置是否正確。



圖 12 台中工業區工廠分佈圖

圖 13 為第一項中六組不同 Eps、Minpts 值所產生的工廠群聚演算結果比較表，表中每一格的顏色直接對映到分佈圖上各點的顏色，同一顏色為一群組，表格中的數字為該群的工廠數量，最下面為各組合計數量。

第一組、(100M/50)的結果沒符合該條件的群聚產生故為空白，其他各組就各有不同群聚結果。

第二組、(200M/50)的結果為四個群聚產生，最大為第四族群 152 家，符合該條件的總數 462 家，為全部 22.4%。

第三組、(300M/50)的結果為六個群聚產生，最大為第五族群 419 家，符合該條件的總數 1,131 家，為全部 54.9%。

第四組、(400M/50)的結果為六個群聚產生，最大為第五族群 766 家，符合該條件的總數 1,361 家，為全部 66%。

第五組、(500M/50)的結果為五個群聚產生，最大為第四族群 773 家，符合該條件的總數 1,413 家，為佔全部 68.6%。

第六組、(600M/50)的結果為六個群聚產生，最大為第五族群 775 家，符合該條件的總數 1,507 家，為全部 73.1%。

Eps/Minpts	100M/5	200M/5	300M/5	400M/5	500M/5	600M/5
群別(cI)	0	0	0	0	0	0
第一群		121	131	134	311	321
第二群		117	240	294	147	65
第三群		72	59	50	122	157
第四群		152	226	60	773	127
第五群			419	766	60	775
第六群			56	57		62
合 計	0	462	1131	1361	1413	1507

圖 13 DBSCAN 群聚演算成果比較表(1)

第二組(200M/50)的結果如圖 14，由圖面可清楚辨識出四個群，第一群為棕色數量 121 家，第二群為綠色數量 117 家，第三群為灰色數量 72 家，第四群為黃色數量 152 家，其中最大的一群是第四群黃色數量為 152 家。

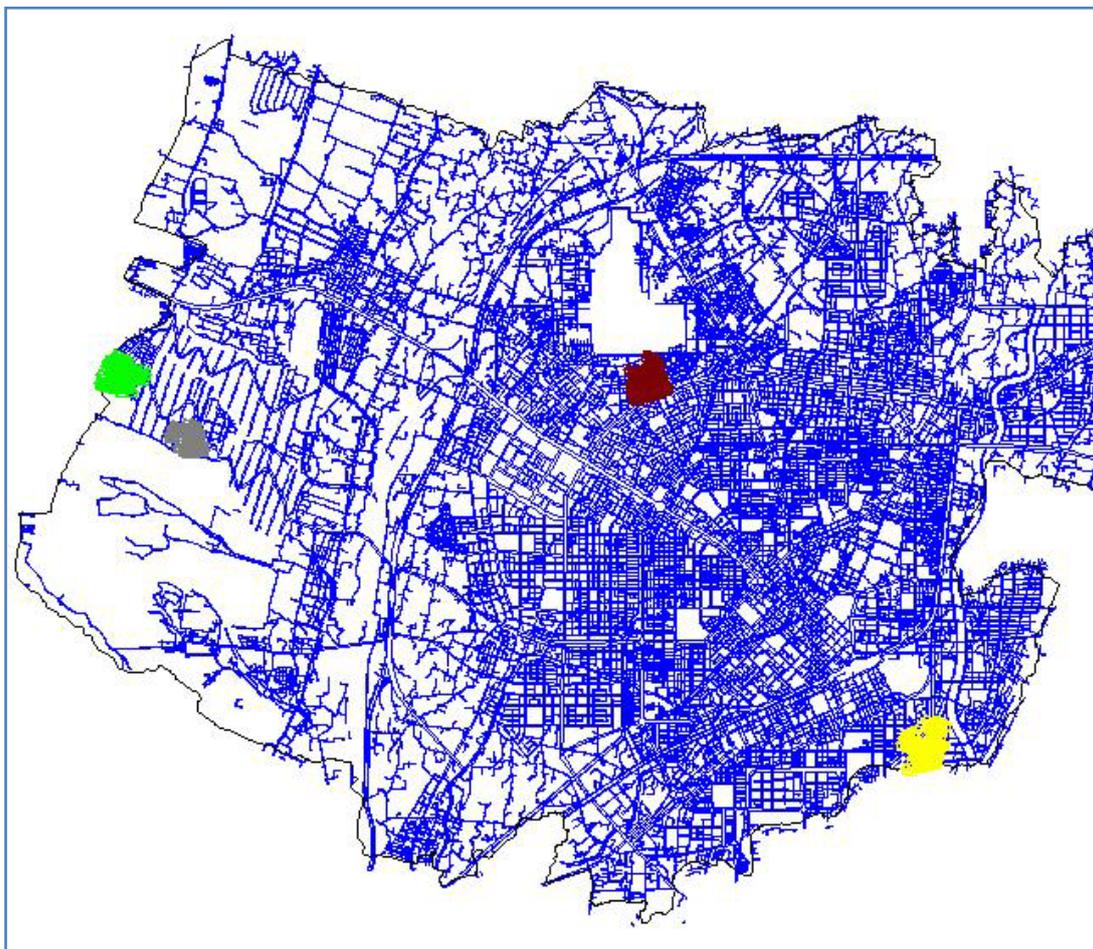


圖 14 Eps=200M Minpts=50

群(c1)	Eps/Minpts 200M/50
第一群	121
第二群	117
第三群	72
第四群	152
合計	462

第三組(300M/50)的結果如圖 15，由圖面同樣可清楚辨識各群，不同於前一組它共分出六個群，第三(灰色)、六(紅色)是新產生的群但數量不多(59、56)在門檻值邊緣，第四(黃色)、五(水藍色)台中工業區部分就增加很多。

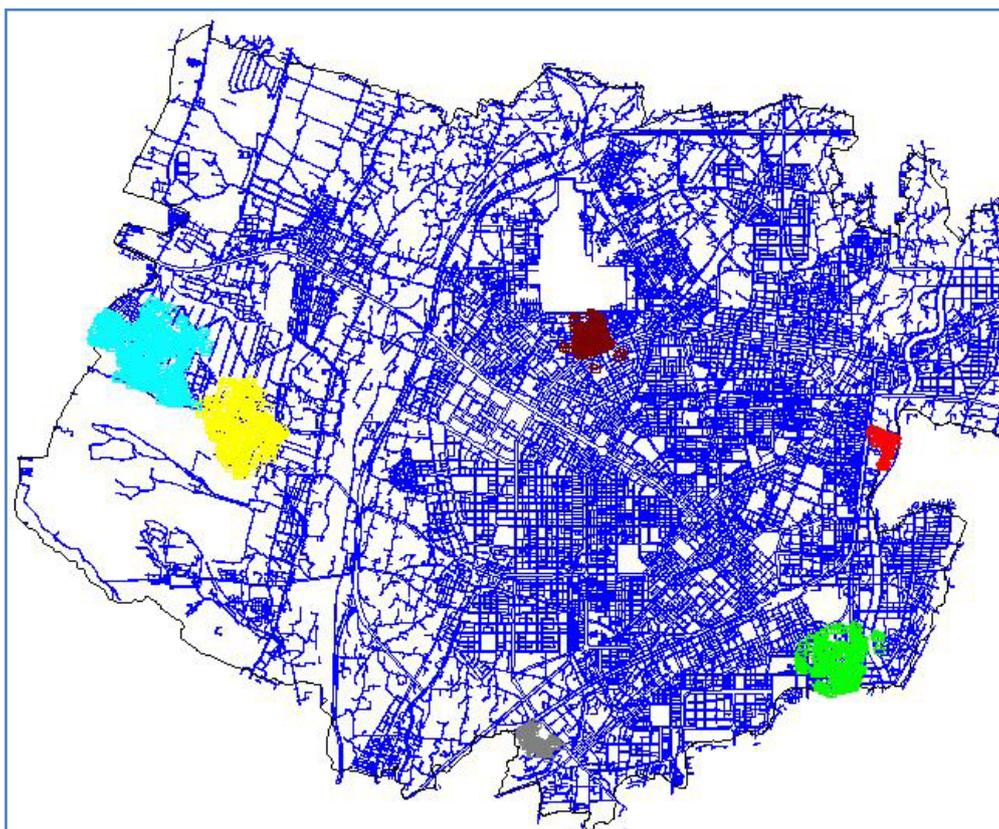


圖 15 Eps=300M Minpts=50

Eps/Minpts 群(c1)	300M/50
第一群	131
第二群	240
第三群	59
第四群	226
第五群	419
第六群	56
合計	1131

第四組(400M/50)的結果如圖 16，由圖面跟前一組比較個群都有增加，台中工業區部分(紅色)則是由兩群合併為一群，但是又出現一群第三群(黑色)門檻值正好等於 50。

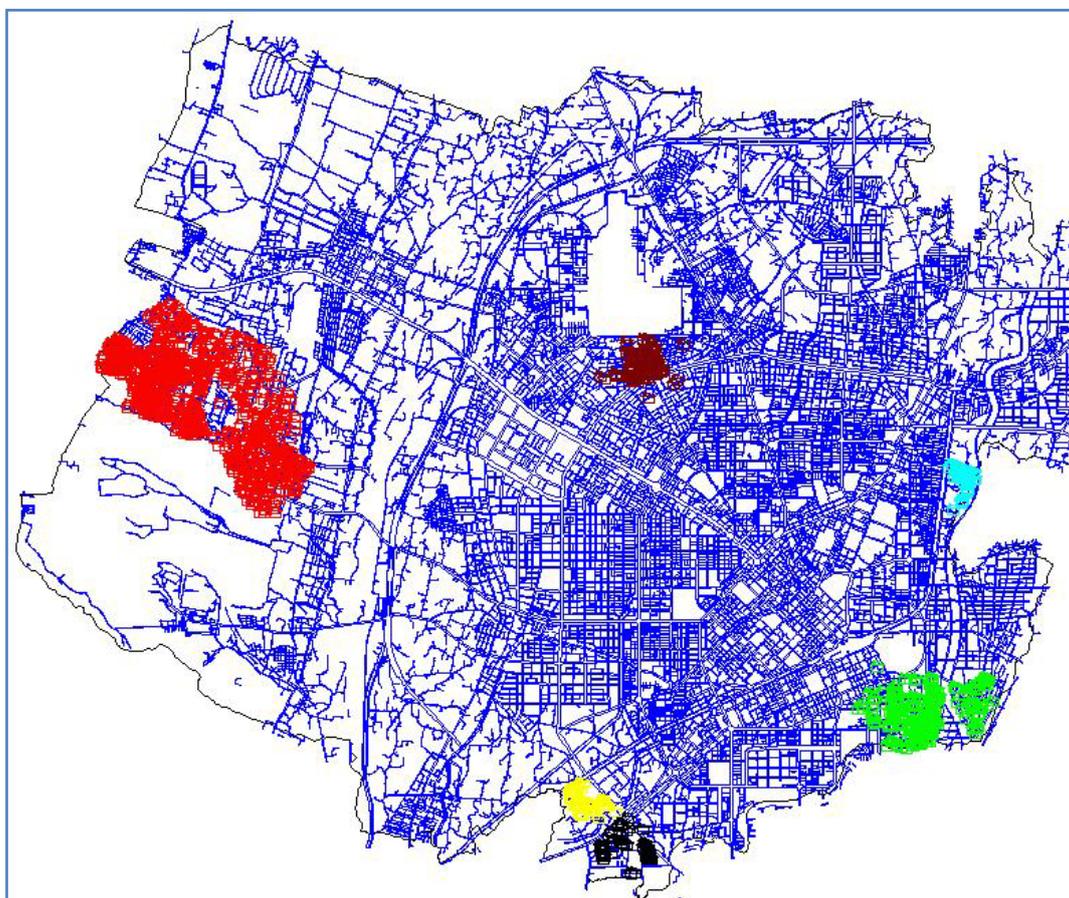


圖 16 Eps=400M Minpts=50

Eps/Minpts 群(c1)	400M/50
第一群	134
第二群	294
第三群	50
第四群	60
第五群	766
第六群	57
合計	1361

第五組(500M/50)的結果如圖 17，由圖面跟前一組比較個群同樣都有增加，前一組三、四群 (黑色、黃色)則是由兩群合併為一群。

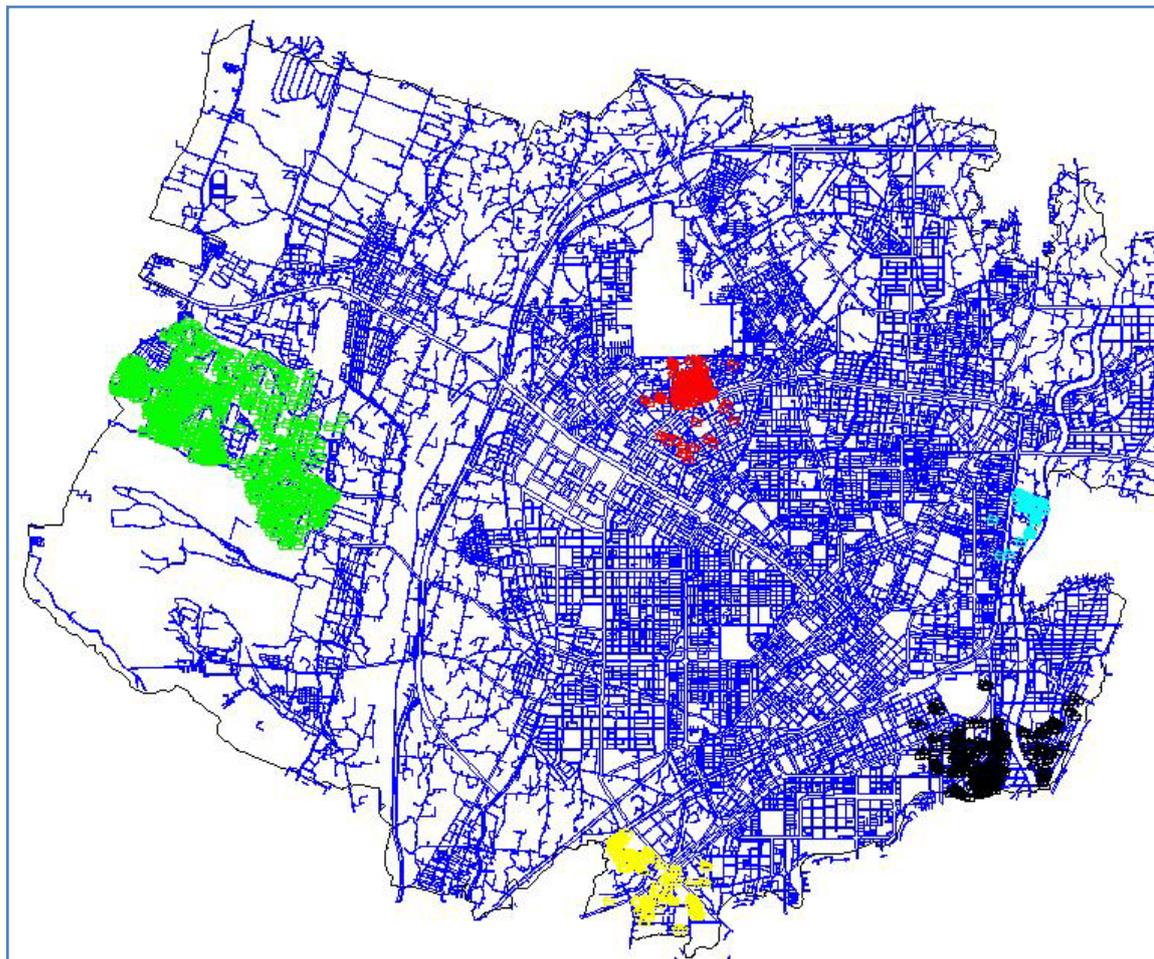


圖 17 Eps=500M Minpts=50

群(c1)	Eps/Minpts	500M/50
第一群		311
第二群		147
第三群		122
第四群		773
第五群		60
合計		1413

第六組(600M/50)的結果如圖 18，由圖面跟前一組比較個群只有增加，群數不變。

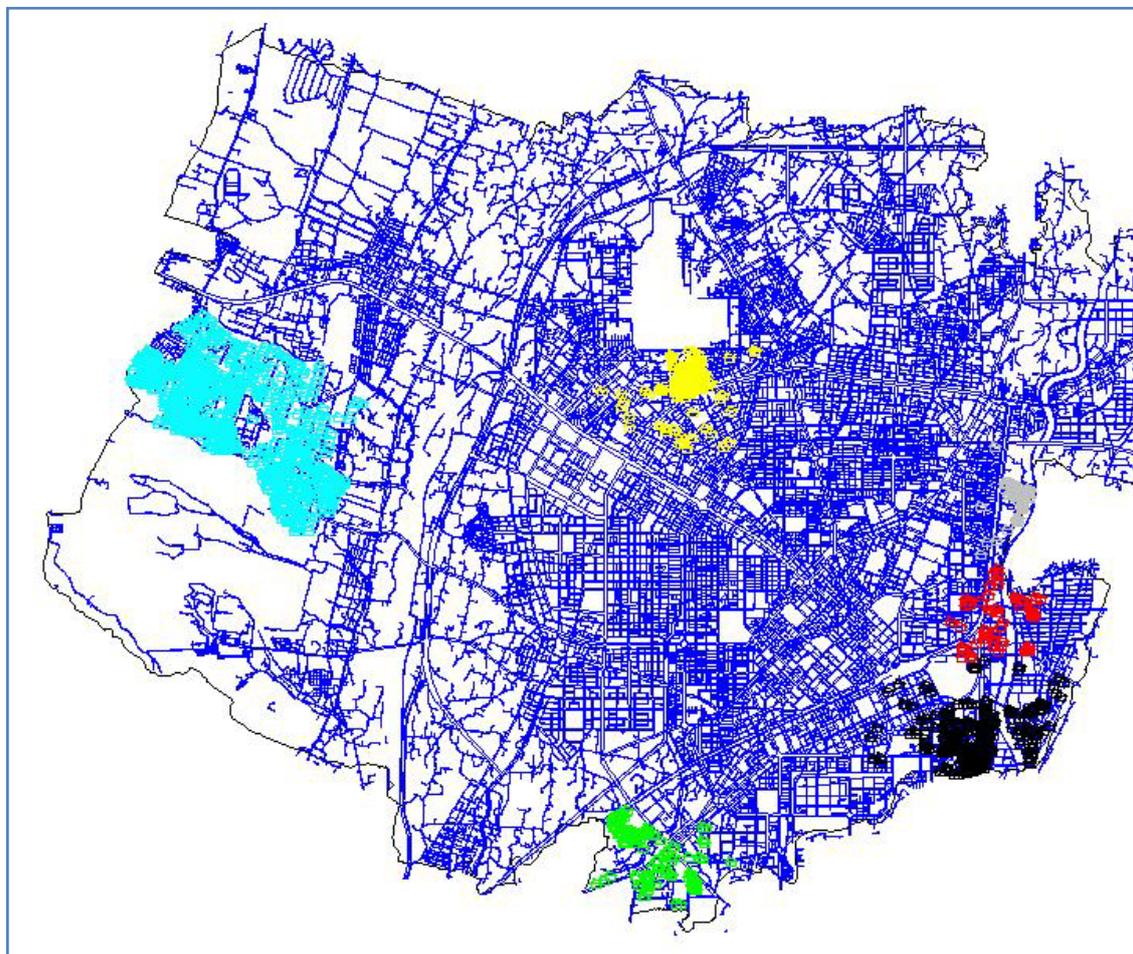


圖 18 Eps=600M Minpts=50

群(c1)	Eps/Minpts	600M/50
第一群		321
第二群		65
第三群		157
第四群		127
第五群		775
第六群		62
合計		1507

透過前面資料的實際運算結果，從工廠位置坐標分佈就能觀察出各群聚之間的關係，群與群之間的距離及分佈，群聚總數與 Eps 值的關係如圖 19，由此圖可看出，群聚總數在門檻值(MinPts)為 50 家時，距離(EPS)與群聚總數為正相關。同樣如果再搭配台中市其他 GIS 圖資如都市計劃圖、區里界圖等，可分析出更多資訊。

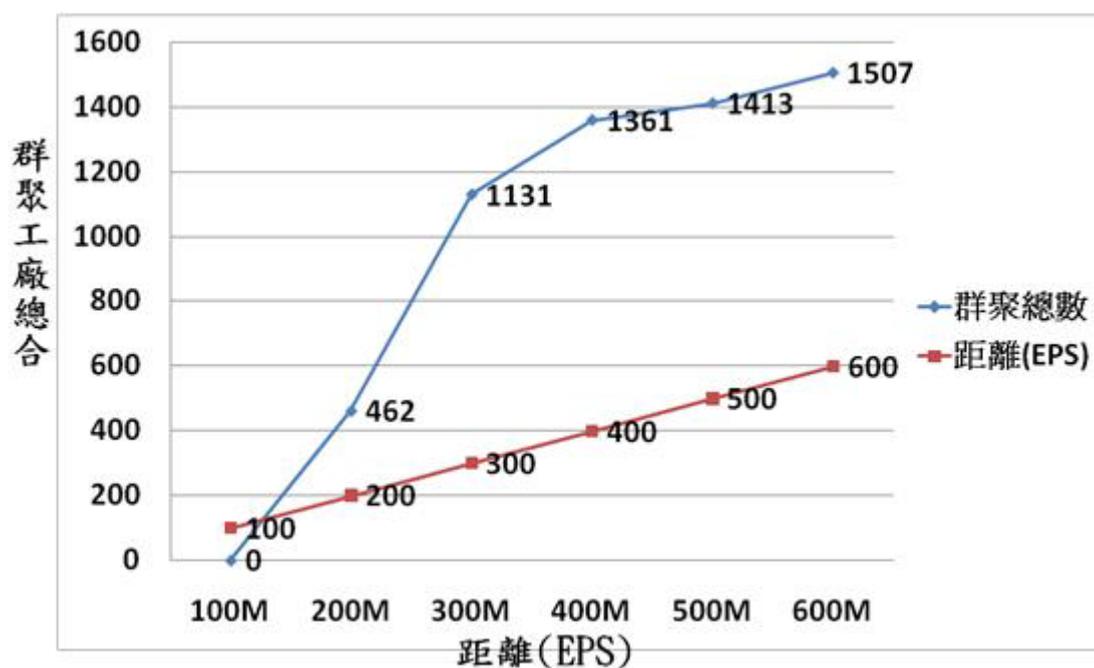


圖 19 群聚數量與Eps值折線圖

第二項共分五組，第一組測試數據是 Eps=500M、Minpts=250，第二組測試數據是 Eps=500M、Minpts=20，第三組測試數據是 Eps=500M、Minpts=150，第四組測試數據是 Eps=500M、Minpts=100，第五組測試數據是 Eps=500M、Minpts=50，所產生的工廠群聚數如圖 20，第一組測試數據是 Eps=500M、Minpts=250 沒有群聚產生為空值。

Eps/Minpts 群別(c)	500M/250	500M/200	500M/150	500M/100	500M/50
第一群		241	259	288	311
第二群		266	438	488	773
第三群			200	243	122
第四群				138	147
第五群					60
合計	0	507	897	1361	1413

圖 20 DBSCAN 群聚演算成果比較表(2)

第一組 Eps=500M/MinPts=250，符合條件為零家，沒有群聚產生，第二組 Eps=500M/MinPts=200，成果為二個群，第一群為黃色數量 241 家，第二群為紅色數量 266 家，如圖 21。

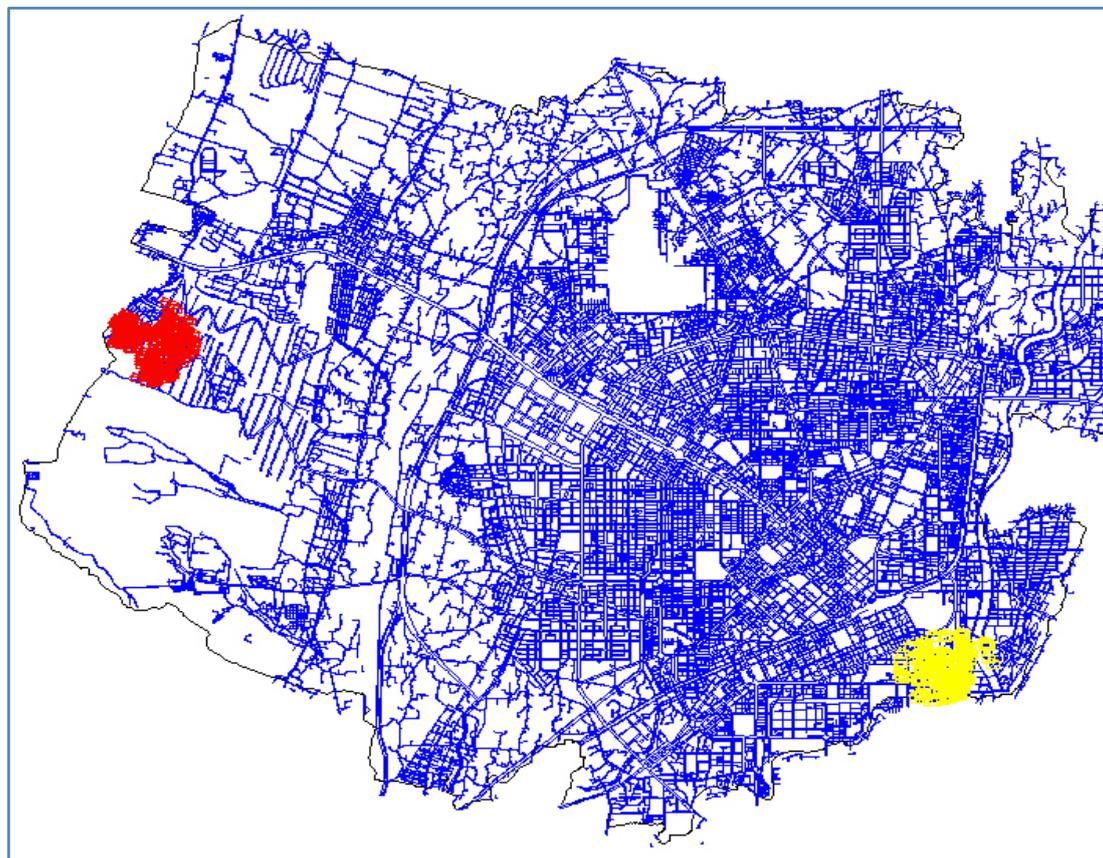


圖 21 Eps=500M Minpts=200

群(c1)	Eps/Minpts	500M/200
第一群		241
第二群		266
合計		507

第二組測試數據是 Eps=500M/Minpts=150 所產生的工廠群聚分佈圖如圖 22，成果為三個群，第一群為玫瑰色數量 259 家，第二群為黃色數量 200 家，第三群為紅色數量 438 家。

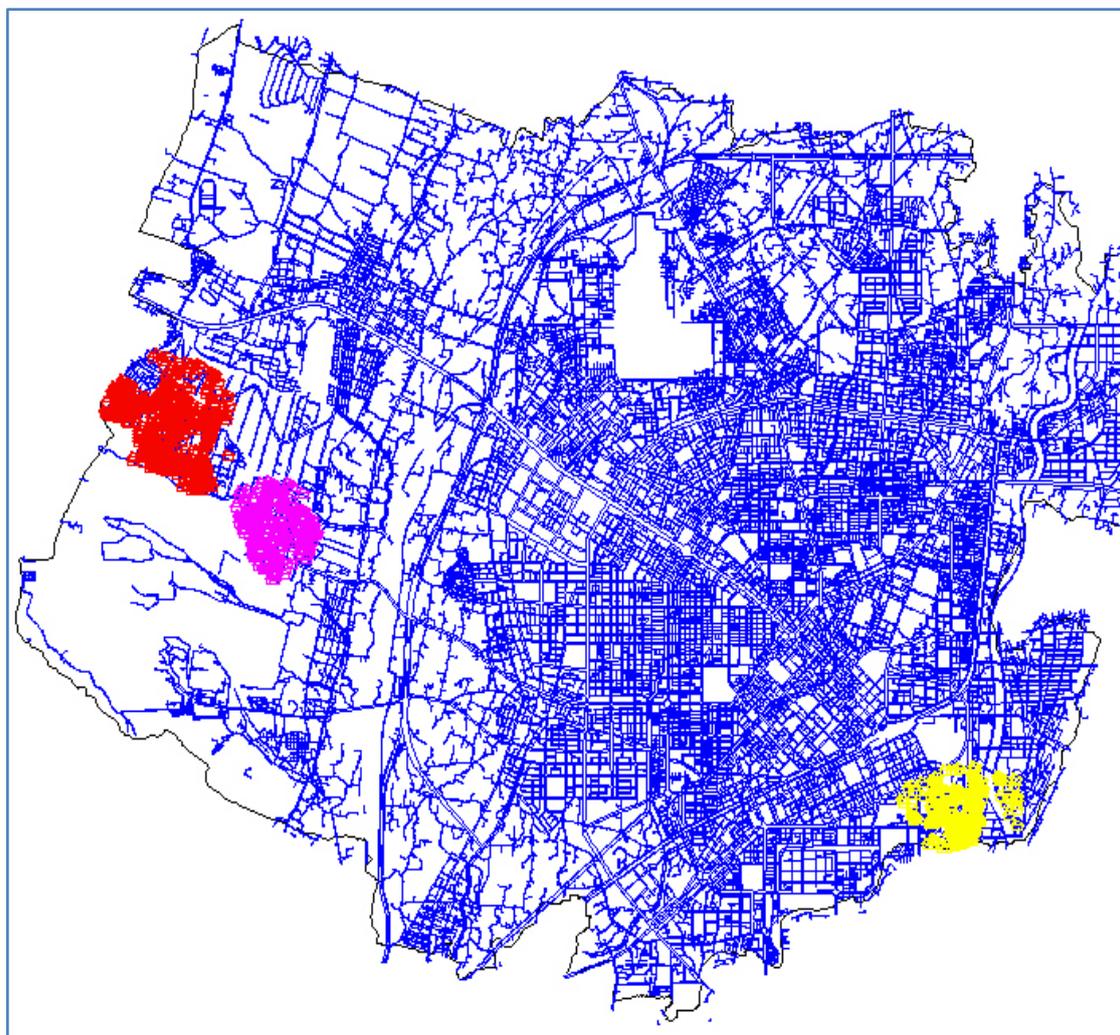


圖 22 Eps=500M Minpts=150

群(c1)	Eps/Minpts	500M/150
第一群		259
第二群		200
第三群		438
合計		897

第三組測試數據是 Eps=500M/Minpts=100 所產生的工廠群聚分佈圖如圖 23，成果為四個群，第一群為綠色數量 138 家，第二群為玫瑰色數量 288 家，第三群為紅色數量 488 家，第四群為黃色數量 243 家。

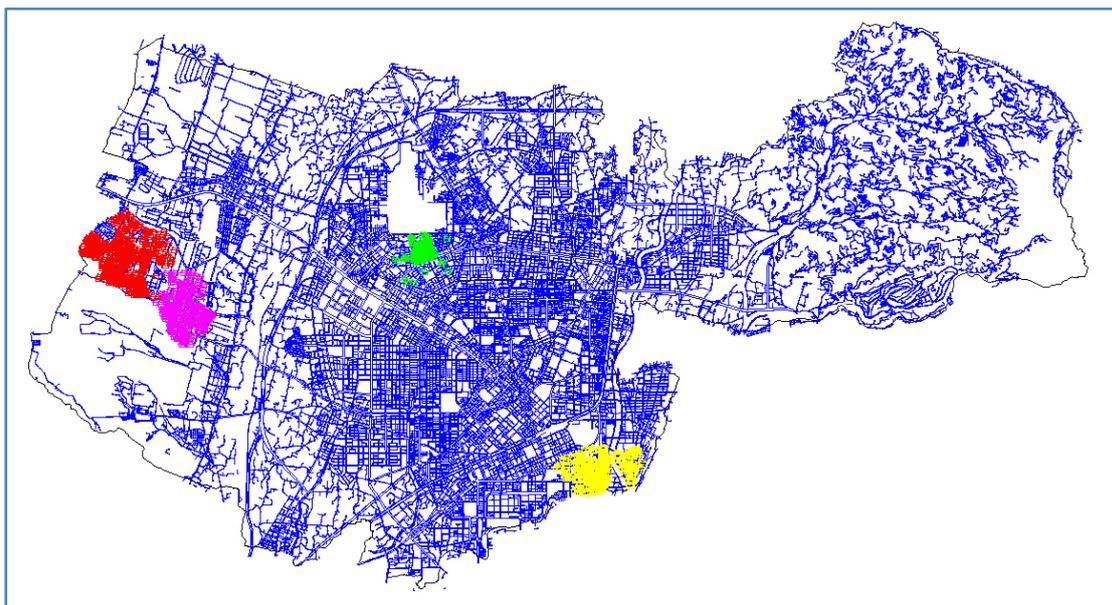


圖 23 Eps=500M Minpts=100

群(c1)	Eps/Minpts	500M/100
第一群		138
第二群		288
第三群		488
第四群		243
合計		1361

第四組測試數據是 Eps=500M/Minpts=50 所產生的工廠群聚分佈圖如圖 24，成果為五個群，第一群為黃色數量 311 家，第二群為灰色數量 147 家，第三群為綠色數量 122 家，第四群為紅色數量 773 家，第五群為黑色數量 60 家。

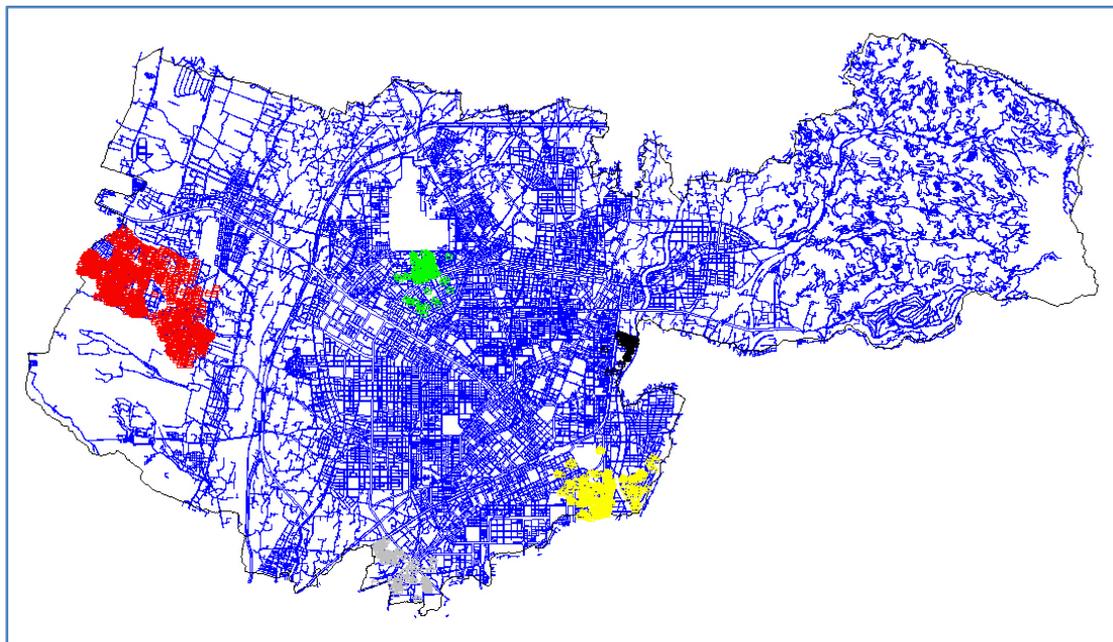


圖 24 Eps=500M Minpts=50

群(c1)	Eps/Minpts	500M/50
第一群		311
第二群		147
第三群		122
第四群		773
第五群		60
合計		1413

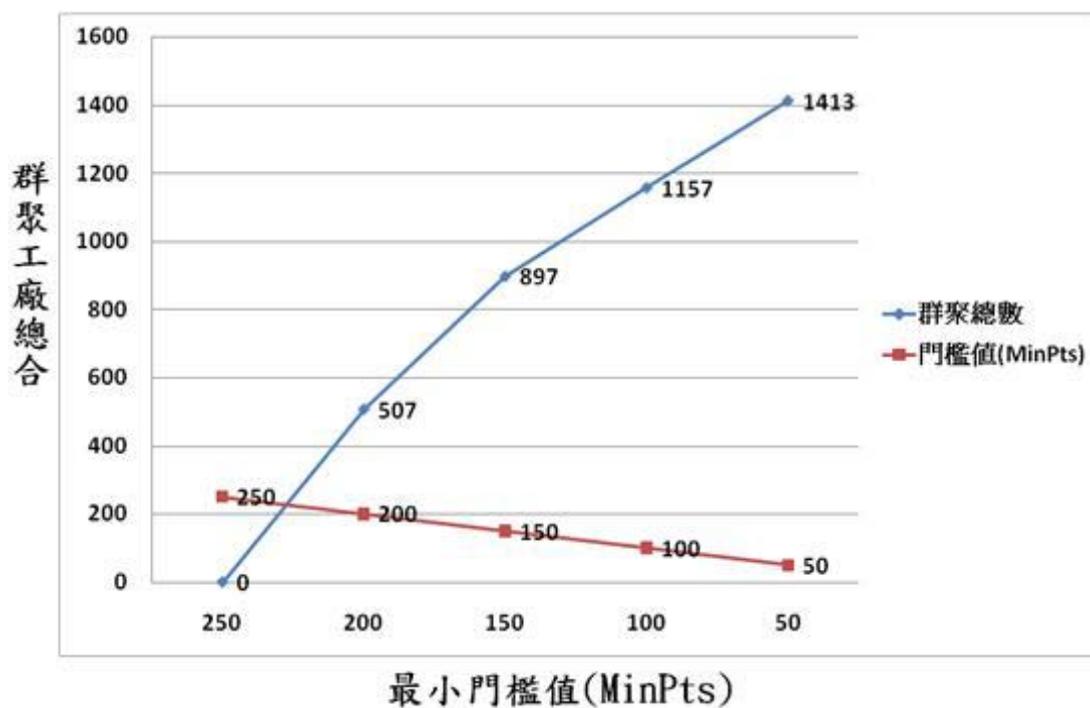


圖 25 群聚數量與MinPts值折線圖

由圖 25 可看出，當距離(Eps)固定為 500M，最小門檻值(MinPts)與群聚總數為負相關，透過以上兩項數據的測驗，可以很容易找出相關性，有了這些數據就可找到較適當的 Eps、MinPts。

第五章、結論與建議

第一節、研究結論

經過本研究實際執行後，可發現 DBSCAN 演算法雖然理論簡單，但是實際使用於程式系統中，必需克服許多問題，包括資料的標準化、欄位的增加、檔案格式的轉換等都是必需克服解決，而當程式完成後，快速且方便的產製出成果檔案，並且可直接匯入大部份的 GIS 系統軟體中分析利用，從前面結果分析就清楚了解，GIS 的分析功能在學術界及業界均以相當成熟，本研究開發之系統，有效率的將 MIS(Management Information Systems)資料庫轉換到 GIS 空間資料庫，同時也解決大部份，原本必需要到現地測量的工作，不但節省經費更節省寶貴的時間。

至於本系統執行效率，以一般筆記型電腦(Intel Core 2 T5600 1.8GHz CPU 1GB RAM)，XP 作業系統(Microsoft Windows XP Service Pack2)，測試資料量為 2,061 筆，測試結果共只花費 8 分 34 秒的時間，對於一般使用者而言是可以接受，目前因為個人電腦的普及且 CPU(Central Processing Unit)運算能力也提升，因此大部份 GIS 系統為方便使用者操作，都有發展個人電腦版本，此版本泛稱”桌上型地理資訊系統”(desktop GIS)，如 Mapinfo、ArcGIS 等，但是桌上型地理資訊系統在個人電腦上執行中系統資源的消耗還是很多，如果再加上其他運算整體效率就會快速下降，本研究針對此問題將資料運算與 GIS 空間資料疊圖分析功能分開執行，使用者可預先使用本研究開發之程式工具，利不同參數進行試算產生成果檔案，

在直接透過 GIS 系統工具一一比對分析，如此不但有效分擔 GIS 系統的執行壓力外更提升整體時間效率，以一般筆記型電腦就可完成 GIS 空間資料群聚分析運算是相當便利且經濟的工具。

第二節、研究建議

本研究使用相當平民化的 xBASE 資料庫(.dbf)，其實大部分的資料庫如：Oracle、Sybase、Informix 等也都是可以使用此概念，利用現有資料庫資源，有效應用，不需額外採購，一樣可以達到此功能，不受 GIS 系統開發語言的限制，運算本身獨立運作，減輕 GIS 系統負荷，對於其它群聚演算法如 K-means[Mac67]、PAM[KR90]、CLARA[KR90]、CLARANS[NH94]、CURE[GRS98]、ROCK[GRS99]、CLIQUE[AGR98]、STING[WYM97]等，同樣可以搭配資料庫輔助來達成，此為後續研究之重點。

第六章、參考文獻

1. 李新延、李德仁，2005，DBSCAN 空間聚類算法及其在城市規劃中的應用，測繪科學 Science of Surveying and Mapping，第 30 卷第 3 期 p51-p53)。
2. 陳榮昌、林育臣，2002，群聚演算法之比較及群聚參數的分析與探討，第三屆網際網路應用與發展研討會。
3. 吳雪琴，2007，DBSCAN 算法研究，軟件專刊 4 月號，p99-p100。
4. 陳禹芳，2004，產業群聚現象之研究—以製藥產業與生物技術產業為例。
5. 閻永祺，2003，產業群聚與區域產業發展關係之研究—以南部區域為例。
6. 楊智盛，2003，臺灣地區製造業產業群聚之研究。
7. 徐欣玉，2003，文化創意產業之空間群聚分析—以中山北路婚紗攝影業為例。
8. 張晉瑞，2006，探討資料採礦技術應用於時空資料分析之研究。
9. 訊光科技，1993，DBTOOLS 5.0 入門指引。
10. 訊光科技，1994，DBTOOLS 技術手冊。
11. 訊光科技，1995，DBTOOLS 函數指令集(DOS 與 WIN)。
12. 林佳生，1991，CLIPPER 程式設計速成。

13. 蕭凱文，薛志達，李政輝，1999，Microsoft SQL Server 7.0 資料倉儲整合應用。
14. 侯捷 譯，2001，C++ Primer 3/e 中文板。
15. [Ester96]Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. KDD 1996: 226-231.
16. [ABK99] M. Ankerst, M. Breunig, H.-P. Kriegel, and J. Sander. "OPTICS: Order-ing points to identify the clustering structure". In Proc. 1999 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'99), pages 49-60, Philadelphia, PA, June 1999.
17. [AGR98] R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan (1998), "Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications". Int. Conf. Management of Data, 94-105, Seattle, Washington.
18. [CFZ99] C. H. Cheng, A. W. Fu, and Y. Zhang(1999), "Entropy-based subspace clustering for mining numerical data". Int. Conf. Knowledge Discovery and Data Mining(KDD'99), pages 84-93.
19. [DLX01] M. Dash, H. Liu, X. Xu. "'1+1>2': Merging Distance and Density Based Clustering". In Proc. 7th Int. Conf. Database Systems for Advanced Applications (DASFAA'01), pages 18-20, Hong Kong, April 2001.

20. [EKSWX98] M. Ester, H.-P. Kriegel, J. Sander, M. Wimmer, X. Xu. “Incremental Clustering for Mining in a Data Warehousing Environment”. In Proc. 24th Int. Conf. Very Large Databases (VLDB'98), pages 24 - 27, New York City, NY, USA, Aug. 1998.
21. [EKSX96] M. Ester, H. -P. Kriegel, J. Sander, X. Xu. “Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In Proc. 1996 Int. Conf. Knowledge Discovery and Data Mining (KDD'96), pages 226-231, Portland, OR, Aug. 1996.
22. [EL00a] V. Estivill-Castro and I. Lee. “AMOEBAs: Hierarchical Clustering Based on Spatial Proximity Using Delaunay Diagram”. In Proc. 9th Int. Spatial Data Handling (SDH2000), pages 10-12, Beijing, China, Aug. 2000.
23. [EL00b] V. Estivill-Castro and I. Lee. “AUTOCLUST: Automatic Clustering via Boundary Extraction for Massive Point Data Sets”. In Proc. 5th Int. Conf. Geo-Computation, pages 23-25, University of Greenwich, Kent, UK. Aug. 2000.
24. [GRS98] S. Guha, R. Rastogi, and K. Shim. “CURE: An efficient clustering algorithm for large databases”. In Proc. 1998 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'98), pages 73-84, Seattle, WA, June 1998.

25. [GRS99] S. Guha, R. Rastogi, and K. Shim. "ROCK: A Robust Clustering Algorithm For Categorical Attribute". In Proc. 1999 Int. Conf. Data Engineering (ICDE'99), pages 512-521, Sydney, Australia, Mar. 1999.
26. [Han00] J. Han and M. Kamber. "Data Mining: Concepts and Techniques". Morgan Kaufmann, 2000.
27. [Hua98] Z. Huang. "Extensions to the K-Means Algorithm For Clustering large Data sets with Categorical values". Data mining and knowledge discovery, 2:283-304, 1998.
28. [KHK99] G. Karypis, E.-H. Han, and V. Kumar. "CHAMELEON: Hierarchical Clustering Using Dynamic Modeling". COMPUTER, 32:68-75, 1999.
29. [KR90] L. Kaufman and P. J. Rousseeuw. "Finding Groups in Data: an Introduction to Cluster Analysis". John Wiley & Sons, 1990.
30. [Mac67] J. MacQueen. "Some Methods for Classification and Analysis of Multivariate Observations". In Proc. 5th Berkeley Symp. Math. Statistics, Prob. 1:281-297, 1967.
31. [NH94] R. Ng and J. Han. "Efficient and Effective Clustering Method for Spatial Data Mining." In Proc. 1994 Int. Conf. Very Large Databases (VLDB'94), pages 144-155, Santiago, Chile, Sept. 1994.
32. [SCZ98] G. Sheikholeslami, S. Chatterjee, and A. Zhang. "WaveCluster: A multi-resolution clustering approach for very large spatial databases". In

- Proc. 1998 Int. Conf. Very Large Databases (VLDB'98), pages 428-439, New York, Aug. 1998.
33. [SEKX98] J. Sander, M. Ester, H.-P. Kriegel, X. Xu. "Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications". In Proc. Int. Conf. Data Mining and Knowledge Discovery, Kluwer Academic Publishers, Vol. 2, No. 2, 1998.
34. [WYM97] W. Wang, Yang, R. Muntz. "STING: A Statistical Information grid Approach to Spatial Data Mining". In Proc. 1997 Int. Conf. Very Large Data Bases(VLDB'97), pages 186-195, Athens, Greece, Aug. 1997.
35. [XEKS98] X. Xu, M. Ester, H.-P. Kriegel, J. Sander. "A distribution-based Clustering Algorithm for Mining in Large Spatial Databases". In Proc. 14th Int. Conf. Data Engineering (ICDE'98), Orlando, Florida, USA, 1998.
36. [ZRL96] T. Zhang, R. Ramakrishnan, M. Livny. "BIRCH: An Efficient Data Clustering Method for Very Large Databases". In Proc. 1996 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'96), pages. 103-114.

附錄一、程式執行原始碼

[INT]直譯程式模組
模組名稱.....:INT.DBSCAN_ALL
程式抬頭.....:'DBSCAN_ALL'

日期:97.03.03
所有權者:蔡宗欣
頁次:1

```
SET EXCLU ON
USE HCDB_0
ZAP
PACK
USE HCDB_ALL
COPY TO HCDB3001
USE HCDB3001
REINDEX
USE
SET EXCLU OFF
M1=0
M2=0
M3=1
SCR("DBSCAN")
MENU TO()
CL_TOTAL=0
DO WHILE .T.
  SELE A
  USE HCDB3001
  SELE B
  USE HCDB_0
  SELE A
  N1=0
  MIN=0
  M1R=0
  XX=0
  YY=0
  ARRAY("SYS")
  ARRAY("SYS2")
  AFILL(SYS,0)
  AFILL(SYS2,0)
```



```
FLOCK
DO WHILE .NOT. EOF()    &&清檔
    REPL CL WITH 0
    SKIP
    N1=N1+1
ENDDO
NO=1
N2=1
GO TOP
    DO WHILE .NOT. EOF()
        XX=X
        YY=Y
        N1=1
        GO TOP
        DO WHILE .NOT. EOF()
            M1R=SQRT((XX-X)*(XX-X)+(YY-Y)*(YY-Y))
            IF M1R<=M1
                IF M1R=0
                    MIN=N1
                ENDIF
                SYS[N1]=RECNO()
                REPL CL WITH M3
                N1=N1+1
            ENDIF
            SKIP
        ENDDO
        IF N1>=M2
            N2=N2+1
            EXIT
        ENDIF
        GO TOP
        DO WHILE .NOT.EOF()    &&CL 歸零
            REPL CL WITH 0
            SKIP
        ENDDO
        NO=NO+1
```

```
GO NO
ENDDO
AR=1
DO WHILE AR<LEN(SYS)+1
  GO SYS[AR]
  XX=X
  YY=Y
  GO TOP
  N11=1
  AFILL(SYS2,0)
  DO WHILE .NOT. EOF()
    M1R=SQRT((XX-X)*(XX-X)+(YY-Y)*(YY-Y))
    IF M1R<=M1                                &&集合 D 中符合檢查
      SYS2[N11]=RECNO()
      N11=N11+1
    ENDIF
    SKIP
  ENDDO
  IF N11>=M2
    ARR=1
    FLOCK
    DO WHILE ARR<LEN(SYS2)+1                &&檢查符合條件集合有無重複
      GO SYS2[ARR]
      IF CL=0
        FF=LEN(SYS)+1
        SYS[FF]=RECNO()
      ENDIF
      FLOCK
      REPL CL WITH M3
      ARR=ARR+1
    ENDDO
    UNLOCK
  ENDIF
  AR=AR+1
ENDDO
UNLOCK
```



```
SUM CL TO CL_TOTAL
IF CL_TOTAL=0
  EXIT
ENDIF
COPY TO OK FOR CL=M3
SELE B
FLOCK
APPEND FROM HCDB3001 FOR CL=M3
UNLOCK
SELE A
DELETE FOR CL=M3
CLOSE ALL
SET EXCLU ON
USE HCDB3001
PACK
USE
SET EXCLU OFF
M3=M3+1
ENDDO
```



附錄二、DBSCAN演算系統模組結構表

模組結構表	
起始模組: INT.MAIN	日期:97.05.09 頁次: 1
模 組 名 稱	
論文資料庫	
工廠資料	
新增	
更改	
查詢	
刪除	
上頁	
下頁	
門牌資料	
新增	
更改	
查詢	
刪除	
上頁	
下頁	
符合資料庫	
新增	
更改	
查詢	
刪除	
上頁	
下頁	
條例矩陣資料庫	
新增	
更改	
查詢	
刪除	
上頁	
下頁	
HCDB O	
新增	
更改	
查詢	
刪除	
上頁	
下頁	
論文程式	
地址對位	
DBSCAN	
DBSCAN3	
DBSCAM_ALL	
EPS 矩陣產生	
公用程式	
PACK"DBF"	
PACK"ALLDBF"	
SYSPACK	
CMPALL	
COMMAND	
資料結構回存	



附錄三、歐基里德距離對稱矩陣

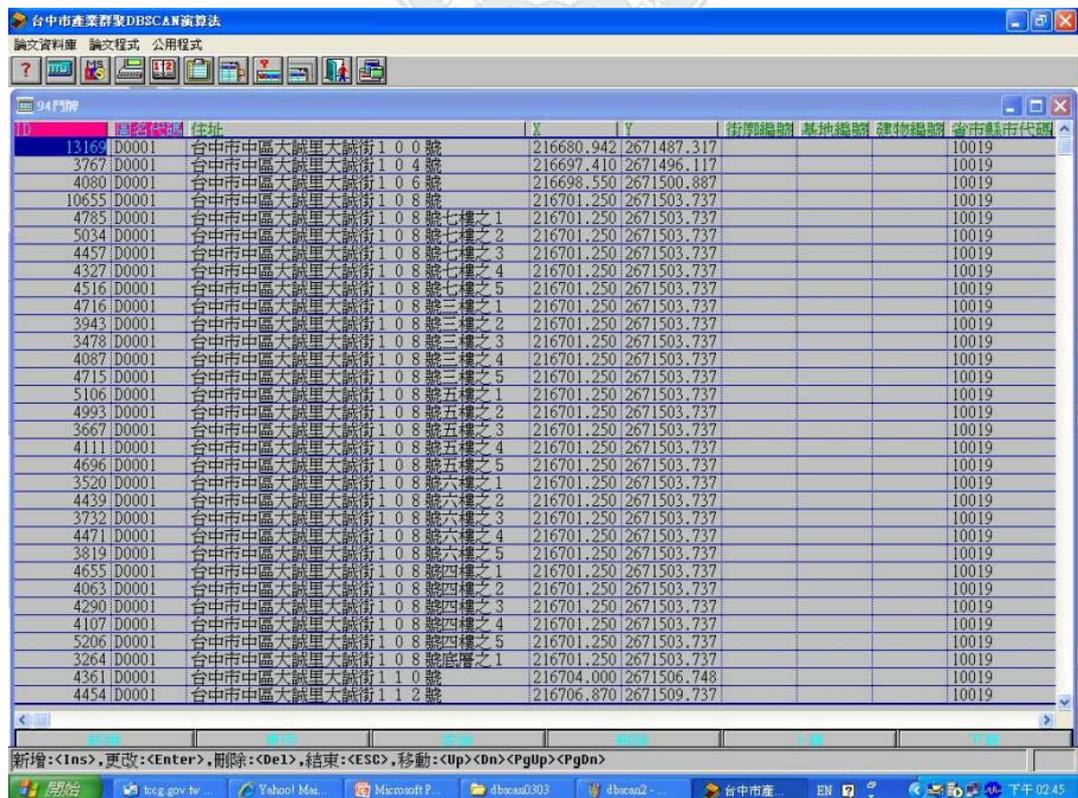
RecNo()	1	2	3	4	5	6	7	8	9	10	...	N
1	0	4307	4324	4406	1856	924	948	899	910	907	...	$\sqrt{(X_1 - X_n)^2 + (Y_1 - Y_n)^2}$
2	4307	0	131	137	3524	3462	3441	3455	3438	3435	...	:
3	4324	131	0	90	3485	3467	3445	3462	3446	3444	...	:
4	4406	137	90	0	3575	3552	3530	3547	3530	3528	...	:
5	1856	5324	3485	3575	0	1179	1166	1255	1267	1286	...	:
6	924	3462	3467	3552	1179	0	24	78	95	115	...	:
7	948	3441	3445	3530	116	24	0	89	102	121	...	:
8	899	3455	3462	3547	1255	78	89	0	22	39	...	:
9	910	3438	3446	3530	1267	95	102	22	0	20	...	:
10	907	34385	3444	3528	1286	115	121	39	20	0	...	:
:	:	:	:	:	:	:	:	:	:	:	0	:
N	$\sqrt{(X_n - X_1)^2 + (Y_n - Y_1)^2}$	0

附錄四、 DBSCAN空間群聚演算系統

程式主畫面



工廠資料匯入瀏覽畫面



台中市產業群聚DBSCAN演算法

論文資料庫 論文程式 公用程式

工廠資料

工廠名稱	區	地址	地址
奇益工業社	台中市 東區	東信里 一心街327巷5號	台中市東區東信里一心街327巷
慶泰工業社	台中市 北屯區	軍功里 新一街15號	台中市北屯區軍功里新一街15
東光實業廠	台中市 北屯區	軍功里 新一街27號	台中市北屯區軍功里新一街27
超達機殼引擎有限公司	台中市 北屯區	軍功里 新一街63號一樓	台中市北屯區軍功里新一街63
銘健企業股份有限公司台中廠	台中市 東區	東勢里 力行路241巷15號	台中市東區東勢里力行路241巷
勁聯鐵製有限公司台中廠	臺中市 東區	十甲里 十甲北街17巷17號1樓	臺中市東區十甲里十甲北街17巷
明功電鍍工業有限公司	台中市 東區	十甲里 十甲北街17巷25號	台中市東區十甲里十甲北街17巷
沅裕實業有限公司工廠	台中市 東區	十甲里 十甲北街41號一樓	台中市東區十甲里十甲北街41號
華森有限公司	台中市 東區	十甲里 十甲北街43號	台中市東區十甲里十甲北街43號
日盛隆有限公司	台中市 東區	十甲里 十甲北街45號	台中市東區十甲里十甲北街45號
長益刀具有限公司	台中市 東區	十甲里 十甲北街59號	台中市東區十甲里十甲北街59號
彩身製業股份有限公司	台中市 東區	十甲里 十甲北街62巷2號	台中市東區十甲里十甲北街62巷
福和實業工廠	台中市 東區	十甲里 十甲北街75巷14號	台中市東區十甲里十甲北街75巷
最育企業社	台中市 東區	東信里 十甲東路116號	台中市東區東信里十甲東路116
邑晉企業有限公司	臺中市 東區	東信里 十甲東路119巷8號	臺中市東區東信里十甲東路119
億發企業股份有限公司台中廠	台中市 東區	東信里 十甲東路235巷13號	台中市東區東信里十甲東路235
運安企業有限公司第一廠	台中市 東區	東信里 十甲東路235巷7號	台中市東區東信里十甲東路235
信和眼鏡廠有限公司	台中市 東區	東信里 十甲東路289號	台中市東區東信里十甲東路289
源成齒輪有限公司	台中市 東區	東信里 十甲東路71號	台中市東區東信里十甲東路71號
巨豐機械股份有限公司	台中市 東區	旱溪里 十甲路106之1號	台中市東區旱溪里十甲路106之
洽興塑膠廠股份有限公司	台中市 東區	旱溪里 十甲路108號	台中市東區旱溪里十甲路108號
久大食品廠	台中市 東區	東南里 十甲路1號	台中市東區東南里十甲路1號
宏隆鐵工廠	台中市 東區	十甲里 十甲路218巷11號	台中市東區十甲里十甲路218巷
祥聖塑膠有限公司一廠	台中市 東區	十甲里 十甲路218巷13號	台中市東區十甲里十甲路218巷
成功印刷廠	台中市 東區	十甲里 十甲路218巷17號	台中市東區十甲里十甲路218巷
資昌塑膠工業股份有限公司	台中市 東區	十甲里 十甲路218巷20號	台中市東區十甲里十甲路218巷
昇得瑞業有限公司	台中市 東區	十甲里 十甲路218巷23號	台中市東區十甲里十甲路218巷
祥聖塑膠有限公司	台中市 東區	十甲里 十甲路218巷2號	台中市東區十甲里十甲路218巷
力梅工業股份有限公司	台中市 東區	十甲里 十甲路218巷6號	台中市東區十甲里十甲路218巷
金宏昌機械有限公司	臺中市 東區	十甲里 十甲路218巷7號	臺中市東區十甲里十甲路218巷
延成實業股份有限公司	台中市 東區	十甲里 十甲路281巷40號	台中市東區十甲里十甲路281巷

新增: <Ins>, 更改: <Enter>, 刪除: , 結束: <ESC>, 移動: <Up><Dn><PgUp><PgDn>

開始 | tcc.gov.tw | Yahoo! Mail | Microsoft P... | dbcan0303 | dbcan6 | 台中市產... | EN | 下午 02:44

門牌點位資料匯入瀏覽畫面





DBSCAN 演算參數輸入畫面

附錄五、 台中市工商資料庫(Microsoft SQL SERVER 2000)

AGENCYCODE	BAN_NO	COMPANY_NAME	ORGN_TYPE	ZONE_CODE	AREA_CODE	COMPANY_ADDR	TEL_NO	CAPITAL_AMT	REAL_AMT	NAME
376590000A	52574175	台名貿易股份有限公司	01		10019	臺中市西屯區大鼎		5000000000000	5000000000000	劉邦
376590000A	84365144	費程建設開發有限公司	02		1001907000	臺中市南屯區永建	04-3843728	2500000000000	2500000000000	王博
376590000A	89808724	晉永有限公司	02		10019	臺中市南屯區文心		1000000000000	0000000000000	洪重
376590000A	86987478	瑞寶建設開發股份有限公司	01		10019	臺中市五權西路1		2500000000000	2500000000000	蔡武
376590000A	86430129	早安超級商店有限公司	02		10019	臺中市西屯區長安		2000000000000	0000000000000	郭全
376590000A	09411004	摩特企業有限公司	02		10019	臺中市文心路4		1000000000000	0000000000000	陳進
376590000A	16507118	朝富建材實業有限公司	02		1001908	臺中市北屯區平康	04-2543945	1000000000000	0000000000000	吳一
376590000A	89905578	紅葉子有限公司	02		10019	臺中市北屯區明德里		1000000000000	0000000000000	陳東
376590000A	23429928	致鼎企業股份有限公司	01	00000	1001904000	臺中市西屯區五權路	04-2251782	5000000000000	5000000000000	彭坤
376590000A	23269491	康勤有限公司	02		10019	臺中市西屯區光朝		5000000000000	0000000000000	廖奇
376590000A	86567659	鈺宇企業管理顧問	01		10019	臺中市西屯區長安		1000000000000	1000000000000	吳俊
376590000A	84651940	康奕貿易有限公司	02		10019	臺中市北屯區昌祥		5000000000000	0000000000000	余俊
376590000A	52917334	利航企業有限公司	02		10019	臺中市北屯區進化		1000000000000	0000000000000	王游
376590000A	86754999	延昌油漆有限公司	02		1001903000	臺中市南屯區美村街	0423714830	1100000000000	0000000000000	黃俊
376590000A	86838858	上品家具股份有限公司	01		1001904000	臺中市西屯區福誠街	0423193839	1400000000000	1400000000000	黃寶
376590000A	86208036	唯來企業有限公司	02		10019	臺中市公園路一		2000000000000	0000000000000	洪振
376590000A	A0186524	立偉機械股份有限公司	01		10019	臺中市台中路1		2400000000000	2400000000000	何斌
376590000A	51681802	隆美貿易股份有限公司	01		1001908000	臺中市大墩二	04-2236266	3000000000000	3000000000000	巫家
376590000A	84558796	天然和園藝有限公司	02		10019	臺中市南屯區南		5000000000000	0000000000000	林秋
376590000A	86585170	鑫欣企業有限公司	02		1001906000	臺中市西屯區何厝	04-3239266	5000000000000	0000000000000	陳銘
376590000A	89384535	壹家室內設計有限公司	02	00000	1001903000	臺中市南屯區東興	04-2625891	1000000000000	0000000000000	蘇錦
376590000A	97205182	永力農牧開發股份有限公司	01		10019	臺中市東區立德		5000000000000	5000000000000	何松
376590000A	52498108	金吉利玩具實業有限公司	01		10019	臺中市西屯區中		5000000000000	5000000000000	林昇
376590000A	23083328	鴻大製模廠有限公司	02		10019	臺中市東區大興		1000000000000	0000000000000	莊林
376590000A	89821986	麗聲空間造型設計有限公司	02		10019	臺中市國光路四		1000000000000	0000000000000	卜世
376590000A	86332149	中正文化企業有限公司	02	00000	1001907000	臺中市南屯區大		5000000000000	0000000000000	周世
376590000A	97386389	環視服裝有限公司	02		1001906000	臺中市西屯區遠	04-3846885	2000000000000	2000000000000	李應
376590000A	23542739	新富木器有限公司	02		1001906000	臺中市西屯區大	04-3115558	5000000000000	0000000000000	劉應
376590000A	22239127	南偉有限公司	02	40300	1001904000	臺中市西屯區文		7000000000000	0000000000000	陳正
376590000A	89303885	恆昇紙品有限公司	02		1001908000	臺中市北屯區大	04-2920015	3000000000000	0000000000000	吳貴
376590000A	22406052	新順成電器有限公司	02		10019	臺中市北屯區陳		1000000000000	0000000000000	劉桂
376590000A	22796235	杰利工程企業有限公司	02		1001908000	臺中市北屯區北		1000000000000	0000000000000	林秋
376590000A	86596649	英發建設有限公司	02		10019	臺中市西屯區公		2500000000000	0000000000000	蔡祥
376590000A	A0316996	明晉網球俱樂部有限公司	02		10019	臺中市南屯區南		1000000000000	0000000000000	鄭木
376590000A	84563453	士達測量工程有限公司	02		10019	臺中市南屯區清		3000000000000	0000000000000	徐士
376590000A	89889487	大汗企業管理顧問	02		10019	臺中市北屯區平		1000000000000	0000000000000	胡炳
376590000A	16253487	巧得實業有限公司	02	408	1001907000	臺中市南屯區田		5000000000000	0000000000000	卓育
376590000A	89546530	松禾建設有限公司	02	00000	1001908000	臺中市北屯區崇	2955585	2500000000000	0000000000000	陳彥
376590000A	51316004	東霖家具裝潢有限公司	02		1001906000	臺中市西屯區福	04-2529100	9000000000000	9000000000000	方東
376590000A	97493304	汗青資訊有限公司	02		10019	臺中市南屯區光		5000000000000	0000000000000	鄭元
376590000A	16416686	宏風國際有限公司	02		1001905000	臺中市北屯區西	0423266157	6000000000000	0000000000000	陳清
376590000A	02998613	全景電腦股份有限公司	01		10019	臺中市北屯區三		1200000000000	1200000000000	許遠
376590000A	22771804	宏工業有限公司	02		1001906	臺中市西屯區福		2000000000000	0000000000000	干基



附錄六、 台中市門牌點位資料(MAPINFO 7.0)

MapInfo Professional

File Edit Tools Objects Query Table Options Browse Window Help

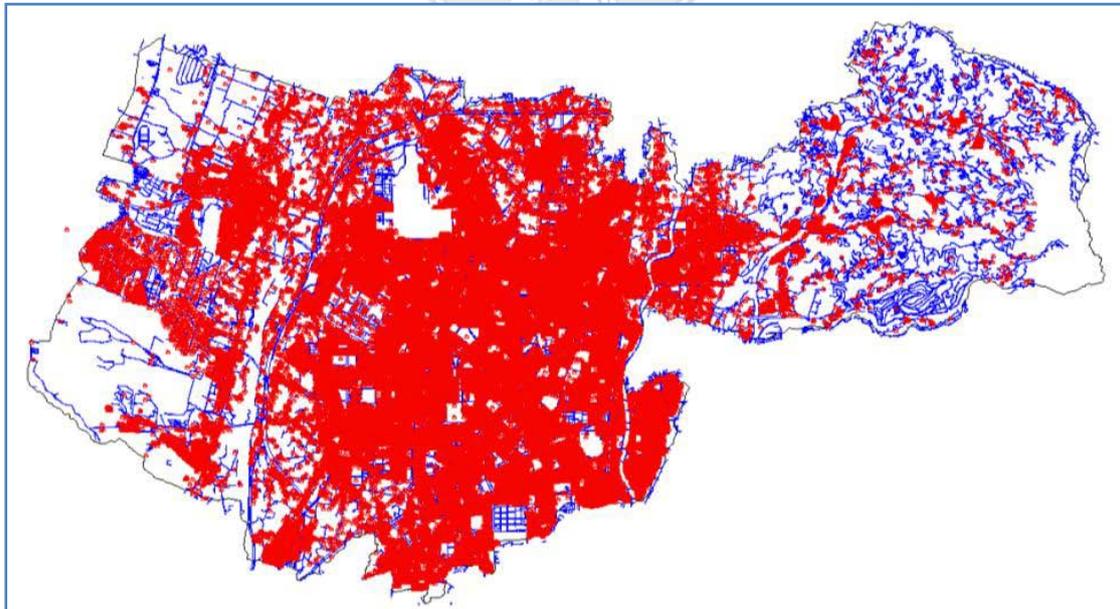
九十四年全市門牌67 Map

九十四年全市門牌67 Browser

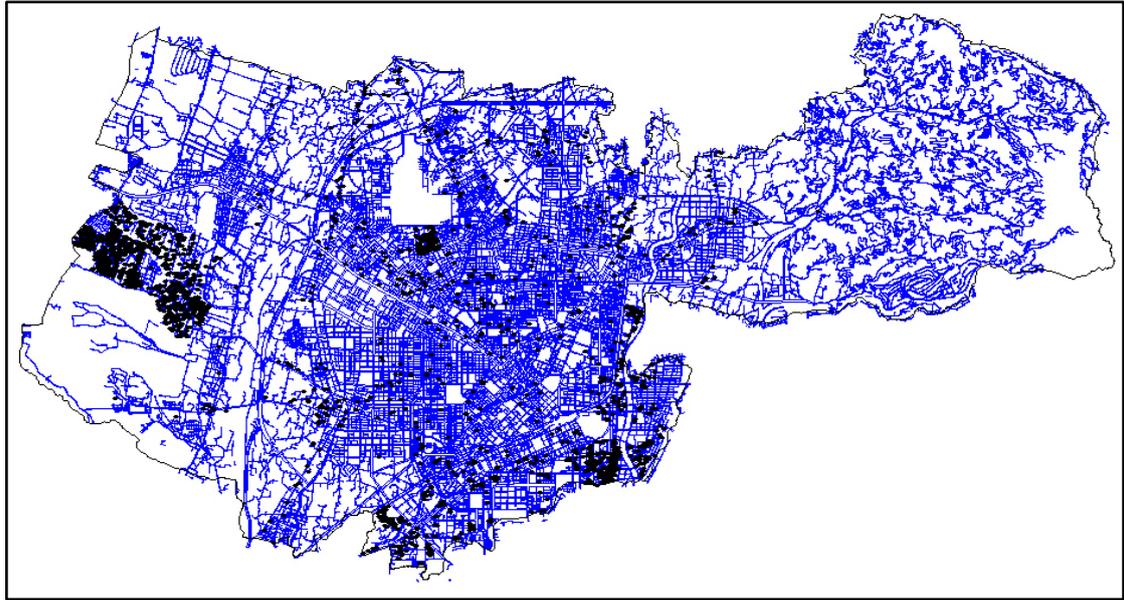
ID	圖名代碼	街序編號	基地編號	建物編號	省市縣市代碼	行政區代碼	村里	鄰	街路段名	地區
494,011	D0001				10019	北屯區	松強里	002鄰	昌平東二路	
494,012	D0001				10019	北屯區	松強里	002鄰	昌平東二路	
494,013	D0001				10019	北屯區	松強里	002鄰	昌平東二路	
494,014	D0001				10019	北屯區	松強里	002鄰	昌平東二路	
494,015	D0001				10019	北屯區	松強里	002鄰	昌平東二路	
494,016	D0001				10019	北屯區	水景里	008鄰	景賢三路	
494,017	D0001				10019	北屯區	平陽里	017鄰	旅順路一段	
494,018	D0001				10019	北屯區	軍功里	001鄰	松竹路一段	
494,019	D0001				10019	北屯區	仁和里	026鄰	崇德十路一段	
494,020	D0001				10019	北屯區	仁和里	026鄰	崇德十路一段	
494,021	D0001				10019	北屯區	仁和里	026鄰	崇德十路一段	
494,022	D0001				10019	北屯區	仁和里	026鄰	崇德十路一段	
494,023	D0001				10019	北屯區	仁和里	026鄰	崇德十路一段	
494,024	D0001				10019	北屯區	仁和里	026鄰	梅川東路五段	
494,025	D0001				10019	北屯區	仁和里	026鄰	梅川東路五段	
494,026	D0001				10019	北屯區	舊社里	005鄰	松竹路一段	
494,027	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,028	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,029	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,030	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,031	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,032	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,033	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,034	D0001				10019	北屯區	同榮里	003鄰	豐樂五街	
494,035	D0001				10019	北屯區	水景里	008鄰	軍福十一路	
494,036	D0001				10019	北屯區	松強里	002鄰	昌平三街	

records 494011 - 494036 of 494036

Mapinfo 開啟門牌檔



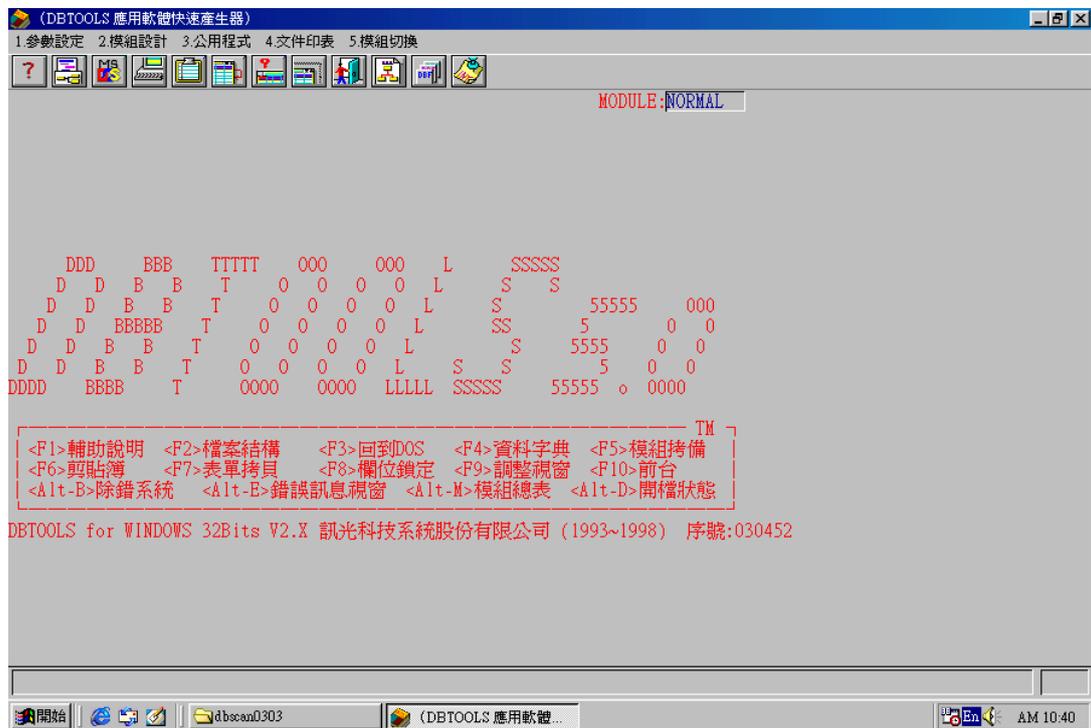
GIS 門牌點位圖



GIS 工廠點位圖



附錄七、開發程式工具主畫面(Dbtools for windows 32bit)



附錄八、台中市政府GIS資料倉儲系統(數位地圖館)

