

# 普查資料之空間資料挖掘研究— 以台北市便利商店區位分析為例

## Spatial Data Mining on Census Data — A Case Study for Location Analysis of Convenience Stores in Taipei City

榮峻德\*

孫志鴻\*\*

李萬凱\*\*\*

Chin-Te Jung

Chih-Hong Sun

Wan-Kai Lee

### 摘 要

由於戶口、工商普查資料可以真實的反應區域裡的社會經濟情況，所以不管是政府部門或者是私人企業都可以用來作為規劃以及分析的依據。本研究因此整合了 2000 年台北市戶口普查資料、2001 年台北市工商普查資料以及台北市的空間資料，並利用台北市土地利用現況調查資料以及台北市門牌地址點資料，將 2001 年工商普查資料從現行最小統計單元的「里」分派到與戶口普查資料對應的「普查區」空間單元中，成為台北市普查區的空間普查資料庫。並且使用資料挖掘技術中的關聯規則方法，應用在台北市便利商店區位分析中，挖掘出便利商店座落區域中的普查資料與其他空間資料的關連性，並將影響便利商店區位因子加以量化描述，找出平常所沒有注意到、有趣的關聯規則。最後整合 GIS，將重要的規則在空間上做呈現，幫助後續決策者，重新評估便利商店的區位，或著在新設置便利商店的區位上，能夠有更進一步的參考。

**關鍵字：**空間資料挖掘、關聯規則、GIS、普查資料、區位分析

---

\* 國立台灣大學地理環境資源研究所碩士班研究生  
Graduate Student, Institute of Geography, National Taiwan University.

\*\* 國立台灣大學地理環境資源學系教授  
Professor, Department of Geography, National Taiwan University.

\*\*\* 國立台灣大學土木工程學系博士候選人  
Ph. D. Candidate, Department of Civil Engineering, National Taiwan University.



## Abstract

Census data often accurately reflect the socio-economic trends and characteristics of the region under survey, they serve as a useful planning tool for both the government and the private enterprises. In this research, we collect the Taipei census data in 2000, the industrial census data in 2001, and the present Taipei geospatial data into the Spatial Census Database. In order to converse the industrial census data in 2001 into the same units of Taipei Census Database, we develop a data conversion model which use the land use data and “Geographic Data Management System of Building Address” developed by Taipei Municipal Government to complete it. Such database is analyzed by the association rules of data mining in the location of the existing convenience stores in the Taipei area. The analysis unveils the interesting association rules among census, demographics and the location of convenience stores. Finally, this research incorporates geographic information system (GIS) technology to map the result of the analysis so as to provide a spatial representation that will be more easily understood.

**Keywords :** Spatial Data Mining, Association Rules, GIS, census data, location analysis.

## 前 言

根據聯合國統計，公元 2000 年前後，全球有九成以上國家同步展開人口普查工作，而普查的目的在於提供基本統計資料作為國家建設的依據，為了使普查資料能真正反映區域的社會經濟狀況，普查資料必須與地理區位建立關係，藉此關係的建立，普查資料所反映的區域社經狀況才能與區域的其他資料如自然環境、天然資源、人為設施等相結合，如此普查資料才能充分被使用，發揮輔助規劃、決策的效能 (朱子豪等，1990)。由於戶口普查資料以及工商普查資料可以真實的反應區域裡的社會經濟情況，所以不管是政府部門或著是私人企業都可以用來規劃以及分析，譬如：分析台灣產業空間結構 (孫嘉玲，2000；許麗蕙，2001；林晶晶，2003)、人口成長與空間變遷 (白仁德，2004)、台北高雄市長期照護設施空間分派之研究 (蔡博文，2004)、研究有車階級的指標 (Klösgen, 2002) 以及以家庭擁有的車輛來探討交通便捷性問題 (Malerba *et al.*, 2002) 的研究等。

對於便利商店的區位選址而言，普查資料也扮演了相當重要的角色，提供大量豐富的社會經濟資料為分析的背景。在過去，傳統分析便利商店區位的方法大致上是以列表法 (Checklist method)、類比法 (Analog approach)、回歸模型 (Regression Model) 以及區位配置

模型 (Location allocation model) 等四種為主 (Kuo *et al.*, 2002)。這些方法雖能有效分析較少量的便利商店區位，但是一旦分析的對象或資料量變多時，這些傳統的方法就不能提供有效的分析。而資料挖掘 (Data Mining) 或資料庫的知識發現 (Knowledge Discovery in Databases) 即是針對大量的資料做分析處理，歸納出有用的、我們不常注意到的知識出來，以改善傳統分析方法的不足 (Fayyad *et al.*, 1996)。

分析便利商店的「區位」，必須要以空間面向的切入，才能完善的分析出有用的成果。而空間資料挖掘 (Spatial Data Mining) 屬於資料挖掘的一個應用領域，專門針對空間或非空間性的資料來找尋有用的空間知識 (Ng, 1994；Koperski, 1995；Chawla *et al.*, 2000)。Ester 在 1999 年提出了五個不同的空間資料挖掘方法：Spatial Characterization、Spatial Discriminate (Outlier) Analysis、Spatial Classification Analysis、Spatial Clustering 以及 Spatial Association Rules，每個方法都針對不同的資料特性或著依研究主題的不同所發展出來的。本研究為了尋找出在便利商店區位中，空間以及普查資料因子間的「關連規則」，使用空間資料挖掘中的空間關聯規則方法 (此方法是最能有效率的獲取到便利商店與其相關因子間的關連規則)，並以台北市為研究區域，訂定 250 公尺為便利商店的服務範圍，彙整台北市 2000 年戶口普查

資料、2001 年工商普查資料以及其他與便利商店相關的空間資料，整合成台北市空間普查資料庫。藉由此資料庫來挖掘分析現今便利商店的顧客來源屬性以及便利商店與其他空間資料間的關係，找出可能的、而且是我們平常沒被發現的規則出來，並整合 GIS，以空間的方式呈現，幫助使用者能夠更有效率的從地圖中的知識歸納出正確的資訊，做正確的決策。

## 文獻探討

### (一) 空間資料挖掘相關研究

近年來資料挖掘的研究與應用已從對屬性資料的挖掘延伸到對空間資料的挖掘 (Bell, 1994; Ester, 1997; Han, 1997; Koperski, 1995; Ng, 1996; Shekhar, 2003; Zhou, 1999)。空間資料挖掘 (Spatial Data Mining) 經 Ester (1999) 的定義為：對於空間資料所進行的資料挖掘研究。空間資料挖掘是針對空間資料庫，包含空間性的資料以及非空間性的資料，來找尋出彼此間的關係 (空間關係、空間和非空間性的關係) 或著具有某種規則的模式。而空間資料具有不同於屬性資料的資料特性以及獨特的特徵 (Miller, 2001)：(1) 空間資料具有拓撲 (Topological) 和距離 (distance) 關係，通常都透過空間資料存取方法存取，所以常常需要空間推理、幾何計算和空間知識表示技術來描述；(2) 屬性資料的資料挖掘演算法都假設資料彼此獨立，而在空間資料中每個物件可能會受其鄰近若干個物件的影響，資料之間也會相互依賴，具有空間自相關性 (Spatial Autocorrelation)，所以使得空間資料挖掘有其獨特性與發展性；(3) 空間資料挖掘的結果包含空間物件，往往是需要圖形或圖像資訊來做展示，不同於一般屬性資料的結果，很難用文字表示清楚，需要涉及到空間知識的視覺化。從空間資料中挖掘潛在的知識需要考慮空間特性是空間資料挖掘的顯著特點，使之具有很多挑戰性的問題 (Mennis, 2003)。所以如果用傳統的資料挖掘方式來對空間資料庫做處理時，必定會造成些許的錯誤，因此在處理空間資料挖掘時必須使用特別的空間運算方法以及特有的空間知識背

景做支撐，才能對空間資料做最好的描述。

空間資料挖掘技術經 Ester (1999) 整理，大致有五種主要的技術：(1) Spatial Characterization：空間特徵分析，針對所做的議題從資料庫中抓取符合議題的相關資料，做空間或是非空間上的描述；(2) Spatial Discriminate (Outlier) Analysis：空間區分分析，是指將「目標類」物件的一般特徵與一個或多個「對比類」物件的一般特徵做比較，近而得到用於區分目標類和對比類的對比度量。譬如：比較不同地方的坡度和坡向、公用設施的地區差異等空間區分分析，方便使用者做區別以及加以判讀；(3) Spatial Classification Analysis：空間分類分析，是指對空間物件基於某一特性或空間關係，利用其他的資料按照一定的分類模型進行歸類，屬於監督性分類的一種方式。譬如分析「山是否會崩塌」，空間分類分析就可以利用其他崩塌相關的資料，進行分類分析，歸納出在符合哪些條件下，可能會造成崩塌；(4) Spatial Clustering：空間分群分析，是將空間資料庫中相似的資料分成一小組，提高每個不同小組之間的異質性，屬於非監督性分類。譬如：可以利用地震的資料庫，將每次地震發生的地點與次數等的資料，做 Spatial Clustering 分析，依據資料的特性找出在哪些地區發生地震次數是較高的，可能屬於斷層帶，應減少人為開發，以避免有重大災情的出現；(5) Spatial Association Rules：空間關聯規則分析，是針對大量的空間以及非空間資料，找出彼此之間的關連性，並且歸納成規則方式來做展現 (Han, 1995)。Klösgen (2002) 以及 Malerba *et al.* (2002) 都曾以空間關聯規則來分析普查資料與空間資料，找出彼此之間的關連規則以提供後續使用者參考。本研究為了尋找出在便利商店區位中，空間以及普查資料因子間的「關連規則」，將採用空間關聯規則來分析台灣的普查資料，應用在台北市便利商店的區位分析中。將在下節更詳細介紹空間關聯規則的概念與方法。

### (二) 空間關聯規則 (Spatial Association Rule)

空間關聯規則是由關聯規則所演變而來的，應用在空間資料上是空間關聯規則與關聯規則最大的不同之

處。關聯規則在 Han (2000) 定義為：是從大量資料項目集合之間，發現有趣的關聯或相關的規則，以幫助決策者做出正確的決策。Agrawal 在 1994 年提出關連規則最基本的演算法 Apriori，透過計算資料庫中每項欄位因子之間的關連程度，以  $A \rightarrow B$  (Support, Confidence) 的規則來做呈現。其中 A 和 B 是屬於資料庫中的某項欄位因子，Support (支持度) 則是代表了 A 和 B 出現在資料庫中的機率，即  $\text{support} = (\text{A 和 B 出現的次數} / \text{資料庫中的總次數})$ ；而 Confidence (信心值) 則是代表了在 A 出現的情況下，又會有 B 出現的機率，即  $\text{confidence} = (\text{A 和 B 出現的次數} / \text{A 出現的次數})$ 。譬如：牛奶  $\rightarrow$  麵包 (40%, 80%)，表示在整個資料庫中出現牛奶和麵包有 40%，並且如果購買牛奶後，還會有 80% 會購買麵包，即透露出牛奶和麵包有高度的購買相關性，所以在商品的擺設上，可以擺在一起增加購買率。我們可以透過設定最小支持度以及最小信心值，來過濾不滿足此限制的規則，進而找出較為強烈的規則來進行分析。而空間關聯規則是針對空間和非空間資料歸納出某種關連程度之規則，規則越強烈代表了此關聯規則出現較為頻繁且是互相影響的 (Han, 1995)。在對於空間資料做空間關

連法則時，我們不能直接的將關聯規則的方法應用在空間資料中，因為空間資料具有空間關係 (Spatial Relation) 且彼此互相影響。所以在處理空間物件與非空間物件的空間資料挖掘前，必須先對於空間關係做處理，譬如：鄰近 (Near by)、在裡面 (Inside)、包含 (Contain)... 等 (Egenhofer, 1991; Koperski, 1995; Clementini, 2000; Appice, 2003)。之後才能交由關聯規則分析空間以及非空間資料。

### (三) 便利商店區位因子

便利商店最大的關鍵成功因素，即是在於便利商店的區位 (Nelson, 1968; Applebaum, 1968; Gruen, 1960)。方國定 (1999) 曾針對國內外相關便利商店文獻，整理出便利商店區位選擇之評估項目，包括有：商圈內的住戶資料、商圈內之住戶水準、商圈內競爭性、商圈內繁榮程度及發展性、商圈內同質店競爭性、店面的所在地、店面環境條件、便利性條件、賣場特性以及其他值得注意的事項。本研究利用方國定 (1999) 所整理出來的便利商店區位因子表加以修改整理，並且加入後續研究所需之資料，如表 1 所示。

表 1 便利商店區位因子表 (方國定, 1999; 本研究修改之)

因 子	說 明	所需資料
商圈內 住戶資料	住戶種類、家庭之個戶數目、居民消費水準、教育程度、所得水準、年齡。	戶口普查資料的家庭組成、住戶職業、教育程度、年齡等相關資料。
商圈內同質店 競爭性	由於同業的競爭將會造成市場的瓜分，故在選擇位置的時候必須考量附近的同業數量。	計算每家便利商店商圈之內其他便利商店數量。
商圈內繁榮程度、 發展性與店面的 環境條件	商圈內的公共設施或著能聚集人潮的設施出現、附近車流量、馬路寬度、商圈內之工作人數。	捷運站、郵局、醫院、文化美術館、寺廟古蹟、公園、小吃、飯店、餐廳、銀行、圖書館、重要道路... 等以及工商普查資料中的工作人數資料。
店面的所在地	是否便利商店位於住宅區、商業區、辦公大樓區等。	使用台北市土地利用資料獲得



## 研究方法

由於本研究採用台北市 2000 年戶口普查以及 2001 年工商普查資料做空間資料挖掘之背景資料庫，所以必須先對戶口、工商普查資料（非空間性的資料）做處理，並且跟空間地圖資料做整合，才能跟其他空間性的資料（譬如：土地利用資料、道路、捷運站...等空間資料）做疊合、分析，產生交易表( Transaction Table)交由資料挖掘工具進行挖掘，歸納出有趣的便利商店區位規則。

### (一) 普查資料庫的建立

#### 1. 台北市 2000 戶口住宅普查資料

根據行政院主計處普查局 (2003)，戶口及住宅普查依照每十年，針對台閩地區二千二百萬人口、七百萬住宅進行全面性普查。調查的內容包括了全國人口之質量、家庭結構、就學就業及住宅使用狀況。並且台北市 2000 年的戶口普查資料為了方便普查員之作業，參考國外的統計普查區概念，建立了比村里還要小的「普查區」空間單元 (羅國華, 2000)。本研究以普查區最小空間單元為基礎，彙整台北市 2000 年戶口普查資料中的性別、年齡、婚姻狀況、就學狀況、教育程度、工作情形、行業代號、有無工作、是否為原住民、族群、住戶類別、戶內人口數、65 歲以上居住狀況、家戶型態、照護程度以及宅內人口數等 16 項欄位。並與之後的工商普查資料做整合。

#### 2. 台北市 2001 工商普查資料

根據行政院主計處普查局 (2003)，工商普查依照每五年辦理一次，舉辦工商普查之目的係為針對台閩地區蒐集工商及服務業經營概況、資源分布、主要設備、資本運用、生產結構、產銷變動及其他有關經濟活動基本資料做全面性的調查，使得政府或民間企業能跟依照此普查來規劃政策、投資策略。本研究將採用西元 2001 年工商普查資料為研究資料，採用的欄位只有：樓地板面積、業別代號、從業員工總數等三個欄位，這是由於工商普查資料只能匯集到里的行政區界，不能整合進

2000 年戶口普查同樣的普查區界中，故本研究必須採用分派模式，將里的工商普查資料分派到普查區中，與戶口普查資料做整合，由於分派模式的限制（假設就業人口數與樓地板面積成正相關，樓地板面積與樓層數成正相關），故只能針對工商普查中的三個欄位項目做分派，整合至普查區的空间單元中。

#### (1) 分派模式

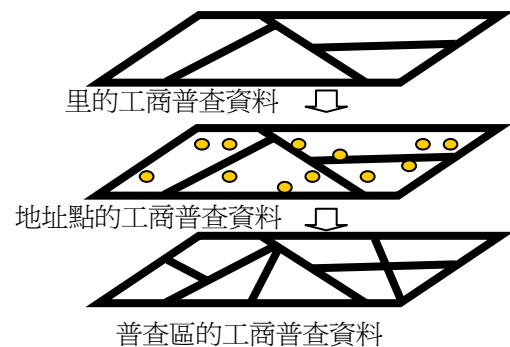


圖 1 工商普查分派流程圖

本研究應用李萬凱 (2003a; 2003b) 的分派模式方法做改良，將台北市每一里的工商普查資料，先分派到每一個地址點中，再由地址點彙整到戶口普查所調查的普查區中，達到以普查區為最小空間單元，來分析戶口與工商普查資料之目的，如圖 1 所示。所需的資料為台北市門牌地址點資料、台北市土地利用資料以及 2001 年工商普查資料。利用式 (1) 計算每個地址點的樓高、所位在的土地利用面積、以及在同一個土地利用中的地址點總數因子，得到每個地址點的樓地板面積。再利用式 (2)，計算每個地址點的樓地板面積、土地利用中的總樓地板面積、樓高、以及地址點數的因子，得到每個地址點在某一里的權重值，之後就可以依照式 (3)，將每一里的工商普查資料乘上地址點的權重，而得到每個地址點的工商普查資料。最後再將同一個普查區地址點集合起來，成為台北市普查區的工商普查資料，並且結合普查區的戶口普查資料，成為一個完整的台北市普查區戶口、工商普查資料庫。

式 (1)

$$Fa = \frac{La * F_{\max} * FA_{\text{工商}}}{\sum \left( \frac{F_{\max} * La}{Lp} \right)}$$

*Fa*：地址點的樓地板面積*La*：土地使用面積*Lp*：位在同一個土地使用之點數*Fmax*：地址點的最高樓層*FA 工商*：工商普查中每里的樓地板面積數量

式 (2)

$$W = \frac{F_{\max} * Lfa}{\sum \left( \frac{F_{\max} * Lfa}{Lp} \right)}$$

*W*：地址點之權重*Lfa*：土地使用中的總樓地板面積*Lp*：位在同一個土地使用之點數*Fmax*：地址點的最高樓層

式 (3)

$$Jp = Jli * W$$

*Jp*：地址點中的工作人數*Jli*：工商普查中一里的工作人數*W*：由公式 (2) 所計算出來的地址點權重

## (二) 空間普查資料庫之建立

1. 加入台北市便利商店點資料與相關空間資料

便利商店超市資料包括 7-11、全家、OK、萊爾富、

福客多、松青超市、頂好超市...等，總共有 1000 筆資料。但本研究只認定 7-11、全家、OK、萊爾富、福客多為便利商店，共 899 筆資料；剩下的 101 筆資料都屬於超市資料。而空間資料，包含了台北市土地利用、警察消防局、郵局、醫院、文化美術館、教堂、寺廟古蹟、圖書館、學校、加油站、銀行、公園、飯店、餐廳、捷運站、道路等可能會與便利商店座落相關的空間資料。

### 2. Transaction Table

在交由關聯規則分析之前，必須先將資料整理成交易表 (Transaction Table) 的形式，譬如：某家便利商店位在哪條道路上、靠近哪所學校、距離是多少以及所位在的普查區中的人口特性、教育程度如何等相關的資訊做整理。而本研究利用 ESRI ArcMap 8.x 軟體中的 Spatial Join 功能，先將 1000 筆台北市便利商店點資料跟上述空間資料之間的空間關係，做距離的計算，找出某家便利商店與上述空間資料最近的空間物件有哪些。並且由於便利商店具有一定的服務範圍，經濟部商業司 (2000) 曾經對於便利商店的服務範圍加以定義：以 300 公尺徒步距離為界，商圈的規模以 3000 人為界。但是由於本研究以台北市為研究區，便利商店數量較為密集，以 300 公尺為服務範圍不太適合，故本研究訂定以 250 公尺為便利商店的服務範圍。並且將便利商店的服務範圍跟之前整合在台北市普查區中的戶口、工商普查資料做疊合，找出每一個便利商店 250m 商圈中所具有的戶口、工商普查資料的特性為何，如表 2 所示。以第一筆資料說明之，代表了位在內湖區金龍里的 7-11 便利商店，跟內湖分局距離有 573.3m、離最近的郵局距離有 563.0m、在 250m 商圈中的男生人數有 1288 人、女生人數有 1364 人、碩士畢業人數有 92 人...等，並且給予唯一代碼 (FID)，利於之後的空間展現。



表 2 便利商店的 Transaction Table

FID	便利商店	便利商店名稱	區	里	警察消防局	警察消防局名稱	警察消防距離 (m)	郵局	郵局距離 (m)	學校	...	男	女	總人數	高中	高職	碩士	博士	...
0	便利商店	7-11	內湖區	金龍里	警察局	內湖分局	573.3341	郵局	563.0142	國小	...	1288	1364	2652	434	227	92	16	...
1	便利商店	7-11	士林區	百齡里	警察局	後港派出所	386.1735	郵局	393.6521	國小	...	2895	3268	6163	849	378	439	121	...
2	便利商店	萊爾富	士林區	承德里	警察局	後港派出所	529.649	郵局	509.274	國小	...	4023	4238	8261	1376	828	317	52	...
3	便利商店	萊爾富	中山區	大直里	消防局	消防團	152.3975	郵局	358.8328	大專學院	...	3492	4041	7533	984	455	522	133	...
4	便利商店	7-11	中山區	大直里	警察局	憲兵隊	82.22532	郵局	203.0394	大專學院	...	137	63	200	26	38	9	13	...
5	便利商店	7-11	中山區	大直里	消防局	消防團	160.7047	郵局	402.4127	大專學院	...	2062	2271	4334	616	279	303	87	...
6	便利商店	全家	內湖區	西湖里	警察局	西湖派出所	1086.15	郵局	136.8576	國小	...	3437	3920	7358	1041	504	448	91	...
7	便利商店	福客多	中山區	成功里	警察局	憲兵隊	225.1777	郵局	80.45494	大專學院	...	3576	4090	7666	1057	446	645	204	...
8	便利商店	7-11	北投區	湖山里	警察局	派出所	1424.382	郵局	177.519	大專學院	...	1029	972	2001	299	257	34	21	...
9	便利商店	OK	北投區	中心里	警察局	派出所	466.1352	郵局	608.8317	大專學院	...	1067	1022	2089	313	257	40	30	...
10	便利商店	7-11	北投區	開明里	消防局	消防隊	458.9172	郵局	388.4018	大專學院	...	1288	1364	2652	839	675	92	4	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
999	便利商店	萊爾富	文山區	試院里	警察局	復興派出所	37.64306	郵局	273.854	國小	...	676	696	1372	163	85	94	33	...

### (三) 空間普查資料挖掘

本研究採用 SPSS 公司所開發的一套強大資料挖掘工具 Clementine，此工具是一套以流程控管的視覺化介面，並將各種分析方法（包括：Classification、Association Rules、Cluster、Neural network...等以及其他相關的統計分析方法）、輸入輸出資料的方式都加以模組化，形成一個一個的組件，使用者就可以很輕易依照自己的研究架構，將所需要的各種工具，加以組合，拼湊成一個分析系統。本研究是採用 Association Rules 的技術來挖掘空間普查資料庫，利用 Clementine 流程化的視覺介面，更有利於本研究結果的產生。

#### 1. 資料預處理

由於 Shekhar 在 2003 年指出關聯規則是針對名目式資料所發展出來的資料來挖掘方法，對於數值性的資料不適合做處理，也就是關聯規則方法不適合處理數值性的資料。故在關聯規則分析前，必須先將數值性的資料分類，

轉換成名目式的資料，譬如距離的遠近，分成 far 或 near 等模糊式的名目式資料來描述原本數值性的資料。而不同的分類方式，將會影響到之後關連法則所挖掘出來的結果 (Mennis, 2003)。本研究曾嘗試過用各種不同的分類方式：等距 (Equal Width)、等量 (Equal Count) 以及 Natual breaks (Jenks, 1963) 分類法，最後以等距分類且分成五級的效果最好 (通過最小支持度、最小信心值的規則數量最多且因子多集中在特定的數值組距中，最有助於本研究的精確性)，故本研究利用等距的分類方式分成五類 (1、2、3、4、5)，來將數值性的資料轉換成名目式資料，並且在每一個分級後面再加上原本組距的範圍，方便解釋規則之用，譬如：郵局的距離分類為 1，實際的分類組距為 <379m，故在將郵局距離這個數值性資料轉換成名目式的資料形式即為 1 (<379)，如表 3 所示。



表 3 數值性資料轉名目式資料表

便利商店	便利商店名稱	...	警察消防距離	郵局距離	醫院距離	...	男	女	...
便利商店	7-11	...	2 (456~1330)	2 (379~1107)	2 (502~1485)	...	2 (941~1930)	2 (996~1985)	...
便利商店	7-11	...	1 (<456)	2 (379~1107)	2 (502~1485)	...	3 (1930~4120)	3 (1985~4326)	...
便利商店	萊爾富	...	2 (456~1330)	2 (379~1107)	2 (502~1485)	...	3 (1930~4120)	3 (1985~4326)	...
便利商店	萊爾富	...	1 (<456)	1 (<379)	1 (<502)	...	3 (1930~4120)	3 (1985~4326)	...

## 2. Association Rule

利用 Clementine 中的 Apriori 模組來進行關聯規則的挖掘，由於關聯規則  $A \rightarrow B (S\%, C\%)$ ，是找出在 A 條件出現的情況下，有多少 C% 的機率下會出現 B 的條件，所以必須先決定出 A、B 的條件為何以及訂定最小支持度 (S%) 以及最小信心值 (C%)，將不符合此閾值 (Threshold) 的規則都過濾掉，以找出強烈的關聯規則出來。由於本研究針對在什麼樣的情況下，會有便利商店的出現，故將 B 的條件鎖定為便利商店，A 的條件即為其他的空間資料以及普查資料；而關於最小支持度與最小信心值的閾值訂定並沒有一定的標準，定的越高代表了所找出來的規則越為頻繁、影響便利商店區位越為重要，相對的規則數量會越少。但是如果定的太高，將有可能忽略掉其他可能影響因子。故本研究考慮這點，為

了找出便利商店區位影響最大的因子，將最小支持度訂為 9%、最小信心值訂為 90%，找出在哪些條件發生下，會有 90% 機率下出現便利商店的分佈。

## 結果與討論

經由 Clementine 的關聯規則分析，將影響台北市便利商店的區位較為重要的規則，以表 4 來說明，並將 A 條件中的因子加以整理方便判讀，符號「^」代表「而且」之意。編號 1 的規則表示：在服務業工作人數介於 2365~4257 人且學齡前兒童有 387~542 人且距離小吃在 942 公尺以內的區域中，有 98% 的機率下，會有便利商店的出現。以此類推其他規則之解釋。

表 4 關聯規則表

ID	A	→	B	(S%, C%)
1	服務業工作人數在 2365~4257 人 ^ 學齡前兒童在 387~542 人 ^ 小吃距離 <942m	→	便利商店	(9.8%, 98%)
2	有工作人數在 6526~9099 人 ^ 主要道路距離 <12m ^ 警察消防局距離 <456m	→	便利商店	(10.3%, 96.3%)
3	學校距離在 316~905m ^ 青壯年 25~65 歲人數在 2130~3387 人 ^ 有工作人數在 6526~9099 人	→	便利商店	(10.4%, 95.4%)
4	服務業工作人數在 2365~4257 人 ^ 主要道路距離 <12 m ^ 圖書館距離 445~1320 m ^ 小吃距離 <942 m	→	便利商店	(12%, 95.2%)
5	學校距離 <316 m ^ 女生人數在 1985~4326 人 ^ 銀行距離 <443 m ^ 學生 6~25 歲人數 1200~2681 人	→	便利商店	(10%, 95.2%)
6	住宅區 ^ 單身戶人數在 324~906 人 ^ 有工作人數在 6526~9099 人 ^ 銀行距離 <443 m	→	便利商店	(9.6%, 93.2%)
7	住宅區 ^ 飯店距離 <657 m ^ 主要道路距離 <12 m	→	便利商店	(9.5%, 93.1%)
8	公園距離在 <443m ^ 商業工作人數在 3555~6365 人 ^ 學校距離 <316 m ^ 學生 6~25 歲人數 1200~2681 人	→	便利商店	(10.6%, 93%)
9	捷運站距離 <605 m ^ 郵局距離 <379 m	→	便利商店	(11.7%, 92.9%)



	^ 主要道路距離<12 m ^ 圖書館距離在 445~1320 m		
10	學校距離<316 m ^ 重要道路 ^ 銀行距離<443 m	→ 便利商店	(10.5% , 92.9%)
11	住宅區 ^ 單身戶人數在 324~906 人 ^ 商業工作人數 3555~6365 人 ^ 餐廳距離<350 m	→ 便利商店	(11.7% , 91.4%)
12	青壯年 25_65 歲人數在 2130~3387 人 ^ 女生人數在 1985~4326 人 ^ 主要道路距離<12 m ^ 餐廳距離<350 m	→ 便利商店	(9.6% , 91.4%)
13	捷運站距離<605 m ^ 青壯年 25~65 歲人數在 2130~3387 m ^ 主要道路距離<12 m ^ 圖書館距離在 445~1320 m	→ 便利商店	(9.6% , 91.4%)
14	服務業工作人數在 2365~4257 人 ^ 有工作人數在 6526~9099 人 ^ 主要道路距離<12 m ^ 餐廳距離<350 m	→ 便利商店	(9.2% , 91.1%)
15	男生人數在 1930~4120 人 ^ 女生人數在 1985~4326 人 ^ 服務業工作人數在 2365~4257 人 ^ 餐廳距離<350 m	→ 便利商店	(9.7% , 90.7%)
16	女生人數在 1985~4326 人 ^ 有工作人數在 6526~9099 人 ^ 工業工作人數<2047 人 ^ 學校距離<316 m	→ 便利商店	(9.6% , 90.6%)
17	有工作人數在 6526~9099 人 ^ 飯店距離<657 m ^ 主要道路距離<12 m ^ 圖書館距離在 445~1320 m	→ 便利商店	(12.5% , 90.6%)
18	單身戶人數在 324~906 人 ^ 主要道路距離<12 m ^ 銀行距離<443 m ^ 餐廳距離<350 m	→ 便利商店	(21.8% , 90.5%)
19	重要道路 ^ 服務業工作人數在 2365~4257 人 ^ 餐廳距離<350 m ^ 附近競爭店數量 2 家	→ 便利商店	(14.5% , 90.1%)
20	青壯年 25~65 歲人數在 2130~3387 人 ^ 主要道路距離<12 m ^ 公園距離<443 m	→ 便利商店	(9% , 90%)

若將規則跟 GIS 整合，以空間面向的方式來展示，即如圖 2 所示，在圖中深色區域即代表了現有便利商店中，符合編號 1 規則的空間分佈，可以看出大部分都是聚集在台北市的中心商業區中。代表中心商業區中由於服務業工作人數、學齡前兒童人數較多以及距離各種小吃較近的因子影響下，便利商店會在此區域中分佈。

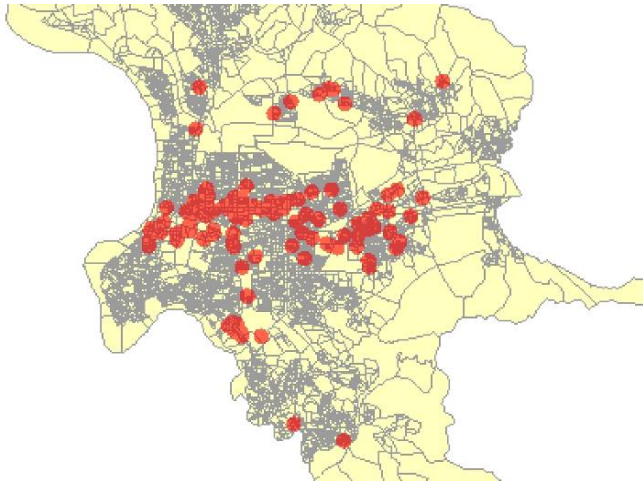
而透過表 4 可以進一步看出，影響便利商店的區位，基本上都是由某幾項最為強烈的基本因子所互相排列組合而成的。若將基本因子的數值做概括化(普查因子部分，人數越多則消費購買機率越大，以大於數值組距中最小的值，四捨五入至百位數，進行概括化；空間因子部分，距離空間因子越近則消費購買機率越大，以小於數值組距中最大的值，四捨五入至十位或百位數，進行概括化)，即可得到影響便利商店區位因子的量化指標，如表 5 所示：

表 5 便利商店區位因子量化指標

普查屬性指標	空間屬性指標
男生人數在 1900 人以上	捷運站距離 <600m
女生人數在 2000 人以上	公園距離 <400m
學齡前兒童人數 在 400 人以上	學校距離 <300 m
學生 6~25 歲人數 1200 人以上	小吃距離 <900 m
青壯年 25~65 歲人數 在 2100 人以上	主要道路距離 <10 m
單身戶在 300 人以上	銀行距離 <400m
工業工作人數在 2000 人以上	飯店距離 <600 m
商業工作人數 在 3600 人以上	附近競爭店數量 小於 2 家
服務業工作人數 在 2400 人以上	住宅區



## 結論與建議



**圖 2 服務業工作人數在 2365~4257 人且學齡前兒童人數有 387~542 人且距離小吃 < 942 公尺情況下有 98% 出現便利商店分佈圖**

將指標與文獻回顧中的表 1 互相對照整理，如表 6 所示，可以發現由空間關聯規則所挖掘出來的區位因子，不但跟前人所找出的因子相呼應，並且本研究還將這些因子量化做為設置便利商店時的參考指標，提供決策者更多評估知識。

**表 6 研究結果與文獻回顧因子對照表**

商圈內住戶資料	<ul style="list-style-type: none"> <li>● 男生人數在 1900 人以上</li> <li>● 女生人數在 2000 人以上</li> <li>● 學齡前兒童人數在 400 人以上</li> <li>● 學生 6~25 歲人數 1200 人以上</li> <li>● 青壯年 25~65 歲人數在 2100 人以上</li> <li>● 單身戶在 300 人以上</li> </ul>
商圈內同質店競爭性	<ul style="list-style-type: none"> <li>● 附近競爭店數量小於 2 家</li> </ul>
商圈內繁榮程度及發展性與店面的環境條件	<ul style="list-style-type: none"> <li>● 工業工作人數在 2000 人以上</li> <li>● 商業工作人數在 3600 人以上</li> <li>● 服務業工作人數在 2400 人以上</li> <li>● 捷運站距離 &lt;600m</li> <li>● 公園距離 &lt;400m</li> <li>● 學校距離 &lt;300 m</li> <li>● 小吃距離 &lt;900 m</li> <li>● 主要道路距離 &lt;10 m</li> <li>● 銀行距離 &lt;400m</li> <li>● 飯店距離 &lt;600 m</li> </ul>
店面的所在地	<ul style="list-style-type: none"> <li>● 住宅區</li> </ul>

隨著科技進步，產出越來越多的空間資料趨勢下，如何利用空間資料挖掘幫助我們對大量的空間資料找尋出有用的知識，以及如何將空間資料挖掘技術與 GIS 的結合，已經成為空間資料挖掘領域以及 GIS 領域所重視的議題。本研究針對便利商店的戶口與工商普查資料以及其他相關的空間資料，進行空間資料挖掘研究，找出了許多隱藏在大量資料背後的便利商店區位規則，並且針對這些規則進行整理、概括化，得到量化過後的便利商店區位因子，如表 6 所示。並且整合 GIS，將規則轉換成空間面向，方便後續決策者加以評估，而這些因子也必須靠著後續決策者之評估，才能進一步轉化成便利商店的區位知識，利於新的便利商店區位分析中。

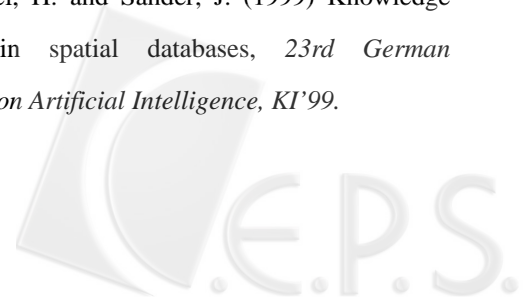
本研究雖然有效的應用資料挖掘技術（關聯規則）在空間普查資料中，但是還有許多地方是值得更進一步研究的，如下各點所列，提供給後進研究者未來研究之方向：

1. 納入更多的普查、空間資料，譬如：台北市每年的統計要覽資料，使得資料庫更加完整。
2. 再加強空間關係，探索彼此空間物件的 Topology 關係。本研究只針對空間物件中的距離概念，來進行研究。對於其他的空間關係，譬如：位向、方向等，並沒有做完整的考量，在未來的研究上，可以參考 Egenhofer 在 1991 年所做的研究，將空間物件的 Topology 關係以 Binary 的方式記錄，有助於往後的空間資料挖掘的空間關係描述上。
3. 數值資料轉成名目資料的方法，利用不同的方法，將會影響到所挖掘出來的規則型態，在本研究中利用等距分類法分成五類，效果對於本研究是最好的。但是對於不同的主題研究必須依照不同的資料特性來加以改變。
4. 最小支持度與最小信心值的訂定，可以依照決策者的領域知識（Domain Knowledge）來決定最小支持度與最小信心值，來獲取更多便利商店因子。

5. 加入 Data Mining 中的 Classification 方法，來建立決策樹 (Decision Tree)，針對空間普查資料庫中不同的欄位 (譬如：距離捷運站的遠近、人口數的多寡等) 來分類，找出在哪種分類下會出現便利商店，也可以形成另外一種的規則出來，輔助關聯規則分析方法所分析出來的規則。

## 引用文獻

- 方國定、許欽嘉 (1999) 便利商店利地選擇之評估—AHP 決策模式，科技學刊，8 (1): 31-43
- 白仁德、黃茹偵 (2004) 北二高沿線人口成長及空間分佈型態變遷之研究，台灣人口學會 2004 年年會暨「人口、家庭與國民健康政策回顧與展望」研討會。
- 朱子豪、孫志鴻、蔡博文 (1990) 內政部地理資訊應用之研究—戶口普查資料運用檢討與規劃，內政部資訊中心。
- 行政院主計處普查局 (2003)  
<http://www.dgbas.gov.tw/census~n/welcome.htm>
- 李萬凱、白仁德 (2003a) 地理資訊系統空間資料庫網格式轉換作業之研究，2003 四學會聯合論文研討會 (都市計畫、住宅、區域科學、地區發展)。
- 李萬凱、孫志鴻 (2003b) 國土規劃與土地利用模擬之研究—以台北都會區為例，地理資訊與國土規劃利用研討會。
- 林晶晶 (2003) 台灣知識密集服務業空間結構之研究，私立中國文化大學地學研究所碩士論文。
- 孫嘉玲 (2000) 台灣產業結構變遷的衡量與生產力之關係，國立台灣大學經濟學研究所碩士論文。
- 許麗惠 (2001) 台中都會區零售服務業空間結構變遷之研究，私立逢甲大學土地管理學系碩士論文。
- 經濟部商業司 (2000) 便利商店經營管理技術實務手冊，台北：經濟部。
- 蔡博文、吳淑瓊、李介中 (2004) 台灣 2000 年戶口住宅普查與門牌地址之整合應用：長期照護設施空間分析，人口學刊，(28)：135-152。
- 羅國華 (2000) 電腦方法在普查上之應用—資料處理與地理發展，公元兩千年人口普查相關問題—實務與研究學術研討會論文集，67-91。
- Agrawal, R. and Srikant, R. (1994) Fast algorithm for mining association rules, *Proceedings of the Twentieth VLDB Conference*, 487-499.
- Appice, A., Ceci, M., Lanza, A., Lisa, F. and Malerba, D. (2003) Discovery of spatial association rules in geo-referenced census data : a relational mining approach , *Intelligent Data Analysis* , 7: 541-566.
- Applebaum, W. (1968) *Guideline to Store Location Research-with Emphasis on Supermarket*, Massachusetts: Addison-Wesley.
- Bell, D. A., Anand, S. S. and Shapcott, C. M. (1994) Data mining in spatial databases, *Proc. International Workshop on Spatio-Temporal Databases*.
- Chawla, S., Shekhar, S., Wu, W. and Ozesmi, U. (2000) Extending data mining for spatial applications : a case study in predicting nest locations, *2000 ACM SIGMOD Workshop on Research Issue in Data Mining and Knowledge Discovery, DMKD 2000*.
- Clementini, E., Felice, D. P. and Koperski, K. (2000) Mining multiple-level spatial association rules for objects with a broad boundary, *Data & Knowledge Engineering* , 34: 251-270.
- Egenhofer, M. (1991) Reasoning about binary topological relation , *Proceedings of the Second Symposium on Large Spatial Databases*, 143-160.
- Ester, M., Kriegel, H. and Sander, J. (1997) Spatial data mining : a database approach, *Proc. Fifth Symposium on Large Spatial Databases (SSD'97)*, 48-66.
- Ester, M., Kriegel, H. and Sander, J. (1999) Knowledge discovery in spatial databases, *23rd German Conference on Artificial Intelligence, KI'99*.



- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (Eds.) (1996) *Advances in Knowledge Discovery and Data Mining*, Menlo Park, CA: AAAI/MIT Press.
- Gruen, V. and Smith, L. (1960) *Shopping Town U.S.A.: the Planning of Shopping Centers*, New York: Van Nostrand Reinhold
- Han, J. and Fu, Y. (1995) Discovery of multiple-level association rules from large databases , *Processings of the 21<sup>st</sup> International Conference on Very Large Databases*, 420-431.
- Han, J., Koperski, K. and Stefanovic, N. (1997) GeoMiner : a system prototype for spatial data mining, *Proc. ACM SIGMOD Conference on the Management of Data*.
- Han, J. and Kamber, M. (2000) *Data Mining : Concept and Techniques*, San Francisco: Morgan Kaufmann Publishers.
- Jenks, G. F., and Coulson, M. R. (1963) Class intervals for statistical maps, *International Yearbook of Cartography*, 3: 119-134.
- Klößgen, W. and May, M. (2002) Census data mining — An Application, *ECML / PKDD'02 Workshop on Mining Official Data*.
- Koperski, K. and Han, J. (1995) Discovery of spatial association rules in geographic information databases , *Proceedings of 4<sup>th</sup> Symposium on Advances in Spatial Databases, SSD'95*, 47-66.
- Kuo, R. J., Chi, S. C. and Kao, S. S. (2002) A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network, *Computers in Industry*, 47: 199-214.
- Malerba, D., Lisi, A. F., Appice, A. and Sblendorio, F. (2002) Mining Spatial Association Rules in Census Data: A Relational Approach, *Proceedings of the ECML/PCDD'02 workshop on mining official data*, 80-93.
- Mennis, J. and Liu, J. W. (2003) Mining association rules in spatial-temporal data, *7<sup>th</sup> International Conference on Geocomputation*.
- Miller, J. H. and Han, J. (2001) *Geographic Data Mining and Knowledge Discovery* , London: Taylor & Francis.
- Nelson, R. L. (1968) *The Selection of Retail Location*, New York: McGraw-Hill.
- Ng, R. T. and Han, J. (1994) Efficient and effective clustering method for spatial data mining , *Proceedings of 1994 International Conference Very Large Data Bases*, 144-155.
- Ng, R. T. (1996) Spatial data mining : discovering knowledge of cluster from maps, *Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*.
- Shekhar, S. and Chawla, S. (2003) *Spatial Databases : A Tour* , Prentice Hall.
- Zhou, X., Truffet, D. and Han, J. (1999) Efficient polygon amalgamation methods for spatial OLAP and spatial data mining, *Proc. 6th International Symposium on Spatial Databases (SSD'99)*.

93 年 12 月 20 收稿

94 年 01 月 18 修正

94 年 03 月 16 接受

