

快速反向關聯法則與調整緊密規則 —促銷商品組合之應用

蔡玉娟、張簡雅文、黃彥文
屏東科技大學資訊管理學系

摘要

企業藉由所建立之專屬會員制度，透過資料探勘技術從龐大的會員交易資料庫發掘消費特徵，實現個人化之服務，有效區隔市場與訂定行銷策略。資料探勘技術之關聯法則的執行程序受限於必須由單一項目集，逐層擴展，經過長時間之重複組合與運算步驟，才能發掘合適之高頻項目集。本研究提出一個新的快速反向關聯法則(Fast-Backward Association Rule, FBAR)以克服上述之缺點，並應用於發掘特定促銷項目之商品組合。FBAR之執行程序反向於 Apriori 關聯法則，執行步驟為：(1)建立促銷目標資料表(table)－掃描交易資料庫一次，將交易資料庫中不符合預定促銷的項目刪除，而保留符合預定促銷的項目，記錄在促銷目標資料表並暫存於主記憶體；(2)分解促銷目標資料表之交易資料－在促銷目標資料表中，由最長交易資料開始逐層分解項目集；(3)發掘符合最小支持度之高頻項目集－當分解至某長度之項目集且已符合最小支持度，則停止該項目集之分解。FBAR 僅需掃描資料庫一次，而將刪減之交易資料記錄在促銷目標資料表並暫存於主記憶體，再透過分解較長交易紀錄，可快速發掘符合最小支持度之高頻項目集。藉由調整緊密規則法(Update-Compact Rule, UCR)可轉換高頻項目集為涵蓋率與緊密程度較高之關聯規則。

關鍵字：資料探勘、關聯法則、快速反向關聯法則

Fast-Backward Association Rule and Update -Compact Rule—Determination of Purchasing Patterns

Yuh-Jiuan Tsay, Ya-Wen Chang Chien, Yen-Wen Hung

Department of Management Information Systems, National Pingtung University of Science
and Technology

ABSTRACT

The discovery of association rules is an important and popular data-mining task, for which many algorithms have been proposed. These solutions are flawed, containing weaknesses that include often requiring repeated passes over the database, and generating a large number of candidate itemsets. In order to overcome the bottlenecks of association rule, we propose a new Fast-Backward Association Rule (FBAR) which is used to discover the specific promotional bundles. The procedure of FBAR is against Apriori algorithm. Efficiency of mining is achieved with three steps: (1) establish promotional table with specific promotional items—delete non-promotional items from database; (2) decompose transactions of promotional table—decomposing from the longest transaction record in promotional table, level by level; (3) find out frequent patterns conform to the minimum support—algorithm terminates when the calculated support is greater than, or equal to, the minimum support. The Update-Compact Rule (UCR) algorithm is obtained by modifying Compact Rule Set to reason the knowledge rules from frequent patterns. FBAR not only prunes considerable amounts of data reducing the time needed to perform data scans and requiring less memory, but also ensures the correctness of the mined results.

Key words: Data Mining, Association Rule, Fast-Backward Association Rule

壹、緒論

企業利用會員交易紀錄資料庫進行資料分析，以發掘有用的特徵或資訊，區隔各類型會員的消費狀況，進而萃取會員的消費模式，建立市場區隔與目標顧客群，提供企業行銷規劃的重要資訊(方世榮，1996；沈維明、翁頌舜，1999；蕭富峰，1993)。近年來，商品項目以組合方式出售的行銷策略，已普遍被應用於各種消費品市場。合適的促銷項目組合與對目標顧客群訂定行銷策略為企業獲利之關鍵因素。因此，企業依不同的產品特色與消費者特性而訂定商品組合策略，以達到吸引顧客及提高銷售量之目的。資料探勘技術可發掘隱藏於大量資料中有用的資訊，提供企業訂定相關策略之參考，以提昇企業之利潤與競爭力。

資料探勘技術之關聯法則被用於發掘資料庫之交易項目間彼此的關聯性，可從大量交易資料中發掘高頻項目集，再轉換為關聯規則。藉由這些規則可幫助企業制定目標行銷策略、執行目標區隔、訂定商品組合策略、分析顧客消費行為與規劃商品之陳設等等。例如：80%的顧客購買牛奶後會接著買麵包，因而將牛奶和麵包的展售櫃檯擺在一起，或組合不同的麵包和牛奶進行促銷(Brin et. al., 1997)。以往關聯法則之相關研究，並未考慮商品之購買數量或商品之價格，且其演算法之執行程序受限於由單一項目組合開始並逐層擴展，經過長時間之重複組合與產生大量之候選項目集，且需多次掃描資料庫進行反覆比對，花費大量的執行與比對時間才能找出合適之高頻項目集。

本研究提出快速反向關聯法則(Fast-Backward Association Rule, FBAR)應用於發掘特定促銷項目之商品組合，其執行程序反向於 Apriori 演算法。FBAR 僅需掃描資料庫一次，同時刪除交易資料庫中不用的項目，以產生目標促銷資料表並儲存於主記憶體中，可降低交易資料量而減少比對時間；再由促銷目標資料表之最長交易記錄開始分解，可快速發掘符合最小支持度之高頻項目集。FBAR 之執行程序反向於 Apriori 關聯法則，執行步驟為：(1)建立促銷目標資料表(table)－掃描交易資料庫一次，將交易資料庫中不符合預定促銷的項目刪除，而保留符合預定促銷的項目，記錄在促銷目標資料表並暫存於主記憶體；(2)分解促銷目標資料表之交易資料－在促銷目標資料表中，由最長交易資料開始逐層分解項目集；(3)發掘符合最小支持度之高頻項目集－當分解之項目集已符合最小支持度，則停止該項目集之分解。藉由本研究之調整緊密規則法(Update-Compact Rule, UCR)可轉換高頻項目集為涵蓋率與緊密程度較高之關聯規則。

貳、文獻探討

企業明確區隔特定產品的客戶群，以決定主要訴求對象的客戶群，即為目標市場的設定；而目標市場的適當與否，對銷售成果有很大影響(張希誠，1987)。企業利用多重通路擴展銷售機會以接觸不同的市場區隔，而為避免通路衝突，其管理方案包括建立會員制度、產品與售價之區隔等。資料探勘技術可發掘隱藏於會員交易資料庫有效的消費行為與模式，以輔助企業決策。Cabena 等五人定義資料探勘為「將事前不知道且有效的資訊從大型資料庫抽出的過程，並將抽出的資訊提供主管做決定性的決策」(Cabena et. al., 1997)。Fu(1997)將資料探勘的技術依資料庫型態、知識類別與應用層次的不同分為分類

(Classification)、趨勢分析(Trend analysis)、關聯法則(Association Rules)、分群法(Clustering)及序列特徵(Sequence pattern)等五種，其中關聯法則被用於發掘資料庫之交易項目間彼此的關聯性。Han & Kamber(2000)指出關聯法則是資料探勘技術中最被廣為討論與使用的方法。

一、*Apriori* 關聯法則

Agrawal 等學者提出關聯法則的數學模式(Agrawal et.al., 1993)，從大型資料庫中發掘高頻項目集以建立關聯規則(Mining Association Rules)，其數學模式如下所述。商品項目集(Itemset) $I=\{i_1, i_2, i_3, \dots, i_m\}$ 為所有可能的商品項目(items)之集合，交易記錄 T 為 I 中商品項目的子集合，D 為所有交易記錄 T 的集合，每筆交易記錄之唯一代號為 TID。設 X 和 Y 為 I 的兩個子集合且包含於交易記錄 T 中，當 $X \cap Y = \emptyset$ ，且同時滿足：(1)支持度(support) $s \geq$ 訂定的最小支持度(minimum support) — s 為 D 中包含 $X \cup Y$ 之交易記錄數目(c_i)佔 D 之全部交易記錄數目(N)的比率，亦即 $s = c_i/N$ ，記為 $\text{Support}(X \cup Y)$ ；(2)信賴度(Confidence) $c \geq$ 訂定的最小信賴度(minimum confidence) — c 為關聯法則的可信程度，其表示在 X 出現的情況下，Y 也會隨著出現的條件機率，亦即 $c = \text{Probability}(Y|X)$ ，記為 $\text{Support}(X \cup Y)/\text{Support}(X)$ ，則此關聯法則表示為 $X \Rightarrow Y$ 。

Apriori 演算法為最具代表性的關聯法則(Agrawal & Srikant, 1994)，主要分為兩步驟：(1)由資料庫中找出所有大於 minimum support 的高頻項目集(Large Itemsets) — 執行程序必須由第一層之單一項目集開始並逐層擴展，利用第 $k-1$ 層所得到的高頻項目集排列組合，而產生第 k 層的候選項目集(Candidate Itemsets)，每個候選項目集再與整個資料庫比對，若其支持度大於或等於所訂的 minimum support，則產生第 k 層之高頻項目集，經過逐層之重複組合與運算步驟，以發掘合適長度之高頻項目集；(2) 建立關聯規則 — 透過分析高頻項目集以建立合適之關聯規則。

二、關聯法則之相關研究

學者專家對於關聯法則之研究，大多以 *Apriori* 關聯法則之精神為主軸，改良其掃描資料庫次數與產生候選項目集的數目，以提昇效率與效能。Pork 等人利用雜湊法建立 Direct Hashing and Pruning(DHP) 關聯法則，有效控制長度 2 之候選項目集的數目，裁剪交易資料庫，以減少產生高頻項目集的計算次數，加快處理速度(Pork et al., 1995)。Han & Fu(1995)以 *Apriori* 演算法為基礎，提出 Multiple-Level(ML) 關聯法則，運用一般化處理的觀念於商品項目，先轉換資料為階層化架構的型式，並應用單層次關聯法則觀念於各層次之關聯規則的發掘，以降低候選商品項目集，減少反覆搜尋資料庫所耗費的時間。Savasere 等人提出 Partition 關聯法則，切割資料庫為數個區塊，先以較低的最小支持度發掘該區塊的可能項目集，再計算可能項目集之支持度，可有效減少資料庫掃描次數，但會產生較多的項目集(Savasere et al., 1995)。Agrawal & Srikant(1995)將時間屬性概念加入關聯法則中，應用於發掘循序特徵；並在 1996 年提出 Mining Sequential Pattern (MSP) 關聯法則(Agrawal & Srikant, 1996)。Toivonen (1996)所提出的 Sampling 關聯法則，僅需掃描資料庫一次，但執行結果有正確性的問題。Carter 等人考慮交易記錄之每一商品項目的購買數量，而提出 Share Based Measures (SBM) 計算最小支持度，改良僅考慮商品項目發生頻率的缺點(Carter et al., 1997)。Brin 等人提出 Dynamic Itemset Count(DIC) 關聯法則，以區塊為單位搜尋取代

整個資料庫的搜尋，可有效減少搜尋資料庫的次數，並應用於分析顧客的購買行為與消費特徵，發掘相關商品之關聯性，做為商品擺設、進貨與儲貨的參考(Brin et. al., 1997)。Liu 等人改良 Apriori 關聯法則中使用者預定之單一最小支持度，而提出多重最小支持度演算法，其可自動產生每個項目集的最小支持度，以發掘更精確且具實用價值之規則(Liu et. al., 1999)。Dunkel & Soparkar(1999)提出 Column-Wise (CW) Apriori 關聯法則，使用欄位交集方式進行比對，以發掘高頻項目集。Han 等人提出 Frequent Pattern Tree(FP-Tree)，將 L1 之高頻項目集依出現頻率排序，並根據交易資料建立 FP-Tree(Han et. al., 2000)。Berzal 等人提出 Tree-Based Association Rule(TBAR) (Berzal et. al., 2001)，兩者之概念為轉換交易資料的結構以節省掃描資料庫的時間，進而提昇資料探勘之效率。綜合上述各關聯法則之特色與缺點，如表 1 所示。

表 1：各關聯法則之特色與缺點比較表

關聯法則	特 色	缺 點
Apriori	最早且具代表性之關聯法則。	產生大量候選項目集，多次掃瞄資料庫。
DHP	建立 C_2 之 Hash Table，減少候選項目集，裁減資料庫。	耗費時間建立 Hash Table
ML	建立多層次關聯法則觀念，以減少候選項目集。	需轉換交易資料為階層式架構
Partition	切割資料庫為多區塊，分別計算各區塊之高頻項目集。	需檢驗所產生之非相關項目集
MSP	關聯法則包含時間屬性之概念，可發掘循序特徵。	演算法較為複雜
Sampling	掃描一次資料庫，依記憶體大小隨機抽取交易資料進行探勘	隨機抽樣可能導致結果不正確。
SBM	關聯法則包含數量顯著性之概念。	耗費較多的運算時間。
DIC	只掃描項目集所屬之資料庫區塊	若項目集分散於各區塊，則需掃描整個資料庫。
Msapriori	關聯法則包含商品項目價值之概念。	耗費較多的運算時間。
CW	掃描一次資料庫，採欄位交集方式取代比對。	不具裁減 large-itemset 的能力
FP-Tree	利用 Tree 結構減少候選項目集。	建立之 Tree 很龐大。
TBAR	掃描兩次資料庫，結合 Hash 與 Tree 結構，減少候選項目集。	Tree 很龐大且 Hash 可能造成 collision。

三、知識規則之推導

關聯法則所發掘之高頻項目集需轉換為關聯規則，以表達其邏輯特性，而常用以推導知識規則的方法為特性規則(Characteristic Rule)、區別規則(Discrimination Rule)與緊密規則(Compact Rule)。特性規則是以概念樹導向的學習方式推導規則，以單一類別資料對於相關屬性往上抽象化而整合出結果，並以數學邏輯方式表達所得之知識規則(Han, 1999)。區別規則為對資料庫中某些類別或資料項的特性比較，以發掘類別或資料項間的區別，其藉由概念樹導向將兩個類別之共同屬性的概念樹屬性值加以抽象化，並可選擇刪除或保留兩個類別特性重複的部分，並以數學邏輯方式表達所得之知識規則(Han et. al., 1992; Han & Fu, 1996; Chen et. al., 1997)。上述兩種規則可將會員資料特徵以機率代表出現比率，歸納各類型會員分佈的狀況。Yen & Chen(1996)應用緊密規則於歸納會員類型，推導流程為：(1)結合相關資料表為單一關聯資料表；(2)在關聯資料表中，若屬性符合概念階層推演，則執行一般化處理，並將不需分析之欄位刪除；(3)計算條件屬性與目標屬性的語意關聯程度值(Semantic Association Degree, SAD)；(4)若 SAD 為 1，則轉換為 IF-THEN 的規則，其代表所推導之規則具有完全可靠度(Full Confidence)。

緊密規則在推導過程中，結合相關資料表為單一關聯資料表，經一般化處理後，有時會發生某些記錄之條件屬性值相同，但目標屬性值卻不同的情形，而無法發掘重要的規則以歸納各類型會員分佈的狀況。蔡秀滿與陳健民(1999)在其研究中，改善緊密規則的方法，以分析多個資料表之消費者特徵與產品特性間的關係，進而推導重要的知識規則，並避免在推導過程將相關資料表作關聯結合。本研究調整緊密規則(UCR)演算法，無需將相關的資料表作關聯結合，以達到資料探勘效率之提升，並應用於歸納會員類型。

參、研究方法

一、研究架構

本研究提出 Fast-Backward Association Rule(FBAR)並應用於發掘特定促銷項目之商品組合，研究架構如圖 1 所示。發掘會員購買商品組合之知識規則推導流程分為四個步驟：(一)會員資料前置處理；(二)會員個人資料屬性一般化處理；(三)以 FBAR 演算法快速發掘高頻項目集；(四)UCR 推導商品組合之知識規則。

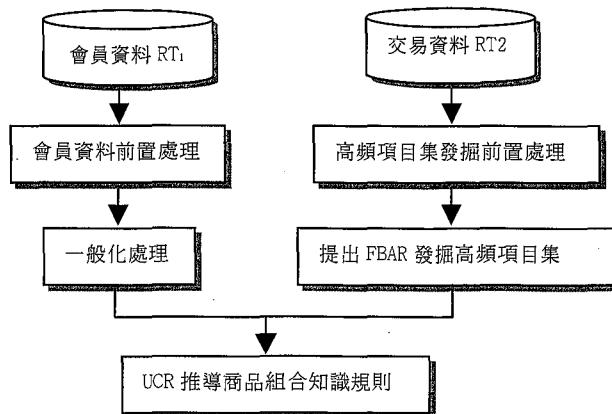


圖 1：發掘會員購買商品組合知識規則之研究架構

二、範例推導

(一)會員資料前置處理

企業所建立交易資料庫儲存大量的日常銷售資料，經過資料前置處理、過濾與篩選，並分析各種產品銷售與會員消費特徵，可加速線上分析處理的效率。表 2 與表 3 為範例推導之部分會員的基本資料及消費記錄。

表 2：會員個人資料記錄

會員編號	年齡	月收入	性別
M1	50	50000	女
M2	25	16000	男
M3	40	100000	女
M4	55	70000	男
M5	65	40000	女

表 3：會員交易資料記錄

TID	會員編號	購買商品項目
T001	M3	003, 002
T002	M4	002, 004
T003	M3	001, 007
T004	M5	002
T005	M2	002 ² , 008, 009
T006	M1	003, 004 ²
T007	M2	004
T008	M4	004, 007, 009
T009	M1	009
T010	M2	008, 009
T011	M3	003, 005
T012	M4	007, 009
T013	M1	003, 007
T014	M2	008
T015	M5	003, 005
T016	M1	001, 007
T017	M3	001

(二)會員個人資料屬性一般化處理

藉由一般化處理以獲得概念階層中往高層次之一般化資訊，概念階層定義屬性值之間的分類關係與階層關係，可以是不同的深度與彈性調整，並可直接應用於各種不同會員消費類型的分析決策活動，以提供較好的決策支援(Han, 1999；Han et. al., 1992；Han & Fu, 1996)。本研究以概念階層表示領域知識，資料表中的每一屬性，都可關聯一概念階層。將表 1 會員個人資料記錄表的部分屬性欄位一般化，如表 4 所示。利用預先設計的概念階層，對會員資料的屬性欄位分別建立完整的概念階層樹，圖 2(a)、2(b)為年齡與月收入兩屬性

欄位的概念階層樹。

表 4：一般化的會員個人資料記錄

會員編號	年齡	月收入	性別
M1	老年	中等收入	女
M2	青年	低收入	男
M3	中年	高收入	女
M4	老年	高收入	男
M5	老年	中等收入	女

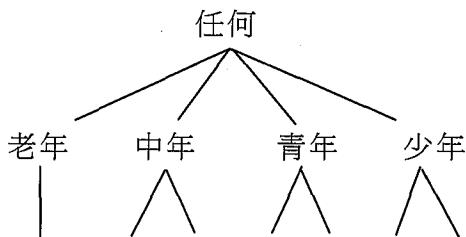


圖 2(a)：會員年齡屬性欄位的概念階層樹

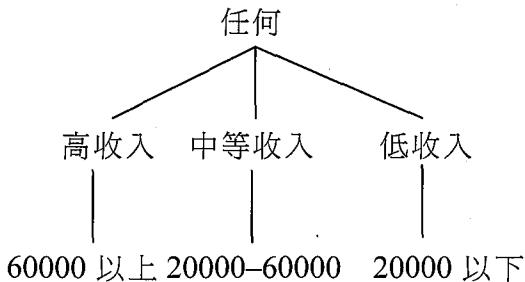


圖 2(b)：會員月收入屬性欄位的概念階層樹

(三) FBAR 發掘高頻項目集

本研究所提出之 FBAR 演算法，執行程序反向於 Apriori 關聯法則，執行步驟為：(1)建立促銷目標資料表(table)—掃描交易資料庫一次，將交易資料庫中不符合預定促銷的項目刪除，而保留符合預定促銷的項目，記錄在促銷目標資料表並暫存於主記憶體；(2)分解促銷目標資料表之交易資料—在促銷目標資料表中，由最長交易資料開始逐層分解項目集；(3)發掘符合最小支持度之高頻項目集—當分解至某長度之項目集且已符合最小支持度，則停止該項目集之分解。FBAR 之執行程序說明如後。

1. 預定促銷之商品項目

本研究假設預定促銷之商品項目編號 PID 為 001、002、003、004、005，刪除表 3 會員交易資料記錄中非促銷的項目，而保留符合預定促銷的項目，記錄在促銷目標資料表並暫存於主記憶體，如表 5 所示。

表 5：促銷目標資料表

會員編號	符合欲促銷商品項目的購買記錄
M1	001, 003 ² , 004 ²
M2	002 ² , 004
M3	001 ² , 002, 003 ² , 005
M4	002, 004 ²
M5	002, 003, 005

註： y^x 表示編號 y 的項目，購買數量為 x

2. 發掘高頻項目集

本研究可依實際應用之需求，選擇是否加入頻率、數量與高頻項目集之最小長度做為參考指標，而各參考指標之計算方式如下。

(1) 計算頻率指標(Support Level)－根據表 5 之促銷目標資料表，統計符合預定促銷商品項目之記錄所包含的會員數量，再計算各項目集佔全部交易記錄的頻率，如表 6 所示。例如：項目集{002, 004}之頻率指標為 40%，計算方式為購買記錄包含項目集{002, 004}之會員有 2 位，除以總會員數 5 人，即為該項目集的頻率。

表 6：商品組合所包含的會員數統計

候選項目集編號	商品組合	會員數量
1	001, 003, 004	1
2	002, 004	2
3	001, 002, 003, 005	1
4	002, 003, 005	1

(2) 計算數量指標(Amount Level)－根據表 5 之促銷目標資料表，統計會員購買各商品項目的數量，如表 7 所示。例如：項目集{002, 004}在數量上的比例為 30%，計算方式為購買記錄包含項目集{002, 004}之會員有 2 位，並加總此二位會員購買 002 之數量與 004 之數量，再除以所有項目分別在全部交易記錄中出現之總次數，即為該項目集之數量比率。

表 7：會員購買各商品項目的數量統計

會員編號	商品項目					合計
	001	002	003	004	005	
M1	1		2	2		5
M2		2		1		3
M3	2	1	2		1	6
M4		1		2		3
M5		1	1		1	3
數量統計	3	5	5	5	2	20

(3)高頻項目集之最小長度—決策者會考量促銷商品項目的類型與價格等因素來設定高頻項目集之最小長度。透過此指標可快速推演超過所設定最小長度之高頻項目集，避免浪費時間發掘不符合需求的高頻項目集。

本研究範例設定頻率指標為 40%、數量指標為 30%及高頻項目集之最小長度為 2，從促銷目標資料表發掘符合條件的高頻項目集，如表 8 所示。

表 8：發掘促銷目標資料表中符合條件的高頻項目集

候選項目集編號	商品組合	頻率指標 = 40%	數量指標 = 30%	是否符合高頻項目集
1	001, 003, 004	20%	25%	不符合
2	002, 004	40%	30%	符合
3	001, 002, 003, 005	20%	30%	不符合
4	002, 003, 005	20%	15%	不符合

設定高頻項目集的最小長度為 2、頻率指標為 40%與數量指標為 30%之條件下，分解表 8 中最長但不符合高頻項目集條件的候選項目集為子集合，再合併其他不符合高頻項目集條件但長度相同的候選項目集，並由合併之候選項目集中發掘符合條件之高頻項目集。本範例中，{001, 002, 003, 005}為最長之不符合高頻項目集條件的候選項目集，將其分解為子集合{001, 002, 003}、{001, 002, 005}、{001, 003, 005}、{002, 003, 005}，並合併此四個子集合到長度 3 之候選項目集，如編號 1~4 之候選項目集分別為{001, 003, 004}、{002, 003, 005}，再從中發掘符合條件之高頻項目集為{002, 003, 005}，如表 9(a)、9(b)所示。

表 9(a)：分解編號 3 候選項目集為子集合

候選項目集與其編號		候選項目集的子集合與其編號	
3	001, 002, 003, 005	3a	001, 002, 003
		3b	001, 002, 005
		3c	001, 003, 005
		3d	002, 003, 005

表 9(b)：合併表 9(a)與表 8 中長度為 3 之候選項目集

候選項目集編號	合併後之候選項目集	頻率指標 = 40%	數量指標 = 30%	是否符合高頻項目集
3a	001, 002, 003	20%	25%	不符合
3b	001, 002, 005	20%	20%	不符合
3c	001, 003, 005	20%	25%	不符合
3d, 4	002, 003, 005	40%	35%	符合
1	001, 003, 004	20%	25%	不符合

重複上述之步驟，繼續對其他不符合高頻項目集條件的最長候選項目集進行分解，以發掘符合條件的高頻項目集，直到無法滿足所設定的高頻項目集之最小長度為止。表 10(a)為分解表 9(b)編號 1, 3a, 3b, 3c 之不符合高頻項目集條件的候選項目集。表 10(b)為合併並

發掘符合的高頻項目集。

3. 彙整符合條件之高頻項目集

因本範例所訂定之高頻項目集的最小長度為 2，因此，分解到表 10(b)即停止，並綜合表 8、表 9(b)及表 10(b)符合條件的高頻項目集為 {001, 003}、{002, 004} 及 {002, 003, 005}。

表 10(a)：分解編號 1、3a、3b、3c 候選項目集的子集合

候選項目集編號	候選項目集	候選項目集的子集合
1	001, 003, 004	001, 003
		001, 004
		003, 004
3a	001, 002, 003	001, 002
		001, 003
		002, 003
3b	001, 002, 005	001, 002
		001, 005
		002, 005
3c	001, 003, 005	001, 003
		001, 005
		003, 005

表 10(b)：發掘長度 2 的高頻項目集

候選項目集 編號	候選項目集的子集合	頻率指標 = 40%	數量指標 = 30%	是否符合高頻項目 集
3a, 3b	001, 002	20%	15%	不符合
1, 3a, 3c	001, 003	40%	35%	符合
1	001, 004	20%	15%	不符合
3b, 3c	001, 005	20%	15%	不符合
3a	002, 003	40%	25%	不符合
3b	002, 005	40%	20%	不符合
1	003, 004	20%	20%	不符合
3c	003, 005	40%	25%	不符合

(四) UCR 推導關聯規則

FBAR 發掘之高頻項目集轉換為類別，如表 11 所示。

表 11：高頻項目集的類別編號

高頻項目集	類別
{001, 003}	T1
{002, 004}	T2
{002, 003, 005}	T3

對應會員購買商品組合至高頻項目集類別，如表 12 所示，對應原則為：(1)若會員購買的商品組合為符合條件之高頻項目集，則直接對應，例如編號 M2 與 M4 會員的購買商品組合為 {002, 004}，則對應到類別 T2；(2)若會員購買商品組合歸納到多個長度不相等的高頻項目集且有項目重疊時，基於找出最長的高頻項目集為目標，選擇最長的高頻項目集為基準作對應，例如編號 M3 會員的購買商品組合為 {001, 002, 003, 005}，符合對應的高頻項目集為 {001, 003}、{002, 003, 005}，重疊項目為 003 但長度不相等，選擇最長的高頻項目集 {002, 003, 005} 對應，因此 M3 購買的商品組合對應到類別 T3。

表 12：對應商品組合至高頻項目集類別

會員編號	高頻項目集類別
M1	{T1}
M2	{T2}
M3	{T3}
M4	{T2}
M5	{T3}

表 4 中一般化會員基本資料之相同資料屬性的會員記錄合併，並轉換會員編號為會員類型代號 C1、C2、C3... 等，如表 13(a) 所示。合併表 12 對應的高頻項目集之類別，如表 13(b) 所示。歸納記錄表 13(a)、13(b) 中屬性值相同的會員類型，如表 14 所示。計算年齡、每月收入、性別與高頻項目集類別之 SAD 值 (Yen & Chen, 1996)，如表 15 所示。

表 13(a)：一般化後的會員類型

會員類型	年齡	每月收入	性別
C1	老年	中等收入	女
C2	青年	低收入	男
C3	中年	高收入	女
C4	老年	高收入	男

表 13(b)、高頻項目集類別合併

會員類型	高頻項目集類別
C1	{T1, T3}
C2	{T2}
C3	{T3}
C4	{T2}

表 14：年齡、每月收入、性別、高頻項目集類別屬性值相同者的會員類型

年齡		每月收入		性別		高頻項目集類別	
青年	{C2}	低收入	{C2}	女	{C1, C3}	{T1, T3}	{C1}
中年	{C3}	中等收入	{C1}	男	{C2, C4}	{T2}	{C2, C4}
老年	{C1, C4}	高收入	{C3, C4}	—	—	{T3}	{C3}

表 15：屬性與高頻項目集類別的 SAD 值計算

屬性	高頻項目集類別			屬性	高頻項目集類別			屬性	高頻項目集類別		
年齡	{T1, T3}	{T2}	{T3}	月收入	{T1, T3}	{T2}	{T3}	性別	{T1, T3}	{T2}	{T3}
青年	0	1	0	低收入	0	1	0	女	1/2	0	1/2
中年	0	0	1	中等收入	1	0	0	男	0	1	0
老年	1/2	1/2	0	高收入	0	1/2	1/2	—	—	—	—

規則 1、2、3、4 與 5 之會員類型分別為 {C2}、{C3}、{C2}、{C1} 及 {C2, C4}，而會員類型 {C2} 符合規則 1、3 及 5，則選擇涵蓋範圍較大的規則 5，並將規則 1、3 刪除，保留規則 2、4 與 5。若推導的規則尚未能涵蓋所有記錄，則繼續計算條件屬性組合與目標屬性的 SAD 值，以推導其他符合條件的規則，直到條件屬性組合包括所有的屬性為止。規則 2、4 與 5 能完全涵蓋所有交易紀錄，因此將所產生的規則對應高頻項目集，最後推導之規則如表 16 所示。

表 16：推導結果之規則內容

	規則內容	
規則 2	IF 年齡 = 中年	THEN {002, 003, 005}
規則 4	IF 月收入 = 中等收入	THEN {{001, 003} OR {002, 003, 005}}
規則 5	IF 性別 = 男	THEN {002, 005}

肆、快速反向關聯法則(Fast-Backward Association Rule, FBAR)

一、快速反向關聯法則演算法

FBAR 演算法中所使用的符號說明，如表 17 所示；FBAR 演算法如圖 3 所示。

表 17 : FBAR 演算法符號說明

RT_2	會員交易資料記錄關聯資料表
N	RT_2 中所包含會員的數量
M_i	第 i 個會員的交易記錄
m_{ij}	第 i 個會員的交易記錄之第 j 個項目
PID	欲促銷之目標商品項目編號
DCB_i	第 i 個會員交易記錄項目中符合促銷項目之項目集(集合內不含相同元素)
$SDCB_i$	分解 DCB_i 為含有 k 個元素的 subsets 所形成的集合
item_diff_count[i]	DCB_i 集合內之不同元素總數
item_count[i]	DCB_i 集合內之所有元素總數 (計算第 i 個會員交易記錄中符合欲促銷之項目數量)
min_support	設訂最小頻率指標
min_amount	設訂最小數量指標
min_item_num	設訂高頻項目集的最小長度

Algorithm Fast_Backward Association Rule(RT_2 , PID, min_item_num, min_support, min_amount)

Input: RT_2 , PID, min_item_num, min_support, min_amount

Output: Large_Itemset (高頻項目集之集合)

Begin

- 1) set PID; (設定預定促銷的目標項目集 PID)
- 2) set min_support; (設定項目組合的最小頻率)
- 3) set min_amount; (設定最小數量)
- 4) set min_item_num; (設定高頻項目集之最小長度)
- 5) **Primary_Bundle**(PID, min_item_num); (從會員交易記錄找出符合預定促銷項目之最初候選項目集)
- 6) **First_Candidate** (Candidate, min_support, min_amount);
- 7) **for** ($i=|PID|$; $i > min_item_num$; $i--$) **do** {
- 8) Tmp_Candidate = \emptyset ;
- 9) **for** ($j=0$; $j < N$; $j++$) **do** {
- 10) $SDCB_j = Decomposition(DCB_j, i-1)$;
- 11) put $SDCB_j$ into Tmp_Candidate;
- 12) }
- 13) **First_Candidate** (Tmp_Candidate, min_support, min_amount);
- 16) }

End

圖 3 : FBAR 演算法

(一) 快速反向關聯法則演算法說明：

FBAR 演算法主要包含四個部份，說明如後。

1. 發掘促銷項目集的前置作業

首先根據實際應用之需求，選擇性加入最小頻率、最小數量與高頻項目集之最小長度做為參考指標，設定所選擇之參考指標的數值；並設定欲促銷的商品項目 PID。

2. Primary_Bundle()為發掘會員交易記錄中符合欲促銷項目之候選項目集 Candidate

首先檢視每一個會員交易記錄中的項目編號是否包含於 PID，分別記錄會員交易中符合 PID 項目的數量 item_count 及不同項目的個數 item_diff_count。若 item_diff_count 小於 min_item_num，表示會員 M_i 的交易記錄長度小於設定高頻項目集的最小長度，則將交易記錄 M_i 從資料表 RT_2 中刪除，否則會員 M_i 交易記錄為候選項目集 DCB_i ，收集所有候選項目集 $DCBi$ 成為候選項目集之集合 Candidate，演算法如圖 4 所示。

Algorithm Primary_Bundle(PID, min_item_num)

Input: PID, min_item_num

Output: Candidate

Begin

- 1) set Answer = \emptyset ;
- 2) total_count = 0;
- 3) for ($i=0; i < N; i++$) do {
 - 4) $DCB_i = \emptyset$;
 - 5) item_diff_count[i] = 0;
 - 6) item_count[i] = 0;
 - 7) for ($j=0; j < n_i; j++$) do { $(n_i$ 為第 i 個會員的交易記錄之項目數量)
 - 8) if ($m_{ij} \in PID$) then { $($ 檢視交易項目是否包含在 PID 中 $)$
 - 9) item_count[i]++; $($ 交易記錄中符合 PID 的項目數量 $)$
 - 10) if ($m_{ij} \neq m_{ij-1}$) then {
 - 11) item_diff_count[i]++; $($ item_count 中不同項目的個數 $)$
 - 12) put m_{ij} into DCB_i ;
 - 13) }
 - 14) }
 - 15) }
 - 16) total = total + item_count[i];
 - 17) if (item_diff_count[i] < min_item_num) then
 - 18) $DCB_i = \emptyset$; $($ 低於所設定的高頻項目集之最小長度 $)$
 - 19) else put DCB_i into Candidate; $($ 將候選項目集儲存於 Candidate $)$

End

圖 4：發掘符合欲促銷項目之演算法

3. First_Candidate()從候選項目集中發掘符合條件的高頻項目集

由 Candidate 中發掘高頻項目集之集合 Large_Itemset。首先檢視候選項目集之集合 Candidate 中的候選項目集 DCB_i，並記錄各 DCB_i 在 Candidate 中出現的次數，再判斷各候選項目集 DCB_i 是否滿足所設定項目集的頻率指標 min_support 及數量指標 min_amount，並合併滿足之候選項目集 DCB_i 置於高頻項目集之集合 Large_Itemset，而其它不滿足的候選項目集 DCB_i，將進一步分解並合併以發掘長度較短但符合條件的高頻項目集，演算法如圖 5 所示。

Algorithm First_Candidate(Candidate, min_support, min_amount)

(從最初候選項目集中找出符合的項目組合)

Input: Candidate, min_support, min_amount

Output: Large_Itemset

Begin

```

1)  for (i=0; i < | Candidate |; i++) do
2)      DCBi.support_count=0;
3)  for (i=0; i < | Candidate |; i++) do {
4)      if (DCBi.support_count=0) then {
5)          tmp = ∅;
6)          for (j=i+1; j < | Candidate |; j++) do {
7)              m = 0;
8)              while ( m < | DCBi|) do {
9)                  if (DCBi[m]=DCBj[m]) then
10)                      m++;
11)                  else
12)                      m = | DCBi |+1;
13)              }
14)              if ( m = | DCBi | ) then {
15)                  DCBi.support_count++;
16)                  put j into tmp;
17)              }
18)          }
19)          while ( tmp ≠ ∅ ) do {
20)              Delete k from tmp;
21)              DCBk.support_count=DCBi.support_count;
22)          }
23)      }
24)  }
25)  for (i=0; i < | Candidate |; i++) do {
26)      frequency[i]= item_count[i]/total_count;
27)  if (DCBi.support_count≥min_support & frequency[i] ≥min_amount) then {
        (檢視 DCBi 是否符合所設定項目組合的頻率指標及數量指標)
28)      put DCBi into Large_Itemset;
29)      set DCBi=∅ in Candidate;
30)  }
31) }
```

End

圖 5：產生候選項目集之演算法

4. 分解並合併候選項目集以發掘長度較短但符合條件的高頻項目集

`Decomposition()`函數如圖 6 所示，分解不符合條件之最長候選項目集，再合併其它具有相同長度之不符合條件的候選項目集 DCB_i ，組成次候選項目集 $Tmp_Candidate$ 。`First_Candidate()`函數將從次候選項目集中，發掘長度較短的高頻項目集。重複此步驟，持續分解其它不符合條件的最長候選項目集以發掘高頻項目集，直到候選項目集之長度小於訂定的高頻項目集之最小長度 \min_item_num 為止。

```

Algorithm Decomposition(S, k)
Input: S, k
Output: Subset
Begin
    1)Subset = { all subsets of S contain
        k elements };
        (將不符合條件的候選項目集分
        解成長度為 k 的子集合)
    End

```

圖 6：分解候選項目集之演算法

二、快速反向關聯法則之模擬

本研究之程式開發工具為 Microsoft Visual Basic 6.0，模擬用的測試資料為 Microsoft SQL Server 2000 所附之 FoodMart 零售市場交易資料庫，在資料庫中分別隨機抽取 5000, 10000, 15000, 20000, 25000 筆交易紀錄進行模擬。本實驗並不考慮數量指標與高頻項目集之最小長度兩參考指標，僅考慮最小支持度並設定為 0.36%、0.38%、0.40% 的情況下，並進行比較 FBAR 與 Apriori 關聯法則的執行效率，實驗結果如圖 7、8、9 所示。FBAR 在不同資料庫 Record 數量與不同最小支持度條件下，執行效能皆優於 Apriori。當最小支持度值越低時，由於其所發掘之高頻項目集長度越長，越能顯現 FBAR 之效能優於 Apriori。

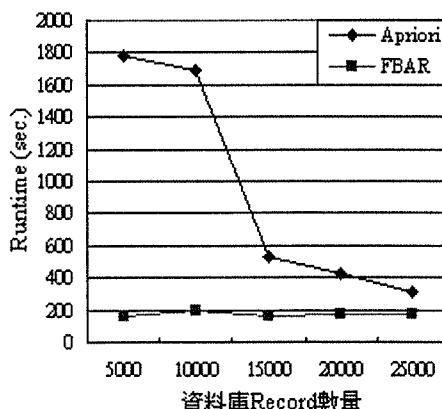


圖 7：最小支持度為 0.36% 之實驗結果

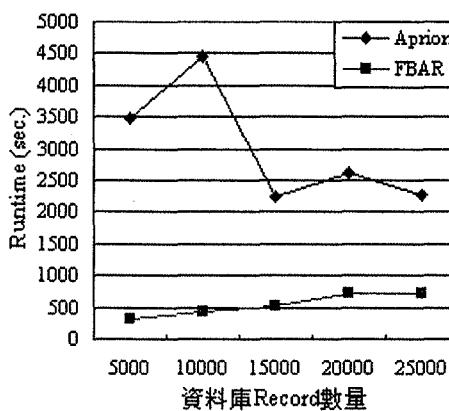


圖 8：最小支持度為 0.38%之實驗結果

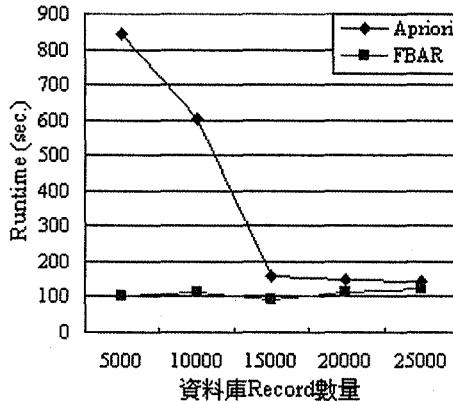


圖 9：最小支持度為 0.40%之實驗結果

伍、調整緊密規則演算法(Update-Compact Rule, UCR)

Yen & Chen(1996)所提出的緊密規則，其推導過程需將相關資料表作關聯結合，在一般化處理時，導致記錄之條件屬性值相同但目標屬性值卻不同之情況。本研究調整緊密規則避免將相關的資料表作關聯結合，以提升資料探勘的效率，並應用於歸納會員類型。

一、調整緊密規則演算法

UCR 所使用之符號說明，如表 18 所示，演算法如圖 10 所示。

表 18：UCR 演算法符號說明

Large Itemset	從所有會員交易記錄中找出之高頻項目集
RT_1	會員個人資料記錄關聯資料表
R_i	對應到 RT_1 及 RT_2 的記錄編號 i
CA_j	RT_1 中第 j 個屬性
m	RT_1 中屬性數目
TA	RT_2 中目標屬性
t_i	RT_2 中 R_i 所對應的目標屬性值

Algorithm UCR(Large_Itemset, RT₁, RT₂)

Input: Large_Itemset, RT₁, RT₂

Output: Derived Rules

Begin

- 1) set each bundle in Large_Itemset which has independent type is contained in TA
- 2) **foreach** record $R_i \in RT_2$ **do** {
- 3) $t_i =$ assemble the fit type from TA for all items in R_i
- 4) }
- 5) **foreach** record $R_i \in RT_1$ **do** {
- 6) **if** (record R_i and the others have the same value in all attributes) **then** {
- 7) retain one record in all identical records and combine t_i with the target
- 8) attribute of the others
- 9) }
- 10) }
- 11) **for** x = 1 to m **do**{
- 12) **foreach** combination $CA_x, CA_{x+1}, \dots, CA_y$ selected from CA_1, CA_2, \dots, CA_m **do** {
- 13) **derive_rules**($CA_x, CA_{x+1}, \dots, CA_y, TA$); (Yen & Chen, 1996)
- 14) }
- 15) **if** (all record $R_i \in RT_1$ characterized by derived rule) **then** {
- 16) terminate
- 17) }
- 18) }

End

圖 10：UCR 演算法

(一)演算法分為三個主要部分說明：

1. 對應會員購買項目至高頻項目集類別

標示類別於 FBAR 所發掘之高頻項目集，並將 RT_2 會員購買記錄對應至高頻項目集類別，以找出合適的類型組合 t_i 。

2. 資料記錄合併

若會員個人資料記錄關聯資料表 RT_1 之某些記錄在各屬性中有相同屬性值，則表示這些記錄相同，因而僅需保留一筆記錄，將其餘相同的記錄刪除，同時以此記錄之各屬性對應會員交易記錄關聯資料表 RT_2 ，並個別合併記錄之屬性與目標屬性。

3. 計算 SAD 以推導關聯規則

使用 Yen & Chen(1996)所設計之 $\text{derive_rules}(CA_x, CA_{x+1}, \dots, CA_y, TA)$ ，在第一循環以單一條件屬性計算 SAD 值，找出會員循序性購買項目的知識規則；在第二循環，以 2 個條件組合發掘會員循序性購買項目的知識規則；依此類推，在第 m 個循環，以 m 個條件組合發掘會員循序性購買項目的知識規則。

依據緊密規則之定義，所有產生的規則必須具有完全可靠度(Full Confidence)，亦即規則之 SAD 值必須為 1。在所發掘的規則中，若有規則的涵蓋資料記錄包含於其它規則所涵蓋的資料記錄，則以涵蓋資料記錄較多者為所求，且刪除涵蓋資料記錄較少之規則。若條件組合與目標屬性的 SAD 值為 0，則刪除其條件組合，以減少運算時間，當所發掘的規則已經涵蓋所有的記錄，則結束規則之推導。所推導之規則即為會員購買項目組合的知識規則。

二、知識規則的評估模式與比較

(一) 知識規則的評估模式

根據 Yen & Chen(1996)提出之規則評估模式，以評估本研究推導會員購買項目組合之知識規則。依據目標屬性以計算推導規則之緊密程度，三個主要的評估方向為：(1)推導規則愈少，緊密程度愈高；(2)推導規則涵蓋的資料記錄愈多，緊密程度愈高；(3)推導規則包含的條件屬性愈少，緊密程度愈高。若推導的規則數量愈少，每個規則涵蓋的記錄數量愈多，則緊密程度與 E 值愈高，規則愈具有代表性。涵蓋記錄比率為規則涵蓋記錄占所有記錄的比率，代表知識規則是否完整涵蓋所有的記錄。評估模式如下所示：

$$E = \frac{1}{r} \sum_{i=1}^r \frac{t_i}{n_i} \times \frac{1}{C_i}, \quad 1 \leq i \leq r$$

r：推導規則數量

t_i ：第 i 個推導規則涵蓋的記錄數量

n_i ：第 i 個推導規則其目標屬性值涵蓋的記錄數量

C_i ：第 i 個推導規則其條件屬性的數量

(二) UCR、緊密規則與 ID3 之比較

藉由實例以瞭解本研究之 UCR 規則推導的效率與可行性，除與 Yen & Chen(1996)提出的緊密規則進行比較，緊密規則推導結果如表 19 所示，並與應用非常廣泛且相當知名的 ID3 分類法(Quinlan, 1986)進行比較，ID3 分類法推導結果如表 20 所示，UCR 推導結果如表 21 所示，三者推導規則方法之評估模式結果如表 22 所示。UCR 所得之 E 值及涵蓋記錄比率最高，表示推導規則的涵蓋記錄數量較多、較完整與緊密程度較高，則規則較具有代表性。由實驗結果可知，在緊密規則之推導過程，若將相關資料表作關聯結合，則進行一般化處理時，導致記錄之條件屬性值相同但目標屬性值卻不同。因此，UCR 演算法避免將相關的資料表作關聯結合，而提升資料探勘的效率。

表 19：緊密規則推導結果

規則 1	IF 年齡=中年 THEN {T3}	(會員編號：M3)
規則 2	IF 性別=男 THEN {T2}	(會員編號：M2, M4)
知識規則之模式評估值 $E = 1/2 [(1/2 * 1/1) + (2/2 * 1/1)] = 0.75$		
涵蓋記錄比率 = 0.6		

表 20：ID3 分類法推導結果

規則 1	IF 月收入=低收入 THEN {T2}	(會員編號 : M2)
規則 2	IF 月收入=高收入 AND 性別=女 THEN {T3}	(會員編號 : M3)
規則 3	IF 月收入=高收入 AND 性別=男 THEN {T2}	(會員編號 : M4)
知識規則之模式評估值 $E = 1/3 [(1/2 * 1/1) + (1/2 * 1/2) + (1/2 * 1/2)] = 0.333$		
涵蓋記錄比率 = 0.6		

表 21：調整緊密規則法(UCR)推導結果

規則 2	IF 年齡 = 中年 THEN {T3}	(會員類型 : C3)
規則 4	IF 月收入 = 中等收入 THEN {T1,T3}	(會員類型 : C1)
規則 5	IF 性別 = 男 THEN {T2}	(會員類型 : C2, C4)
知識規則之模式評估值 $E = 1/3 [(1/1 * 1/1) + (1/1 * 1/1) + (2/2 * 1/1)] = 1$		
涵蓋記錄比率 = 1		

表 22：推導規則方法之評估模式的結果

	緊密規則	ID3 分類法	UCR
模式評估值(E)	0.58	0.333	1
涵蓋記錄比率	0.6	0.6	1

六、結論

本研究提出新的快速反向關聯法則(Fast-Backward Association Rule, FBAR)，以快速發掘高頻項目集，其執行程序反向於 Apriori 關聯法則，執行步驟為：(1)建立促銷目標資料表—掃描交易資料庫一次，將交易資料庫中不符合預定促銷的項目刪除，而保留符合預定促銷的項目，記錄在促銷目標資料表並暫存於主記憶體；(2)分解促銷目標資料表之交易資料—在促銷目標資料表中，由最長交易資料開始逐層分解項目集；(3)發掘符合最小支持度之高頻項目集—當分解至某長度之項目集且已符合最小支持度，則停止該項目集之分解。

FBAR 僅需掃描資料庫一次，將刪減之資料記錄在促銷目標資料表並暫存於主記憶體，再透過分解較長交易紀錄，可快速發掘符合條件之高頻項目集。模擬結果顯示 FBAR 與 Apriori 關聯法則所產生的高頻項目集是相同，但 FBAR 有較高之執行效率。

在知識規則推導上，經 FBAR 發掘高頻項目集後，藉由調整緊密規則演算法，利用會員交易記錄及會員個人資料，發掘客戶特徵，避免將相關的資料表作關聯結合，推導涵蓋面較完整的規則，建立目標行銷的區隔市場。透過實例分別使用 UCR、緊密規則法與 ID3 分類法推導知識規則，顯示 UCR 所推導的規則涵蓋率及緊密程度高於緊密規則與 ID3 分類法。

參考文獻

- 1.方世榮譯，Philip Kotler 著，1996，行銷管理學—分析、計劃、執行、與控制，第二版，台北，東華書局。
- 2.沈維明、翁頌舜，1999，「資料挖掘之關聯式法則架構以零售業目標行銷為例」，中華民國科技管理論文集。
- 3.張希誠編譯，1987，行銷實務：掌握企業行銷的竅門書，台北：書泉出版社。
- 4.蔡秀滿、陳健民，1999，「多重資料來源之消費特徵分析」，第五屆資訊管理暨實務研討會。
- 5.魏尚敬，1992，台灣市場環境行銷管理，中國文化大學出版部。
- 6.蕭富峰，1993，如何進行促銷，台北：遠流出版社。
- 7.Agrawal, R., Imilienski, T. and Swami, A., "Mining Association Rules between Sets of Items in Large Database," In Proc. of the ACM SIGMOD International Conference on Management of Data, 1993, pp.207-216.
- 8.Agrawal, R. and Srikant, R., "Fast Algorithms for Mining Association Rules," Proc. of the 20th Int'l Conference on Very Large Databases, Sept. 1994. Expanded version available as IBM Research Report RJ9839, June 1994.
- 9.Agrawal, R. and Srikant, R., "Mining Sequential Patterns," Proc. of the Int'l Conference on Data Engineering(ICDE), Taipei, Taiwan, March 1995. Expanded version available as IBM Research Report RJ9910, October 1994.
- 10.Berzal, F., Cubero, J.C., Marin, N. and Serrano, J.M., "TBAR: An efficient method for association rule mining in relational databases," Elserier, Data & Knowledge, Engineering 37, 2000, pp. 47-64.
- 11.Berry, M. and Linoff, G., Data Mining Techniques: For Marketing, Sales, and Customer Support, New York: John Wiley and Sons, 1997.
- 12.Brin, S., Motwani, R. and Silverstein C. "Beyond Market Baskets: Generalizing Association Rules to Correlations," In Proc. of ACM SIGMOD International Conference on Management of Data, 1997, pp. 265-276.
- 13.Brin, S., Motwani, R., Ullman, J.D. and Tsur, S., "Dynamic Itemset Counting and Implication Rules for Market Basket Data", in Proc. of ACM SIGMOD International Conference on Management of Data, 1997, pp. 255-264.
- 14.Cabena, P. et. al., Discovering Data Mining: From Concept to Implementation, Prentice Hall, 1997.
- 15.Calantone, R. J. and Sawyer, A. G., "The stability of benefit segments," Journal of

- Marketing Research, vol. 15, Aug. 1978, pp. 395-404.
- 16.Carter, C., Hamilton, H. and Cercone, N., "Share Based Measures for Itemsets," Principles of Data Mining and Knowledge Discovery, Komorowski, J. and Zytkow, J. (Eds.), 1997, pp. 14-24.
- 17.Chen, M. S., Han, J. and Yu, P. S., "Data Mining: An Overview from Database Perspective," IEEE Transaction on Knowledge and Data Engineering, 1997, pp.866-883.
- 18.Dunkel, B. and Soparkar, N., "Data Organization and Access for Efficient Data Mining," ICDE, Sydney, Australia, 1999.
- 19.Fu, Y., Data Mining: Tasks, Techniques, and Applications, IEEE Potentials, Vol. 16, No. 4, 1997, pp. 18-20.
- 20.Han, J., "Characteristic Rules," in Kloesgen, W. and Zytkow, J. eds., Handbook of Data Mining and Knowledge Discovery, Oxford University Press, 1999.
- 21.Han, J., Cai, Y. and Cercone, N., "Knowledge Discovery in Databases: An Attribute-Oriented Approach," Proc. of 1992 Int'l Conf. on Very Large Data Bases (VLDB'92), Vancouver, Canada, August 1992, pp. 547-559.
- 22.Han, J. and Fu, Y., "Exploration of the Power of Attribute-Oriented Induction in Data Mining," in Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. eds., Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996, pp.399-421.
- 23.Han, J. and Fu, Y., "Discovery of Multiple-Level Association Rules from Large Databases," Int. Conf. Very Large Databases, Sept. 1995, pp. 400-431.
- 24.Han, J. and Kamber, M., Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2000.
- 25.Han, J., Pei, J. and Yin, Y., "Mining Frequent Patterns without Candidate Generation," Proc. of the ACM SIGMOD on management of data, 2000, pp. 1-12.
- 26.Liu, B., Hsu, W. and Ma, Y., "Mining Association Rules with Multiple Minimum Supports," the Proc. of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1999.
- 27.Mannila, H. and Ronkainen, P., "Similarity of Event Sequences," In Proceedings of the Fourth International Workshop on Temporal Representation and Reasoning (TIME'97), 10th - 11th May, 1997, pp. 136-139.
- 28.Pork, J.S., Chen, M.S. and Yu, P.S., "An Effective Hash Based Algorithm for Mining Association Rules," ACM SIGMOD, May 1995, pp. 175-186.
- 29.Quinlan, J. R., Induction of Decision Tree, Machine Learning, 1986, pp. 81-106.

- 30.Savasere, A., Omiecinski, E., Navathe, S., "An efficient algorithm for mining association rules in large databases," In Proc. of VLDB 95, 1995.
- 31.Srikant, R. and Agrawal, R., "Mining Generalized Association Rules," In Proc. of the 21st Int'l Conference on VLDB, 1995.
- 32.Toivonen, H., "Sampling large databases for association rules," In 22nd VLDB Conference, 1996, pp.134-145.
- 33.Yen, S.J. and Chen, A.L.P., "The Analysis of Relationships in Databases for Rule Derivation," Journal of Intelligent Information Systems, 1996, pp. 1-25.