

顧客交易資料庫之探勘 ——以網路電話公司之非契約型顧客為例

周世玉

中興大學行銷學系

蕭登泰

交通大學管理科學系

摘要

顧客交易資料庫蘊含對實施顧客關係管理不可或缺的重要資訊。本研究利用國內某網路電話公司顧客交易資庫的非契約型顧客實際交易資料，估計個別顧客關係危險率，進而分析顧客關係長度與顧客終身價值，以及建構各類顧客之顧客關係危險率預測模式。研究結果指出，利用倒傳遞網路估計個別顧客關係危險率時，訓練集合的 MSE 與錯誤率分別為 0.000077 與 1.30631%，而測試集合的 MSE 與錯誤率則分別為 0.002469 與 3.636028%，皆顯示出不錯的績效。估計顧客關係長度中位數的平均與剩餘關係長度中位數的平均分別為 9.14 個月與 1.97 個月。在顧客終身價值方面，平均每位顧客的價值約為 964 元，共可產生約 615 萬元的價值。以各期估計自我相關係數與偏自我相關係數為自變數對顧客進行分群，將顧客分為兩群，集群一的顧客顧客危險率預測適合 AR(1) 模式，而集群二的顧客顧客危險率預測則較適合白噪音模式。

關鍵字：顧客關係管理、資料庫、類神經網路

A Data Mining Study on Customer Transaction Data Base – Using Non-contracted Customers of a Domestic I-Phone Company as an Example

Shihyu Chou

Department of Marketing, National Chung Hsing University

Teng-Tai Hasio

Department of Management Sciences, National Chao Tung University

Abstract

Customer transaction databases contain relevant information that is related to the implementation of customer relationship management practice. In this study, we use transaction records of non-contract type customers of a domestic internet phone company to estimate individual customer relationship hazard rates, analyze customer lifetime length and lifetime value, and construct customer relationship hazard forecast models for different types of customers. The empirical results indicate that the MSE and error rate are 0.000077 and 1.30631% for the training set, and 0.002469 and 3.636028% for the testing data set by using the back propagation neural network technique to estimate customer relationship hazard rates. Both highlight good performance. The medians of average customer relationship length and average remaining relationship length are 9.14 and 1.97 months respectively. The average customer value is about 964 dollars which is equivalent to 615 million dollars in total. Customers are separated into two groups by using autocorrelation functions and partial autocorrelation functions. The customer relationship hazard rate forecasting models for the two groups are AR(1) and white noise models, respectively.

Keywords: Customer Relationship Management, Database, Neural Network

壹、導論

資訊科技爆炸性的發展，對於企業經營具有深遠的影響。完整蒐集顧客交易資料並非難事，愈來愈多企業將顧客交易資料的記錄與分析視為主導企業經營的重要工作之一。從龐大的顧客交易資料庫分析顧客交易特性，是目前顧客關係管理的重要課題。

顧客關係管理的本質在於長期關係的建立，企業為強化與顧客間的關係程度，會發展顧客管理方案來吸引顧客持續且密集地與企業進行交易。基於長期關係的建立，如何衡量顧客關係的程度便成為一項重要的管理議題。衡量顧客關係的指標大多為交易金額、交易發生頻率、或交易產品數量等構面，但只從這些指標較無法得知顧客在未來可能與企業終止交易關係的機率。Drew et al.(2001)及 Helsen & Schmittlein(1993)利用顧客關係危險率衡量顧客關係程度。顧客關係危險率指顧客交易關係延續到 t 期期初，顧客於 t 期終止交易關係的機率。企業使用危險率衡量顧客關係程度，有助於追蹤可能流失的顧客，以利顧客管理活動的執行。

許多研究所使用的危險率估計方法中，以參數估計模型、無參數估計模型及 Cox 的比例危險迴歸模型 (proportional hazard model) 較常見，例如 Drew et al. (2001), Helsen & Schmittlein (1993)，及 Lawless (1982)等。大部分的方法應用於估計顧客關係危險率時，會受到限制，尤其是顧客異質程度較高的產業。企業欲針對顧客交易特性來發展顧客管理方案，必須將顧客異質性納入顧客分析中。

企業若能估計出個別顧客終止交易關係的機率（即顧客關係危險率），將有助於顧客管理決策的制定。Drewet et al.(2001)認為一般估計危險率的方法並未將顧客異質性納入考慮，因而提出結合統計分析與資料探勘技術的方法來估計顧客關係危險率。本研究應用倒傳遞神經網路 (back-propagation neural network, BPN) 分析技術來獲得個別顧客關係危險函數。倒傳遞神經網路的運作是透過不斷調整鍵結值來學習輸出變數與輸入變數之間的關係，不需要對資料假設特定機率分配，並可將顧客異質性納入考慮。倒傳遞神經網路架構中的隱藏層，可用以表現變數間的交互作用，以捕捉複雜且非線性的函數關係。

顧客關係長度的估計對顧客管理決策的制定具有重要意義。若企業能估計顧客關係長度，將有助於選擇適當執行顧客管理的時機。顧客終身價值的計算，有助於分類不同重要程度的顧客，以有效分配行銷資源於顧客關係管理。顧客關係危險率的預測使顧客關係管理具有未來導向，而顧客關係危險率預測模型的建構將有助於預測觀察期間外顧客終止交易關係的機率。基於以上的研究動機，本研究之研究目的可歸納為下列三點：

1. 用倒傳遞神經網路的分析技術來估計個別顧客關係危險率，使危險率的估計能考慮個別顧客的交易特性。
2. 以估計之顧客關係危險率為基礎，進行顧客關係長度分析與顧客終身價值分析，以協助顧客關係管理決策之制定。

3. 建構顧客關係危險率之預測模型，以利制定適當的顧客管理決策來維持顧客關係。

本文之組織架構共分為五節。在第一節中，我們敘述研究動機與研究目的，第二節為相關文獻之回顧，第三節為研究方法之介紹，第四節為資料分析結果與解釋，而最後一節則為研究結論。

貳、文獻探討

一、顧客關係管理與顧客終身價值

由於市場競爭越來越激烈，資訊越來越透明化，顧客關係管理逐漸成為企業重視的議題。隨著行銷概念的演進，越來越多企業將焦點放在維持長久且忠誠的顧客關係上，期望透過顧客關係的維持與改善來增加獲利。Wyner (1996)指出顧客關係的強化與維持，能夠為企業帶來充沛的利潤與持續性的競爭優勢。顧客關係管理指「企業為獲取策略發展的優勢，深入分析及探索顧客資料以了解顧客與企業的關係型態，並藉此發展有效的行銷活動來建立、維持或強化長期的顧客關係，以提高顧客忠誠度與利潤貢獻的策略性分析方法」。顧客關係管理的首要工作在於深入分析顧客交易資料，以依據所獲得的資訊發展適用於各類型顧客的行銷策略。

量測顧客與企業之間的交易關係是一項重要的管理議題。衡量顧客關係的指標大多為交易金額、交易發生頻率、或是交易品項數目等；然而這些指標並不能讓企業獲知顧客持續維持與企業間關係的可能性。顧客關係終止機率可作為測量顧客關係強度的有效指標，它能讓企業在顧客流失之前，適時針對可能離開的顧客進行顧客關係管理的措施，以延續及鞏固顧客關係。

許多研究以顧客關係危險率來衡量顧客關係強度或顧客關係終止機率，例如 Drew et al.(2001)與 Helsen & Schmittlein (1993)等。企業大多想知道顧客與公司之間的交易往來維持一段時間後，顧客會流失的可能性。本研究提議利用危險率做為顧客關係強度的衡量指標，以有助於追蹤可能流失的顧客，而能幫助企業決定顧客關係管理活動執行的時間點，使投入顧客關係管理的資源能有效的被使用。

隨著一對一行銷概念的出現及資訊科技的蓬勃發展，企業越來越重視且越有能力分析個別顧客的終身價值。Hansotia & Wang(1997)指出藉由量化分析顧客所產生的價值貢獻，可使行銷策略的規劃與執行更有依據。顧客終身價值分析可協助企業依重要性分類顧客，而能將有限的行銷資源依照顧客重要性來分配使用。Hansotia & Wang (1997)認為企業在執行顧客關係管理活動時，應對單次購買量大、購買頻率頻繁且與企業能維持長期關係的顧客投入較多的行銷資源。Mulhern(1999)指出大多數企業的利潤集中於少數有重大價值貢獻的顧客，所以顧客終身價值的有效計算是顧客關係管理的非常重要的工作之一。

Berger & Nasr (1998)指出施行關係行銷的成本很高，不宜對每位顧客投入相同的行銷資源，而應針對高價值貢獻的顧客提供較佳的服務。Blattberg & Deighton (1996)指出顧客價值最大化的八項顧客關係管理之建議，並提出顧客權益模型來計算被獲取的潛在顧客終身價值及被維持的現有顧客的終身價值。基於行銷資源有限的觀點，Mulhern (1999)認為較多行銷資源應分配於能為企業產生較大價值貢獻的顧客，他提出顧客歷史價值模型與顧客未來價值模型以衡量顧客過去以及未來的價值貢獻。

二、存活分析理論與顧客關係危險率

存活分析理論使用統計方法，分析觀察個體在未來某時間點仍然會持續存在的機率；並藉由存活函數與危險函數的推導，研究觀察個體存活於系統的生命長度等性質。Drye et al. (2001)與 Harrison & Ansell(2002)均指出存活分析的資料型態牽涉時間因素，因此它是一個探索資料動態關係的方法。在存活分析中，危險函數是存活事件的重要函數。危險率函數是指已知觀測個體至少存續到 t 期的機率下，該個體在 t 期瞬間死亡或未能發揮預定機能的機率。應用於顧客關係管理，危險率越高表示顧客與企業之間會繼續持續往來的可能性越低，而越有可能終止與企業的關係。

危險率函數的估計方法大致有三種：參數估計模型、無參數估計模型與比例危險迴歸模型。Drew et al. (2001)認為上述估計危險率的方法不適用於分析顧客關係管理的危險率。參數估計模型須假設顧客關係存續服從某種機率分配，例如指數分配(exponential distribution)、韋伯分配(Weibull distribution)或伽瑪分配(Gamma distribution)，並依據假設的分配估計危險率。若資料不能以某種分配描述時，以參數估計模型估計危險率並不適當。無參數估計方法包括 Kaplan-Meier 估計法與生命表估計法，這些方法在分析前並不假設事件符合特定存活機率分配，而是透過事件過去的經驗來建立模式，非常適用於事件分配未知的情況，但所估計的危險率是以所有觀測樣本為基礎而獲得的共同危險率，故無法求得個別顧客之顧客關係危險率。Lawless(1982)指出 Cox 的比例危險迴歸模型假設分析事件具有共同的基本危險函數，再根據個體間的差異比例調整基本危險函數以獲得個體之危險函數，雖然將個體異質性考慮到模型中，但卻須假設所有顧客具有相同的基礎危險函數，再依個體差異做比例調整而估計每位顧客的危險函數。然而個別顧客危險函數的型態不一定完全相同，因此比例危險迴歸亦無法符合真實的狀況。

顧客間的異質性應納入顧客關係管理的決策制定。Drew et al.(2001)認為個別顧客應該具有不同交易特性。Thomas(2001)指出在分析企業顧客交易資料時，若不將顧客的異質性考慮到分析過程中，所制定的顧客關係管理決策將無法達成其目標。

「契約型顧客」是指與企業具有契約關係的顧客。Drew et al.(2001)針對「契約型顧客」提出估計顧客關係危險率的估計方法，將契約因素納入分析方法中。對於「契約型顧客」，企業較能明確掌握發生交易的時機與感應到終止交易關係的訊號；然而有些顧客並非契約型顧客，對於這類型顧客，企業較無法迅速察覺顧客是否終止交易關係及無法獲得他們發出終止交易關係的訊息，因此採用 Drew et al.(2001)的方法來估計

非契約型顧客之危險率函數並不十分恰當。為估計非契約型顧客的顧客關係危險率函數，本研究針對非契約型顧客的交易特性，提出能估計非契約型顧客個別危險函數的方法，以延伸 Drew et al.(2001) 的所提出的估計方法。

參、研究方法

一、研究分析資料說明

本研究使用的顧客交易資料庫，是國內一家以網路電話業務為主的公司之網路電話卡訂購交易記錄。交易資料記錄的資料項目包括交易日期、交易時間、產品編號、交易金額、交易數量、交易產品、身分證字號、聯盟介紹人與購買方式等項目。其中交易日期、交易時間、產品編號、交易金額、交易數量、交易產品、和身分證字號等資料項目為該筆記錄之顧客完成交易的日期、時間、購買金額與購買電話卡張數、所購買電話卡的類型、及其身分識別等。該公司採取聯盟銷售策略，其他網站可志願掛其橫幅廣告，若顧客經由點選此橫幅廣告而被轉移至此公司，且因而完成一筆交易，則此公司將付一定比例的佣金給轉介網站。此公司稱志願掛其橫幅廣告的網站為聯盟廠商，並在每一筆成交記錄中記載此聯盟介紹人。購買方式指信用卡付款或自動櫃員機轉帳付款的方式。

由於本研究使用存活分析理論來分析顧客與此公司保持交易關係的時間長度，因此所需使用的資料項目僅包括交易日期、交易金額、與交易數量等。由於該公司並未對於顧客蒐集有關年齡、職業、甚至個人風險指標等資料，使本研究變數選用上受到限制。若能有年齡、職業、個人風險指標等其他額外資料，分析結果將更能精確。分析資料擷取自民國 89 年 5 月 1 日到民國 91 年 11 月 30 日的日交易記錄。以月為單位將資料聚合成 31 期，若某顧客與公司往來的時間越長，表示該顧客與公司之間的關係長度越長。在去除無意義的資料後（例如交易金額為零或身分證資料不正確的交易記錄），總交易筆數為 24653 筆，總交易人數為 6383 人。

二、個別顧客關係危險率之倒傳遞神經網路建構

本研究應用倒傳遞神經網路分析技術於顧客交易資料庫，以學習並尋找顧客交易資料與顧客關係危險率之間的複雜函數關係，從而可估計個別顧客的各期危險率。相對於傳統分析方法（如迴歸分析）而言，Mehra & Benjamin(1992)與 Kosko(1992)接指出倒傳遞類神經網路分析技術可利用隱藏層的建構表達變數間的交互作用，以捕捉非線性的函數關係，並能依個別顧客之交易資料，推估其顧客關係危險率函數，即考慮顧客的異質程度。

(一) 網路輸入層之建構

Drew et al. (2001)指出顧客在每期與企業的交易狀況(例如交易金額、交易數量、交易次數等)皆有可能影響顧客於未來繼續與此企業交易的機率(即顧客關係持續一段時間後，顧客不會斷絕與此企業繼續交易的可能性)。如果某位顧客交易金額、交易數量、或交易次數逐期減少，或是連續數期內未發生交易，都有可能顯示該顧客逐漸疏遠此企業，而於未來斷絕與企業的交易關係的可能性大幅提升。為估計個別顧客關係危險率函數，本研究選取「每期交易金額」、「每期交易數量」、「每期交易次數」與「當期是否有交易發生」交易記錄資料項目做為倒傳遞神經網路的輸入變數。令 X_{pt} 表示第 p 個資料項目的第 t 期之輸入層處理單元，其中 $p=1,2,3,4$, $t=1,2,\dots,31$ ，共計建構 124 個輸入變數。

(二) 網路輸出層之建構

由於倒傳遞神經網路屬於監督式學習網路，故模式須包含目標輸出變數來「監督」學習效果。本研究先利用生命表估計法來獲得顧客各期的共同危險率，並以其做為網路的目標輸出變數及網路訓練效果判斷及調整的依據。顧客關係危險率為已知顧客與企業間的交易關係持續到 t 期期初，顧客在 t 期期末與企業終止交易關係的可能性，如下式所示：

$$h_i(T,t) = \Pr(t \leq T \leq t+1 | T \geq t) \quad (1)$$

其中 T 為顧客與企業交易關係期間之持續長度， $h_i(T,t)$ 為第 i 個顧客的在時間點 t 的危險率。

由於個別顧客的交易習性不盡相同，若使用參數方法之模型來描述交易特性並不適當，因此本研究建議使用無參數之生命表法之技術來估計共同危險率。根據生命表法，各期危險率的估計式如式(2) (推導過程參考附錄)。

$$\hat{h}(t) = -\log_a \left(1 - \frac{D_t}{N_t} \right) \quad (2)$$

其中 D_t 表示在區間 I_t 觀察個體發生失效的總數； $N_t' = N_t - \alpha W_t$ ，表示區間 I_t 可觀察到的個體總數； N_t 表示在區間 I_t 期初個體仍然存活或遭到設限¹ (censored) 的數目； W_t 表示在區間 I_t 個體發生設限的數目； α 為一係數，表示設限在區間內發生的時間點²。

Drew et al.(2001)之契約型顧客的目標危險率函數如下式：

$$h_i(t) = \begin{cases} 0 & 1 \leq t \leq TENURE \\ 1 & CHURN = 1 \& TENURE < t \leq T \\ \frac{d_t}{n_t} & CHURN = 0 \& TENURE < t \leq T \end{cases} \quad (3a)$$

¹ 設限指個體的生命長度在研究時間內無法觀察所做的假設。對於發生設限的資料須假設生命長度至少大於研究期間，並在估計危險率時將設限資料剔除。

² 若設限發生在期初，則 α 值為 0；若發生在期末，則 α 值為 1。由於從資料庫之交易記錄中無法得知設限發生時間，故設 α 值為 0.5。

其中 $h_i(t)$ 表示顧客 i 在時間點 t 的目標危險率；TENURE 表示顧客的關係長度； $CHURN=1$ 表顧客發出終止契約的訊號， $CHURN=0$ 表顧客沒有發出終止契約的訊號； d_i/n_i 為 Kaplan-Meier 估計法之各期危險率估計值， d_i 表示第 t 期終止顧客關係的人數， n_i 表示第 t 期顧客關係仍然延續的人數。

由於本研究資料庫中的顧客屬於非契約型，所以無法明確得知其發出終止交易關係訊號的時間點，使得 Drew et al.(2001)的目標危險函數並不適用於本研究。針對非契約型顧客的特性，本研究建議以式(3b)取代(3a)做為非契約型顧客之目標危險率函數：

$$h_i(t)=\begin{cases} -\log\left(1-\frac{D_i}{N_i}\right) & i \in W \text{ 或 } (i \in D \quad 1 \leq t < d) \\ 1 & i \in D \quad d \leq t \leq T \end{cases} \quad (3b)$$

其中 $-\log_a(1-D_i/N_i)$ 表示生命表法所估計之各期共同顧客關係危險率，其中 N_i 表示第 t 期顧客關係仍然延續的人數， D_i 表示第 t 期終止顧客關係的人數； W 表示設限資料集合， $i \in W$ 意指顧客 i 為設限資料，即未來關係長度無法觀察的顧客； D 表示失效資料集合， $i \in D$ 指顧客 i 為失效資料，即該顧客在觀察期間已與公司終止關係； d 表示顧客終止交易關係的時間點³； T 表示研究觀察的最終時間。

目標危險率函數估計完成後，即可用來建構倒傳遞神經網路之輸出層變數。令 Y_t 表示第 t 期之輸出層處理單元，即目標危險率函數且 $t = 1, \dots, 31$ 。倒傳遞神經網路之輸出層總共計建構 31 個目標輸出變數。

三、個別顧客關係平均長度估計方法

假設 t_i^* 為顧客 i 之顧客關係長度， t_{0i} 為顧客 i 在研究觀察期間內的關係存續長度，則 $t_i^* - t_{0i}$ 為顧客 i 剩餘的關係長度。若顧客屬於失效資料，則 $t_{0i} = t_i^*$ 。 $P(t_i^*, t_{0i})$ 為已知顧客關係存續到 t_{0i} 期，可再存續到 t_i^* 期的機率。因為存活函數 $\hat{s}_i(t)$ 可藉由利用交易資料推算(3b)而估算出，所以已知關係持續至 t_{0i} 之下，再持續至 t_i^* 的累積機率 $P(t_i^*, t_{0i}) = \hat{s}_i(t_i^*) / \hat{s}_i(t_{0i})$ 即可由資料推算出。對於 $t_i^* - t_{0i}$ 的機率分配若假設其為某種型式，則須利用資料驗證其適當性，因本研究缺乏足以驗證此分配適當性的資料，所以無法求算 $t_i^* - t_{0i}$ 的期望值。Drew et al. (2001)指出若不知危險函數的機率分配，可以使用 $t_i^* - t_{0i}$ 之中位數代替其期望值來表示 $t_i^* - t_{0i}$ 的集中趨勢，以進而估計 t_i^* 之中位數；換言之，即從 $P(t_i^*, t_{0i}) = \hat{s}_i(t_i^*) / \hat{s}_i(t_{0i}) = 0.5$ 解出 $t_i^* - t_{0i}$ 之中位數，再推算 t_i^* 之中位數。

四、終身價值計算模型

本研究使用 Mulhern (1999) 所提出的模型來計算顧客的終身價值。Mulhern (1999) 建議將顧客關係存續期間分為成兩部分，即已存在的關係期間 $[0, t_0]$ 與未來剩餘的關係期間 $[t_0+1, t_i^*]$ ，再分別算出式(4)的歷史終身價值與式(5)的未來終身價值。本研究以

³ 顧客是否終止關係是以「顧客是否連續一段時間（本研究假設 6 個月）未與公司發生交易」為條件決定，意指若顧客連續 6 個月未與公司發生交易，則該顧客將被視為與公司終止顧客關係。

存款加權平均利率做為顧客價值的幣值調整因子。

$$HV_i = \sum_{t=0}^{i_j} [(1 - c - m)R_t] (1 + r)^{i_j - t} \quad (4)$$

$$FV_i = \sum_{t=t_0+1}^{i_j} \frac{(1 - c - m)E(R_t)}{(1 + r)^{k-t_0}} \quad (5)$$

其中 HV_i 為顧客 i 的歷史價值， FV_i 為顧客 i 的未來價值， $E(R_t)$ 為顧客 i 帶來的平均收益， r 為折現因子。若顧客屬於失效集合，則顧客終身價值等於歷史價值；若顧客屬於設限集合，則顧客終身價值等於歷史價值與未來價值的加總。

五、顧客關係危險率預測模型建構方法

為取得觀察期間外的顧客關係危險率，須建構顧客危險率預測模式。首先計算個別顧客危險率的自我相關函數(autocorrelation function, ACF)與偏自我相關函數(PACF, partial autocorrelation function)，再利用集群分析將顧客依據 ACF 與 PACF 相似程度分群，然後以各集群中心的各期 ACF 與 PACF 來建構各集群之危險率預測模式。本研究使用 Box & Jenkins(1976)的 ARIMA 模型建構預測模式，Wei (1990)建議分析者可使用以下之 ARIMA 模式建構步驟：

- (1)辨認候選模式：根據各集群 ACF 與 PACF 圖研判及挑選可能的 ARIMA 模式。
- (2)估計模式參數：根據步驟(1)所得的候選模式，利用樣本資料估計其參數。
- (3)模式診斷分析：參數估計後，應進行模式診斷分析，包括估計參數顯著性檢定及殘差分析。若有參數估計值不顯著或殘差不符合假設，則回到步驟(1)與步驟(2)調整候選模式，並重新估計參數；否則前進至步驟(4)
- (4)使用預測模式預測觀察期間外的關係危險率。

肆、資料分析

一、生命表估計法分析

本研究利用生命表估計法來估計各期顧客關係之共同危險率。圖 1 為各期失效人數、設限人數與顧客總數之趨勢圖，共計 31 期； $D(t)$ 為失效人數， $W(t)$ 為設限人數， $N(t)$ 為顧客關係仍然存續的人數。由圖 1 可知，觀察開始時的總顧客數為 6383 人，隨著期間經過，顧客數逐漸減少。在失效人數方面，共有 4069 位（約 64%）顧客在觀察期間內終止交易關係，其中第八期顧客終止交易關係的人數劇增（2920 人），顯示個案公司將近半數的顧客屬於短期交易型顧客。在設限人數方面，共有 2314 位（約 36%）顧客的關係長度無法觀察而遭設限，然而前面幾期的設限人數較後面幾期為多，顯示設限顧客有較大比例為新顧客。

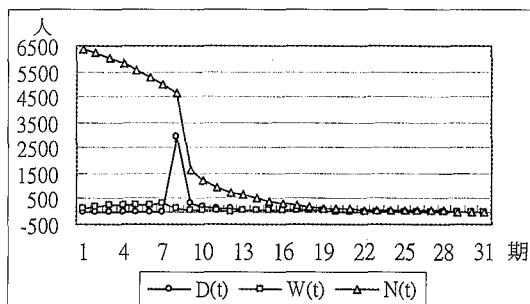


圖 1：失效人數、設限人數與顧客總數之趨勢圖

圖 2 為各期顧客關係之估計共同危險率與存活率，其中 $h(t)$ 為各期顧客關係之估計共同危險率， $s(t)$ 為各期顧客關係之估計共同存活率。由圖 2 可知，第八期的估計危險率 (0.43647) 為所有期間最高，顯示大多數的顧客在第八期與公司終止交易關係。一般而言，顧客對於長期往來的企業較熟悉且信任，較不容易終止交易關係，故顧客關係危險率應隨時間經過而降低。從圖 2 危險率的趨勢變化可知，隨著顧客關係的長度越長，顧客關係危險率逐漸下降，表示個案公司的顧客會隨往來時間增長，而與企業維持穩定顧客關係。此外，從存活率的趨勢變化可知，隨著顧客關係長度增長，顧客關係存續的機率將遞減，遞減速度隨時間經過會趨緩。

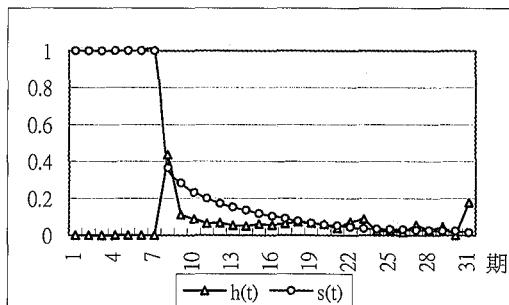


圖 2：各期顧客關係之估計共同危險率與存活率

表 1 為生命表估計法之各期估計存活率與危險率。 $I_j = [a_{j-1}, a_j)$ 表示第 j 期； $\hat{q}_j = D_j / N_j$ 表示個體至少存活到 I_{j-1} ，而在 I_j 失效的條件機率； $\hat{p}_j = 1 - \hat{q}_j$ ，表示個體至少存活到 I_{j-1} ，繼續存活到 I_j 的條件機率； $\hat{P}_j = \hat{p}_1 \hat{p}_2 \cdots \hat{p}_j$ 表示個體存活到 I_j 的機率。我們以 $Q = h(t)/s(t)$ 表示當期危險率與當期存活率的比值，以做為進行顧客管理時機選擇之參考。若 $Q > 1$ ，則表示當期顧客終止交易關係的機率大於顧客持續維持交易關係的機率，在表 1 中，第 8、19、20、22、23、27、28、29、31 期的 Q 值皆大於 1，管理者可於這些期間執行顧客關係管理方案以強化顧客關係。

表 1：生命表估計法之各期估計存活率與危險率

I_j	\hat{q}_j	\hat{p}_j	\hat{P}_j	\hat{h}_j	Q
[0,1)	0.00000	1.00000	1.00000	0.00000	0.00000
[1,2)	0.00000	1.00000	1.00000	0.00000	0.00000
[2,3)	0.00000	1.00000	1.00000	0.00000	0.00000
[3,4)	0.00000	1.00000	1.00000	0.00000	0.00000
[4,5)	0.00000	1.00000	1.00000	0.00000	0.00000
[5,6)	0.00000	1.00000	1.00000	0.00000	0.00000
[6,7)	0.00000	1.00000	1.00000	0.00000	0.00000
[7,8)	0.63396	0.36604	0.36604	0.43647	*1.19241
[8,9)	0.22538	0.77462	0.28354	0.11091	0.39116
[9,10)	0.18291	0.81709	0.23168	0.08773	0.37867
[10,11)	0.13662	0.86338	0.20003	0.06380	0.31895
[11,12)	0.14324	0.85676	0.17138	0.06714	0.39176
[12,13)	0.11589	0.88411	0.15151	0.05350	0.35311
[13,14)	0.10950	0.89050	0.13492	0.05037	0.37333
[14,15)	0.13148	0.86852	0.11718	0.06122	0.52244
[15,16)	0.11931	0.88069	0.10320	0.05518	0.5347
[16,17)	0.13514	0.86486	0.08926	0.06305	0.70636
[17,18)	0.15267	0.84733	0.07563	0.07195	0.95134
[18,19)	0.13986	0.86014	0.06505	0.06543	*1.00584
[19,20)	0.12785	0.87215	0.05673	0.05941	*1.04724
[20,21)	0.08284	0.91716	0.05203	0.03756	0.72189
[21,22)	0.15603	0.84397	0.04392	0.07367	*1.67737
[22,23)	0.18349	0.81651	0.03586	0.08804	*2.45510
[23,24)	0.04878	0.95122	0.03411	0.02172	0.63676
[24,25)	0.05479	0.94521	0.03224	0.02447	0.75900
[25,26)	0.03333	0.96667	0.03117	0.01472	0.47225
[26,27)	0.12000	0.88000	0.02743	0.05552	*2.02406
[27,28)	0.06250	0.93750	0.02571	0.02803	*1.09024
[28,29)	0.10000	0.90000	0.02314	0.04576	*1.97753
[29,30)	0.00000	1.00000	0.02314	0.00000	0
[30,31)	0.33333	0.66667	0.01543	0.17609	*11.4122

* 表示 Q 值大於 1

二、倒傳遞神經網路分析

本研究在倒傳遞類神經網路分析上，是採用 NeuroSolutions 4.21 版之軟體。我們以單層隱藏層來建構網路，即足以捕捉網路輸入與輸出變數間的關係。在處理單元數方面，輸入層建構 124 個處理單元，輸出層建構 31 個處理單元。一般而言，隱藏層處理單元數的決定並無一定公式，應該以試誤方式以取得一個較適當的數目，亦即隱藏層處理單元數應視問題而定，太少則可能無法有效處理問題的關係，太多則可能導致過度訓練。我們先以輸入層與輸出層的處理單元總數之半(即 77)做為隱藏層處理單元數的起始測試值，經不斷試誤比較，發現以使用 93 個處理單元為較佳，因此最後決定為隱藏層建構 93 個處理單元。層與層之間的連結方式以「後層第 t 期處理單元組不與前層第 $t+i$ 期變數相連結」為原則。在網路參數設定方面，以動態調整學習速率演算法(Delta-Bar-Delta algorithm, DBD algorithm)進行網路學習。經過不斷的試誤，我們將學習速率設為 0.9，加法項設為 0.1，乘法項設為 0.5，平滑因子設為 0.5，而慣性則項設為 0.7。

我們從資料庫中隨機選取 5107 人(約 80%)做為訓練集合進行網路學習，其餘 1276 人(約 20%)則做為測試所訓練的網路是否具有一般性。進行 10000 次學習循環後，倒傳遞神經網路的訓練績效如表 2 所示。對於訓練所得之網路固定其參數，並代入測試集合的資料，不再做學習調整，其測試績效亦顯示於表 1。由表 2 可知訓練集合的均方誤(MSE)為 0.000077 ($\sqrt{MSE} < 0.01$)、錯誤率為 1.30631% ($< 5\%$)，顯示網路訓練具有不錯的效果。測試集合的 MSE 為 0.002469 , ($\sqrt{MSE} < 0.05$)，錯誤率為 3.636028% ($< 5\%$)，顯示模式應具備一般性。不論是訓練或測試集合，網路輸出變數與目標輸出變數間皆具有高度相關性 ($r > 0.7$)。

表 2：倒傳遞神經網路之訓練績效與測試績效衡量表

	MSE	% Error	r
訓練集合	0.000077	1.30631	0.774072
測試集合	0.002469	3.636028	0.767346

三、平均顧客關係長度之分析

表 3 是由所有交易資料所估計之三種個別顧客關係長度之敘述統計，其中 T_0 為從資料所能觀察到的顧客關係長度， TII 為利用第參節所提方法所估計的顧客關係長度之中位數， $LIFE$ 為由估計之顧客關係長度的中位數減去所能觀察到的顧客關係長度所得之顧客剩餘關係長度中位數。「失效」表示顧客在觀察期間已終止交易關係，「設限」表示顧客在觀察期間未終止交易關係。由表 3 得知，估計顧客關係長度中位數的平均與剩餘關係長度中位數的平均分別為 9.14 與 1.97，表示個案公司的顧客大多僅剩下 1 個多月的平均關係長度，公司可加強進行顧客關係管理，以減少顧客流失。

若顧客區分為設限顧客與失效顧客，則設限顧客的估計關係長度中位數的平均為 9.62，平均還剩餘約 3 個月的關係長度，公司可提早規劃適合的行銷方案或促銷活動，來延長顧客關係長度。失效顧客的估計關係長度中位數的平均為 8.87，平均剩下不到一個月的關係長度。在所有顧客當中，有 6 位顧客的關係長度無法估計，進一步分析這些顧客，發現其與公司往來超過 31 期，對公司而言是忠誠度極高的客戶。

表 3：三種個別顧客關係長度估計中位數之敘述統計

		N(人數)	平均	最小值	最大值
T0	失效	4069	8.00	8	8
	設限	2314	6.62	1	31
	合計	6383	7.50	1	31
THI	失效	4069	8.87	8.56	11.81
	設限	2308	9.62	7.00	30.83
	合計	6377	9.14	7.00	30.83
LIFE	失效	4069	0.87	0.56	3.81
	設限	2308	3.07	0.56	6.01
	合計	6377	1.67	0.56	6.01

四、顧客終身價值分析

表 4 為利用式(4)與式(5)和 4.3 節所估計之顧客關係長度所計算之個別顧客價值之敘述統計，其中 HV 為顧客歷史價值，FV 為顧客未來價值，而 CLV 為顧客終身價值(歷史價值與未來價值之和)。在表 4 中，失效人數為 4069 人，而設限人數為 2314 人。大約 63.7% 的顧客在兩年半之內流失，此現象反映 2000 至 2002 年間網路電話卡業者競爭激烈的事實，對個案公司而言，是思索如何留住現有顧客的一項重要警訊。

由表 4 可知，若不將顧客分為失效與設限時，則個別顧客的歷史價值從 54 元到 25284 元不等，平均每位顧客帶來約 769 元的利潤，共產生約 490 萬元的價值。最有價值的顧客與最不俱價值的顧客之歷史價值比約 468 倍，平均歷史價值僅數百元，由此可見歷史價值的貢獻是集中於較少數的重要顧客上，個案公司應該對這些少數顧客提供更多增值服務，以防止這類顧客的流失。此外，設限顧客平均產生約 1177 元，共創造約 272 萬元的價值；失效顧客平均貢獻約 537 元，共創造約 219 萬元的價值。設限顧客之平均歷史價值高於失效顧客約 2 倍，而且最俱價值的顧客出現在設限顧客

群，此顯示現有及兩年半內的新顧客有高於流失顧客的消費，個案公司應該把握這種現象，開發新商品及為顧客提供更多加值服務，以繼續吸引這些顧客。

在未來價值方面，僅考慮顧客關係未終止的顧客(2314 位)，則個別顧客價值從 0 元到 10367 元不等，平均每位顧客帶來約 554 元的價值，合計產生約 124 萬元的價值。從平均價值僅數百元亦可反映出大部份的未來價值是集中於少數重要顧客上。在顧客終身價值方面，個別顧客價值從 55 元到 29943 元不等，平均每位顧客帶來約 964 元的價值，共產生約 615 萬元的價值。設限顧客的終身價值平均約 1714 元，共創造約 397 萬元的終身價值；失效顧客終身價值平均約 537 元，共創造約 219 萬元的終身價值。從顧客終身價值也可看出，設限顧客(原有及兩年內新加入)的價值高於失效顧客(流失)，而且大部份顧客價值是集中於少數重要顧客，這個觀察也為個案公司的顧客關係管理策略指出一個方向。

表 4：顧客價值之敘述統計

		人數	平均價值	總價值	最小值	最大值
HV	失效	4069	537.43	2186783	54.88	12151.08
	設限	2314	1176.59	2722633	54.00	25284.01
	合計	6383	769.14	4909416	54.00	25284.01
FV	失效	4069	0.00	0	0.00	0.00
	設限	2314	537.62	1244061	0.00	10367.48
	合計	6383	194.90	1244061	0.00	10367.48
CLV	失效	4069	537.43	2186783	54.88	12151.08
	設限	2314	1714.22	3966694	62.36	29942.93
	合計	6383	964.04	6153477	54.88	29942.93

為進一步強調顧客終身價值是集中於少數重要顧客上，我們製作顧客終身價值累積百分比對顧客人數累積百分比之表示圖，參見圖 3。從圖中可發現，前 80% 的顧客僅為公司創造 30% 的價值，其餘 20% 的顧客卻能貢獻約 70% 的價值，公司應對具有較高價值的顧客分配較多顧客關係管理的投入資源。

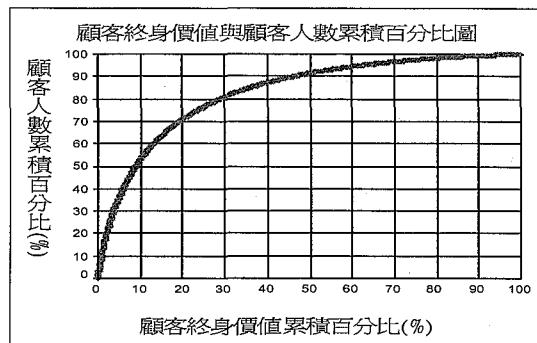


圖 3：顧客終身價值與顧客人數之累積百分比圖

五、顧客關係危險率預測模式之建構

利用資料庫全部交易記錄，我們以各期估計自我相關函數與偏自我相關函數為自變數對顧客進行分群。集群分析之結果顯示可將顧客分為兩群，集群一的 ACF 在前四期顯著不為零，PACF 僅第一期顯著不為零，故推斷其顧客關係危險率預測模式為 AR(1) 模式，共 4112 人屬於該集群。集群二的各期 ACF 與 PACF 皆不顯著，故推斷其顧客關係危險率預測模式為白噪音(white noise)模式，共 2271 人屬於該集群。

我們從各集群中分別任意抽出一位顧客來示範模式建構的過程。集群一的示範顧客的關係危險率 ACF 呈現指數遞減，而 PACF 在落遲一期後呈現截斷，故適合 AR(1) 模式：

$$h_t = \delta + \phi \cdot h_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

其中 h_t 為第 t 期之顧客關係危險率， δ 為一常數， ϕ 為自我迴歸係數， ε_t 為誤差項， σ_ε^2 為誤差項之變異數。利用此顧客關係危險率資料估計模式之參數，可得其估計平均危險率 $\hat{\mu}$ 為 0.7468，估計危險率變異數 $\hat{\sigma}_\varepsilon^2$ 為 0.1416，自我迴歸估計係數為 $\hat{\phi}$ 為 0.8457，誤差項估計變異數 $\hat{\sigma}_\varepsilon^2$ 為 0.0403，估計常數項 $\hat{\delta}$ 為 0.1152，故其顧客關係危險率預測模式為

$$\hat{h}_t = 0.1152 + 0.8457 \cdot \hat{h}_{t-1}.$$

集群二的示範顧客關係危險率 ACF 與 PACF 各期皆不顯著，故推斷其適合白噪音模式，其模式為

$$h_t = \mu + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2).$$

利用此顧客關係危險率資料估計模式之參數，可得其估計平均危險率 $\hat{\mu}$ 為 0.1022，而其估計危險率變異數 $\hat{\sigma}_\varepsilon^2$ 為 0.005，誤差項估計變異數 $\hat{\sigma}_\varepsilon^2$ 為 0.005，故其顧客關係危險率預測模式可表示為 $\hat{h}_t = 0.1022$ ，即各期危險率皆為固定常數。

伍、結論

為展現分析與挖掘顧客資料庫所蘊含重要資訊的方法，本研究利用國內某網路電話公司顧客交易資料庫的非契約型顧客實際交易資料，以倒傳遞神經網路分析技術來估計個別顧客關係危險率，進而分析顧客關係長度與顧客終身價值，以及建構各類顧客之顧客關係危險率預測模式。本研究將 Drew et al.(2001)的方法延伸至可以分析非契約型顧客之危險率。研究結果顯示，利用倒傳遞網路估計個別顧客關係危險率時，訓練集合的 MSE 與錯誤率分別為 0.000077 與 1.30631%，而測試集合的 MSE 與錯誤率則分別為 0.002469 與 3.636028%，皆顯示出不錯的績效。估計顧客關係長度中位數的平均與剩餘關係長度中位數的平均分別為 9.14 個月與 1.97 個月，其中設限顧客關係長度中位數的平均為 9.62 個月，而失效顧客關係長度中位數的平均為 8.87 個月。在顧客終身價值方面，平均每位顧客的價值約為 964 元，共可產生約 615 萬元的價值，其中設限顧客平均產生約 1177 元，共創造約 272 萬元的價值，而失效顧客平均貢獻約 537 元，共創造約 219 萬元的價值。以各期估計自我相關函數與偏自我相關函數為自變數對顧客進行分群，可將顧客分為兩群，集群一的顧客顧客危險率預測適合 AR(1) 模式，而集群二的顧客顧客危險率預測則較適合白噪音模式。

本研究以企業實際交易資料庫為例，探討顧客關係危險率之估計，以及利用所估計之危險率進行各項有關顧客關係管理之資料分析，為有效顧客管理決策的規劃與執行之前置資料庫分析工作提供一可行示範，有助於企業顧客關係管理實務之強化。

參考文獻

1. Berger, P. D. and Nasr, N. I., "Customer Lifetime Value: Marketing Models and Applications," *Journal of Interactive Marketing*, (12:1) 1998, pp:17-30.
2. Blattberg, R. C. and Deighton, J., "Manage Marketing by the Customer Equity Test," *Harvard Business Review*, (74:July/August) 1996, pp:136-144.
3. Box, G. E. and Jenkins, G. M., *Time Series Analysis Forecasting and Control*, Second Edition, San Francisco: Holden-Day, 1976.
4. Drew, J. H., Mani, D. R., Betz, A. L. and Datta, P., "Targeting Customers with Statistical and Data-Mining Techniques," *Journal of Service Research*, (3:3) 2001, pp:205-219.
5. Drye, T., Wetherill, G. and Pinnock, A., "Donor survival analysis: An alternative perspective on lifecycle modeling," *International Journal of Nonprofit and Voluntary Sector Marketing*, (6:4) 2001, pp:325-334.
6. Hansotia, B. J. and Wang, P., "Analytical Challenges in Customer Acquisition," *Journal of Direct Marketing*, (11:2) 1997, pp:7-19.

7. Harrison, T. and Ansell, J., "Customer retention in the insurance industry: Using survival analysis to predict cross-selling opportunities," *Journal of Financial Services Marketing*, (6:3) 2002, pp:229-239.
8. Helsen, K. and Schmittlein, D. C., "Analyzing Duration Times in Marketing: Evidence for the Effectiveness of Hazard Rate Models," *Marketing Science*, (11:4) 1993, pp:395-414.
9. Jain, D. and Singh, S. S., "Customer Lifetime Value Research in Marketing: A Review and Future Directions," *Journal of Interactive Marketing*, (16:2) 2002, pp:34-46.
10. Kosko, B., *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice-Hall, Inc, 1992.
11. Lawless, J. F., *Statistical Models and Methods for Lifetime Data*, John Wiley, New York, 1982.
12. Mehra, P. and Benjamin, W. W., *Artificial Neural Networks: Concepts and Theory*, IEEE Computer Society Press, 1992.
13. Mulhern, F. J., "Customer Profitability Analysis: Measurement, Concentration, and Research directions," *Journal of Interactive Marketing*, (13:1) 1999, pp:25-40.
14. Thomas, J. S., "A Methodology for Linking Customer Acquisition to Customer Retention," *Journal of Marketing Research*, (38:May) 2001, pp:262-268.
15. Wei, W. W. S., *Time Series Analysis: Univariate and Multivariate Methods*, Addison-Wesley Publishing Company, Inc, 1990.
16. Wyner, G. A., "Customer Profitability: Linking Behavior to Economics," *Marketing Research*, (8:2) 1996, pp:36-38.

附錄

假設 $H(t)$ 為累積危險函數， $s(t)$ 為存活函數，其關係式為 $H(t) = -\log_a s(t)$ 。根據生命表估計法與泰勒展開式，估計累積危險函數如下式：

$$\hat{H}(t) = -\log_a \hat{s}(t) = -\sum_{j:t_j < t} \log_a \left(1 - \frac{D_t}{N_t} \right)$$

其中 $\hat{H}(t)$ 為估計累積危險函數， $\hat{s}(t)$ 為估計存活函數， D_t 為各期失效人數， N_t 為各期存活人數。若危險函數為離散型機率函數，則 $H(t) = \sum h(t)$ ， $t = 1, 2, \dots, T$ ，故估計危險函數為 $\hat{h}(t) = -\log_a (1 - D_t / N_t)$ 。