

利用分類分析發掘消費者最適性之 產品項目

陳垂呈*

(收稿日期：96 年 3 月 13 日；第一次修正：96 年 4 月 10 日；
第二次修正：96 年 9 月 3 日；接受刊登日期：96 年 10 月 26 日)

摘要

在本論文中，我們以消費者之交易資料為探勘的資料來源，每一筆交易資料除了記錄有消費者曾經購買過的產品項目，也記錄著其購買的次序性，以某一消費者為探勘的目標，利用分類分析 (classification analysis) 分別從以下兩方面發掘此消費者最適性的產品項目：一是只考量產品項目是否出現在交易資料中，文中設計一個建構決策樹的方法，以顯示那些產品項目與此消費者之間有關聯性的傾向特徵，藉此做為發掘此消費者最適性之產品項目的依據；二是考量產品項目具有購買次序性，文中設計一個建構決策樹的方法，以顯示那些產品項目與此消費者之間具有購買次序的關聯性傾向特徵，藉此做為發掘具有購買次序之此消費者最適性的產品項目的依據。我們根據所提出的方法，設計與建置一個發掘消費者最適性之產品項目的探勘系統。

關鍵詞彙：交易資料，分類分析，次序性

壹· 簡介

隨著資訊技術迅速的發展，企業可以更加輕易及方便儲存消費者曾經交易的記錄 (Laudon & Laudon, 2000)。這些交易記錄可能來自於消費者的信用卡交易、大賣場的收銀機、填寫的特徵資料或是網頁的瀏覽記錄等。因此，如何利用這些大量交易記錄，深入分析消費者的交易行為，以改善與顧客之間的關係，並提供最貼切的產品服務，進而提昇顧客的滿意度與忠誠度，是企業經營者必須思考的問題之一。

資料探勘 (data mining) 是從大量資料中找出有用的資訊與知識，目前已廣泛應用在許多的領域中 (Han & Kamber, 2006)，並已被證明可以有效應用在產品行銷、銷售及顧客服務上，是企業提昇經營優勢與競爭力的重要工具之一 (Berry & Linoff, 2004; Hui & Jha, 2000)。在本論文中，我們以消費者之交易資

* 作者簡介：陳垂呈，南台科技大學資訊管理系副教授。

料為探勘的資料來源，每一筆交易資料記錄有消費者曾經購買的產品項目及其購買的次序，並以某一消費者為探勘的目標，利用分類分析 (classification analysis) 分別從以下兩方面發掘此消費者最適性的產品項目：

- 1. 只考量產品項目是否出現在交易資料中：**假設此消費者的交易資料為 X ， X 為一個或以上產品項目所形成的項目組，在探勘的過程中，若交易資料包含 X ，則設定交易資料與此消費者之間的關聯性為「有」；否則設定其關聯性為「無」。我們視此消費者未曾購買過的产品項目為欲分類的影響屬性，然後利用 ID3 演算法對交易資料進行分類計算，並建立決策樹。從決策樹的路徑 (paths) 中，可找出那些屬性項目的狀態會與此消費者之間的關聯性為「有」，即定義其中狀態是「曾經購買」的屬性項目，為此消費者最適性的產品項目。
- 2. 考量產品項目具有購買次序性：**假設此消費者的交易資料為 Y ， Y 為一個或以上產品項目所形成具有次序性的項目組，在探勘的過程中，若交易資料包含 Z ， Z 為一個或以上產品項目所形成具有次序性的項目組，當 Z 包含的產品項目等於 Y 包含的產品項目，則設定交易資料與此消費者之間的關聯性為「有」；否則設定其關聯性為「無」。若交易資料包含 Z ，則必須依序分解 Z 之後的產品項目成各項目組，我們視分解後的項目組為欲分類的影響屬性，利用 ID3 演算法對交易資料進行分類計算，並建立決策樹。從決策樹的路徑中，可找出那些屬性項目的狀態會與此消費者之間的關聯性為「有」，我們即定義其中狀態是「曾經購買」的屬性項目，為具有購買次序之此消費者最適性的產品項目。

我們根據所提出的方法，設計與建置一個探勘系統，以發掘消費者最適性的產品項目。本文各章節介紹如下：第貳節介紹資料探勘技術、及其產品行銷管理的相關研究；第參節以某一消費者為探勘的目標，說明利用分類分析發掘此消費者最適性之產品項目的探勘過程；第肆節以某一消費者為探勘的目標，說明利用分類分析發掘具有購買次序之此消費者最適性的產品項目的探勘過程；第伍節根據所提出的方法，設計與建置一個探勘系統，以發掘消費者最適性的產品項目；最後在第陸節中做一結論。

貳· 相關研究

資料探勘是從大量資料中挖掘出潛在有用的資訊與知識，發現專家尚且未知的新關係，以提供企業專業人員的決策參考。資料探勘可完成以下的工作：關聯規則 (association rules)、分類分析 (classification analysis)、分群 (clustering)、次序相關分析 (sequential pattern analysis) 及預測 (prediction) 等 (Chen et al., 1996; Han & Kamber, 2006)，利用資料探勘於行銷決策及市場預測等活動，可以提供非常有價值的參考資訊 (Berry & Linoff, 2004; Hui & Jha, 2000)。

在交易資料庫中，Agrawal et al. (1993) 首先利用關聯規則表現產品項目之間的關聯性，例如項目組 X 與 Y 之間有一關聯規則被表示成 $X \rightarrow Y$ ， X 、 Y 為包含一個或以上項目所形成的項目組，且 $X \cap Y = \emptyset$ ，則關聯規則所顯示出的傾向特徵為：若消費者曾經購買 X ，則也會有購買 Y 的傾向特徵。就消費者的產品行銷而言，可將 Y 之產品項目推薦給曾經購買 X 的消費者。目前，探勘關聯規則及其應用是資料探勘最重要的研究主題之一，已有許多相關研究被陸續提出 (Agrawal & Srikant, 1994; Srikant & Agrawal, 1995; Park et al., 1997; Han et al., 2000, 2004; Tsay & Chiang, 2005; Pei et al., 2006; Xu & Wang, 2006)。

以上探勘產品項目之間關聯性的過程中，會忽略考量項目出現的購買次序性，但在現實消費者購物的交易中，除了記錄購買的產品項目，也會記錄購買的時間。因此，若將消費者購買產品的時間也列入考量，則可發掘消費者購物的次序性。例如，消費者在購買「洗衣機」之後，也會有購買「烘衣機」的傾向。在探討產品項目之間次序性的研究中，次序相關分析是最常被使用來分析消費者購物之次序性的方法之一，其將每一位消費者曾經購物的交易資料，視為一群有次序性之產品項目的集合，次序相關分析的目的就是在擷取最常出現具有次序性的項目組，並且包含最大的項目個數，其相關研究可參考 (Agrawal & Srikant, 1995; Srikant & Agrawal, 1996; Chen et al., 2003; Yu & Chen, 2005; Zaki, 2001; Wang & Wang, 2006; Yun & Leggett, 2006; Tan et al., 2006)。

藉由資料探勘技術所找出的規則，是反映出其當時所探勘之資料庫中資料的傾向特徵，在探勘的過程中，不同的探勘技術需要設定不同參數值。例如找出的關聯規則 $X \rightarrow Y$ 是否有意義或是有用，必須判斷是否符合所設定的「最小支持度」及「最小信賴度」，最小支持度可表示產品 $X \cup Y$ 的市場大小，最小信賴度可表示產品 X 與 Y 之間關聯的強度，因此各參數值的設定，有賴於管理者的經驗、智慧及需求。

至於找出的規則如何有效應用，必須有賴於管理人員擬定相關的行銷策略。例如關聯規則「便當」→「飲料」成立，其顯示的傾向特徵為：若消費者購買便當，則也會有購買飲料的傾向。管理人員擬定的行銷策略可能為：同時購買便當與飲料可打 95 折的優惠以進行促銷；或是進行消費者適性化的產品推薦，管理人員擬定的行銷策略可能為：把飲料放置於便當的旁邊，以方便消費者拿取，叮嚀門市人員當有人購買便當時，詢問其是否要購買飲料，因為這些消費者是最有可能會購買的。

分類分析是從已知的物件群中，根據所訂定的屬性條件進行分類，決策樹 (decision trees) 與決策法則 (decision rules) 是分類分析最常被使用的兩種表示法。例如，若對消費者曾經購買過的产品項目進行分類，把消費者對某一产品的購買意願分為「高」與「低」兩種類別，再將消費者之交易資料中其他的產品項目、或是人口特徵項目視為影響屬性，經由分類計算所建構的決策樹中，可得知影響購買此產品之意願高低的關鍵屬性。

資料進行分類分析時，一般可以產生不只一種的分類模式，但期望得到的分類模式是越精簡越好。以決策樹為例，若決策樹的高度 (height) 愈小，其路徑 (paths) 包含的節點數目也較少，則表示可用愈少的屬性能分類出所有物件。因此，一個好的分類技術，應該具有精簡及預測能力佳的特性，目前常被利用的分類技術有 ID3 (Quinlan, 1986; Gaddam et al., 2007; Wu et al., 2006; Liu & Zhou, 2006; Xu, 2005)、CN2 (Clark & Niblett, 1989)、倒傳遞類神經網路 (back-propagation) (Rich & Knight, 1991) 等。

Berry & Linoff (2004) 指出對物件的分類及預測而言，利用分類分析所建構的決策樹是強大且常被使用的工具，由其導出的規則可以很容易被人類瞭解、及用來粹取描述物件類別的模型或是對趨勢的預測。目前分類分析已廣泛應用在許多的領域中，例如信用核證 (credit approval)、醫療診斷 (medical diagnosis)、銷售 (sales)、顧客服務、及選擇性行銷 (selective marketing) 等 (Berry & Linoff, 2004; Han & Kamber, 2006)。因此，若能從交易資料中，利用分類分析建構產品項目之間關聯性的決策樹，對產品銷售及行銷預測必可提供相當有用的資訊。

本研究以消費者之交易資料為探勘的資料來源，每一筆交易資料包含曾經購買的產品項目及其次序，並以某一消費者之交易資料 X 為探勘的目標，探討在分類技術中，是否可以解決此消費者與其他項目之間關聯性的問題，並延伸至包含有購買次序之項目間關聯性的問題。即一是只考量產品項目是否出現在交易資料中，發掘此消費者最適性的產品項目；二是增加考量產品項目的

次序性，發掘具有購買次序之此消費者最適性的產品項目。例如 X 與 Y 之間有一關聯規則 $X \rightarrow Y$ ， X 、 Y 為包含一個或以上項目所形成的項目組，且 $X \cap Y = \emptyset$ ，則關聯規則所顯示出的傾向特徵為：*若消費者曾經購買 X ，則也會有購買 Y 的傾向特徵*。此問題在探勘關聯規則之研究領域中，應是最典型、最重要及最廣泛的問題，而本研究目標即在探討如何利用分類技術有效地解決以上的問題。

參· 發掘消費者最適性之產品項目

在消費者曾經購買過的產品項目中，除了可反映出消費者本身的需求傾向，也可顯示出產品項目之間的關聯性，若能從消費者曾經購買過的產品項目中，找出產品項目之間的關聯性，對企業擬訂消費者個人化的產品行銷策略，必可提供相當有用的參考資訊。此一章節中，我們以消費者之交易資料為探勘的資料來源，每一筆交易資料包含有消費者曾經購買過的產品項目，並以某一消費者為探勘的目標，對於此消費者曾經購買的產品項目，視之為已知的適性化產品，對於未曾購買而有傾向會購買的產品項目，則利用分類分析做為發掘此消費者最適性之產品項目的方法依據。

一、探勘方法

在分類分析幾種較具有代表性的方法中，如 CHAID (Hartigan, 1975)、CART (Breiman et al., 1984)、ID3 (Quinlan, 1986)、C4.5 (Quinlan, 1993)、及 Neural Networks (Widrow et al., 1994) 等，高淑珍 (2004) 曾整理 Katharina & Dirk (1999)、Mak & Munakata (2002)、及 Ohmann et al. (1996) 的研究，比較以上幾種不同的分類方法，指出由 ID3 演算法建構之決策樹所展開的決策規則，其具有較佳的分類預測準確率、解釋能力、及精簡度。高淑珍 (2004) 曾以壽險業為例，利用 ID3 演算法評量顧客回應模式的研究；Xu (2005) 曾利用 ID3 演算法管理顧客對軟體產品的滿意度。因此，本論文將利用 ID3 演算法做為建構決策樹的方法依據，並從決策樹所顯示的傾向特徵，發掘消費者最適性的產品項目。

在 ID3 演算法建構決策樹的過程中，必須選擇最佳的屬性做為節點，以建構出最簡單狀態、或是接近最簡單狀態的決策樹。選擇最佳節點是依據節點所產生的熵值 (entropy) (或期望值) 做為判斷的依據，其計算方式說明如下。

假設 C 為某一物件集合，其物件分屬於 i 個不同類別，以 R_1, R_2, \dots, R_i 表示之，則此物件集合的熵值 $E(C)$ 為：

$$E(C) = - \sum_{k=1}^i P_k \log_2 P_k \quad (1) \text{ (Quinlan, 1986)}$$

i 表示全部類別的數量。

n 表示 C 的物件總數量。

$P_k = (\text{屬於類別 } P_k \text{ 的物件總數量})/n, 1 \leq k \leq i$ 。

設定 $\log_2 0 = 0$ 。

選擇某一屬性 d 做為決策樹的節點，若屬性 d 有 m 個可能的屬性值，則以 d_1, d_2, \dots, d_m 表示之，然後在此節點下建立 m 個子節點，並將原本屬於節點的所有物件，分配至最適當的子節點中。若物件被分配至相同的子節點中，則表示其 d 的屬性值必相同。計算以屬性 d 為節點之子決策樹的熵值 $E(d)$ 為：

$$E(d) = \sum_{j=1}^m |x_j| (n_j/n) \times E(C_j) \quad (2) \text{ (Quinlan, 1986)}$$

C_j 表示屬性值 d_j 的物件子集合， $1 \leq j \leq m$ 。

$E(C_j)$ 表示物件子集合 C_j 的熵值。

n_j 表示物件子集合 C_j 所包含的物件總數量。

計算資訊收益 (information gain) 的公式為：

$$G(d) = E(C) - E(d) \quad (3) \text{ (Quinlan, 1986)}$$

其表示原物件集合的熵值與以屬性 d 為子節點的熵值之間的差值。

根據熵值及資訊收益，以下說明利用 ID3 演算法建構決策樹的計算過程 (魏志平、董和昇，2002; Quinlan, 1986)：

1. 將決策樹的根節點設定為 C ，此時所有物件都屬於 C 的物件集合。
2. 若 C 中的物件都屬於相同類別，則定義 C 節點為此類別並停止計算，否則繼續執行。
3. 計算熵值 $E(C)$ 。
4. 從根節點 (root) 至目前節點的路徑 (paths) 中，若有尚未當過節點的屬性，則分別計算以這些屬性為節點之子決策樹的資訊收益，以做為對物

件集合 C 進行分割的依據。

5. 在步驟(4)中找出最大資訊收益的屬性，將之做為節點 C 的分類屬性。
6. 在步驟(5)中所找出做為節點的分類屬性，假設共有 m 個屬性值，則在節點 C 下建立子節點分別為 C_1 、 C_2 、 \dots 、 C_m ，並依據屬性值來分派節點 C 中的所有物件集合，將具有相同屬性值的物件歸屬於同一子節點中。
7. 對每個子節點 C_i 視之為節點 C ， $1 \leq i \leq m$ ，並跳至步驟(2)重覆執行。

在建構決策樹的過程中，選擇具有最大資訊收益的屬性作為當時節點的分割屬性，此屬性能使得分割後之物件分類所需要的資訊量最小，並能反映出分割的最小隨機性或是不純性 (impurity)。這種資訊理論方法 (information theoretic approach) 對於物件分類所需要的期望分割數目能達到最小，並確保建構出最簡單狀態、或是接近最簡單狀態的決策樹 (Quinlan, 1986; Han & Kamber, 2006)。

我們以某一消費者為探勘的目標，假設其交易資料為 X ， X 為一個或以上產品項目所形成的項目組。在探勘的過程中，若一交易資料包含有 X ，則設定其關聯性為「有」，表示此交易資料中非 X 的產品項目與 X 之間有關聯性，如此非 X 的產品項目與 X 才會同時出現在此交易資料中；否則設定其關聯性為「無」。在建構決策樹的過程中，視非 X 的產品項目為欲分類的影響屬性，然後利用 ID3 演算法對消費者之交易資料進行分類計算，並建立決策樹。在決策樹中，從根節點開始往下尋找到關聯性為「有」之類別的路徑，其可顯示出那些屬性項目之狀態會導致與此消費者之間有關聯性。我們定義這些屬性項目與其狀態為「關聯因子」，藉此做為發掘此消費者最適性之產品項目的依據，其定義如下：

此消費者最適性的產品項目=「關聯因子」中狀態為曾經購買的產品項目。「關聯因子」中的屬性項目與其狀態，可導出與此消費者有關聯性的決策規則，因此可推導出其中狀態為曾經購買的產品項目，為此消費者最適性的產品項目。

經由以上決策樹找出的規則若為：產品 Y 與消費者之間有關聯性， Y 為一個或以上產品項目所形成的項目組，表示產品 Y 與 X 會出現在同一交易資料中的傾向，因此可推導出此消費者很有可能會購買產品 Y ，稱之為此消費者最適性的產品。找出產品之間的關聯性、及利用決策樹於行銷的有效性與有用性，已被許多文獻所探討 (Agrawal & Srikant, 1994; Han et al., 2004; Agrawal & Srikant, 1995; Quinlan, 1986; Berry & Linoff, 2004; Han & Kamber, 2006)，企業

若想利用分類分析來找出消費者個人最適性的產品，以上所提出建構決策樹的方法、及決策樹所推導出的結果，其有效性應是肯定的。

二、範例說明

我們以一實例說明發掘消費者最適性之產品項目的探勘過程。假設{A, B, C, D, E, F, G, H}表示全部產品項目的集合，表一為交易資料庫 D_1 其包含有 16 筆消費者的交易資料。若目前欲探勘之消費者的交易資料為 A，發掘此消費者最適性之產品項目的探勘過程說明如下。

表一 交易資料庫 D_1

交易資料編號	產品項目
1	ABCDEFGH
2	G
3	ACE
4	ABCDEFG
5	EGH
6	ABCDEF
7	ABCDEGH
8	BGH
9	C
10	BD
11	BEGH
12	C
13	BG
14	BDF
15	ACEGH
16	D

首先檢查各交易資料中是否包含 A，若有則設定關聯性為有，否則設定關聯性為無，其結果如表二。其中 B、C、D、E、F、G 及 H 等為非 A 的產品

項目，視之為欲分類的影響屬性。將表二的資料轉換成表三，其中標示為“1”者表示曾經購買該欄的產品項目，標示為“0”者表示未曾購買該欄的產品項目。

表二 交易資料與此消費者之間的關聯性

交易資料編號	產品項目	關聯性
1	ABCDEFH	有
2	G	無
3	ACE	有
4	ABCDEFG	有
5	EGH	無
6	ABCDEF	有
7	ABCDEGH	有
8	BGH	無
9	C	無
10	BD	無
11	BEGH	無
12	C	無
13	BG	無
14	BDF	無
15	ACEGH	有
16	D	無

利用公式(1)計算物件集合的熵值，在表三中有 6 筆交易資料與此消費者之間有關聯性、及 10 筆交易資料與此消費者之間無關聯性，計算所有交易資料的熵值如下：

$$E(\text{全部交易資料}) = -(6/16) \log_2(6/16) - (10/16) \log_2(10/16) = 0.954$$

接著利用公式(2)計算出各屬性下其子節點的熵值。以 B 為例，其屬性值的熵值計算如下：

$$E(\text{曾經購買 B}) = -(4/9) \log_2(4/9) - (5/9) \log_2(5/9) = 0.991$$

$$E(\text{未曾購買 B}) = -(2/7) \log_2(2/7) - (5/7) \log_2(5/7) = 0.863$$

$$E(B) = (9/16) \times 0.991 + (7/16) \times 0.863 = 0.577 + 0.378 = 0.935$$

再利用公式(3)計算 B 的資訊收益：

$$G(B) = 0.954 - 0.935 = 0.019$$

表三 未包含 A 的交易資料與關聯性

編號 \ 屬性項目	B	C	D	E	F	G	H	關聯性
1	1	1	1	1	1	0	1	有
2	0	0	0	0	0	1	0	無
3	0	1	0	1	0	0	0	有
4	1	1	1	1	1	1	0	有
5	0	0	0	1	0	1	1	無
6	1	1	1	1	1	0	0	有
7	1	1	1	1	0	1	1	有
8	1	0	0	0	0	1	1	無
9	0	1	0	0	0	0	0	無
10	1	0	1	0	0	0	0	無
11	1	0	0	1	0	1	1	無
12	0	1	0	0	0	0	0	無
13	1	0	0	0	0	1	0	無
14	1	0	1	0	1	0	0	無
15	0	1	0	1	0	1	1	有
16	0	0	1	0	0	0	0	無

註：1：曾經購買；0：未曾購買

依據以上相同的計算方法，可求出其他各產品項目屬性的資訊收益如下：

$$G(C) = 0.548$$

$$G(D) = 0.094$$

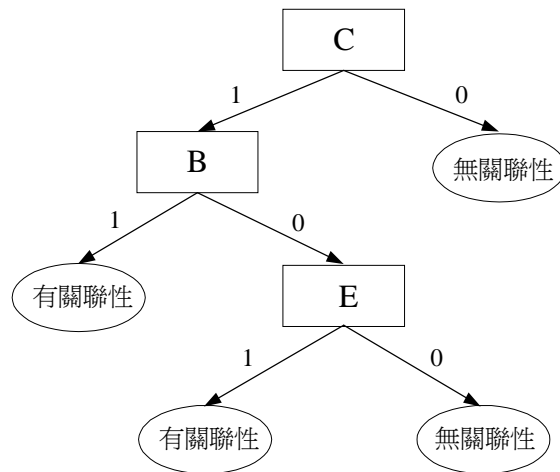
$$G(E) = 0.476$$

$$G(F) = 0.143$$

$$G(G) = 0.001$$

$$G(H) = 0.028$$

其中 C 具有最大的資訊收益，因此把 C 做為該決策樹的根節點，再把其餘的 6 個產品項目加入 C 之後。然後再往下反覆地進行分類的計算，直到節點中的交易資料都已屬於相同類別為止。最後將所有分類屬性依照先後順序連接起來，可形成圖一的分類決策樹。



圖一 分類決策樹

從圖一之分類決策樹的路徑中，其傾向特徵可以歸納出的決策規則為：

有購買 C and 有購買 B then 與此消費者之間有關聯性。

有購買 C and 未購買 B and 有購買 E then 與此消費者之間有關聯性。

在決策規則中若顯示有關聯性，其表示在交易資料中會有同時出現的傾向特徵。根據以上決策規則可以找出的「關聯因子」有：有購買 C+有購買 B、及有購買 C+未購買 B+有購買 E。因此，可發掘此消費者最適性的產品項目為：BC 或 CE。

肆 發掘具有購買次序之消費者最適性的產品項目

在消費者購物的交易記錄中，除了記錄購買的產品項目，也伴隨記錄購買的時間。因此，若考量消費者購物的時間性，則從消費者之交易資料中，可顯示消費者是依本身的需求及產品的特性來有次序性地購買產品項目。在此一章節中，我們以消費者之具有購買次序的交易資料為探勘的資料來源，並某一消費者為探勘的目標，對於此消費者曾經購買的產品項目，視之為已知的適性化產品，對於未曾購買而後續有傾向會購買的產品項目，則利用分類分析做為發掘具有購買次序之此消費者最適性的產品項目的方法依據。

一、探勘方法

我們仍以 ID3 演算法做為建構決策樹的方法依據，並以某一消費者為探勘的目標，假設其交易資料為 X ， X 為包含一個或以上產品項目且具有次序性的項目組。在探勘的過程中，若一交易資料包含 Z ， Z 為包含一個或以上產品項目且具有次序性的項目組，且 Z 包含的產品項目等於 X 包含的產品項目，則設定其關聯性為「有」，表示此交易資料中非 X 的產品項目與 X 之間有關聯性，如此非 X 的產品項目與 X 才會同時出現在此交易資料中；否則設定其關聯性為「無」。例如一交易資料為 ABCDE，若 $X=CD$ 或是 DC ，則表示 X 與此交易資料之間有關聯性。利用上述關聯性的定義方式，可增加交易資料與此消費者之間有關聯性的彈性與數量，有助於發掘較多的傾向特徵。

在建構決策樹的過程中，必須對包含 Z 之交易資料中的產品項目進行以下的處理：

假設 $B_1B_2B_3\dots B_i$ 依序分別為 Z 之後所購買的產品項目， $i \geq 1$ ，其是會被 Z 所影響是否購買的後次序產品。我們依序分解這些產品項目成為以下項目組： B_1 , B_1B_2 , ..., $B_1B_2\dots B_{i-1}$, $B_1B_2B_3\dots B_i$ ，在分解的過程中忽略 Z 之前所購買的產品項目。

在以上分解的過程中，我們考量產品項目只會影響是否購買其相鄰後續之產品項目，例如某一交易資料為 ABC，其表示的購買次序為產品 A→產品 B、及產品 B→產品 C，即目前購買的產品只會影響到相鄰後續購買的產品。

經由對全部交易資料進行上述的處理，並將分解後的項目組視為影響屬性，若交易資料中有出現影響屬性的項目組，則設定為曾經購買此項目組，否則設定為未曾購買此項目組。然後利用 ID3 演算法對消費者之交易資料進行分類計算，並建立決策樹。在決策樹中，從根節點開始往下尋找到關聯性為「有」之類別的路徑，其可顯示出那些屬性項目之狀態會導致與此消費者之間有關聯性。我們定義這些屬性項目與其狀態為「後次序關聯因子」，並藉此做為發掘具有購買次序之此消費者最適性的產品項目，其定義如下：

具有購買次序之此消費者最適性的產品項目=「後次序關聯因子」中狀態為曾經購買的產品項目。

「後次序關聯因子」中的屬性項目與其狀態，可導出與此消費者有關聯性的決策規則，因此可推導出其中狀態為曾經購買的產品項目，為此消費者最適性的後次序產品項目。

經由以上決策樹找出的規則若為：產品 Y 與消費者之間有關聯性， Y 為包

含一個或以上產品項目且具有次序性的項目組，表示產品 Y 會在 X 之後出現在同一交易資料中的傾向，因此可推導出此消費者後續很有可能購買產品 Y ，稱之為此消費者最適性的後次序產品。找出產品之間的次序關聯性、及利用決策樹於行銷的有效性與有用性，已被許多文獻所探討 (Agrawal & Srikant, 1995; Yu & Chen, 2005; Quinlan, 1986; Berry & Linoff, 2004; Han & Kamber, 2006)，企業若想利用分類分析來找出消費者個人最適性的後次序產品，以上所提出建構決策樹的方法、及決策樹所推導出的結果，其有效性應是肯定的。

二、範例說明

我們以一實例來說明發掘具有購買次序之消費者最適性的產品項目的探勘過程。假設{A, B, C, D, E, F, G, H}表示全部產品項目的集合，表四為交易資料庫 D_2 其包含 16 筆消費者的交易資料，每一筆交易資料分別記錄消費者曾經購買的產品項目與其次序。若目前欲探勘之消費者的交易資料為 A，發掘具有購買次序之此消費者最適性的產品項目，其探勘過程說明如下。

表四 交易資料庫 D_2

交易資料編號	產品項目
1	EABDE
2	BGED
3	FACDG
4	CED
5	BCF
6	EACDF
7	BACDF
8	CEDG
9	BCE
10	EBACD
11	BHD
12	BGHE
13	GBACD
14	HDABD
15	FACDB
16	BFEG

首先檢查交易資料與此消費者之間的關聯性，若交易資料包含 A，則設定關聯性為有；否則設定關聯性為無，其結果如表五。

表五 交易資料與此消費者之間的關聯性

交易資料編號	產品項目	關聯性
1	EABDE	有
2	BGED	無
3	FACDG	有
4	CED	無
5	BCF	無
6	EACDF	有
7	BACDF	有
8	CEDG	無
9	BCE	無
10	EBACD	有
11	BHD	無
12	BGHE	無
13	GBACD	有
14	HDABD	有
15	FACDB	有
16	BFEG	無

若交易資料中包含 A，則分解 A 以後所購買之產品項目成為各有次序的項目組，而不考量 A 之前所購買的產品項目。經由對所有交易資料進行前述的分解過程，並將分解後的項目組視為影響屬性，若交易資料中有出現影響屬性的項目組，則表示曾經購買此項目組，並以“1”作標示；否則設定為未曾購買此項目組，並以“0”作標示。交易資料經由上述的處理資之後，可轉換成如表六的交易資料庫。

表六 包含影響屬性之項目組的交易資料與關聯性

項目組 編號	BDE	BD	B	CDG	CD	C	CDF	BC	CDB	關聯性
1	1	1	1	0	0	0	0	0	0	有
2	0	0	1	0	0	0	0	0	0	無
3	0	0	0	1	1	1	0	0	0	有
4	0	0	0	0	0	1	0	0	0	無
5	0	0	1	0	0	1	0	1	0	無
6	0	0	0	0	1	1	1	0	0	有
7	0	0	0	0	1	1	1	0	0	有
8	0	0	0	0	0	1	0	0	0	無
9	0	0	1	0	0	1	0	1	0	無
10	0	0	0	0	1	1	0	0	0	有
11	0	0	1	0	0	0	0	0	0	無
12	0	0	1	0	0	0	0	0	0	無
13	0	0	0	0	1	1	0	0	0	有
14	0	1	1	0	0	0	0	0	0	有
15	0	0	0	0	1	1	0	0	1	有
16	0	0	1	0	0	0	0	0	0	無

註：1：曾經購買；0：未曾購買

利用公式(1)計算物件集合的熵值，在表 6 中有 8 筆交易資料與此消費者之間有關聯性，有 8 筆交易資料與此消費者之間無關聯性，計算所有交易資料的熵值如下：

$$E(\text{全部交易資料}) = -(8/16) \log_2(8/16) - (8/16) \log_2(8/16) = 1$$

接著計算出各個屬性下其子節點的熵值，以 BDE 為例，其屬性值的熵值計算如下：

$$E(\text{曾經購買 BDE}) = -(1/1) \log_2(1/1) - (0/1) \log_2(0/1) = 0$$

$$E(\text{未曾借閱 BDE}) = -(7/15) \log_2(7/15) - (8/15) \log_2(8/15) = 0.996792$$

$$E(\text{BDE}) = (1/16) \times E(\text{曾經購買 BDE}) + (15/16) \times E(\text{未曾購買 BDE}) = 0.934492$$

再利用公式(3)來計算 BDE 的資訊收益：

$$G(\text{BDE}) = 1 - E(\text{BDE}) = 0.065508$$

依此類推，可求出所有其餘屬性項目的資訊收益如下：

$$G(\text{BD}) = 0.065508$$

$$G(\text{B}) = 0.188722$$

$$G(\text{CDG}) = 0.065508$$

$$G(\text{CD}) = 0.548795$$

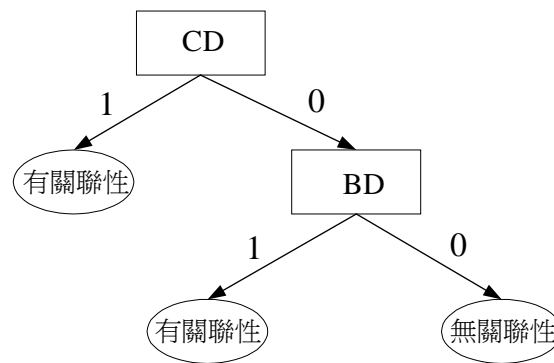
$$G(\text{C}) = 0.048795$$

$$G(\text{CDF}) = 0.137925$$

$$G(\text{BC}) = 0.018791$$

$$G(\text{CDB}) = 0.065508$$

其中 CD 具有最大的資訊收益，因此將 CD 做為該決策樹的根節點，再把其餘的屬性項目加入 CD 之後。然後再往下反覆地進行分類的計算，直到節點中的交易資料都已屬於相同類別為止。最後將所有分類屬性依照先後順序連接起來，可形成圖二的分類決策樹。



圖二 具有購買次序之分類決策樹

從圖二之分類決策樹所顯示的傾向特徵，可以歸納出的決策規則為：

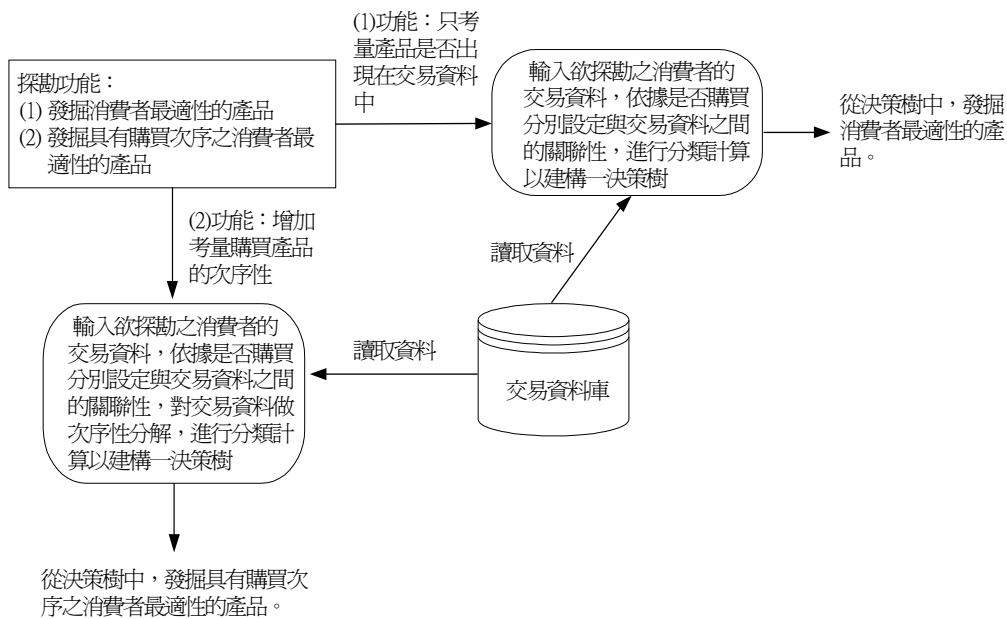
後續有購買 CD then 與此消費者之間有關聯性。

後續未購買 CD and 後續購買 BD then 與此消費者之間有關聯性。

在決策規則中若顯示有關聯性，其表示在交易資料中會有同時出現的傾向特徵。根據以上決策規則可以找出的「後次序關聯因子」有：有購買 CD、及未購買 CD+有購買 BD。因此，可發掘具有購買次序之此消費者最適性的產品項目為：CD 或 BD。

伍· 探勘系統之設計與實作

我們將前面章節所描述的探勘方法，應用到發掘消費者最適性之產品項目的系統實作上，系統探勘過程的模型如圖三。



圖三 系統探勘模型

消費者之交易資料由 IBM Data Mining 網站 (<http://www.almaden.ibm.com/>) 下載資料模擬程式，以產生探勘系統所需要的交易資料 (Agrawal. & Srikant, 1994)，並以 C# 為撰寫的程式語言。由模擬程式產生的資料中，每一筆交易資料包含的項目視為曾經購買的產品項目，然後再以亂數隨機的方式調整項目購買的次序。本系統共產生 12 個交易資料庫，如表七，其主要參數值的意義分別為： n 為項目的數量、 $ntran$ 為交易資料筆數的數量、 np 為型樣組合的數量、 tl 為每筆交易資料的平均項目個數、其餘參數以預設值表示之。

表七 交易資料庫與其參數

參數	n	$ntran$	np	tl
交易資料庫 D0	50	100000	10000	10
D1	50	10000	10000	10
D2	55	10000	10000	10
D3	60	10000	10000	10
D4	65	10000	10000	10
D5	70	10000	10000	10
D6	75	10000	10000	10
D7	80	10000	10000	10
D8	85	10000	10000	10
D9	90	10000	10000	10
D10	95	10000	10000	10
D11	100	10000	10000	10

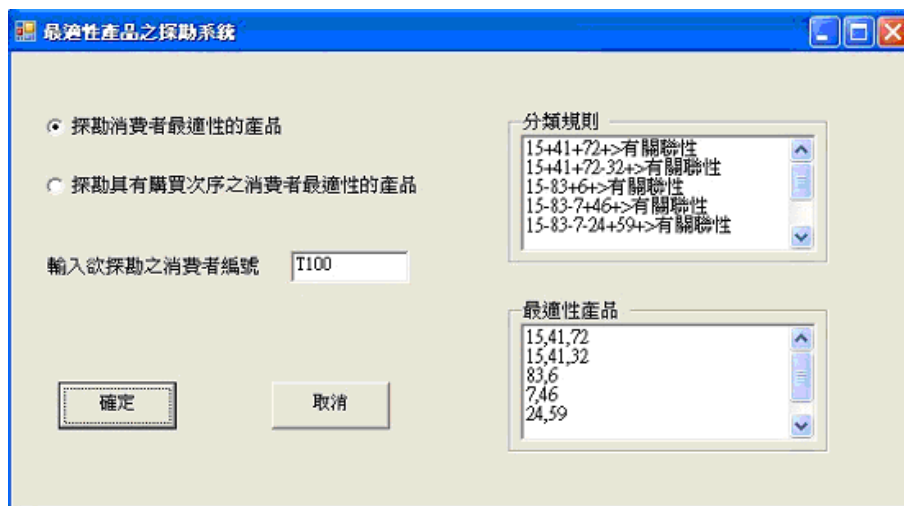
SN	ID	Items
1	T1	23,1,47,13,88,75,28,71,47,68
2	T2	49,73,6,25
3	T3	18,52,5,93,29,87
4	T4	1,54,86,13,65
5	T5	31,9,76,34,81,96,67
6	T6	8,37,3,15,82,100,64,49
7	T7	24,52,12,73,61,94
8	T8	40,17,61,29,85,75,99
9	T9	11,75,3,87,34,38,91,52,68
10	T10	70,57,28,92,34
11	T11	24,63,15,41,90,86,90
12	T12	8,54,19,81,99,30,67
13	T13	74,5,89,28,31,62,47,51
14	T14	25,3,80,69,58,100,37,91
15	T15	13,84,25,95,41,39,68
16	T16	82,11,41,29,78,53
17	T17	18,44,31,77,62,89
18	T18	1,80,33,16,97,68,47
19	T19	68,9,84,45,24
20	T20	13,41,28,85,53,74,66,97
21	T21	3,57,26,88,37
22	T22	28,61,15,40,98,83,58
23	T23	10,38,71,69,45,58,95,82,24
24	T24	38,51,1,72,23,93,49,60
25	T25	6,32,19,80,67,54,91
26	T26	79,4,99,29,41,83,55,18,65
27	T27	15,69,38,74,51,86,7

圖四 交易資料庫 D11

我們以交易資料庫 D11 做為系統探勘的資料來源，分別以編號 1, 2, 3, ..., 100 來表示產品項目，各交易資料之消費者編號分別以 T1, T2, ..., T10000 來

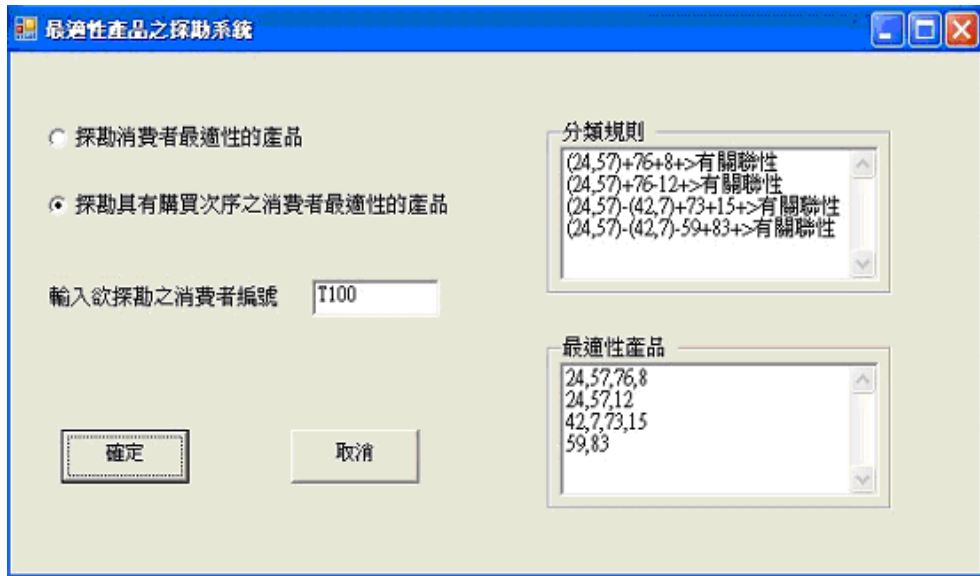
表示之，如圖四，利用前面章節所提出的方法，設計與建置一個探勘系統。以下說明此一系統之探勘過程的執行介面。

圖五為此探勘系統的執行畫面，其中包含有兩項功能選項：「探勘消費者最適性的產品」及「探勘具有購買次序之消費者最適性的產品」。假設目前點選「探勘消費者最適性的產品」功能，因此在探勘的過程中將忽略產品購買的次序性，在「輸入欲探勘之消費者編號」欄位中輸入“T100”為例，根據第參節所描述之探勘方法所建構的決策樹，可在「分類規則」欄位中顯示有那些屬性項目狀態與此消費者之間有關聯性。例如「15+41+72+>有關聯性」表示決策樹中有一條路徑為：產品編號 15、41、72 與此消費者之間有關聯性，其中“+”表示曾經購買其前面的產品項目。然後，在「最適性產品」欄位中可顯示所推導出的探勘的結果，如圖五。

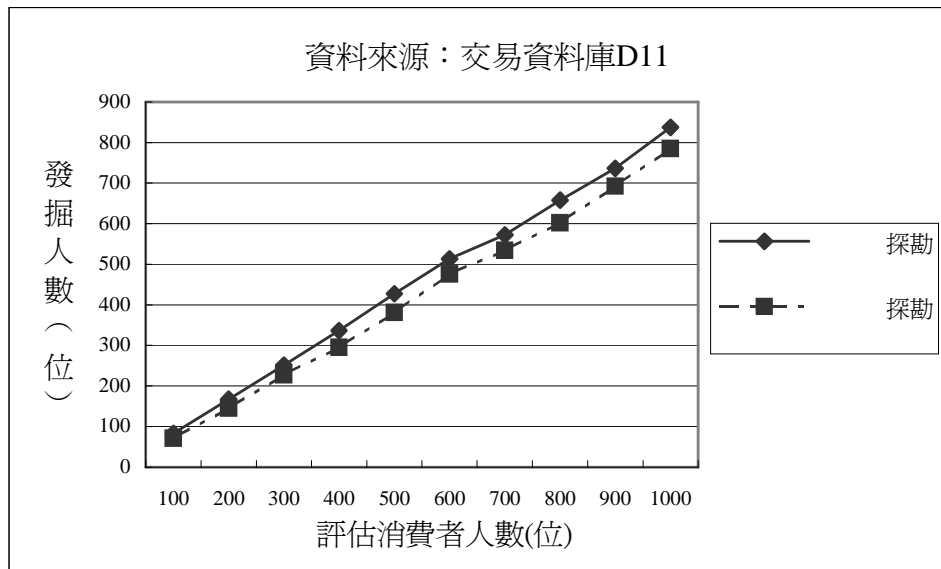


圖五 探勘消費者最適性之產品的執行畫面

圖六為點選「探勘具有購買次序之消費者最適性的產品」功能，在「輸入欲探勘之消費者編號」欄位中仍輸入“T100”為例，經由第肆節所描述之探勘方法所建構的決策樹，可在「分類規則」欄位中顯示有那些屬性項目與此消費者之間有關聯性。例如「(24, 57) +76+8+>有關聯性」表示決策樹中有一條路徑為：產品編號 (24、57)、76、8 (後次序購買) 與此消費者之間有關聯性。然後，在「最適性產品」欄位中可顯示所推導出具有購買次序的探勘結果，如圖六。



圖六 探勘具有購買次序之消費者最適性的產品的執行畫面



圖七 評估探勘方法發掘消費者最適性之產品的結果

我們從交易資料庫 D11 中，分別隨機挑選出 100, 200, 300, ..., 1000 位消費者為評估對象，以挑選後剩餘之消費者的交易資料為訓練資料，評估文中所描述之探勘方法的準確性：一是只考量產品是否出現在交易資料中，其中隨機挑選每位消費者之交易資料中的 1/3 項目為探勘目標（採用四捨五入，若不足一個則以一個計算），其餘 2/3 個項目為驗證資料，以 Y 表示之，利用第參節的探勘方法，以發掘每位消費者最適性的產品項目，若只要其中有任一項目

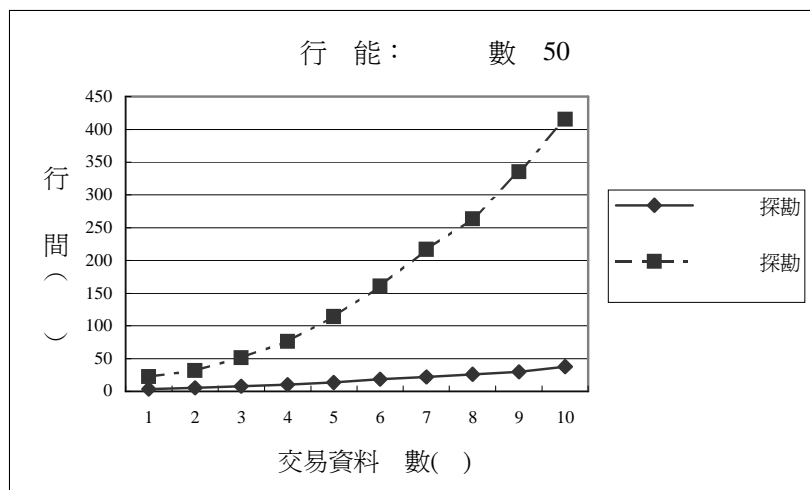
$\subseteq Y$ ，則表示可成功發掘這位消費者最適性的產品項目，其結果如圖七；二是考量產品出現在交易資料中的次序性，其中挑選每位消費者之交易資料中的前 1/3 個項目為探勘目標（採用四捨五入，若不足一個則以一個計算），其餘後 2/3 個項目為驗證資料，以 Z 表示之，利用第肆節的探勘方法，以發掘具有購買次序之每位消費者最適性的產品項目，若只要其中有任一項目 $\subseteq Z$ ，則表示可成功發掘具有購買次序之這位消費者最適性的產品項目，其結果如圖七。

接下來，我們實驗評估前面章節所描述之探勘方法的執行效能，其實驗平台說明如表八。

表八 實驗平台

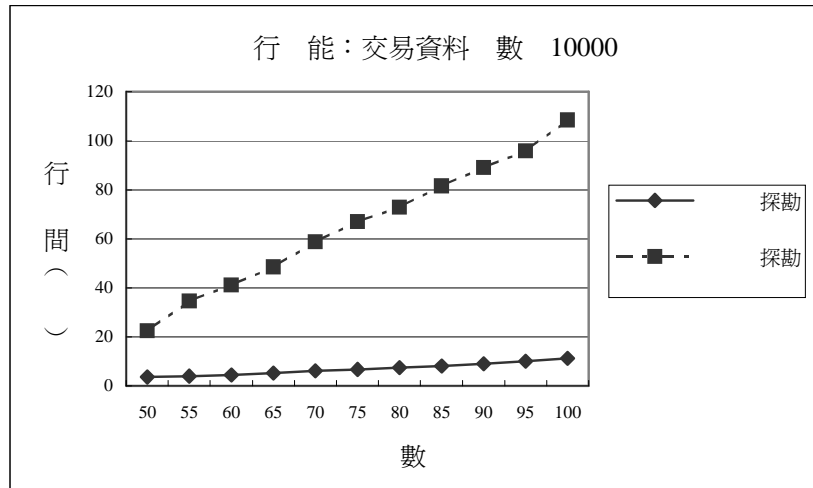
CPU	CPU-Pentium IV 3.2G Hz
Main memory	512 Mbytes
作業系統	Windows XP Professional Edit
使用語言	C#
探勘資料來源	表 7 之交易資料庫

我們先將交易資料庫 D0 分割成 10 個分別包含有 10000 筆的交易資料，然後再依次累加其交易資料的數量分別成爲 10000, 20000, 30000, ..., 100000，並分別以 D0.1, D0.2, D0.3, D0.4, D0.5, D0.6, D0.7, D0.8, D0.9, D0 來標示其所代表的交易資料庫。在探勘計算的過程中，在不失一般性的條件下，探勘某一消費者之交易資料爲單一產品爲例（例如編號 20），在發掘此消費者最適性之產品項目的情況下，分別評估第參節、及第肆節之探勘方法的執行效能。



圖八 不同交易資料數量之兩個探勘方法的執行時間

在圖八中，我們分別以交易資料庫 D0.1, D0.2, D0.3, D0.4, D0.5, D0.6, D0.7, D0.8, D0.9, D0 為探勘的資料來源，分別評估在不同交易資料數量的條件下，第參節之探勘方法及第肆節之探勘方法的執行時間。在圖九中，我們分別以交易資料庫 D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11 為探勘的資料來源，分別評估在不同項目個數的條件下，第參節之探勘方法及第肆節之探勘方法的執行時間。從圖中顯示當交易資料或是項目個數增加時，探勘的執行時間會明顯地增加。



圖九 不同項目個數之兩個探勘方法的執行時間

陸· 結論與未來研究

目前企業的經營環境中，資訊技術已扮演著非常重要的角色，其功能除了協助支援企業平常的交易處理，對企業決策也提供非常有用的資訊與知識。在消費者曾經購買產品的記錄中，可反映出消費者對產品需求的傾向特徵、及不同產品項目之間的關聯性及次序性。在資料探勘技術中，考量產品是否出現在交易資料中，探勘產品項目之間的關聯性的相關方法有關聯規則 (Agrawal et al., 1993; Agrawal & Srikant, 1994; Han et al., 2000, 2004; Tsay & Chiang, 2005; Pei et al., 2006; Xu & Wang, 2006)。對增加考量產品購買的次序性，探勘產品項目之間具有購買次序關聯性的相關方法有次序相關 (Agrawal & Srikant, 1995; Srikant & Agrawal, 1996; Chen et al., 2003; Yu & Chen, 2005; Zaki, 2001; Wang & Wang, 2006; Yun & Leggett, 2006; Tan et al., 2006)。

本研究以某一消費者為探勘的目標，提出兩個利用 ID3 演算法建構決策樹的方法，以分別找出此消費者最適性的產品、及找出具有購買次序之此消費者最適性的產品。本論文的研究價值與貢獻說明如下：

- 1.文中所探討產品項目之間的關聯性、及次序性，其重要性與有用性已被許多研究（關聯規則及次序相關）所證實。
- 2.而利用 ID3 演算法建構決策樹的優越性、及應用於行銷的有效性也被許多研究所闡述。
- 3.本研究利用 ID3 演算法探討某一消費者與產品項目之間的關聯性、及次序性，由決策規則所推導出的結果，以發掘此消費者最適性的產品項目，在個人化行銷上的價值、有效性及有用性，其理論基礎應是可顯見。

目前，本研究僅考量欲探勘之消費者與交易資料間是否有關聯性、及產品只影響其相鄰後續之產品的購買做探討，對於未來繼續從事之相關研究有：

- 1.探討利用其他資料探勘技術在此研究問題的可行性，例如分群化 (clustering)。
- 2.延伸關聯性的適用範圍，例如增加機率的表示法，以配合不同的資料型態及應用。
- 3.延伸考量產品影響其後續之產品的購買條件，例如符合有前後次序即可，並非一定要連續相鄰。
- 4.改良探勘系統的設計與建置，將所提出之方法應用於實際個案資料中，並探討所找出之規則的有效驗證。
- 5.提昇探勘過程的執行效能。

參考文獻

- 高淑珍，「應用資料探勘於顧客回應模式之研究-以國內 A 壽險公司為例」，國立成功大學企業管理研究所博士論文，2004 年。
- 魏志平、董和昇，「電子商務理論與實務」，2 版，華泰書局，2002 年，頁 167-205。
- Agrawal, R. and Srikant, R., "Fast Algorithms for Mining Association Rules in Large Database", Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487-499.
- Agrawal, R. and Srikant, R. "Mining Sequential Patterns", Proceedings of the International Conference on Data Engineering (ICDE), 1995, pp. 3-14.

- Agrawal, R., Imielinski, T. and Swami, A., "Mining Association Rules between Sets of Items in Large Database", *Proceedings of the ACM SIGMOD Conference on Management of Data*, 1993, pp. 207-216.
- Berry, M. J. A. and Linoff, G. S., "Data Mining Techniques for Marketing, Sales, and Customer Support", 2nd ed., New York: John Wiley, 2004.
- Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., "Classification and Regression Trees", Wadsworth, Belmont, California, 1984.
- Chen, M. S., Han, J. and Yu, P. S., "Data Mining: An Overview from a Database Perspective", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8(6), 1996, pp. 866-883.
- Chen, Y. L., Chiang, M. C. and Ko, M. T., "Discovering Time-Interval Sequential Patterns in Sequence Databases", *Expert Systems and Applications*, Vol. 25(3), 2003, pp. 343-354.
- Clark, P. and Niblett, T., "The CN2 Induction Algorithm", *Machine Learning*, Vol. 3, 1989, pp. 261-283.
- Gaddam, S. R., Phoha, V. V. and Balagani, K. S., "K-Means+ID3: A Novel Method for Supervised Anomaly Detection by Cascading K-Means Clustering and ID3 Decision Tree Learning Methods", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19(3), 2007, pp. 345-354.
- Han, J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd ed., Morgan Kaufmann, 2006.
- Han, J., Pei, J. and Yin, Y., "Mining Frequent Patterns without Candidate Generation", *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data Table of Contents*, 2000, pp. 1-12.
- Han, J., Pei, J., Yin, Y. and Mao, R. "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach", *Data Mining and Knowledge Discovery*, Vol. 8(1), 2004, pp. 53-87.
- Hartigan, J. A., "Clustering Algorithms", Wiley, New York, 1975.
- Hui, S. C. and Jha, G., "Data Mining for Customer Service Support", *Information and Management*, Vol. 38, 2000, pp. 1-13.
- IBM, <http://www.almaden.ibm.com/software/projects/hdb/resources.shtml>.
- Katharina, D. C. S. and Dirk, U. P., "The Application of Non-Parametric Techniques to Solve Classification Problems in Complex Data Sets in Veterinary Epidemiology - an Example", *Intelligent Data Analysis*, Vol. 3(1), 1999, pp. 23-35.
- Laudon, K. C. and Laudon, J. P., "Management Information Systems", 6th ed., Upper Saddle River, NJ: Prentice Hall, 2000.
- Liu, D. and Zhou, Y., "Distributed ID3 Model Based on Mobile Agent", *Proceedings of the 1st International Symposium on Systems and Control in Aerospace and Astronautics*, 2006, pp. 1316-1320.
- Mak, B. and Munakata, T., "Rule Extraction from Expert Heuristics: A Comparative Study of Rough Sets with Neural Networks and ID3", *European Journal of Operational Research*, Vol. 136(1), 2002, pp. 212-229.

- Ohmann, C., Moustakis, V., Yang, Q. and Lang, K., "Evaluation of Automatic Knowledge Acquisition Techniques in the Diagnosis of Acute Abdominal Pain - Acute Abdominal Pain Study Group", *Artificial Intelligence in Medicine*, 1996, Vol. 8(1), 1996, pp. 23-36.
- Park, J. S., Chen, M. S. and Yu, P. S., "Using a Hash-Based Method with Transaction Trimming for Mining Association Rules", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 9(5), 1997, pp. 813-825.
- Pei, M., Taniguchi, S., Hara, T. and Nishio, S., "Discovering Important Rules and Loyal Customer by Considering the Repetition in Association Rules", *Proceedings of the Innovations in Information Technology*, 2006, pp. 1-5.
- Quinlan, J. R., "Induction of Decision Trees", *Machine learning*, Vol. 1, 1986, pp. 81-106.
- Quinlan, J. R., "C4.5: Programs for Machine Learning", Morgan Kaufmann, 1993.
- Rich, E. and Knight, K., "Learning in Neural Network", 2nd ed., McGraw-Hill, New York, 1991.
- Srikant, R. and Agrawal, R., "Mining Generalized Association Rules", *Proceedings of the 21st International Conference on Very Large Data Bases*, 1995, pp. 407-419.
- Srikant, R. and Agrawal, A., "Mining Sequential Patterns: Generalizations and Performance Improvements", *Proceedings of the Fifth International Conference on Extending Database Technology*, 1996, pp. 3-17.
- Tan, X., Yao, M. and Zhang, J., "Mining Maximal Frequent Access Sequences Based on Improved WAP-tree", *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications*, 2006, pp. 616 – 620.
- Tsay, Y. J. and Chiang, J. Y., "CBAR: An Efficient Method for Mining Association Rules", *Knowledge-Based Systems*, Vol. 18, 2005, pp. 99-105.
- Wang, H. and Wang, S., "A Purchasing Sequences Data Mining Method for Customer Segmentation", *Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics*, 2006, pp. 883-886.
- Widrow, B., Rumelhart, D. E. and Lehr, M. A., "Neural Networks: Applications in Industry Business and Science", *Communication of the ACM*, Vol. 37(3), 1994, pp. 93-105.
- Wu, S., Wu, L. Y., Long, Y. and Gao, X. D., "Improved Classification Algorithm by Minsup and Minconf Based on ID3", *Proceedings of the International Conference on Management Science and Engineering (ICMSE '06.)*, 2006, pp. 135-139.
- Xu, B., "Managing Customer Satisfaction in Maintenance of Software Product Family via ID3", *Proceedings of International Conference on Machine Learning and Cybernetics*, 2005, pp. 1820-1824.
- Xu, W. X. and Wang, R. J., "A Fast Algorithm of Mining Multidimensional Association Rules Frequently", *Proceedings of the International Conference on Machine Learning and Cybernetics*, 2006, pp. 1199-1203.
- Yu, C. C. and Chen, Y. L., "Mining Sequential Patterns from Multi-Dimensional Sequence Data", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17(1), 2005, pp. 136-140.

Yun, U. and Leggett, J. J., "Wspan: Weighted Sequential Pattern Mining in Large Sequence Databases", Proceedings of the 3rd International IEEE Conference on Intelligent Systems, 2006, pp. 512-517.

Zaki, M., "SPADE: An Efficient Algorithm for Mining Frequent Sequences", *Machine Learning*, Vol. 40, 2001, pp. 31-60.

Using Classification Analysis to Find Consumers' Most Adaptive Product Items

CHUI-CHENG CHEN *

ABSTRACT

In this paper, we use consumers' transaction data as the source data of mining. Each transaction data contains a consumer ever bought product items with sequence. We let one consumer as the target of mining and use classification analysis to find the consumer's most adaptive product items from two aspects. One is to consider whether product items are contained in transaction data. A method is presented to construct a decision tree that shows whichever product items to have association with the consumer. The consumer's most adaptive produce items are found from the decision tree. The other is to consider extra the sequences of buying product items. A method is presented to construct a decision tree that shows whichever product items to have sequential association with the consumer. The consumer's most adaptive produce items with sequence are found from the decision tree. We design and construct a mining system to find consumers' most adaptive product items according to the above methods.

Keywords: transaction data, classification analysis, sequence

* Chui-Cheng CHEN, Associate Professor, Department of Information Management, Southern Taiwan University.