

以決策樹分析台北市便利商店區位之適宜性研究

Model Suitable Locations for Convenience Stores in Taipei City with Decision Tree Analysis

Chin-Te Jung^{*a}, Chih-Hong Sun^{**b}

^aPh.D. student, Dept. of Geography, National Taiwan University, Taipei, Taiwan

^bProfessor, Dept. of Geography, National Taiwan University, Taipei, Taiwan

摘要

普查資料可以充分的反應某地區的社會經濟的指標與特色。隨著科技的快速發展，有越來越多的(空間)資料可以被數位化，並記錄在資料庫中。而這些(空間)資料必須加以分析，粹取出更深層的資訊或知識，並且利用較佳的展示方式加以展示，以利於進一步的應用與分析。在本研究中，將針對(1)台北市的便利商店區位分佈議題，整合相關空間以及非空間的資料至空間資料庫中；(2)利用資料挖掘中的決策樹 C5.0 演算法，有效率的分析這些大量的資料，找出在哪些條件之下，利於便利商店之分佈，並演化出便利商店區位分析模式，預測分析其他潛在的便利商店區位是否適合，最後(3)進一步整理出便利商店區位決策指標，以利於決策者進行決策。本研究結果顯示，此便利商店區位分析模式之準確度可以達到 80%，說明本研究所提出的方法與架構，可以有效的支持與分析台北市便利商店之區位。

Abstract

Census data often accurately reflect the socio-economic trends and characteristics of the region under survey. As booming technologies, it produces a lot of digital (spatial) datasets or databases in our surroundings. It is essential to be able to analyze and utilize these large datasets, convert them into useful knowledge and finally use a friendly visualization to present the useful knowledge. In this research, we would achieve the following goals: (1) collect the Taipei census data in 2000, the industrial census data in 2001, and spatial data in Taipei City into a spatial census database which is focused on location suitability of convenience stores (CVS) in Taipei City; (2) such database is analyzed by decision tree used C5.0 algorithm of conventional data mining method to discover spatial or non-spatial patterns of CVS and (3) build a

* d94228001@ntu.edu.tw

** chsun@ntu.edu.tw

location suitability model for CVS in Taipei City. The model after validation achieves the correct percentage of 80 to predict the location whether is suitable or not for CVS in Taipei City. Finally, this research incorporates geographic information system (GIS) to be a visualization tool for decision-makers to evaluate the result of the analysis.

Keyword: Spatial Data Mining, Decision Tree, Convenience Stores, Location

1. 緒論

隨著空間科技的進步，越來越多的空間資料可以被自動化的蒐集，儲存在空間資料庫中，而這些大量的資料，已經遠遠超過我們的能力加以有效的分析與利用(Chelghoum, 2002)。而空間資料挖掘 (Spatial Data Mining, SDM) 或空間知識發現 (Geography Knowledge Discovery, GKD) 卻可以補強此一缺點，有效且自動化的將大量的空間資料加以分析，粹取出有用的空間知識，提供我們作進一步參考。而空間資料挖掘有許多不同的方法，譬如：關連規則 (Association Rules)、決策樹方法 (Decision Tree)、偏差分析 (Outlier analysis)、叢聚分析 (Cluster analysis) 以及特徵分析 (Characteristic analysis) 都可以應用在不同的空間相關課題中 (Ester *et al.*, 1999)。

而在台北市，超過 1000 家的便利商店 (譬如：7-11、全家、OK 等) 彼此競爭的非常激烈，每一家便利商店都想先找到適合便利商店之區位，搶奪最好的商機，因為影響便利商店成功的關鍵因素即是「區位」(Nelson, 1968; Applebaum, 1968; Grune, 1987)。但是，過去所採用的研究方法，並不能針對目前大量資料庫作有效的分析。並且在我們過去的研究中(Jung, 2005)曾經嘗試用關連規則分析此台北市的便利商店資料庫，找出相當多有用的規則以及因子，並且進一步的整理出影響便利商店重要的指標，以利於決策者做決策時之參考，證明以資料挖掘之方法，可以有效的分析出相當好之結果。但是關連規則方法，並不能演化模擬出模式，加以預測且分析。故本研究，將採用決策樹的方法，另外一種資料挖掘之方法，找出哪些影響便利商店的因子是最具有決策力，以及因子間相關階層組成是如何...等的重要問題，進而模擬出便利商店區位之模式，加以預測且分析未來潛在良好的便利商店區位。

本篇文章結構將在第二章的文獻回顧中，簡單的敘述決策樹的方法以及前人所回顧出影響便利商店的重要因子有哪些，作為收集本研究資料的依據；並在第三章中，提出本研究架構，如何整合空間以及非空間性的資料，以及利用決策樹

的方法加以分析，演化出一個準確預測之模式；在第四章中，我們將討論決策樹之結果、有哪些因子影響便利商店的區位以及比較我們之前利用關連規則的結果，互相比較以找出共通性；最後，提出結論以及未來可以發展之方向，提供後人研究之參考。

2. 文獻回顧

2.1 決策樹

決策樹是一個類似樹狀結構的流程圖，層級性的方式，將知識做有結構化的表現，並且依據不同的樹狀流程整理出不同的決策規則，如圖 1 所示 (Han and Kamber, 2001; Chelghoum, 2002; Moran and Bui, 2002; Pal and Mather, 2003)。而決策樹是屬於監督性的分類演算法，由使用者決定要分類哪一個欄位（稱為 classes，分類欄位），以及利用其他欄位（稱為 explanatory，解釋欄位）針對此一分類欄位作分類的解釋與描述，將混雜的資料做有效的分類，整理架構出結構化的知識 (Moran and Bui, 2002)。決策樹分類的好壞將會隨著不同解釋欄位的分類效果好壞而有所影響，亦即我們都不希望決策樹的決策規則太過於繁瑣，希望有簡單的決策規則，就可以達到良好的分類效果，如圖 1 的 A 與 B。故如何評量哪些解釋欄位是好或壞，是決策樹很重要的議題之一。

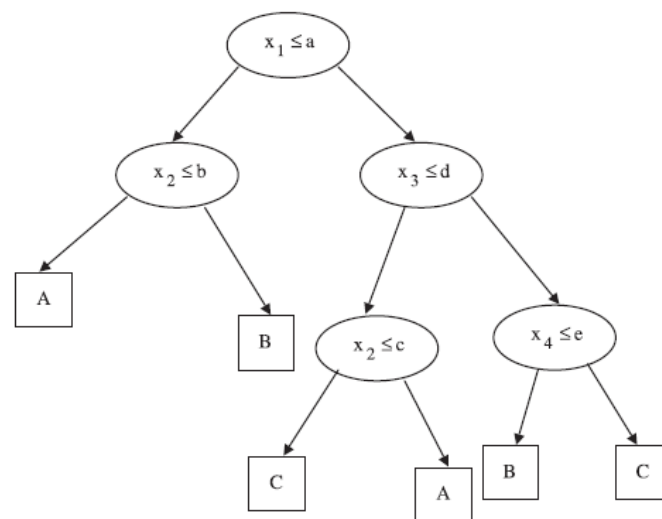


圖 1: 決策樹結構圖 (Pal and Mather, 2003)

在過去，有許多決策樹演算法的提出，如 ID3、C4.0、C5.0、CART...等，藉由 Entropy (熵度)以及 Information Gain (訊息獲利) 利用不斷的檢驗以及測試的統計方式，來評量哪些解釋欄位具有最好的分類效果(Quinlan, 1993; Han and

Kamber, 2001; Moran and Bui, 2002)。在本研究中，我們將採用 C5.0 的決策樹演算法進行分類，由於 C5.0 是經由 ID3 以及 C4.0 不斷的改良測試之後，最具有效率以及準確性的演算法 (Quinlan, 2003)，針對台北市便利商店的空間資料庫，進行分類分析，找出關鍵影響便利商店的決策規則與因子。

2.2 台北市便利商店相關因子回顧

便利商店成功最關鍵的成功因素即在於「區位」(Nelson, 1968; Applebaum, 1968; Grune, 1987)。Fang (1999) 以 Kuo *et al.*(2002) 都曾經利用 AHP (Analytic Hierarchy Process)方法，針對台灣的便利商店做過研究，也蒐集了需多影響便利商店的重要因子，譬如競爭店的數量、是否靠近一些具有良好吸引人潮的公共設施以及附近人口的社會經濟背景如何等相關因子，並將因子整理分成兩部分：(1) 屬性因子(Attribute factors)：便利商店消費者數量會跟某些因子的距離遠近有關，譬如：學校、圖書館等；(2)變異因子(Variable factors)：這些變異因子包括道路寬度、經濟收入、消費能力、人口數量等，不屬於屬性因子的即為變異因子，如圖 2 所示。但是在這些過去研究中，都只描述出有哪些因子會影響便利商店的區位分佈，並沒有量化數值的描述，也就是距離主要道路多少距離以內，是較適合便利商店的設置，而本研究將針對這點，以決策樹的方式加以歸納出，以完整便利商店區位因子的質性與量性的描述。

	Attribute factors		Variable factors	
Government organization	Market	Supermarket	The number of visiting customers	Road width
Office building	Park	Hypermarket	Store front width	Sidewalk width
Post office & economic organization	Theater	Station	Store front area	Parking convenience
School	Department store	Bus station	Store front door width	Population
Studying center	KTV	Pedestrian crossing	Store front visibility	Population density
Library	Recreation center		Store located at road intersection	Population growth rate
Hospital	Garage		Competing store area	Income level
Restaurant	Parking lot		Competition	Consumption level
Hotels	Gas station		People flow	The number of families
Temples	CVS		Car flow	

Figure 2: 便利商店因子 (Kuo *et al.*, 2002)

3. 建立便利商店區位資料庫與決策樹分析

在這章中，我們將敘述如何整合便利商店區位相關的空間和非空間性資料到空間資料庫中，並且先找出空間關係，之後再交由決策樹方法，針對此便利商店空間資料庫進行挖掘，找出潛在、有用的知識，以利於決策者分析。圖 3 為本研究的架構圖，分成三大部分：(1) 資料(data)；(2) 決策樹分析(decision tree)，訓

練模式 (3) 驗證模式並討論結果(validate and result)。

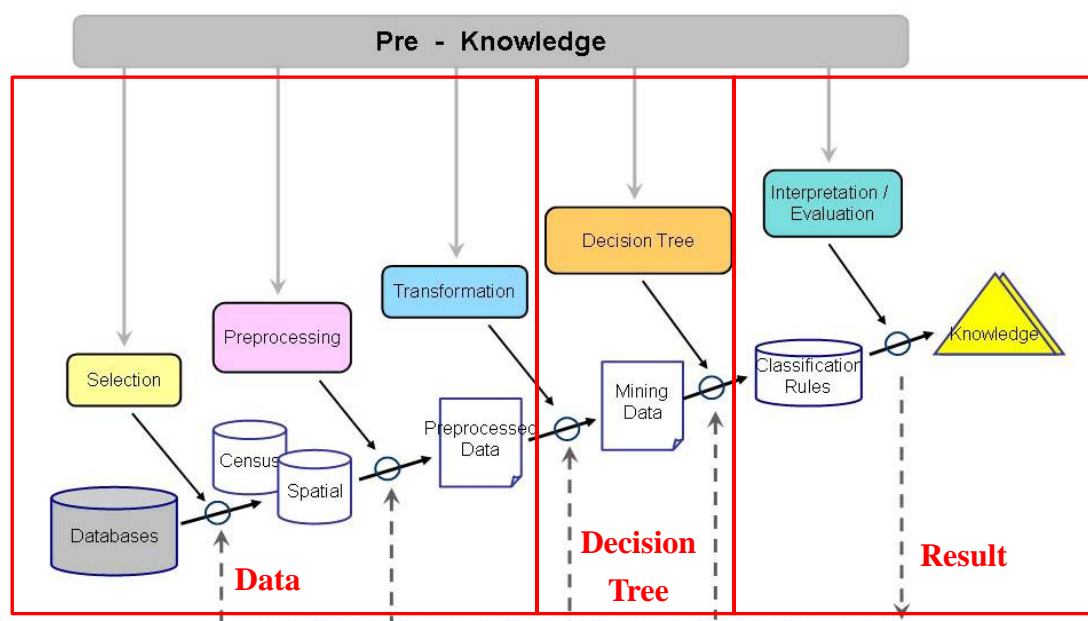


圖 3：研究架構圖

3.1 資料

3.1.1 空間資料

根據前人對於台灣便利商店的研究，找出影響便利商店區位相關的因子。在空間資料方面，我們蒐集了便利商店的點位資料、捷運站、加油站、土地利用資料...等，如表 1 所示。

Table 1: spatial data and census data

	資料類型	資料
空間資料	點	便利商店、捷運站、加油站、警察局、郵局、餐廳、飯店、醫院、銀行、小吃、美術館、圖書館、教堂、寺廟、公園、學校。
	線	主要道路
	面	台北市普查區、土地利用資料(包括住宅區、商業區、工業區..等)。
普查資料 (非空間資料)	屬性	男性人數、女性人數、總人數、年齡 0~6 歲人數、年齡 6~25 歲人數、年齡 25~65 歲人數、年齡>65 歲人數、工業工作人數、商業工作人數、服務業工作人數、單身戶人數...等。

3.1.2 普查資料(非空間性資料)

由於普查資料可以充分反應某地區的人口社會經濟結構，所以本研究將採用以普查區為範圍的台北市 2000 年戶口普查、2001 工商普查資料，作為便利商店商圈人口的社會經濟資料基礎。並且根據前人的研究，找出影響便利商店區位重要的因子，譬如：男性人數、女性人數、總人數、年齡在 0~6 歲的人數（兒童人數）、年齡在 6~25 歲人數（學生人數）、年齡在 25~65 歲人數（青壯年人數）、年齡超過 65 歲人數（老年人數）、單身戶人數.....等相關的普查資料，如表 1、表 2 所示。但是，由於這些非空間性的資料爲了要跟空間資料相互整合，必須將非空間資料的資料轉換成空間性的資料。故本研究透過與空間資料中的普查區範圍圖（如圖 3 所示）做結合（Join），先將非空間性的資料轉換成空間性的資料，以利於後續的分析使用。

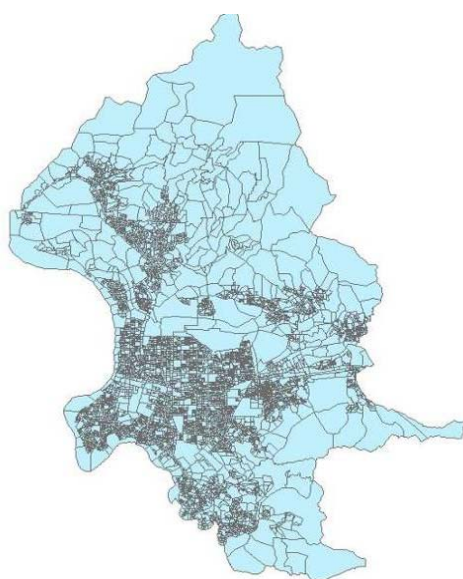


圖 3：台北市普查區範圍圖

表 2：普查資料表

ID		戶口普查資料							工商普查資料		
里 ID	普查區 ID	男性 人數	女性 人數	總人數	學生人數	碩士人數	博士人數	...	工業 工作人數	商業 工作人數	服務業 工作人數
6312008	6312008023	282	264	546	106	107	81	...	47	102	28
6312009	6312009023	154	155	309	65	45	36	...	38	68	17
6312017	6312017023	204	204	408	114	81	55	...	51	80	22
6312020	6312020023	125	139	264	69	46	22	...	35	74	20
6312021	6312021023	179	177	356	80	57	48	...	28	99	249
6312038	6312038023	192	199	391	105	81	52	...	47	74	20
6312006	6312006024	146	169	315	66	43	30	...	35	55	15

6312007	6312007024	162	188	350	88	65	52	...	38	59	16
...

3.2 前處理

由於空間資料具有空間關係（譬如：距離關係、位相關係以及方向關係）的特殊性，再加上便利商店具有商圈範圍，在商圈範圍中的人口社會經濟是我們必須獲取到的，所以將針對本研究資料做以下幾點前處理之步驟：

3.2.1 粹取空間關係

空間資料之所以不等同於非空間性資料，在於空間資料具由空間關係的特性，譬如距離關係 (Distance)、位相關係 (Topology)、方位關係 (Direction) (Shekhar and Chawla, 2003)。故再交由決策樹分析之前，必須先將空間關係加以記錄，才能完整的分析出便利商店空間區位的特性。但是由於本研究時間之限制，不能全盤的針對各項空間關係作粹取記錄，所以本研究只針對距離關係，將便利商店與其他空間資料彼此之間的距離加以記錄，譬如：便利商店到學校的距離、便利商店到捷運站的距離等，如表 3 所示。

3.2.2 便利商店商圈

由於便利商店具有商圈特性，在此商圈範圍中的人口、社會以及經濟特性，將會影響便利商店是否能夠成功的經營。然而，不同的商業活動以及不同的地區，商圈的範圍也會有所不同。本研究根據台灣經濟部商業司在 2003 年定義了大約的台灣便利商店商圈範圍：以 300 公尺徒步距離為界，商圈的規模以 3000 人為界，為便利商店之商圈範圍。但本研究以台北市為研究範圍，便利商店競爭較於激烈，故本研究採用 250 公尺為便利商店的商圈範圍，找出在此便利商店商圈範圍的人口、社會以及經濟特性。

3.2.3 資料推估 (Area interpolation)

然而便利商店商圈範圍與之前整合在台北市普查區中的戶口、工商普查資料做疊合(Intersect)，來整合出此商圈範圍中的人口、社會經濟特性的同時，由於一家便利商店 250 公尺的服務範圍可能會跨到很多普查區中，如圖 4 所示，圖中 1、2、3、4、5 都是屬於同一個便利商店的服務範圍所跨到不同的普查區並且普查區中的人數並不是平均的分佈在此普查區的範圍中，因此本研究假設「普查區內人口數都平均分佈此區域中的建地範圍內（包括住宅區、商業區、工業區）；如果商圈中沒有建地面積，則不分配普查人數」，並利用建地面積與普查區的面積的比例關係（如公式 1 所示），將普查區的資料分配至落在商圈範圍的建地中，最後統整起來，以做為此商圈的人口、社會經濟特性的基礎資料，如表 3 所示。

在表 3 中，紀錄了便利商店與其他空間資料的距離數值以及商圈範圍內的普查人數，以做為決策樹分析的基礎資料。

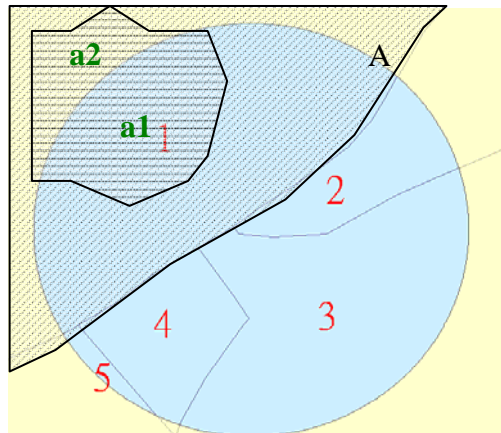


圖 4：便利商店跨普查區示意圖

區域 1,2,3,4,5：為便利商店商圈所跨的普查區範圍。

區域 a1：落在區域 1 的建地範圍。

區域 a2：沒有落在區域 1 的建地範圍。

區域 A：普查區

$$P_I = (BA_I / BA_A) * P_A \quad (\text{公式 1})$$

P_I ：區域 1 的普查人數

BA_I ：落在區域 1 的建地面積 (a1)

BA_A ：落在普查區 A 的總建地面積 (a1+a2)

P_A ：普查區 A 的普查人數

Table 3: Summarized census and spatial table

Buffer ID	CVS	Distance to Police Office	...	Male	Female	Sum People	...
0	YES	573.3341	...	1288	1364	2652	...
1	YES	386.1735	...	2895	3268	6163	...
2	YES	529.649	...	4223	4538	8761	...
3	NO	152.3975	...	3492	4041	7533	...
4	YES	82.22532	...	137	63	200	...
5	NO	160.7047	...	2062	2271	4334	...
6	NO	1086.15	...	3437	3920	7358	...
7	YES	1424.382	...	1029	972	2001	...
8	YES	466.1352	...	1067	1022	2089	...
...

3.3 模式訓練與驗證

本研究將研究資料分成兩大部分做模式的訓練與驗證：(1) 70%(707 筆)的資料做模式的訓練，交由決策樹 C5.0 的演算法，以「是否為便利商店」為分類欄

位(classes field)，以其他資料做為解釋欄位 (explanatory field)，來找出便利商店的區位特性；(2) 30%(290 筆)的資料做模式的驗證，來檢驗 70%所訓練出來的模式是否準確。驗證的結果達到 81.4%的準確性，說明本研究藉由決策樹方法所建立的模式具有相當良好的預測能力。

4. 結果討論

本研究利用決策樹 C5.0 的演算法，針對台北市便利商店區位分析出不同的決策規則，準確率達到 81.4%，以”If...then...”的形式表示。由於決策樹屬於分類的方法，將混雜的資料分類到相近類似的組別中，最好的分類情況即一分為二，也就是信心值 (Confidence)達到 100% (在 If 出現的條件之下，會有 100%的機率出現便利商店)，譬如規則 1：假如在高度住宅區以及離博物館的距離小於 846.8 公尺的區域內，會有 100%的機率出現便利商店，而有 9 條規則符合此一決策規則，也就是代表在此條件的區位下，是非常適合經營便利商店的。但是，如果每條規則都要求到 100%的機率，將會使得決策規則非常繁雜。所以本研究採用信心值大於 80%的規則篩選出來，做為影響便利商店區位較為強烈的決策規則，如表 5 所示。

表 5：決策規則表 (I：規則符合數量, C：信心值)

ID	決策規則	I.	C.(%)
1	If 土地利用屬於高層住宅區 and 距離博物館<846.8m then CVS	9	100
2	If 土地利用屬於中層住宅區 and 工業工作人數< 1148.7 人 and 年齡大於 65 歲<1717 人 and 距離教堂< 710.3m and 距離小吃< 302.9m and 距離學校< 158.5m then CVS	2	100
3	If 土地利用屬於中層住宅區 and 工業工作人數< 1148.7 人 and 年齡大於 65 歲<1717 人 and 距離教堂<710.3m and 距離小吃> 302.9m then CVS	19	100
4	If 土地利用屬於連鎖商業區 and 距離郵局< 713.4m and 距離教堂< 710.3m and 工業工作人數< 1150 人 then CVS	31	100
5	If 土地利用屬於連鎖商業區 and 距離郵局< 713.4m and 距離教堂< 710.3m and 工業工作人數> 1150 人 and 距離公園< 267.9m then CVS	8	100
6	If 土地利用屬於中層住宅區 and 工業工作人數> 1148.7 人 then CVS	53	98.7
7	If 土地利用屬於低層住宅區 and 距離寺廟> 161.3m and 附近競爭的便利商店數量<4 家 and 距離郵局< 1066.9m then CVS	134	97.8
8	If 土地利用屬於低層住宅區 and 距離寺廟> 161.3m and 附近競爭的便利商店數量> 4 家 and 女性人數< 8370 人 then CVS	15	93.3
9	If 土地利用屬於一層住宅區 or 學校 or 停車場 or 公園 or 服務業區 or 醫院等地方 then CVS	349	92.8
10	If 土地利用屬於低層住宅區 and 距離寺廟< 161.3m and 距離飯店< 284.9m and 距離	13	92.3

	捷運站> 354.8m then CVS		
11	If 土地利用屬於低層住宅區 and 距離寺廟> 161.3m and 附近競爭的便利商店數量< 4 家 and 距離郵局> 1066.9m and 距離主要道路> 7m then CVS	8	87.5
12	If 土地利用屬於未使用區 and 距離加油站< 540.2m then CVS	6	83.3
13	If 土地利用屬於建地區 and 附近競爭的便利商店數量< 3 家 then CVS	6	83.3
14	If 土地利用屬於連鎖商業區 and 距離郵局< 713.4m and 距離教堂< 710.3m and 距離小吃> 2150.3m then CVS	6	83.3

由於決策樹是屬於階層性的知識結構，利用解釋欄位(explanatory fields)的資訊增值(information gain)高低比較，來決定分類的先後順序。由本研究的決策規則結果中，可以發現土地利用型態是為資訊增值最高的解釋欄位，順序放在第一個進行便利商店區位好壞的分類。之後，在計算各個不同決策樹分支中的資訊增值，再進行分類，所以解釋欄位會隨著不同的分支而有所不同。但是，可以進一步的從決策規則中歸納出影響台北市便利商店區位較為重要的因子，並且將每個因子之後的數值做簡化，做為決策的指標，提供給相關便利商店決策者做為決策之參考，如表 6 所示。

表 6：影響台北市便利商店區位指標

空間因子指標	空間因子指標
距離博物館 < 800m 距離餐廳 < 280m 距離教堂 < 700m 距離小吃 < 300m 距離學校 < 150m 距離郵局 < 700m 距離警察局 < 700m 距離捷運站 < 350m 距離主要道路 < 7m 距離加油站 < 500m 附近其他便利商店競爭數量< 4 家	土地利用型態屬於「低層住宅區」、「中層住宅區」、「高層住宅區」、「連鎖商業區」、「學校」、「公園」、「停車場」、「服務業區」、「醫院」等。
	普查因子指標
	工業工作人數 < 1150 人 老年人數 < 1700 人 女性人數 < 8400 人

5. 結論與建議

本研究針對台北市便利商店區位問題，蒐集相關的空間以及非空間性的因子，並劃設便利商店 250 公尺的商圈，整合進空間資料庫中，提供便利商店商圈

人口、社會經濟以及與其他空間因子的完善分析背景，並且採用資料挖掘中的決策樹分析方法，找出台北市便利商店區位的結構性決策規則，並進一步的整理成便利商店區位決策指標，以提供決策者之參考。由決策樹的方法，不但可以找到哪些因子，影響便利商店區位最為強烈，並且還可以找出因子的屬性與數值，提供更進一步的資訊，是過去研究中所不具備的地方。因此，藉由本研究可以互補過去研究的結果，將因子的數值與屬性納入，以完善便利商店區位的決策指標。並且可藉由決策樹模擬出的模式，來預測分析在未來潛在的地點，是否適合設置便利商店。

在未來，本研究將會繼續(1)粹取更多的空間關係，譬如：拓樸關係、方位關係，以強化空間關係；(2)將便利商店區位指標訪談專業人士，評量這些指標有助於決策之參考；(3)繼續利用其他資料挖掘之方法來輔助、強化便利商店區位研究之領域。

引用文獻

- Applebaum, W.** (1968) *Guideline to Store location research-with emphasis on supermarket*, Massachusetts: Addison-Wesley.
- Chelghoum, N., Zeitouni, K. and Boulmakoul, A.** (2002) A decision tree for multi-layered spatial data, *Symposium on Geospatial Theory, Processing and Applications*, Ottawa.
- Ester, M., Kriegel, H., Sander, J.** (1999) Knowledge Discovery in Spatial Databases, *Lectures Notes in Computer Science* : 64-71.
- Fang, K. and Hsu, C.** (1999) The site selection for convenient store- AHP decision model, *Journal of Science and Technology*, 8(1): 31-43.
- Gruen, V. and Smith, L.** (1996) *Shopping Town U.S.A.: the Planning of Shopping Centers*, New York: Van Nostrand Reinhold.
- Han, J. W. and Kamber, M.** (2001) *Data Mining: Concepts and Techniques*, San Francisco: Morgan Kaufmann.
- Jung, C.** (2005) *Spatial data mining on census data — a case study for location analysis of convenience stores in Taipei city*, M.A. dissertation, National Taiwan University.

- 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.
- Li, D., Di, K. and Li, D.** (2000) Land use classification of remote sensing image with gis data based on spatial data mining techniques, *International Archives of Photogrammetry and Remote Sensing*, XXXIII (part B3): 238-245.
- Moran, C. J. and Bui, E. N.** (2002) Spatial data mining for enhanced soil map modeling, *International Journal of Geographical Information Science*, 16(6): 533-549.
- Nelson, R. L.** (1968) *The selection of retail location*, New York: McGraw-Hill.
- Pal, M. and Mather, P. M.** (2003) An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment*, 86: 554-565.
- Shekhar, S. and Chawla, S.** (2003) *Spatial Databases : A Tour*, Prentice Hall
- Quinlan J.R** (1993) C4.5: programs for machine learning. Morgan Kaufmann, San Mateo, CA.
- Quinlan J.R** (2003) C5.0 online tutorial, <http://www.rulequest.com> (access 2005)