

運用多模式表決提升預測效益之研究

陳彤生

東南技術學院企業管理系講師

E-mail:tschen@mail.tnit.edu.tw

蔣台程

東南技術學院企業管理系副教授

E-mail:tcchiang@mail.tnit.edu.tw

摘要

顧客關係管理重視何種客戶為忠誠者。為瞭解每位獨立客戶，首先必須建置完整的消費者資料庫，資料庫中應記載每位客戶的互動資料，包括客戶基本資料、客戶交易資料、客戶服務資料、活動因應資料及其他相關互動記錄。

由整體資料倉儲中，找出客戶交易軌跡，以預測客戶購買偏好，並以 RFM 模式--即購買頻率 (frequency) 購買近期性 (recency) 和購買金額 (monetary) 及貝氏分類法加權 RFM 模式、決策樹歸納法加權 RFM 模式、吉尼指標法加權 RFM 模式、多模式表決法來評估客戶的效益。本研究收集並分析相關提升顧客購買文獻，同時介紹多模式表決，結果顯示由收納法 (bagging) 導出之合成模式比單一模式好 1% 至 10%。

關鍵詞：多模式表決、收納法、貝氏分類法、決策樹歸納法、吉尼指標法

一、緒論

1.1 研究背景與動機

國際化及科技的潮流加上網路交易的發達，企業越來越容易收集到大量的顧客資料，為有效的運用這些資料，資料探勘 (data mining) 即成為熱門的顧客關係管理資料分析工具。

為了長期吸引顧客，持續改善顧客滿意度、增進競爭力、維持同業領先地位，以達成顧客關係管理的目標，是各服務業刻不容緩的努力方向。

為何要重視顧客關係管理？研究顯示，20% 的好顧客貢獻了利潤的 150%，而最差的 40% 顧客，使利潤縮減 50%，顧客區隔的重要性不言而喻。

1.2 研究目的

為使行銷效益最佳化，吾人皆知 20/80 法則，亦即在企業實際營運中，常發生百分之二十的客戶能創造百分之八十的營業額或利潤。因此若能找出這些客戶，並有效運用行銷資源，以增加其交易次數，必可為企業創造更高之營運效能。[6]

1.3 研究方法

為瞭解每位獨立客戶，首先必須建置完整的消費者資料庫，資料庫中應記載每位客戶的互動資料，包括客戶基本資料、客戶交易資料、客戶服務資料、活動因應資料及其他相關互動記錄。

本研究由型錄業者 (郵購公司) 近五年來之二十萬筆交易資料，除去缺貨、退貨及無客戶基本資料之現場購買者之外，所剩餘 139715 筆交易資料，萃取出用的特徵，找出客戶交易軌跡，以預測客戶購買偏好。並進行 RFM 分析，來評估客戶的效益。

資料勘查可得兩變數間之關係，根據過去的經驗，建立資料探勘預測模型，來預測誰可能會對此商品有興趣。再用資料探勘，找出可能對商品有興趣的人。

這些方法的共同點是它們都會給每一位顧客設定一個分數。分數高的顧客是促銷活動鎖定的目標，而分數低的顧客則排除在促銷活動之外。

本研究使用的資料探勘技術有 RFM 法則、貝氏分類法、決策樹歸納法、吉尼指標法以及結合前三種模式而成的多模式表決資料探勘方式。

二、文獻探討

Schmittlein & Peterson (1994) 以為行銷資料庫要能利用觀察到的過去顧客的購買行為來了解顧客們目前和未來可能的購買模式，廠商可以由資料庫得知：1、消費者目前仍在活動的人數；2、這些仍在活動的顧客會隨時間如何變動；3、有那些消費者未來仍是活動的；4、預測未來期間有多少購買。[7]

Roberts (1992) 指出 RFM 分析法是直銷業者最常用來找出有價值顧客的方法，也有學者特別針對顧客績效作探討，更有相當多的學者整合影響因素並對其績效作探討。Bob Stone 建議採用 RFM 分析法 (期間 recency、頻率 frequency、金額 monetary)，來評估客戶的效益。[9, 10]

檢視傳統 RFM 分析法，直銷業者決定以五分法或十分法進行分群，並無一定衡量標準，而使用五分法或十分法所產生的結果又不同，在十分法中不具價值的顧客可能在五分法中是具有價值的，因衡量標準不同使結果遺漏

許多具有價值的顧客，另外業者也只能憑著過去的經驗來決定顧客未來的反應率(David Shepard Associates, 1995)。[5]

Kalakota & Robinson(1999)指出顧客生命週期分獲得(acquire)加強(enhance)保留(retain)三階段。在獲得階段應採創新及便利策略以吸引客戶。在加強階段應採降低成本及利用交叉銷售策略以加強客戶。在保留階段應採傾聽客戶需求及研發新產品策略以保留客戶。[11]

Quinlan(1986, 1987, 1993)的 ID3 歸納學習演算法是採用資訊理論中的不純度函數為基礎，最終形成決策樹[12, 13, 14]，此方法已有效運用在許多領域。決策樹在建構過程中，可以發掘出一般所了解的分類規則[3, 4]。決策樹的演算過程簡單，不必繁雜運算即可完成決策樹的動作。在處理分類過程時能夠掌握連續性以及屬於某一範疇的變數。在預測及分類上，決策樹之分析結果能提供極重要資訊。

2.1 顧客市場區隔

Michael J.A. Berry & Gordon S. Linoff (2000) 以為顧客市場區隔是將顧客區分成互斥(mutually exclusive)群集的程序。在同一群集內，都是某些特性上極為相似的顧客。市場區隔相當有用，因為這可以在資料探勘的過程，加入特定領域的專業知識，針對不同的顧客族群，推出不同的促銷活動，也建立各自獨立的預測模型。這樣區隔使得資料探勘不需重新發現已知的事實，因此會發展出較好的預測模型。

資料探勘演算法可用來協助客戶市場區隔。這些資料探勘分類方法，不同於商業上熟知的分類。原因是這些分類並非掌控企業的重要因素。或者這些分類可能是掌控企業的因素，但管理者不熟悉這些分類，無法善加利用。

顧客生命週期是指顧客和企業間關係的不同階段。顧客生命週期很重要，因其直接影響長期顧客的價值。首先，顧客生命週期是一個架構，可以用來了解顧客消費行為。資料探勘可以在顧客生命週期的不同階段中切入，因此了解顧客生命週期可以讓資料探勘的過程更有效率。除此之外，某些顧客生命週期所發生的事件是很重要的。若能預測這些變化，將會帶來很大的好處。

整個顧客生命週期中，不同的階段都有其重要事件。每個階段都有資料探勘及顧客關係管理可以切入的點。

已確定顧客的行為，通常都隱藏在記錄每次交易的詳細資料中。要善用這些資料，必須學會如何從中萃取出有用的特徵。

顧客早期的消費及使用率資訊非常有價

值。在某些公司中，利用顧客早期的消費行為，就可以準確預測其未來使用率。

顧客關係管理的法則，可以歸納成：在適當的時間，以適當的價格，對適當的顧客提供適當的商品或服務。

2.2 貝氏分類法

貝氏分類法(Bayesian classification)是統計分類法，它可預測類別機率，如一已知樣本屬於某一特定類別之機率。貝氏分類乃基於貝氏理論，分類演算法之研究顯示簡單貝氏分類法之功能似決策樹及類神經網路分類，在應用於大型資料庫時亦有高正確性及快速。[1, 2]

簡單貝氏分類法假設某一類別之一屬性值與其他屬性值無關，稱為類別條件獨立，以簡化相關計算。

貝氏理論乃令 X 為樣本，其輸出類別為未知，令 H 為假設： X 屬於一特定類 C ， $P(H|X)$ 表示樣本 X 中假設 H 成立之機率。

$P(H|X)$ 為後驗機率，即 X 條件下之 H 機率。例如：樣本含水果，以其顏色及形狀描述，若 X 為紅色圓形， H 表示 X 為蘋果，則 $P(H|X)$ 表示信心 - 當我們看見 X 為紅色圓形時， X 為蘋果。反之， $P(H)$ 為 H 之先驗機率(Prior probability)。例如表樣本為蘋果之機率，後驗機率 $P(H|X)$ 比先驗機率 $P(H)$ 需更多背景知識。

$P(X|H)$ 為 X 在 H 條件下之後驗機率，即已知 X 為蘋果，則 X 為紅色圓形之機率； $P(X)$ 為 X 之先驗機率，表水果樣本中為紅色圓形的機率。

如何估計機率？ $P(X)$ ， $P(H)$ ，及 $P(X|H)$ 可由已知資料估計，貝氏理論為

$$P(h|x) = \frac{P(x|h)p(h)}{p(x)}$$

與其他分類法比較，貝氏分類之誤差最小，但實際上因我們假設條件獨立之不正確，故與其他方法比較之實證研究顯示其稍遜競爭力。

2.3 決策樹歸納法

決策樹為像流程圖之樹形狀結構，內部節點表示一屬性之測試，分支表示測試之結果，葉節點表示分類或類別分佈。最上層稱根節點。(1)

為了分類，將樣本之屬性值加以測試，由根至葉節點沿著樣本中含分類值之路徑追蹤，決策樹易於轉換為分類規則。

描述學習決策樹之基礎演算法。建立決策樹後，分支反映訓練資料中之雜質或離群值，修剪以除去這些分支，以達改進新資料正確分

類之目標。決策樹可應用於商業、遊戲及醫學，它為多種商業規則歸納之基礎。

樹中每一節點所選之測試屬性皆以資訊增益加以衡量，這種衡量稱為屬性選擇量度或量度一完善的分割。含最高資訊增益之屬性（或最大量之減少）被選為該節點的測試屬性。這屬性能將樣本分類 所需資訊最小化，且使分類所含雜質最小化。這種資訊理論方法使分割物件所需測試最小化且保證可得一簡化樹。

令 S 為含 s 資料樣本之集合，假設標記 m 個不同屬性表示 m 個不同類， C_i (其中 $i=1, \dots, m$) 令 S_i 為 S 在 C_i 類之樣本數目，欲分類某樣本所需期望資訊為

$$I(s_1, s_2, \dots, s_m) = -\sum_{i=1}^m p_i \log_2(p_i), \quad (1)$$

其中 P_i 為任意一樣本屬於類 C_i 之機率且其值為 S_i/S ，注意因資訊以 bits 編碼，故以基底 2 之對數表示。

令 A 屬性含 $\{a_1, a_2, \dots, a_v\}$ 不同值，屬性 A 用來將 S 分成 v 個次集合 $\{S_1, S_2, \dots, S_v\}$ ，其中 S_j 含 S 中具 A 中 a_j 值之樣本，若 A 被選為測試屬性（即為最佳分割屬性），則其子集合對應於 S 集合中含此節點之分支。令 s_{ij} 為子集合 S_j 中類 C_i 之樣本數，這系統亂度或由分

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j}, \dots, s_{mj}) \quad (2)$$

其中 $\frac{s_{1j} + \dots + s_{mj}}{s}$ 為第 j 項子集合之權重，為 S 中全部樣本分出之子集合（含 A 中 a_j 值）樣本數。系統亂度值愈小，則子集合之純淨度愈高，對於一已知子集合 S_j ，

$$I(s_{1j}, s_{2j}, \dots, s_{mj}) = -\sum_{i=1}^m P_{ij} \log_2(P_{ij}) \quad (3)$$

其中 $P_{ij} = \frac{s_{ij}}{|S_j|}$ 且為 S_j 樣本屬於類 C_i 之機率。 A 分支 $|S_j|$

之編碼資訊為

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A) \quad (4)$$

換言之， $Gain(A)$ 為 A 屬性值而減少系統亂度值。演

演算法計算每一屬性之資訊增益，選含最大資訊增益之屬性為 S 集合之測試屬性，建立此節點並標記為該屬性，以每一屬性值建立分支，並據以切割樣本。

綜合言之，決策樹推論演算法可用於許多應用領域之分類，這些系統並未使用領域知識，決策樹推論之學習及分類步驟非常快速。整體說來，鎖定顧客的回應率較高 這是用 RFM 分析無法預測到的結果。

2.4 吉尼指標法

多樣性 (diversity) 的概念 (反之則為純

淨度 purity) 是決策樹法之核心，包括系統亂度值 (Entropy) 及 吉尼指標 (Gini Index)。吉尼指標是以一位義大利經濟學家命名，計算吉尼指標時，可以設想取出母體裡的某一成員，且在下次取出之前先將它放回。此多樣性指標是第二次取出的成員與第一次取出的，分屬不同類別的機率。多樣性指標值高表示集合裡幾種類別的分佈平均；指標值低表示其中有一種類別較他類多出許多。最好的分割條件是能夠讓記錄集合之多樣性指標值降低最多的。[4]

2.5 多模式表決

在資料探勘領域，由機械學習產生的模式可靠性是依靠訓練資料之量及品質，與學習演算法是否適合目前的問題而定，一種比較可靠作決策的方法是合併幾種不同模式的輸出，合併由機械學習產生的模式的比較有效方法是收納法 (bagging) 加權收納法 (boosting)。它們比使用單一模式更能提升預測效益。[1,2]

Bagging 及 boosting 皆使用集合多模式法，作數值預測及分類。

合併不同模式之決策意即合併不同的輸出成單一預測，在分類上此法最簡單者即投票（或加權之投票），在數值預測時求其平均（或加權之平均）Bagging 及 boosting 皆採用此法，但它們以不同的方法導出個別模式，bagging 模式是同等加權，boosting 給較有影響力之模式較大加權，正如一位經理依專家群之不同經歷分別給所提建議不同的價值。

先介紹 bagging，假設由問題定義域隨機選出幾個相同大小的訓練資料集。假想用一特定機器學習技巧以建立各資料集之一決策樹，你可想像這些樹實際相同且作每一新測試集之相同預測。但這些假設是錯的，特別當訓練資料集很小。

若思考由機械學習產生的模式為個別決策樹，我們可將每一測試案例投票以合併決策樹，若某類獲較多投票，則取此正確者，通常以投票預測時，獲較多票者會被列入考慮，若作用於新訓練集，投票預測之決策不會變壞，合成分類通常會比僅由一資料集所建立的決策樹更正確。

可由偏差一變異分解 (bias-variance decomposition) 之理論說明合成多假說之效果，假設我們有大量相同長度之獨立訓練集，使用其產生不定數目之分類，一測試案例由所有分類法作處理再由多數投票決定答案，在理想的情況，由於學習設計不是完美故仍有誤差，誤差率依所用機械學習方法是否符合題目而定。

一分類法之所有期望誤差乃是由偏差及變異之加總而得，即偏差一變異分解。合併多分類法可以減少變異降低期望誤差。包含愈多的分類法，愈可減少變異。

通常僅有一組訓練集，不可能得到更多資料。Bagging 試著將上述方法用在已知訓練集，以消除學習方法之不穩定，每一次並不是用新樣本，而是將原訓練資料刪除某些案例且重複其餘者，隨機取樣，替代以建立一新資料集，因 bootstrap 法估計學習法之概化誤差，bagging 代表 bootstrap aggregating。Bagging 將學習設計用於每一人工推導之資料集，由投票產生之分類法以供預測。

Bagging 及理想程序之差異在於導出訓練集之方法，bagging 並非由定義域得到獨立資料集，而是由原訓練資料中再取樣，再取樣之各資料集皆不同，但由 bagging 程序導出之合成模式一定比單一模式好很多。

三、個案研究

M 公司是一家型錄郵購公司，打算花費 2,000,000 元來促銷，並希望儘可能有效地運用這筆錢。促銷活動的預算必須控制在某個範圍之內，因此必須儘量針對可能會購買的顧客來加以促銷。如何得知誰可能會有興趣？有四種常用的方法可用：

- 從名單中，隨機選擇顧客，或推測有興趣的人會是誰。
- 進行 RFM 分析，鎖定最近曾消費多次、花費大額金錢的顧客。
- 根據過去的經驗，建立資料探勘預測模型，來預測誰可能會對此商品有興趣。
- 運用資料勘查(data exploration)以及非監督性資料探勘，找出可能對商品有興趣的人。

3.1 累計獲益圖

傳統促銷活動中，M 公司每寄出一件商品，需要花 20 元。這 20 元包含了型錄製作、印刷、郵寄的成本。且最少需寄出 20,000 封信。

M 公司有 2,000,000 促銷預算，因此可以寄給 100,000 名顧客。可以輕易地發現，100,000 大於 20,000，因此符合此促銷活動的最低需求。最簡單的方法，是將所有的顧客加以評分，針對分數最高的前 100,000 名加以促銷。

圖 1 之累計獲益圖(cumulative gains chart)是觀察資料探勘結果常用的圖表。此圖的目的，是提升促銷績效，如：依表 2RFM 法得知只要針對所有顧客的 10%促銷，就可以接觸到所有對該產品有興趣的顧客中的 17.03%。依表 4dti 加權 RFM 法得知只要針對所有顧客的 10%促銷，就可以接觸到所有對該產品有興趣的顧客中的 20.39%。

在此例中，可以利用資料探勘來將促銷活動最佳化。建立模型來評估顧客對某產品的偏

好，可以讓我們接觸更多對該產品有興趣的顧客(從本來的 10%加到 17.03%)。

3.2 促銷成本評估

累計獲益圖可以提供令人驚異的因素，卻無法讓使用者得知該模型的量化評估。促銷活動可能真的符合預算，而該模型也可能藉此找到更多對產品有興趣的顧客，但此促銷活動會為公司帶來收益嗎？這是在預算最佳化之後，下一步需要了解的。要進行量化評估，需先發展淨利模型(net revenue model)。淨利模型需要一些額外的資訊才能進行：

在這次的促銷活動中，對該商品有興趣的顧客，要花費大約 2000 來訂購該商品。在這 2000 元中，包含該商品的成本、運送及搬運費 200 元。

以上是發展一套簡易淨利模型所需要的資訊。資訊中的兩個關鍵項目是：

郵寄型錄的成本是 20 元。b. 賣出一件商品的淨利是 2000-200=1800 元。

通常會將這兩個項目表示成本矩陣(cost matrix, 見表 1) [4]。例如，預期會購買產品，且真的有購買的顧客，會為公司帶來 1780 元的收入。為何是 1780 元？是由其淨利減去郵寄目錄的成本而得來。預期會購買產品，但沒有購買的顧客，使公司損失 20 元，也就是郵寄型錄的成本。

表 1 M 公司成本矩陣

| 預測是否購買 | 實際有購買 | 實際沒有購買 |
|--------|-------|--------|
| 有 | 1780 | -20 |
| 沒有 | 0 | 0 |

在此必須注意的是，在淨利模型中，只有用到成本矩陣中的預測有購買那一行。這麼做是正確的。促銷活動所接觸到的顧客，是公司預期對該產品有興趣的顧客。由於預測不會購買的顧客，並非促銷活動鎖定的目標，因此在該行中，就沒有成本或淨利。換句話說，在淨利模型中，沒有用到預測不會購買該產品的顧客。

要預測收益情況，還需要其他三項資訊。第一是預估一般大眾中，會購買該產品的比例。累計獲益圖可以告訴我們，執行一個特定模型可以讓結果變好多少，但並沒有說到底有多少人會購買該產品。

第二項資訊是促銷活動的固定成本。有時，經常性的成本也會用來當成促銷活動規模的下限。之前曾經提到，M 公司促銷活動的下限是 20,000 名顧客。這代表如果該活動最低促銷人數 20,000 人都對該產品沒興趣的話，該活動的經常性成本是 400,000 元。

最後一項資訊是郵寄名單的潛在規模。此項資訊必須轉換成一個數字，就好像 3%的購

買率轉換成金額，如 450,000 元一樣。對 M 公司來說，我們假設有 100 萬人收到郵購型錄。

表 2 顯示根據以上假設，所計算出促銷活動帶來的收益(每名購買商品顧客的淨利為 1780 元、沒有購買商品的顧客需花費成本 20 元、經常性成本為 400,000 元、整體人口中，會購買此商品的比例為百分之一、共有一百萬名顧客)。第一欄是百分之幾的顧客，下三欄是此模型成效的測量。GAIN 代表在該百分比中，有購買該商品顧客的比例，其範圍是從 0% 到 17.03%。CUM 是累計收益(cumulative gain)，為所有截至該百分比為止，所有收益的綜合。前百分之十的顧客內，含有所有購買該產品顧客的 17.03%。百分之十至百分之二十之間的顧客，則含有所有購買該產品顧客的 11.98% 以上。總和來說，前百分之二十的顧客，含有全部購買該產品顧客中的 29.01%。LIFT 這一欄，則顯示截至目前的百分比，使用此模型的成果較隨機選取模型好多少。

GAIN、CUM、以及 LIFT 的數字皆可互相交換。意思是說，給定其中一個欄，就可以算出其他兩欄的值。很重要的兩個欄位是人數(有)和人數(無)，分別代表截至目前的百分比，預測出來會購買和不會購買該商品的顧客數目。因此，在百分之十到二十之間，預測出來會購買商品的人數是 2,901，不會購買的人數是 197,099，總和是 200,000 人。利用這些數字，再加上之前敘述過的收益矩陣(profit matrix)，就可計算出公司的收益。

必須注意的是，收益從 -400,000 元開始計算，在前百分之二十的時候，到達最大值。之後便開始減少。此為這類曲線的正常狀態。另一方面，也可由表中看出，將型錄目錄寄給所有的顧客是不划算的。若只寄給前百分之二十的顧客，則是划算的。了解收益曲線可以讓公司得知促銷活動虧損的原因，並加以改善。然而，相同的模型，如果採用的數字不同，產生出來的收益曲線會有很大的不同。

表 3 顯示 RFM 積分經貝氏法加權後所產生的收益矩陣。表 4 顯示 RFM 積分經決策樹歸納法加權後所產生的收益矩陣。表 5 顯示 RFM 積分經吉尼指標法加權後所產生的收益矩陣。

表 6 顯示 RFM 積分經多模式表決法加權後所產生的收益矩陣。產生的步驟為：

a. 給每一位顧客設定三個分數(即三模式各設定一個分數，共三個分數)。

b. 依下式計算各模式在十一點平均反查法中各點的成功率。

$$TP + FP + TN + FN$$

c. 以各點的成功率分別加權各模式所設定的一個分數。

d. 以各模式的 F 量度分別加權各模式所

設定的各個分數。

e. 將每一位顧客的三個分數加總，依分數排序以找出可能對商品有興趣的人。

四、實證分析

M 郵購擁有大約 200 萬的顧客名單，這些人會收到定期的型錄。其而與 32 家長期配合銀行每年要寄送約 60 期，400 萬份的型錄一年就超過 400 萬份的型錄。像 M 這樣的公司，若每件郵寄成本下降一點點，或者顧客意願度上升一點點，就會增加數百萬的營收。因而資料探勘在此便可大展所長。

在一次行銷活動中，M 郵購擬由 44385 筆有效客戶中篩選 rfm 值最高的某部份以寄出型錄；先求 44385 筆有效名單之 rfm 三變數之變異係數，再以本研究加權 rfm 三變數而得每一客戶之 rfm 值，製作效益評估表。[2]

表 7 至 11 為加權 RFM 模型的效益評估。由表 12 各 RFM 模型的 F 量度知 M 郵購應採多模式表決法加權 RFM 模式由 44385 筆有效客戶中篩選 rfm 值最高的 30%，即 13315 筆以達利潤最大化，其他的效益包括透過剩餘產能再執行 3 檔篩選；能在同樣銷售資源中執行更多檔的銷售活動；找出顧客需求以減少不必要的銷售接觸；增長銷售週期；提昇顧客再銷售率，提昇顧客忠程度，提昇銷售效率。

五、結論

傳統 RFM 分析法是依 RFM 三項指標給分將分數加總即可得到各別客戶的相對價值。但缺乏 RFM 間權重之分析技術，使 RFM 喪失許多關於客戶的資訊。本研究將多模式表決法應用於 RFM 間權重之分析，以改良傳統 RFM 分析法，使直銷業者不會因衡量標準不同使結果遺漏任何具有價值的顧客，另外業者也能憑著過去的經驗來決定顧客未來的反應率。

多模式表決法比使用單一模式更能提升預測效益，由表 12 各 RFM 模型的 F 量度知多模式表決法的 F 量度決策值為 0.5042，比其他三種單一模式的 F 量度值(0.4597、0.5002、0.5009)分別提升預測效益 9.68%、0.78%、0.66%，故 M 郵購應採多模式表決法加權 RFM。

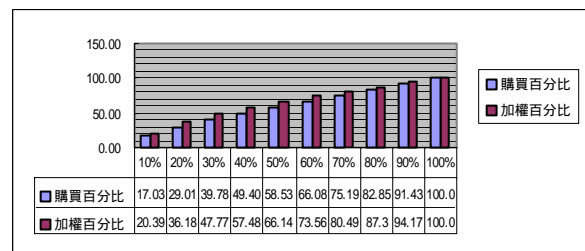


圖 1 累計獲益圖

六、參考文獻

- [1] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. USA. pp. 284-289. 2000
- [2] L. H. Witten and Eibe. Frank, Data Mining: Practical Machine Tools and Techniques with Java Implementations, Morgan Kaufmann. 2000
- [3] Dorian Pyle, "Data Preparation for Data Mining", Morgan Kaufmann. 1999.
- [4] Michael J.A. Berry and Gordon S. Linoff, "Mastering Data Mining", John Wiley & Sons, Inc. pp. 114-117. 2000.
- [5] David Shepard Associates, The New Direct Marketing: How to Implement a Profit-Driven Database Marketing Strategy, New York: Richard D. Irwin, Inc. 1995.
- [6] D.C. Schmittlein, L.G. Cooper and D.G. Morrison, "Truth in Concentration in the Land of (80/20) Laws", Marketing Science, (Spring), 12, pp. 167-183. 1993.
- [7] D.C. Schmittlein, and R.A. Peterson, "Customer Base Analysis: An Industrial Purchase Process Application", *Marketing Science*, 13, pp. 41-67. 1994.
- [8] A. M. Hughes, Strategic Database Marketing, Probus Publishing. 1994.
- [9] Bob Stone, Successful Direct Marketing Methods, 5th ed., Lincolnwood, IL. NTC Business Books. 1994.
- [10] M. L. Roberts, "Expanding the Role of the Direct Marketing Database," *Journal of Direct Marketing*, 6, pp.51-60. 1992.
- [11] Kalakota, Ravi and Marcia Robinson, E-Business: Roadmap for Success, Mass.: Addison-Wesley. 1999.
- [12] J. R. Quinlan, "Induction of Decision Trees", *Machine Learning*, vol. 1, no. 1, pp. 81-106. 1986.
- [13] J. R. Quinlan, "Simplifying Decision Trees," *International Journal of Man-Machine Studies*, 27, 3, pp. 221-234. 1987.
- [14] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann. 1993

表 2 M 公司收益矩陣(RFM 模式)

| 百分比 | GAIN | CUM | LIFT | 規模 | 人數(有) | 人數(無) | 收益 |
|------|--------|---------|-------|-----------|--------|---------|--------------|
| 0% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | -\$400.000 |
| 10% | 17.03% | 17.03% | 1.703 | 100.000 | 1.703 | 98.297 | \$665.400 |
| 20% | 11.98% | 29.01% | 1.451 | 200.000 | 2.901 | 197.099 | \$821.800 |
| 30% | 10.77% | 39.78% | 1.326 | 300.000 | 3.978 | 296.022 | \$760.400 |
| 40% | 9.62% | 49.40% | 1.235 | 400.000 | 4.940 | 395.060 | \$492.000 |
| 50% | 9.13% | 58.53% | 1.171 | 500.000 | 5.853 | 494.147 | \$135.400 |
| 60% | 7.55% | 66.08% | 1.101 | 600.000 | 6.608 | 593.392 | -\$505.600 |
| 70% | 9.11% | 75.19% | 1.074 | 700.000 | 7.519 | 692.481 | -\$865.800 |
| 80% | 7.66% | 82.85% | 1.036 | 800.000 | 8.285 | 791.715 | -\$1,487.000 |
| 90% | 8.58% | 91.43% | 1.016 | 900.000 | 9.143 | 890.857 | -\$1,942.600 |
| 100% | 8.57% | 100.00% | 1.000 | 1,000.000 | 10.000 | 990.000 | -\$2,400.000 |

表 3 M 公司收益矩陣二(貝氏法加權 RFM 模式)

| 百分比 | GAIN | CUM | LIFT | 規模 | 人數(有) | 人數(無) | 收益 |
|-----|--------|--------|-------|---------|-------|---------|-------------|
| 0% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | -\$400.000 |
| 10% | 19.03% | 19.03% | 1.903 | 100.000 | 1.903 | 98.097 | \$1,025.400 |
| 20% | 13.08% | 32.11% | 1.606 | 200.000 | 3.211 | 196.789 | \$1,379.800 |
| 30% | 11.79% | 43.90% | 1.463 | 300.000 | 4.390 | 295.610 | \$1,502.000 |
| 40% | 10.19% | 54.09% | 1.352 | 400.000 | 5.409 | 394.591 | \$1,336.200 |

| | | | | | | | |
|------|-------|---------|-------|-----------|--------|---------|--------------|
| 50% | 9.56% | 63.65% | 1.273 | 500.000 | 6.365 | 493.635 | \$1,057,000 |
| 60% | 7.73% | 71.38% | 1.190 | 600.000 | 7.138 | 592.862 | \$448,400 |
| 70% | 7.04% | 78.42% | 1.120 | 700.000 | 7.842 | 692.158 | -\$284,400 |
| 80% | 7.09% | 85.51% | 1.069 | 800.000 | 8.551 | 791.449 | -\$1,008,200 |
| 90% | 7.09% | 92.60% | 1.029 | 900.000 | 9.260 | 890.740 | -\$1,732,000 |
| 100% | 7.40% | 100.00% | 1.000 | 1,000.000 | 10.000 | 990.000 | -\$2,400,000 |

表 4 M 公司收益矩陣三(決策樹歸納法加權 RFM 模式)

| 百分比 | GAIN | CUM | LIFT | 規模 | 人數(有) | 人數(無) | 收益 |
|------|--------|---------|-------|-----------|--------|---------|--------------|
| 0% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | -\$400,000 |
| 10% | 20.39% | 20.39% | 2.039 | 100.000 | 2.039 | 97.961 | \$1,270,200 |
| 20% | 15.79% | 36.18% | 1.809 | 200.000 | 3.618 | 196.382 | \$2,112,400 |
| 30% | 11.59% | 47.77% | 1.592 | 300.000 | 4.777 | 295.223 | \$2,198,600 |
| 40% | 9.71% | 57.48% | 1.437 | 400.000 | 5.748 | 394.252 | \$1,946,400 |
| 50% | 8.66% | 66.14% | 1.323 | 500.000 | 6.614 | 493.386 | \$1,505,200 |
| 60% | 7.42% | 73.56% | 1.226 | 600.000 | 7.356 | 592.644 | \$840,800 |
| 70% | 6.93% | 80.49% | 1.150 | 700.000 | 8.049 | 691.951 | \$88,200 |
| 80% | 6.81% | 87.30% | 1.091 | 800.000 | 8.730 | 791.270 | -\$686,000 |
| 90% | 6.87% | 94.17% | 1.046 | 900.000 | 9.417 | 890.583 | -\$1,449,400 |
| 100% | 5.83% | 100.00% | 1.000 | 1,000.000 | 10.000 | 990.000 | -\$2,400,000 |

表 5 M 公司收益矩陣四(Gini 指標法加權 RFM 模式)

| 百分比 | GAIN | CUM | LIFT | 規模 | 人數(有) | 人數(無) | 收益 |
|------|--------|---------|-------|-----------|--------|---------|--------------|
| 0% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | -\$400,000 |
| 10% | 20.39% | 20.39% | 2.039 | 100.000 | 2.039 | 97.961 | \$1,270,200 |
| 20% | 15.79% | 36.18% | 1.809 | 200.000 | 3.618 | 196.382 | \$2,112,400 |
| 30% | 11.65% | 47.83% | 1.594 | 300.000 | 4.783 | 295.217 | \$2,209,400 |
| 40% | 9.65% | 57.48% | 1.437 | 400.000 | 5.748 | 394.252 | \$1,946,400 |
| 50% | 8.66% | 66.14% | 1.323 | 500.000 | 6.614 | 493.386 | \$1,505,200 |
| 60% | 7.42% | 73.56% | 1.226 | 600.000 | 7.356 | 592.644 | \$840,800 |
| 70% | 6.93% | 80.49% | 1.150 | 700.000 | 8.049 | 691.951 | \$88,200 |
| 80% | 6.81% | 87.30% | 1.091 | 800.000 | 8.730 | 791.270 | -\$686,000 |
| 90% | 6.87% | 94.17% | 1.046 | 900.000 | 9.417 | 890.583 | -\$1,449,400 |
| 100% | 5.83% | 100.00% | 1.000 | 1,000.000 | 10.000 | 990.000 | -\$2,400,000 |

表 6 M 公司收益矩陣四(多模式表決法)

| 百分比 | GAIN | CUM | LIFT | 規模 | 人數(有) | 人數(無) | 收益 |
|------|--------|---------|-------|-----------|--------|---------|--------------|
| 0% | 0.00% | 0.00% | 0 | 0 | 0 | 0 | -\$400,000 |
| 10% | 20.87% | 20.87% | 2.087 | 100.000 | 2.087 | 97.913 | \$1,356,600 |
| 20% | 15.71% | 36.58% | 1.829 | 200.000 | 3.658 | 196.342 | \$2,184,400 |
| 30% | 11.57% | 48.15% | 1.605 | 300.000 | 4.815 | 295.185 | \$2,267,000 |
| 40% | 9.68% | 57.83% | 1.446 | 400.000 | 5.783 | 394.217 | \$2,009,400 |
| 50% | 8.68% | 66.51% | 1.330 | 500.000 | 6.651 | 493.349 | \$1,571,800 |
| 60% | 7.38% | 73.89% | 1.232 | 600.000 | 7.389 | 592.611 | \$900,200 |
| 70% | 6.82% | 80.71% | 1.153 | 700.000 | 8.071 | 691.929 | \$127,800 |
| 80% | 6.78% | 87.49% | 1.094 | 800.000 | 8.749 | 791.251 | -\$651,800 |
| 90% | 6.78% | 94.27% | 1.047 | 900.000 | 9.427 | 890.573 | -\$1,431,400 |
| 100% | 5.73% | 100.00% | 1.000 | 1,000.000 | 10.000 | 990.000 | -\$2,400,000 |

表 7 RFM 模型的效益評估

| | | | | | | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 反查 | 0.1703 | 0.2901 | 0.3978 | 0.4940 | 0.5853 | 0.6608 | 0.7519 | 0.8285 | 0.9143 | 1.0000 |
| 反應率 | 0.5608 | 0.4775 | 0.4366 | 0.4067 | 0.3854 | 0.3626 | 0.3537 | 0.3410 | 0.3345 | 0.3293 |
| 名單縮減 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| F 量度 | 0.2613 | 0.3609 | 0.4163 | 0.4461 | 0.4648 | 0.4683 | 0.4810 | 0.4831 | 0.4898 | 0.4954 |

表 8 貝氏法加權 RFM 模型的效益評估

| | | | | | | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 反查 | 0.1903 | 0.3211 | 0.4390 | 0.5409 | 0.6365 | 0.7138 | 0.7842 | 0.8551 | 0.9260 | 1.0000 |
| 反應率 | 0.6275 | 0.5293 | 0.4824 | 0.4458 | 0.4196 | 0.3922 | 0.3693 | 0.3524 | 0.3392 | 0.3297 |
| 名單縮減 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| F 量度 | 0.2921 | 0.3998 | 0.4597 | 0.4887 | 0.5058 | 0.5062 | 0.5021 | 0.4991 | 0.4965 | 0.4959 |

表 9 決策樹歸納法加權 RFM 模型的效益評估

| | | | | | | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 反查 | 0.2039 | 0.3618 | 0.4777 | 0.5748 | 0.6614 | 0.7356 | 0.8049 | 0.8730 | 0.9417 | 1.0000 |
| 反應率 | 0.6723 | 0.5964 | 0.5249 | 0.4738 | 0.4360 | 0.4042 | 0.3791 | 0.3597 | 0.3449 | 0.3297 |
| 名單縮減 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| F 量度 | 0.3129 | 0.4504 | 0.5002 | 0.5194 | 0.5256 | 0.5217 | 0.5154 | 0.5095 | 0.5049 | 0.4959 |

表 10 Gini 法加權 RFM 模型的效益評估

| | | | | | | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 反查 | 0.2039 | 0.3618 | 0.4783 | 0.5748 | 0.6614 | 0.7356 | 0.8049 | 0.8730 | 0.9417 | 1.0000 |
| 反應率 | 0.6723 | 0.5964 | 0.5256 | 0.4738 | 0.4360 | 0.4042 | 0.3791 | 0.3597 | 0.3449 | 0.3297 |
| 名單縮減 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| F 量度 | 0.3129 | 0.4504 | 0.5009 | 0.5194 | 0.5256 | 0.5217 | 0.5154 | 0.5095 | 0.5049 | 0.4959 |

表 11 多模式表決法的效益評估

| | | | | | | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 反查 | 0.2087 | 0.3658 | 0.4815 | 0.5783 | 0.6651 | 0.7389 | 0.8071 | 0.8749 | 0.9427 | 1.0000 |
| 反應率 | 0.6881 | 0.6030 | 0.5292 | 0.4766 | 0.4385 | 0.4060 | 0.3801 | 0.3605 | 0.3453 | 0.3297 |
| 名單縮減 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| F 量度 | 0.3203 | 0.4554 | 0.5042 | 0.5226 | 0.5286 | 0.5241 | 0.5168 | 0.5107 | 0.5054 | 0.4959 |

表 12 多模式表決的 F 量度比較

| | | | | | | | | | | |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 名單縮減 | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
| RFM 法 | 0.2613 | 0.3609 | 0.4163 | 0.4461 | 0.4648 | 0.4683 | 0.481 | 0.4831 | 0.4898 | 0.4954 |
| 貝氏法加權 | 0.2921 | 0.3998 | 0.4597 | 0.4887 | 0.5058 | 0.5062 | 0.5021 | 0.4991 | 0.4965 | 0.4959 |
| dti 法加權 | 0.3129 | 0.4504 | 0.5002 | 0.5194 | 0.5256 | 0.5217 | 0.5154 | 0.5095 | 0.5049 | 0.4959 |
| Gini 法加權 | 0.3129 | 0.4504 | 0.5009 | 0.5194 | 0.5256 | 0.5217 | 0.5154 | 0.5095 | 0.5049 | 0.4959 |
| 多模式表決 | 0.3203 | 0.4554 | 0.5042 | 0.5226 | 0.5286 | 0.5241 | 0.5168 | 0.5107 | 0.5054 | 0.4959 |