

資料庫中空間性週期關聯規則之發掘— 以便利商店交易資料為例

Data Mining of Spatial Cyclic Association Rules in Databases — A Convenience Store Transaction Data Example

楊亨利

Heng-Li Yang

國立政治大學教授

郭家佑

Jia-You Guo

國立政治大學

email : yanh@mis.nccu.edu.tw

摘要

資料發掘目前在傳統關聯式資料庫相關議題上已有不少研究，但如果能再整合空間和時間要素進來，將可從資料中發掘出更多有趣的知識。以往常使用統計分析方法來分析空間資料，不幸的是，統計分析方法仍有許多問題亟待解決。而 Han 等人利用概念樹發掘「多層次關聯規則」的技術已相當成熟，值得學習。在時間方面，另外有學者提出「週期關聯規則」的觀念。本研究試著將「空間特性」和「週期關聯規則」結合，提出「空間性週期關聯規則」的想法。首先從相關文獻中分別瞭解目前空間、時間資料發掘領域的研究現況，進而提出研究架構，試圖結合地理之空間分析與週期關聯規則之資料發掘，同時強調時、空、商品種類三維度之多層次發掘。再以全球

資訊網動態網頁技術配合假想的台北市便利商店交易資料庫，發展出一套雛型系統，結合地理資訊系統 ArcView 與傳統的 Access 資料庫（目前只能作單一項目之間的關聯）。最後提出進一步的研究建議，以供後續研究參考。

關鍵字：週期空間關聯規則、資料發掘、交易資料庫、時空維度

Abstract

There have been a number of research of data mining in relational database. Considering spatial and temporal dimensions, we might obtain more useful knowledge in transaction databases. The statistical spatial analysis has been one common technique for analyzing spatial data. However, many problems still remain. Han et al. have used concept hierarchies to mine multiple-level association rules. On the other hand, some researchers presented the idea of cyclic association rules. In this research, we try to integrate the research of spatial associations with cyclic association and suggest the idea of spatial cyclic association rules. First, we conduct a literature survey in the fields of spatial and temporal data mining. A framework is then proposed to integrate geographical spatial analysis with the data mining of cyclic association rules, and emphasize the multiple levels of time, space, and commodity dimensions. A prototype system is implemented to combine the traditional relational database (Access) with a geographical information system (ArcView) in a WWW environment (though only the association between one commodity itemset can be found). Finally, we suggest some possible directions for further research.

Keywords: Cyclic Spatial Association Rules, Data Mining, Transaction Database, Spatio-Temporal Dimensions

壹、導論

資料發掘為依據使用者需求自資料庫中選擇合適資料，並加以處理、轉換、發掘至評估的一連串過程，期能找出真實世界運行時隱含於其內的運作現象，以輔助解決問題之用（林幸怡,民 86）。資料發掘目前在傳統關連式資料庫相關議題上已有不少研究（Agrawal

and Srikant, 1994；Fayyad et al., 1996；Han et al., 1993；Piatetsky-Shapiro and Frawley, 1991），但其他如空間資料庫、時間資料庫、物件導向資料庫以及多媒體資料庫中的資料發掘仍是個亟待開發的領域。

從交易資料中發掘隱含的知識一直是資料發掘的主要研究項目之一，但傳統的關聯規則中，並沒有加入時間和空間的要素，這樣並不能代表真實世界的狀況。所以如何改進傳統的關聯規則發掘方法，並加入時間、空間等相關條件，以發掘出更有價值的「新關聯規則」，就成為本研究的動機。

在現實生活中，時間與空間本來就是並存的，如果以時間和空間的變動情形互相排列組合，來劃分這整個世界的話，可概略分為四個部分：靜態（Static）—不考慮時間，只考慮空間相對關係；運動態（Kinematic）—考慮物體的某一個量沿著時間軸變動的軌跡，但不考慮其變動的速度或變化量；動態（Dynamic）—考慮物體隨時間變化的情形，且考慮到變動的速度或變化量及變動的原因，通常也探討 2 個以上不同變數之間的變化；混沌（Chaos）—非線性變化的時空運動。而本研究的範圍只能屬於運動態這一類，尚無法涵蓋動態。

本研究在對於先前相關的時空資料發掘研究深入瞭解後，提出一個概念性系統架構，期望能有效的從時間、空間資料庫中發掘出有用的週期空間關聯知識。另外，本文將報導一個運用所提出的架構在便利商店交易資料庫上進行實作之雛型。

貳、文獻探討

由於篇幅限制，在此無法對過去之研究逐一介紹，而僅整理如表 1。其中與本研究觀念相關的為 Han 等人之屬性導向歸納（Attribute-oriented induction）演算法。這是以一般化為基礎的知識發掘，需要領域知識（Background Knowledge）的輔助，這些知識通

常是以概念階層（Concept Hierarchies）的形式儲存於知識庫。在空間資料庫中，地理資料之特性可能同時存在空間與非空間兩種概念階層。如蘋果和梨子都可以提升至水果的階層，水果又可以提升至經濟作物。同樣的，木柵區屬於台北市，台北市又屬於台灣省。

綜合表 1 而言，不論是空間資料發掘或時間資料發掘，都各自有許多方法，可發掘出各種不同特性的規則，其間的取捨，可能有些考慮。首先是資料來源或資料型態，某些特別的資料只能限定用某些特定的方法來發掘，例如由外太空傳回來的星雲照片，可能就必須使用影像資料發掘才行；其次是輔助資訊的有無，例如以一般化為基礎的資料發掘都必須使用到概念樹，如果在進行資料發掘時缺乏概念樹等的先前知識的輔助，可能就必須改用統計方法才行；最後則必須視所想發掘的知識規則來選擇發掘方法，如果想瞭解一公司歷年來的營運績效，趨勢分析是個不錯的選擇。

表 1 各種時間、空間資料發掘方法與相關的條件及規則範例整理表

作者、條件及規則範例		資料型態	輔助資訊	欲發掘規則之性質
資料發掘方法	相關文獻作者			
以統計為基礎的空間資料發掘	(Fotheringham and Rogerson, 1994) ; (Shaw and Wheeler, 1994)	空間資料	不需要	空間現象 統計樣式推演
以一般化為基	以空間資料	(Lu et al.,	空間概念樹	各區域的

礎的空間資料 發掘	爲主的一般化	1993)		一般概念樹	特性
	以非空間資料 爲主的一般化			某一特性所屬 的 空間範圍	
利用分群技巧 的空間資料發 掘	以空間資料 爲主的分群	(Ng and Han, 1994)		一般概念樹	空間物件群集 的特性
	以非空間資料 爲主的分群				某一特性之 空間群集情形
發掘空間關聯規則的方法		(Koperski and Han, 1995)		空間概念樹 一般概念樹 述語概念樹	與空間有關之 關聯規則
利用 CRH 演算法的 空間資料發掘		(Knorr and Ng, 1996)		不需要	找出空間群集 周圍物件的距 離及特性
影像資料發掘		(Fayyad et al., 1993);	影 像		影像資料的自 動辨識及判讀

		(Fayyad and Smyth, 1993); (Smyth et al., 1994)	資料		
	其他空間資料發掘方法	(Major and Mangano, 1993)	空間資料		其他相關議題，如熱帶風暴
以概念樹導向歸納分析技巧為基礎的時間資料發掘	發掘演進類的特性規則	(Han et al., 1995)	時間資料	一般概念樹	演進類的資料特性
	發掘演進類的區別規則				演進類的資料區別
	發掘資料演進趨勢			時間概念樹 一般概念樹	瞭解某一屬性之長期情形
	樣式(Pattern)資料發掘方法	(Dietterich and Michalski, 1985); (Agrawal, Faloutsos and Swami, 1993)		不需要	相似性分析

以統計為基礎的時間資料發掘	(Spiegel, 1996)		求取趨勢線的理想方程式
週期關聯規則的資料發掘方法	(Ozden et al., 1998); (Han et al., 1999)		關聯規則加入時間週期因素考量

參、概念性系統架構

空間與時間在資料發掘時似乎有共同存在的可能，這就形成時空（Spatio-Temporal）資料發掘。以週期關聯規則而言，若加入空間因素，例如：「過去一年以來，靠近政大的便利商店，每天中午 11 點到 1 點之間，有 70% 以上的交易中有飲料。」是一條空間性的週期關聯規則。當然也可能可考慮更複雜，如以政大為中心，每隔 10 公里之便利商店的商品關聯購買型態有何差異；或進一步考慮空間對空間的週期關聯或考慮非空間對空間之週期關聯規則，但可能形成時間太久遠、變數太多或很難找出其週期。在此，本研究只在時間結合的領域上，跨出一小步，考慮空間性的週期關聯規則，也就是先選擇空間資料，再考慮非空間資料之週期關聯。又可分為 A、B 兩類，A 類較簡單，為發掘多數項目，如：「過去一年以來，靠近政大的便利商店，每天中午 11 點到 1 點之間，有 70% 以上的交易中有飲料。」；B 類較複雜，如上述之同時，買報紙與飲料之間的關即為 B 類。此外，在有關週期的部分，只討論完整週期。如果類似 01001 這種複合週期，本研究視為是分開的 2 個週期，分別是 (5,1) 和 (5,4)。

一、概念性系統架構介紹

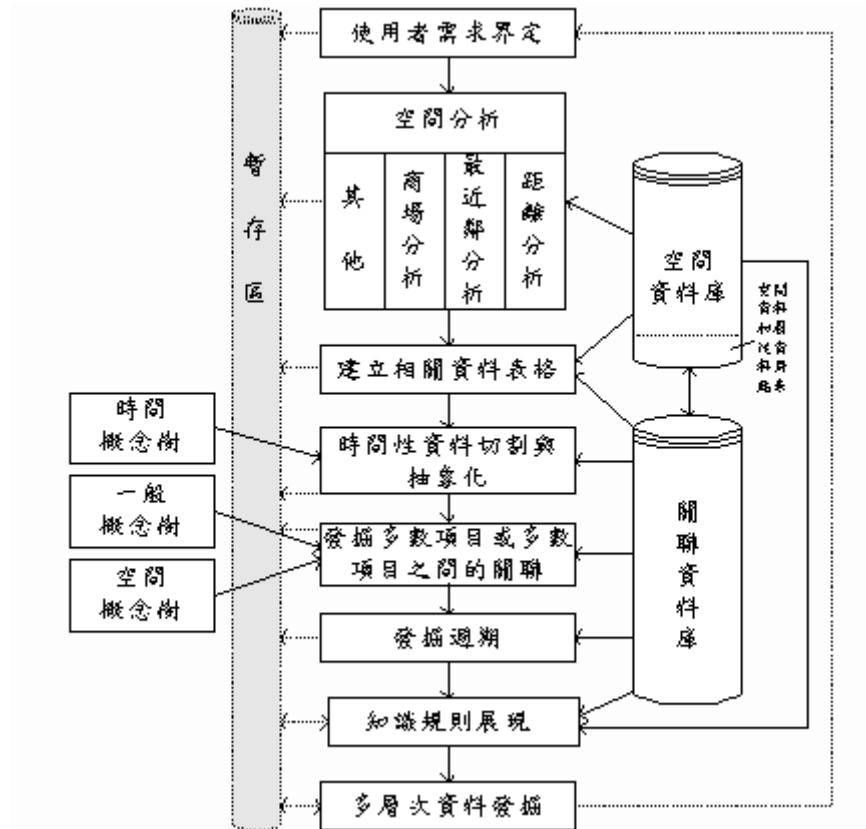


圖 1 本文概念性系統架構圖

本研究提出一個概念性系統架構（如圖 1）,以下將依其組成介紹此一架構。

（一）使用者需求界定

1.使用者界面

本研究的使用者界面希望能讓一般社會大眾也能隨時隨地、輕易的使用資料發掘技術，所以讓使用者以滑鼠、軌跡球等點選工具在螢幕上選擇有興趣的目標，而不是以鍵盤輸入艱深難懂的查詢語言，這點可利用網際網路和圖形化使用者界面來達成。

2.輸入參數

本研究共分為數個模組來進行，每個模組皆須使用者設定一些選項。這些參數包括（1）使用者必須根據問題種類選定適當的研究對象，例如便利商店、書店、百貨公司等。（2）使用者必須選定一段適當的研究期間，也就是一個時間起點和一個時間終點（這裡都是指實際時間而言）。（3）選定適當的抽樣間隔，以便進行週期的發掘。值得注意的是，由於太小樣本所找出的週期可能較不具代表性，例如選定研究期間為 2 日，抽樣間隔為 12 小時，想找出每半天有何週期，這樣找出的週期可能較不具說服力。所以實作系統可能會要求設定研究期間 ÷ 抽樣間隔必須 \geq 某個預設值，不過使用者還是可以自行視情況調整這些參數的。（4）本架構支持 A、B 兩種不同性質的關聯規則的發掘，使用者可以自由選擇發掘 A 類或 B 類的規則，當然也可同時選擇 A 和 B，表示兩類的規則都要發掘。（5）此外，由於本研究所採用的方法有利用到屬性導向歸納分析技巧，所以尚必須指定所選的抽樣間隔內所發生的事件的最小支持度門檻值和最低信賴度門檻值，以便找出有力的規則。（6）使用者還可更進一步選定理想的最大週期長度，以找出更符合其需求的週期。但因為週期數太少並無意義，所以實作系統可能也必須要求研究期間 ÷ 最大週期長度必須 \geq 某個預設值，當然使用者還是可以自行調整這些參數的。

最後，由於資料發掘為不斷的循環過程，本研究希望能藉由圖形化的界面及內建的判斷機制，經由使用者對於問題的反覆發掘，能漸漸使其問題明朗化，並可在反覆過程中，新增或刪除一些條件設定，以找出更多的隱含資訊。

(二) 資料庫

在圖 1 中右邊為資料庫，此處假設所處理的資料庫為完整資料庫，並不考慮其空白值 (Null) 或雜訊 (Noise) 的情況。而在資料庫種類上，包含空間資料庫及關連資料庫兩大部分。空間資料庫可儲存向量式 (如表 2) 或網格式地圖資料。不過，假設在研究期間空間資料大致維持不變，所以此部分資料並無時間特性。

表 2 向量式地圖資料範例

ID	座標
N00101	(3,5)
N00001	(7,6)
L00002	(-1,-1) (-3,-1)
P00001	(-1,1) (-1,5) (-5,5) (-5,1)

註：ID 為辨識子，由模組自動產生

關連資料庫又包含儲存地圖的屬性資料 (如表 3) 及一般研究對象之交易資料，後者存了大量歷史資料，所以，可能要有時間性的新欄位。表 4 是個由相關表格合併後的可能例子，在該表中之「日期」、「時間」均為交易實際發生時間。而空間資料庫和其屬性資料的連接在邏輯上是透過一個隱含在空間資料庫裡的「空間資料和屬性資料對應表」(如表 5) 來運作。另外，在圖 1 左邊之「暫存區」為儲存各模組中間步驟結果之用。

表 3 關連式資料庫資料範例—儲存地圖的屬性資料範例

Key	類 型	名 稱	主管單位
N01	點	消防栓	消防局
N02	點	巴士站	交通局
L01	線	中山路	養工處
P01	面	XX 便利商店	經濟部

表 4 關連式資料庫資料範例—交易資料範例

XX 便利商店銷售紀錄表

交易編號	產品	產品編號	單價	數量	單位	小計	日期	時間
001	牛奶	0001	15	1	盒	15	98/03/20	7 : 35 : 30
001	麵包	0501	20	1	個	20	98/03/20	7 : 35 : 33
002	牛奶	0001	15	1	盒	15	98/03/20	7 : 36 : 10
003	汽水	0003	20	3	罐	60	98/03/20	7 : 41 : 56

表 5 空間資料和屬性資料對應表

ID	Key
N00101	N01
N00001	N02

L00002	L01
P00001	P01

(三) 概念樹

表 6 概念樹的定義及儲存範例（以空間概念樹為例）

超型態 (Super Type)	次型態 (Sub Type)	層級 (Level)
台北市	木柵區	1
北部	台北市	2

爲了各種資料的抽象化和特殊化，在圖 1 中用了三種概念樹，分述如下：

1. 空間概念樹

定義了空間資料抽象化和特殊化所需使用的概念樹，以做爲空間資料抽象化之用，如表 6。例如，木柵區可抽象化爲台北市，再抽象化爲北部。

2. 時間概念樹

本研究所使用的時間資料最小單位爲秒，但在「時間性資料切割與抽象化」階段，將會應用到時間資料的抽象化，例如：秒 ⊂ 分 ⊂ 時 ⊂ 日 ⊂ 月 ⊂ 年 ⊂ 十年 ⊂ 世紀，所以需要建立相關的時間概念樹。

3.一般概念樹

在「發掘多數項目模組」需要利用一般概念樹來做資料的抽象化，例如：果汁、汽水可抽象化為冷飲，冷飲又可再抽象化成飲料。

(四) 資料發掘方法

圖 1 架構的主體為資料發掘方法，乃分兩階段，先作空間分析，再進行時間資料發掘，說明如下。

1.空間分析模組

以前文獻中的空間資料發掘，大多仍停留在與空間相關之屬性資料發掘上，很少直接對空間資料進行空間分析，頂多只是行政區的分割、合併而已。此處特別針對此一缺點加以改進，直接對空間資料進行空間分析動作。

首先使用者必須從空間資料庫中選擇有興趣的研究區域，也就是選定一張現有的圖檔。接著選擇所欲進行的空間分析種類(如下)，系統會自動根據分析所需，轉換地圖資料模式，其中除了最近鄰分析需要網格式資料模式地圖配合之外，其餘分析皆使用向量式資料模式地圖。

- 距離分析：選定地圖上某些重要設施、建築物，和指定一段距離 D ，就可分別找出距離這些設施距離 D 的範圍。（周學政、周天穎，民 86）。
- 最近鄰分析：選定地圖上某些重要設施，本模組會依各設施之間的距離劃定界限，將整個地圖分成數個區域，離區域內每一點最近的設施皆為其中心設施。（周學政、周天穎，民 86）。
- 商場分析：這裡所謂的“商場”，並不單指發生商業行為的空間而言，而是一種服務範圍的概念。例如由使用者在地圖上點選幾處相同性質的重要設施或場所，如公園、學校、醫院、運動場、消防局、百貨公司等，並假設地圖上每一個地點的需求，都

由其路程最近的設施來提供服務。因此，此分析會根據道路系統的通行難易情形及所需路程，來決定每一個重要設施其服務區域、學區或商品消費圈的大小（周學政、周天穎，民 86）。

- 其他空間分析：其他空間分析尚有：網路分析、服務區域分析、計算坡向、坡度、產生等高線或陰影圖等。

此（1）至（3）的三種空間分析，雖然乍看之下差不多，但其實還是有其不同的地方。距離分析所使用的距離是由使用者指定的，是一種主觀的距離、絕對的距離；相反的，最近鄰分析所使用的距離則是經由計算產生的，是一種各設施之間的相對距離。至於商場分析的最大優點就是計算道路的長度，而不是單純計算距離而已，隱含了將地理障礙（山川、河流）、公共道路分佈考慮進來的精神，不過同時也有運算較複雜的缺點。

2.建立相關資料表格

- 尋找區域內的研究對象：在空間分析模組，不論選擇哪一種分析，最後皆會產生一張新的圖層，圖的範圍與原地圖一樣大小，並以不同的顏色來表示各設施的影響區域。本步驟的目的，就是要分別從這些區域中，找出位於其範圍內的研究對象（此為一開始即由使用者選定）。方法是利用疊圖的方式，將上一階段的結果圖，套疊在研究對象分佈圖上，即可分別找出各區域範圍內的研究對象。舉例來說，如果上一步驟已經找出距離木柵國小 10 公里的區域，此階段即可更進一步找出此區域內的便利商店。利用此方法，我們就可以將台北市的便利商店分成數個群集，例如分別有靠近學校或商業區的。
- 找出相對應的屬性資料：接下來，我們要從關連資料庫中找出這些研究對象分別對應的屬性資料。利用前述的空間資料和屬性資料對應表，分別找出各群集內之研究對象在某連續一段時間所對應的屬性資料。至於研究期間的起點和終點，則在一開始時就由使用者選定。
- 合併資料表格：由於每一個研究對象群集裡，可能有 1 個以上的研究對象，且可能來自不同的組織、分屬不同的表格，所以需要將這些不同來源的資料表格，各自依其所屬的研究對象群集，合併成新資料表格。此步驟隱含了一個抽象化的意義，就是同一個群集裡的個別研究對象，去掉其各別辨識鍵，抽象化為一個整體。例如：將政大附近的 A、B 便利商店抽象化為靠近政大的便利商店。

3.時間性資料切割與抽象化

- 依抽樣間隔切割新資料表格：本架構一開始時，曾經要求使用者選定適當的抽樣間隔。本步驟即依此抽樣間隔，利用簡單的關連表格操作，分別將上述的新資料表格切割成一段一段的連續時間區塊。這些時間區塊就是藉以發掘週期的抽樣間隔。
- 進行時間抽象化：此處乃將上述的每一段時間區塊中所發生的事件，根據事先定義的時間概念樹抽象化成同一個上層時間單位，如此實則將同一時間片段內所發生的事件可視為是發生在同一個時間單位裡。

4.發掘多數項目或多數項目之間的關聯

- 找出多數 (Large) 的項目：對每一研究對象群集的資料，分析其各時間區塊中，以事先定義的一般概念樹為基礎，利用 Han 等人所發展的屬性導向歸納分析技巧，對資料進行抽象化的動作，並分別計算各個項目在此時間區塊內的支持度，如果低於之前使用者選擇的最小支持度門檻值則刪除之，剩下來的項目即為此時間區塊中多數項目。
- 找出所有多數項目之間的關聯：對每一研究對象群集的資料，分析其各時間區塊中，考慮所有 2 項目的組合，分別計算 2 個項目的支持度，若其高於最小支持度門檻值，且其信賴度高於最低信賴度門檻值，則此 2 項目間的關聯成立，然後依序找出 3 個項目的、4 個項目等的關聯，直到找出所有的多數項目之間的關聯為止。
- 產生序列資料：對每一研究對象群集建立一 0 與 1 支連續系列，此序列之長度即為其時間區塊總數；而序列之第 i 個字元為 1 或 0 則視是否在第 i 時間區塊找出多數項目（或多數項目關聯）而定。

5.發掘週期

對每一研究對象群集所建立之 0 與 1 的連續序列，先從最短的週期長度開始找起，也就是週期長度 = 1，看看是否能從此序列當中找出週期長度為 1 的週期。然後依序增加週期長度，反覆尋找，直到週期長度增加到和最大週期長度一樣為止，以找出此序列的所有大小週期。然後整理此序列所有的週期，如果有某一大週期可從一小週期延伸而來，則將此大週期刪除。

（五）知識規則展現

將上述資料發掘方法在各階段所發掘出來的特性組合（有些存放在暫存區），並轉化成口語的方式來表達。此外，對於道路、山岳、河流分佈的資訊，可再加上圖形或地圖配合規則一起呈現。

（六）多層次資料發掘

由於資料發掘本身即為一不斷重覆的過程，使用者可能需要在第一次下達問題的結果呈現後，再修正其問題以找出更合適之解答，因此在本架構，讓使用者有機會修正其問題，盼能找出更適切的知識規則，並不斷重覆此過程，此為回饋的機制。

本研究提供了四種選項供使用者參考選擇，分別說明如下：

1. 區域層級的提升：提供空間概念樹供使用者參考、選擇。如原本使用者有興趣的是「文山區裡面的便利商店」，由於文山區屬於台北市，所以就可提供使用者如「位於台北市內所有的便利商店」的選擇。然後系統就可根據使用者選擇的新參數，從建立相關資料表格模組重新出發。
2. 節點層級的提升：同樣提供空間概念樹中之節點供使用者參考、選擇。例如原本使用者有興趣的是「文山區附近的便利商店」，這時就可列出往上、往下各一層級之節點供使用者選擇，如「台北市附近的便利商店」、「政大附近的便利商店」。然後系統就可根據使用者選擇的新節點，從空間分析模組重新出發。
3. 抽樣間隔的變更：提供時間概念樹的相關層級供使用者選擇。例如當使用者發掘出每天的週期之後，可能會想發掘每週、每月或是每半天、每小時的週期，這時就可從時間性資料切割與抽象化模組重新出發。
4. 商品多層級多數項目或多數項目之間的關聯：提供一般概念樹的相關層級供使用者選擇。例如：「飲料可特殊化為冷飲；冷飲又可再特殊化為汽水」。此時系統就可回到發掘多數項目模組，依照一般概念樹，針對項目進行特殊化的動作。

二、與以前研究的比較

（一）在空間資料發掘方面

以前的研究主要是針對與空間資料相關的屬性資料做資料發掘，本研究對空間資料發掘侷限於空間關聯規則，但強調對地圖資料直接先進行空間分析，而 Koperski 與 Han (1995) 則利用空間述語來進行空間物件關係的推演，以尋找空間關聯規則，Lu、Han 與 Ooi (1993) 之研究在找空間特性規則，Ng 與 Han (1994) 則在作空間分群，不過其分群的依據是依距離來分，而非本研究之空間分析尚可以道路長度來分析。

（二）在週期關聯規則發掘方面

Ozden 等人曾經提出兩套發掘週期關聯規則的演算法。本研究參考其中的交錯演算法。不同之處在於本研究將原本的關聯規則部分改為空間關聯規則，而且加入 Han 多層級知識發掘的概念，在規則的表達上，分別加入了研究期間和最低信賴度以上的說明，但 Ozden 等人在效率上做了不少努力，例如：週期的簡化 (Cycle-Pruning)、週期計算的省略 (Cycle-Skipping)、週期的刪除 (Cycle-Elimination)。這些在本研究目前仍未加入。

（三）時間和空間資料發掘的整合

以往研究中也有企圖同時對時空資料進行發掘，如 Stolorz 等人曾經分析大型氣候資料庫中的氣候資料。不過他們是對影像資料找出大氣物件的演進規則。本研究在時空資料的整合上，是採二階段式，先以地圖之空間分析挑出研究對象，再尋找交易資料之週期關聯型態，故可稱為空間性週期關聯規則。

（四）在查詢語言方面

以往的資料發掘查詢語言，通常沿襲許多 SQL 語法或類似的語言為工具，本研究讓使用者可直接在地圖上點選物件，盡量不讓使用者輸入查詢語言，而是在使用者點選之後，才由系統自動產生查詢語言。

（五）多層次關聯規則

以往 Han 等人的研究中也有多層次關聯規則的概念，但只考慮商品種類這個維度。但在本研究中加入了時間與空間另外兩個重要維度的多層次。

（六）知識規則的表達

知識規則的表達也是資料發掘領域中一個相當重要的課題，本研究之口語化表達較易讓人明瞭，而且能夠讓使用者在真實地圖上，看出與查詢相關空間物件之實際空間分佈及相對關係，或是周邊的交通、道路情形等，提供一些額外的資訊以增加知識規則的可讀性。

肆、系統雛型實作

一、系統環境

本研究假想有台北市 12 家便利商店，於 1999 年 3 月 1 日凌晨 0 點至 1999 年 3 月 10 日深夜 24 點之銷售資料共 9621 筆，並配合真實的最新台北市地圖（但假想資料分佈只有大安、中正、內湖、文山四區）完成一雛型系統實作（本應用資料庫商品及交易資料欄位如表 6、7）。其中商品編號已隱含其分類，如 6021 之第一碼「6」代表「麵食」，第二、三碼「02」代表「辣味涼麵」，第四碼「1」代表「義美」。此種代碼的設計，實已隱含概念樹觀念，將有助於抽象層次的提昇。「間隔序數」乃將交易依時間分類，代表該交易發生在使用者選定期間之第幾個時間片段。

表 6 商品一覽表之例

商品名稱	商品編號	單價
統一低脂牛奶	1010	17
光泉低脂牛奶	1011	17
統一全脂牛奶	1020	17
光泉全脂牛奶	1021	17
統一草莓麵包	2010	20

表 7 交易資料之例

商店編號	交易編號	商品編號	數量	時間	間隔序數
H1	H1301001	6021	2	199903010025.00	1
H1	H1301002	2020	1	199903010033.00	1

雛型開發的環境，發展瀏覽器圖形化使用者界面(Browser-Based GUI)，並利用 HTTP 作為通訊協定，以便讓使用者能透過網路由任何一台個人電腦，在極具彈性的視覺化互動環境中進行資料發掘。至於程式部分，主要使用標準的 HTML 標記語言和 Microsoft 發展的 Active Server Pages (ASP) 語言，並輔以其他瀏覽器能支援的 Script 語法，來撰寫本雛型（如圖 2）。

本雛型的軟體開發環境與技術如圖 2 所示，在伺服器端作業系統方面將使用 Microsoft NT Server 4.0，其上將架設 Microsoft Internet Information Server 4.0 (IIS) 以提供 WWW 的網路服務（雛型程式主要放在此）。空間資料庫將採用 ArcView 所內建的雙元架構空間資料庫，並另外使用能與 ODBC(Open Database Connector)界面溝通的開放式資料庫，此處用 Microsoft Access Server。在空間分析應用程式的部分則使用 ESRI 的 ArcView 和其 Web 應用模組 IMS Server。

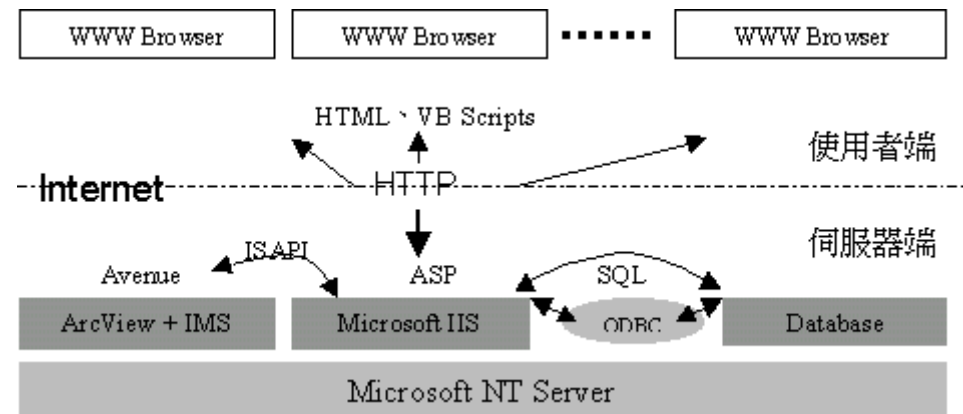


圖 2 本雛型開發環境與技術示意圖

在伺服器端方面，爲了借重 ArcView 強大的空間分析功能，本系統利用 ISAPI 作爲 IIS 與 ArcView 之間溝通的介面。而 ArcView 又必須經由 Avenue 程式撰寫才能應用，所以 ISAPI、Avenue 和 IMS 皆是 ASP 和 GIS 整合不可或缺的元素。在資料庫連結方面，則是使用微軟的 ODBC 程式介面，並輔以 SQL 作爲查詢的利器。

在用戶端方面，除了 HTML 之外，由於本雛型系統需要使用者輸入資料發掘相關參數，例如研究日期、最大週期長度等等，所以需要資料檢查，目前是利用 VB Script 來撰寫，所以使用者只需使用有支援 VB Script 的 WWW 瀏覽器，不需另外再安裝任何軟體即可執行本雛型系統。至於伺服器端和用戶端之間訊息的傳遞，則是利用 HTTP 作爲兩者之間溝通的橋樑。

二、雛型系統介紹

本雛型系統共有 15 個模組。各個模組以資料流圖說明如圖 3。目前雛型之空間分析並未包含商場分析模組，而區域選擇是直接在地圖上點選，以 Image Map 連接至已預建其相關便利商店之網頁檔建成。最近鄰分析則採事先 ArcView 之離線作業找出所有可能地點之鄰近便利商店，直接寫入程式內。只有距離分析真正連結 ArcView 地理資訊系統，其選擇畫面如圖 4。

當使用者按「下一步」時，系統會呼叫 ArcView 和其 Web 應用模組 IMS Server，以便在伺服器端進行距離分析，其結果將輸出至用戶端的瀏覽器，使用者可利用 ArcView 之地圖瀏覽工具來瀏覽該畫面。如圖 5。

建立起始表格之畫面如圖 6，由於雛型資料庫目前只有 10 天時間，所以設定使用者最小研究期間為 4 天。時間資料歸類之選擇畫面如圖 7，由於週期數太少並無意義，目前要求研究期間（以小時計）除以最大週期長度必須大於 6。使用者必須選擇支持度與信賴度之門檻值，如圖 8。最後知識規則之展現，如圖 9。使用者可選擇要進行多層次資料發掘，如圖 10。其中節點層級的提升，乃將地點之節點抽象提升，由「點」至「區」、「市」、「省」，如圖 11。而商品多層次關聯規則之選擇，乃依商品之概念樹直接選擇，如圖 12。

三、與以前研究之系統雛型比較

本研究之雛型與以往研究之系統雛型比較，如表 5，目前雛型並未從交易資料庫中找出所有商品組合之關聯規則，而只限於單一商品之關聯。也未採用 Han 與 Fu (1995) 中所提之特性來改進效率。在時間複雜度上，主要的複雜度在關聯規則發掘及週期偵測整理上。前者為 $O(n \log(n))$ ，後者為 $O(r \times I \times L_{\max})$ （後者與 Ozden (1998 年) 同），其中 n 、 r 、 I 、 L_{\max} 分別表示交易資料庫中總資料筆數、找到的規則數、時間間隔總數（即研究期間除以時間間隔）、使用者選擇的最大週期長度。所以時間長短及交易筆數均會對時間複雜度有重大影響。

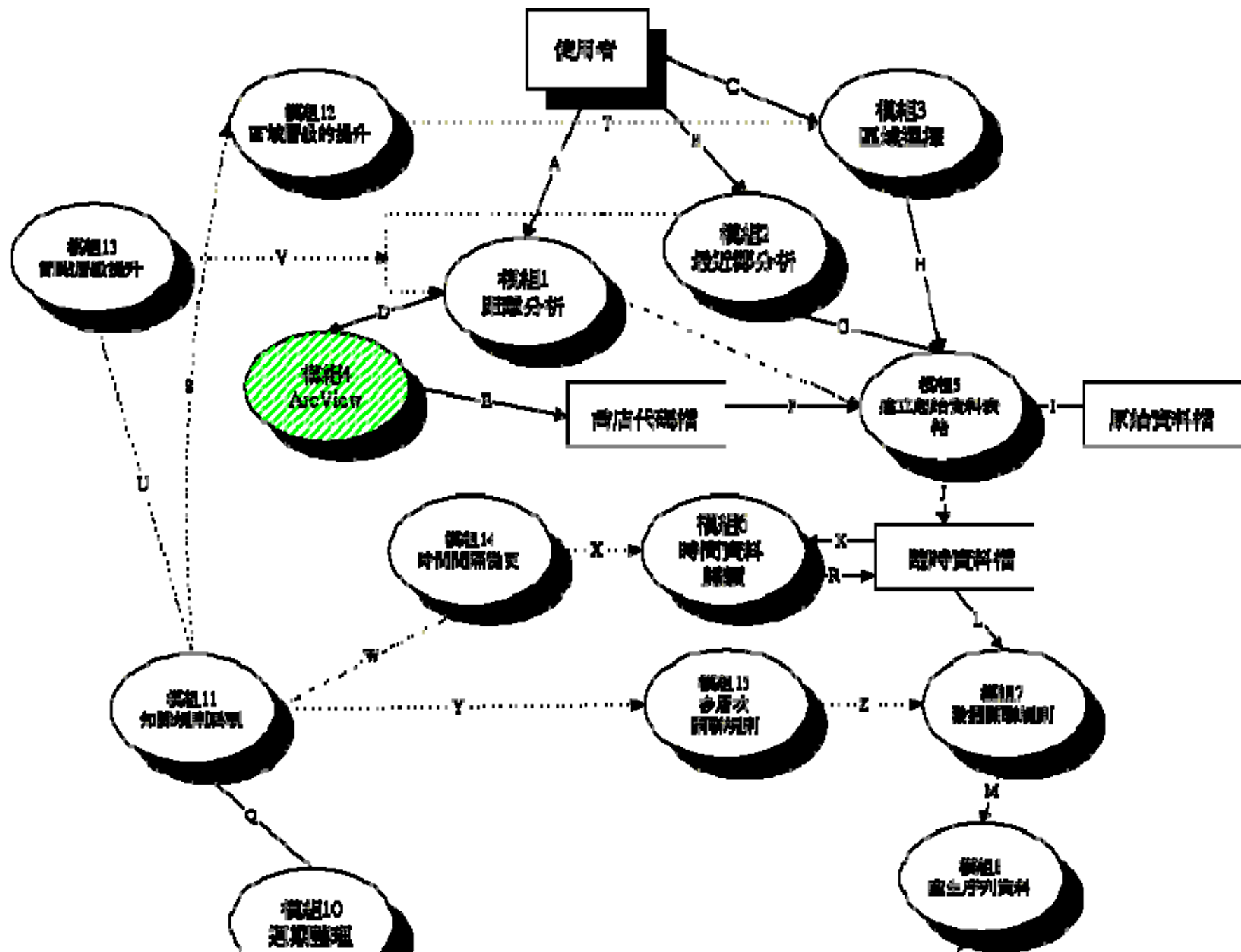
由表 5 中可見，本雛型系統與以往之研究最大差異是在於：（1）為 WWW 上之瀏覽器環境、（2）以個人電腦代替工作站、（3）實際執行效率較差、（4）先作空間分析再作時間之兩階段模式資料發掘。

表 5 雛型系統比較表

	雛型系統列表		
比較項目	本研究	Koperski 與 Han (1995)	Ozden 等人 (1998)
程式語言	HTML、ASP、 VB Script、Avenue	C、SR	未知
系統開發 及 操作環境	Intel Celeron 300a PC	SUN SPARCstations 5	SUN SPARCstations 20
作業系統	Microsoft Windows NT Server 4.0	Solaris	Solaris 2.5.1
資料來源	假想之便利商店 交易資料	實際的 美國人口普查資料	程式產生的 交易資料
發掘規則	空間性週期關聯規則	空間關聯規則	週期關聯規則
網路版	是	否	否
效率	$O(n \log(n)) + O(r \times I \times L_{\max})$	$O(C_c \times n_c + C_r \times n_r + C_{\text{nonspatial}})$	$O(r \times I \times L_{\max})$

實際執行效率 (相對而言)	低	高	高
------------------	---	---	---

- 註：其中 C_c 為粗略計算每一個空間述語的平均成本， n_c 為粗略計算的空間述語總數； C_f 為詳細計算每一個空間述語的平均成本， n_f 為詳細計算的空間述語總數；而 $C_{\text{nonspatial}}$ 則是在屬性資料庫中發掘規則的總成本。 r 是所發掘的關聯規則總數； I 則代表所有的時間單位個數總和； L_{max} 則是由使用者選定的最大週期長度。



- 註： A：使用者選擇的距離、地點、輸出圖面大小
- B：使用者選擇的地點
- C：使用者點選的台北市行政區
- D：距離、地點、輸出圖面大小
- E1：距離分析圖
- E2：選出的商店代碼
- F：商店代碼
- G：地點、商店代碼
- H：台北市行政區、商店代碼
- I：原始交易資料
- J：起始資料表格
- K：起始資料表格之時間欄位
- R：時間單位序數
- L：歸類之後的起始資料表格
- M：商品出現次數
- N：各商品組合的 01 序列
- O：各商品組合的 01 序列
- P：每一串序列所有的週期
- Q：發掘資料結果（含原先使用者所選擇的發掘條件）
- S：使用者選擇的新區域
- T：新區域
- U：使用者選擇的新節點
- V：新的影響區域

W：使用者選擇的新時間間隔和新最大週期長度以及商店代碼

X：時間間隔、最大週期長度和商店代碼

Y：使用者選擇的商品特殊化層級和商店代碼

Z：符合此商品特殊化層級的商品名稱和商品代碼

α ：研究期間

β ：時間間隔、最大週期長度

γ ：規則類別、門檻值

δ ：時、空或商品多層次的選擇及其資料發掘條件

(畫斜線的模組代表外部模組；灰階的模組代表多層次資料發掘；
虛線則代表多層次資料發掘的輸入、輸出。)

圖 3 雛型系統資料流圖 (Level 1)

距離分析



距離： 150公尺 200公尺 500公尺 750公尺 1000公尺

地點： 中正紀念堂 政治大學 台灣大學 大安公園 大湖公園

輸出圖面大小： 解析度480*320 解析度640*480 解析度800*600

[回首頁](#)

[上一步](#)

[下一步](#)

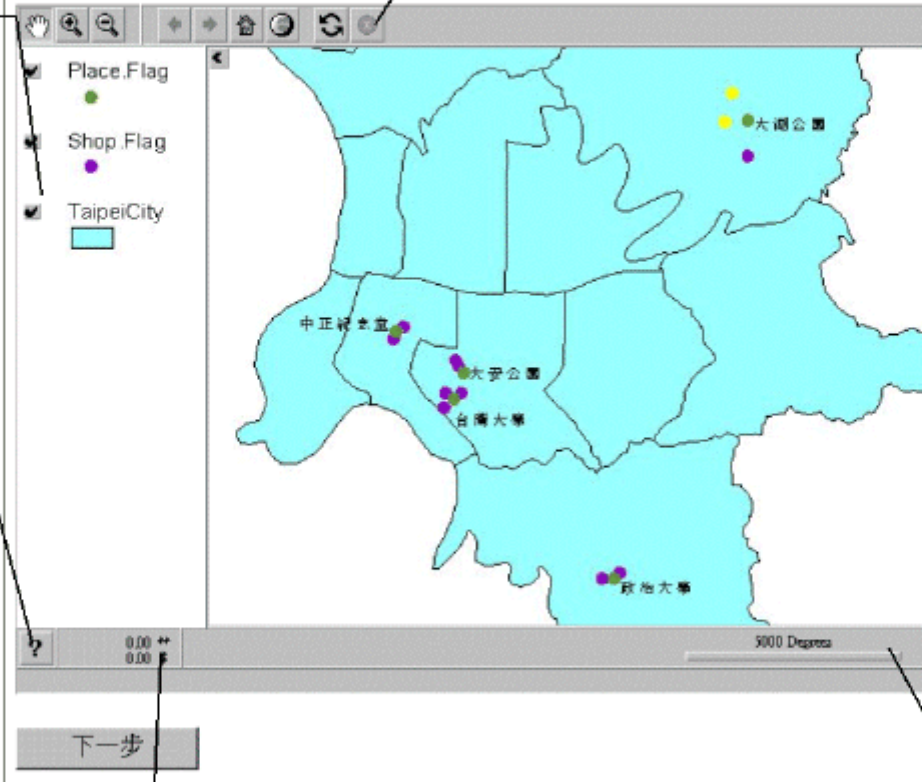
圖 4 距離分析選擇畫面

地圖瀏覽工具列

距離分析結果

黃色點為分析後被選取之商店，綠色點為分析中心位置，紫色點為其他未被選取商店位置。

選擇圖層



關於此地圖

X、Y 座標

比例尺

- 註：1.圖中頂端有 9 個地圖瀏覽工具由左至右分別為：移動地圖、放大、縮小、上一個畫面、下一個畫面、回到系統預設的縮放比例、觀看整張地圖、重新傳送地圖和停止傳送地圖
- 2.由於選擇大湖公園 750 公尺以內，故右上方之兩家店應為黃色，其他便利商店超出距離未被選出，應呈現紫色，唯此圖因黑白印刷無法表現出顏色。

圖 5 距離分析結果輸出畫面

建立起始表格

您選擇的地區為文山區，接下來請選擇研究期間

研究期間：

開始日期 1999 年 3 月 1 日 00時00分

結束日期 1999 年 3 月 4 日 24時00分

開始建立 重新選擇

系統說明：在您選擇了研究期間之後，本系統將會從交易資料庫中，找出符合您所選定條件的所有便利商店，從中選擇（Select）研究期間的交易記錄並建立起始表格以供後續資料發掘之用。

圖 6 建立起始資料表格畫面

您總共選擇2家商店，系統正在挑選符合條件的資料，並建立起始表格，請稍待...

起始資料表格建立完成！

列出起始資料表格

接下來，請選擇時間間隔與最大週期長度：

時間間隔：小時 最大週期長度：小時

下一步

系統說明：在您選擇了時間間隔和最大週期長度之後，本系統將會從起始資料表格中，

將所有交易紀錄依時間先後順序和您所選擇的時間間隔作歸類。

以時間間隔等於1為例，則所有發生在研究期間開始第一個小時之內的所有交易都將在其時間單位序數欄位中登記為1，第二小時登記為2，以此類推。

圖 7 時間資料歸類選擇畫面

B類規則—門檻值選擇

您選擇的規則種類為B類規則，請選擇以下門檻值：

最小支持度門檻值 最低信賴度門檻值

例如：「從1999年3月1日到1999年3月10日，靠近政大的便利商店，每天中午11點到1點之間，在資料庫裡，有60%以上，發生飲料交易（20%）也會發生牛奶的交易。」

其中20%即為最小支持度門檻值；而60%即為最低信賴度門檻值。

發掘關聯

圖 8 B 類規則—門檻值選擇畫面

知識規則展現

共找到以下4條B類規則。

「從1999年03月01日00時到1999年03月8日24時，在文山區裡的便利商店，每14個小時的第7個小時開始到第8個小時結束之間，在資料庫裡，有5%以上，發生牛奶的交易也會發生麵包的交易。」

「從1999年03月01日00時到1999年03月8日24時，在文山區裡的便利商店，每16個小時的第13個小時開始到第14個小時結束之間，在資料庫裡，有5%以上，發生牛奶的交易也會發生麵包的交易。」

「從1999年03月01日00時到1999年03月8日24時，在文山區裡的便利商店，每20個小時的第9個小時開始到第10個小時結束之間，在資料庫裡，有5%以上，發生牛奶的交易也會發生麵包的交易。」

「從1999年03月01日00時到1999年03月8日24時，在文山區裡的便利商店，每22個小時的第1個小時開始到第2個小時結束之間，在資料庫裡，有5%以上，發生牛奶的交易也會發生麵包的交易。」

[回首頁](#)

[多層次資料發掘](#)

[結束](#)

圖 9 B 類規則展現畫面

多層次資料發掘

請選擇您要進行的多層次資料發掘：

註：由於您之前選擇的空間分析種類為"區域選擇"，所以目前您無法選擇節點層級的提升。

〈請參見系統說明〉

- 區域層級的提升
- 時間間隔的變更
- 商品多層次關聯規則

回首頁

下一步

結束

系統說明

圖 10 多層次資料發掘選擇畫面二

節點層級的提升



您之前所選擇的地點是政治大學，現在您可將其抽象化成區

(例如：政大附近可抽象化成文山區附近，文山區附近又可再抽象化成台北市附近)

下一步

圖 11 節點層級的提升—細部選擇畫面

商品多層次關聯規則

您之前所發掘的規則為「B類規則」，現在您可以更進一步發掘「多層次B類規則」。

（例如可以將牛奶從商品概念樹的根部向下特殊化，如果選擇「商品細類層」，則可特殊化成低脂牛奶或全脂牛奶兩類；

而在「商品細類和製造廠商層」，低脂牛奶又可再特殊化成統一的低脂牛奶或光泉的低脂牛奶兩類。）

※※ 請點選您想發掘商品特殊化的層級 ※※

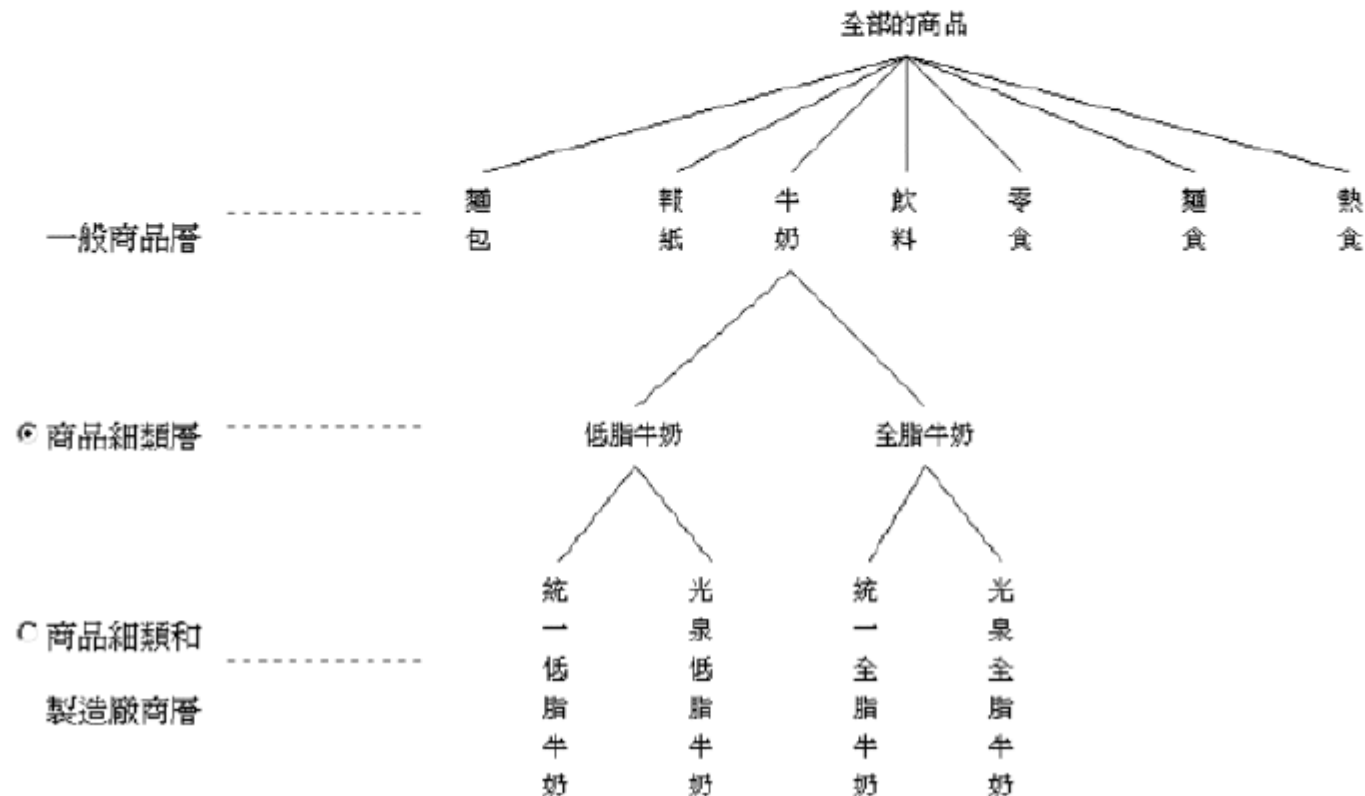


圖 12 商品多層次關聯規則—細部選擇（部分畫面）

伍、結論與建議

一、結論

關聯規則的發掘一直是資料發掘的主要研究項目之一，其目的不外乎是想從龐大的交易資料庫中，發掘出有用的知識規則，除了可以更深入瞭解企業本身的銷售情況之外，也可根據這些知識規則擬定因應的對策，進而提高競爭力。後來有學者提出「週期關聯規則」的觀念，認為除了關聯規則本身之外，關聯規則的週期也是值得考慮，例如除了能夠知道「買牛奶也會買麵包」之外，如果能夠知道「每天早上 8 點到 9 點之間，買牛奶也會買麵包」的話，那這樣的關聯規則似乎具有更高的商業價值。此外，在商業上區位的考量也是相當重要的因素，例如許多便利商店喜歡開在學校附近，如果能夠瞭解符合某種空間特性之便利商店具有什麼特殊知識規則的話，那企業就更能夠因地制宜了。

本研究結合了以上各種相關研究，提出了「空間性週期關聯規則」的構想，期望能先考慮空間分析再將關聯規則加入週期特性，以找出更有價值的知識規則。因此，本研究具體達成的成果包含：

1. 整理「空間性週期關聯規則」相關的時間、空間資料發掘文獻。
2. 提出一個發掘「空間性週期關聯規則」的架構，為一個二階段的模式，先直接針對空間資料進行分析，再作時間資料發掘以及考慮時、空、商品種類三維度多層次的觀念。
3. 雛型系統實作方面

- 查詢介面的改進：改進了以往命令列式的查詢介面，使用者只要用滑鼠點選即可完成整個資料發掘動作，並且還可直接在地圖上點選物件。
 - 嘗試結合地理資訊系統與資料發掘:雖然有許多研究是與空間資料發掘有關，但並無直接結合地理資訊系統。本研究嘗試將空間之距離分析交由地理資訊系統來執行。
 - 多層次空間性週期關聯規則：本研究除了發掘一般的「空間性週期關聯規則」之外，並嘗試進行時間、空間及商品種類三個維度的「多層次空間性週期關聯規則」的發掘，希望能讓使用者找出更深一層的規則。
 - 知識規則的口語化表達:本研究嘗試使用較口語化的表達方式來展現發掘的成果，希望能讓使用者更瞭解所發掘的知識規則。
 - 整合時間與空間資料發掘：本雛型先作空間距離分析、區域選擇及最近鄰分析，再進行時間資料發掘，雖未作到時空資料之全面整合，但已是此領域的一個初步嘗試。
 - 瀏覽器的資料發掘環境：以往的資料發掘多是單機的環境，本研究嘗試將資料發掘系統架在全球資訊網之上，希望能提供一個更便利的環境給資料發掘使用者。
4. 在實務上，本系統除了便利商店相關應用如促銷（可找出哪些商品在某段時間之週期關聯以加強促銷）、選址（可找出哪些地區銷售情況較佳，作為選址考量）之外，本系統還可應用於如：區域規劃（可收集不同行政區在各時間之週期性活動資料，以作為區域規劃參考）；動物行為研究（可觀察不同地區動物在各時間之週期性行為）等。

但是，本研究系統雛型仍有一些限制包括了：

1. 本系統只能發掘單一商品項目之間的關聯：目前本系統只能發掘單一商品項目之間的關聯，但實際上，關聯規則也可以是多商品項目之間的關聯。
2. 只利用單一屬性發掘關聯規則：在交易資料庫中，一筆交易其實有許多屬性，本研究目前只針對「商品編號」部分發掘關聯規則，但其他如價格、製造商、數量等屬性，仍然有發掘的價值，理論上應利用更多屬性來發掘關聯規則。

3. 本系統只能發掘「嚴格週期」：在現實生活中，有許多週期的發生機率並不是 100%，但本研究目前卻只能發掘會 100%再發生的「嚴格週期」。
4. 發掘週期前需指定週期條件：本研究在發掘週期之前，需由使用者選擇研究期間、時間間隔及最大週期長度才能進行週期偵測，但實際上，如果能由系統自動發掘出所有週期則將會更便利。
5. 與地理資訊系統的結合不夠緊密：雖然本研究嘗試結合地理資訊系統與資料發掘，但目前只有距離分析真正結合地理資訊系統。而且只能發掘出空間性週期關聯規則，無法找出其他的時空規則。
6. 實際執行效率較差：本研究爲了整合空間分析、多層次關聯規則及週期發掘，並將系統架在全球資訊網上，限於研究時間及人力、經費，所以在雛型系統效率上並未做特別考量，導致實際系統執行時效率不彰。

二、後續研究的建議

未來加強的研究主題包括上述所提之各種限制，在文獻上不是沒有提出解決方案，例如，Han 與 Fu（1995）就曾提出一次發掘多商品項目之間關聯的作法，也曾考慮到在此過程中利用一些特性來提升效率，Ozden（1998）也對週期發掘中效率問題提出三種解決方案。而在 Ng 等人（1998）及 Lakshmanan 等人（1999）的研究也提出在尋找商品種類關聯規則時考慮加入其他屬性限制的架構。在 Han, Dong 與 Yin（1999）的研究中也提出演算法來一次發掘所有週期（而不需要使用者指定參數）及考慮非嚴格週期的問題。然而這些研究均有待進一步結合在一起才能有效地解決目前時間、空間、商品種類等維度同時考慮的問題，此外，如何發掘出空間對空間之週期關聯規則或其他種類的時空關聯限制，這些均有待日後的研究突破。

參考文獻

1. 林幸怡，民 86，擴充先前知識以輔助資料發掘，政治大學資訊管理研究所碩士論文。
2. 周學政、周天穎，民 86，ArcView 透視 3.X，台北：松岡電腦圖書資料股份有限公司。

3. 許建志、傅志雄，民 87，精通 Active Server Pages，靖宇資訊科技股份有限公司。
4. Agrawal, R., Faloutsos, C. and Swami, A. 1993. Efficient Similarity Search in Sequence Databases. Proc. 4th Int. Conf. on Foundations of Data Organization and Algorithms.
5. Agrawal, R., Imielinski, T. and Swami, A. 1993. Mining Association Rules Between Sets of Items in Large Databases. Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data , Washington, D. C.: 207-216.
6. Agrawal, R. and Srikant, R. 1994. Fast Algorithms for Mining Association Rules. Proc.1994 Int. Conf. VLDB, Santiago, Chile:487-499.
7. Cai, Y., N. Cercone. and J. Han. 1993. Data-driven Discovery of Quantitative Rules in Relational Databases. IEEE Trans, Knowledge and Data Eng., Vol. 5:29-40.
8. Chen, M. S., J. Han. and P. S. Yu.1996. Data Mining : An Overview from Database Perspective. IEEE Transactions Knowledge and Data Engineering:866-883.
9. Dietterich, T. G. and Michalski, R. S. 1985. Discovering Patterns in Sequences of Events. Artificial Intelligence, Vol. 25.
10. Fayyad, U., et al. 1993. Automated Analysis of a Large Scale Sky Survey : The SKICAT System. Proc. 1993 Knowledge Discovery in Databases Workshop, Washington, D. C.:pp.1-13.
11. Fayyad, U. M. and Smyth, P. 1993. Image Databases Exploration : Progress and Challenges. Proc. 1993 Knowledge Discovery in Databases Workshop, Washington, D. C.:14-27.
12. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R., editors. 1996. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, Menlo Park, CA.
13. Fotheringham, S. and Rogerson, P. 1994. Spatial Analysis and GIS, Taylor and Francis.
14. Han, J., Cai, Y. and Cercone, N. 1993. Data-Driven Discovery of Quantitative Rules in Relational Databases. IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 1:29-40.
15. Han, J., Dong, G., and Yin, Y. 1999. Efficient Mining of Partial Periodic Patterns in Time Series Database. Proc. 1999 Int. Conf. on Data Engineering (ICDE'99),Australia: 106-115.
16. Han, J. and Fu, Y. 1995. Discovery of Multiple-Level Association Rules from Large Databases. Proc. of 1995 Int. Conf. on Very Large Data Bases (VLDB' 95) , Zich, Switzerland:420-431.
17. Knorr, E. and Ng, R. T. 1995. Applying Computational Geometry Concepts to Discovering Spatial Aggregate Proximity Relationships. Technical Report, University of British Columbia.

18. Koperski, K., Adhikary, J., and Han, J. 1996. Spatial Data Mining: Progress and Challenges. SIGMOD'96 Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'96), Montreal, Canada.
19. Koperski, K. and Han, J. 1995. Discovery of Spatial Association Rules in Geographic Information Databases. Proc. 4th Int'l Symp. on Large Spatial Databases (SSD' 95), Portland, Maine:47-66.
20. Lakshmanan, L.V.S., Ng, R., Han, J. and Pang, A. 1999. Optimization of Constrained Frequent Set Queries with 2-Variable Constraints. Proc. 1999 ACM-SIGMOD Conf. on Management of Data, Philadelphia, PA, June:157-168.
21. Lu, W., Han, J. and Ooi, B. C. 1993. Discovery of General Knowledge in Large Spatial Databases. Proc. Far East Workshop on Geographic Information Systems, Singapore.:275-289.
22. Major, J. and Mangano, J. 1993. Selecting among Rules Induced from a Hurricane Database. Proc of 1993 KDD Workshop, Washington, D. C.:pp.28-47.
23. Michalski, R. S., Carbonnel, J. M. and Mitchell, T. M., editors. 1983. Machine Learning : An Artificial Intelligence Approach, Morgan Kaufmann, Los Altos, CA.
24. Mitchell, T. M. 1982. Generalization as Search. Artificial Intelligence, Vol. 18:203-226.
25. Ng, R. and Han, J. 1994. Efficient and Effective Clustering Method for Spatial Data Mining. Proc. 1994 Int. Conf. Very Large Data Bases, Santiago, Chile, September:144-155.
26. Ng, R., Lakshmanan, L.V.S., Han, J. and Pang, A. 1998. Explorator Mining and Pruning Optimizations of Constrained Association Rules. Proc. of 1998 ACM-SIGMOD Conf. On Management of Data, Seattle, Washington, June:13-24.
27. Ozden, B., S. Ramaswamy. and A. Silberschatz. 1998. Cyclic Association Rules. Proc. of 1998 Int. Conf. Data Engineering (ICDE' 98) :412-421.
28. Piatetsky-Shapiro, G. and Frawley, W. J., editors, 1991. Knowledge Discovery in Databases, AAAI/MIT Pres, Menlo Park, CA.
29. Shaw, G. and Wheeler, D. 1994. Statistical Techniques in Geographical Analysis, London, David Fulton.
30. Smyth, P., Burl, M. C., Fayyad, U. M. and Perona, P. 1994. Knowledge Discovery in Large Image Databases : Dealing with Uncertainties in Ground Truth. Proc. of AAAI-94 Workshop on KDD, Seattle, WA:109-120.
31. Snodgrass, R. T. and I. Ahn. 1985. A Taxonomy of Time in Databases. Proc. ACM-SIGMOD Int. Conf. Management of Data, Austin, TX, USA:236-246.
32. Snodgrass, R. T. and I. Ahn. 1986. Temporal Databases. IEEE Computer, Vol. 19, No 9:35-42.
33. Spiegel, M. 1996. Schaum' s Outline Series of Theory and Problems of Statistics, McGraw Hill.

[[回目錄](#) | [Top](#)]