

# 應用類神經網路技術於顧客價值分析\*

## Applying Neural Network to Customer Value Analysis

趙景明  
東吳大學資訊科學系  
chao@cis.scu.edu.tw

楊清潭  
東吳大學資訊科學系  
nms8910@cis.scu.edu.tw

### 摘要

企業依據顧客價值的高低，提供顧客不同的服務及行銷策略，有助於吸引更多高價值的顧客成為忠實顧客，進而使企業獲得最大的經營效益。本研究改良傳統 RFM 分析模型，提出以 RFM 屬性個別差異為權重，建立量化的顧客價值指標。此外，本研究應用資料探勘中的類神經網路技術，將過去期間的顧客交易資料做為輸入源，以目前期間的顧客價值做為學習目標，建立顧客價值的預測模型。最後將實際產業的顧客交易資料導入顧客價值的計算及預測，並針對分析結果提出顧客價值的區隔建議。研究結果發現本研究所提出之方法可有效識別及預測顧客的價值。

**關鍵詞：**顧客價值、資料探勘、類神經網路、RFM 分析模型

### 一、簡介

近年來隨著消費市場型態的轉變，行銷方式也有大幅改變，由原來以整體大眾為行銷對象，逐漸朝向為「關係行銷」(Relationship Marketing)、「一對一行銷」(One-to-One Marketing)及「區隔行銷」(Segment Marketing)。在做顧客區隔策略時，必須先界定所謂「顧客價值」(Customer Value)。研究顧客價值的學者很多，主要區別為：一、顧客內心所在乎的價值，又稱為顧客認知的價值[5,8,9]。二、顧客對企業的貢獻價值[4]。

本研究之目的是建立通用於各行業的顧客價值指標，以及應用類神經網路學習顧客價值的變化以建立顧客價值的預測模型。研究目的及方法說明如下：

#### 1. 建立量化的顧客價值指標

目前實務界廣泛使用的 RFM 分析模型，受到許多學者的研究及企業的採用。但因行業別的差异，對「交易期間」、「交易頻率」及「交易金額」有不同的重要性，且無法有效量化為單一的顧客價值指標。本研究提出以 RFM 屬性個別差異為權重值，將 RFM 屬性轉換成三度空間向量，計算絕對距離做為顧客價

值，建立量化的顧客價值指標。

#### 2. 應用類神經網路建立預測顧客價值變化模型

應用資料探勘技術進行知識發掘的方法中，類神經網路具有自動學習的能力。應用倒傳遞類神經網路技術，將過去一段固定期間的顧客交易資料做為輸入源，以目前期間的顧客價值做為學習目標。經由學習所得的類神經網路模型，做為未來顧客價值的預測。最後將實際產業(健康檢查業)的顧客交易資料導入顧客價值的計算及預測，並針對分析結果提出顧客價值的區隔建議。

### 二、文獻探討

#### (一) 類神經網路(Neural Network)

人類的大腦持續地從外界輸入訊息並處理產生適當的輸出反應，它有數十億個神經元相互連接而產生複雜的神經網路。類神經網路的組成可分為三個部份，包括最小單位的人工神經元、由許多人工神經元組成的層及由許多層組成的網路。相關研究請參考[1,2]。

#### (二) 倒傳遞類神經網路(Back-propagation Neural Network)

倒傳遞類神經網路屬於類神經網路中的監督式學習網路，是目前類神經網路模式中最具代表性、應用最廣的模型。類神經網路是多階層向前連接式設計，每一個連接點需連接到下一層。圖 1 是三層倒傳遞類神經網路架構圖。

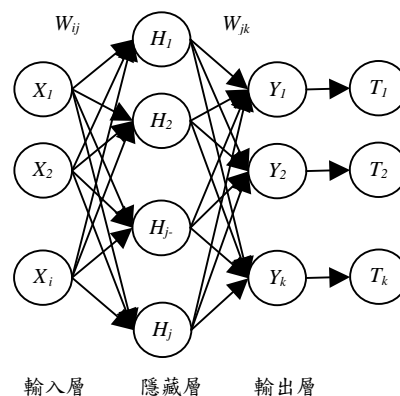


圖 1 三層倒傳遞類神經網路架構圖

\*本研究接受國科會研究計劃(計劃編號 NSC 92-2213-E-031-006)之部分補助

其中

$X_i$  是第  $i$  個輸入層人工神經元的輸入變數。

$H_j$  是第  $j$  個隱藏層神經元的輸出變數。

$Y_k$  是第  $k$  個輸出層神經元的輸出變數。

$T_k$  是第  $k$  個對應的目標輸出變數。

$W_{ij}$  是輸入層至隱藏層間連結的權重變數。

$W_{jk}$  是隱藏層至輸出層間連結的權重變數。

倒傳遞類神經網路的學習過程說明如下：

1. 選出具代表性的學習資料，將資料轉成輸入向量(一般為 0~1 或 -1~+1)做為網路的輸入變數( $X$ )。
2. 經隱藏層與輸出層的計算，得到網路的輸出變數( $Y$ )。
3. 比較學習資料的目標輸出值( $T$ )與網路之計算輸出值( $Y$ )，求出誤差值。
4. 根據誤差值重新調整網路層間的權重值( $W_{ij} \cdot W_{jk}$ )。
5. 重覆步驟 1 至 4，直到網路的誤差值低於一個可以容忍的範圍或達到最大重複次數，即表示網路訓練完成。

### (三) 顧客價值(Customer Value)

#### 1. 顧客認知價值

消費者購買商品或服務時心中自有衡量標準，皆期待能夠「物超所值」，也就是所獲得的價值大於所付出的價值。而這所謂「價值」，Zeithaml [10]認為價值是顧客對於付出與獲得的知覺態度，這種主觀態度影響消費者對於商品的整體評價。

顧客認知的價值是非常主觀意識的看法，在某人感受是有價值的，但另一個人的感受則未必如此。因此有學者提出顧客價值測量模型，例如，Keeney [5]提出價值焦點思考(Value-Focused Thinking)、Woodruff [9]提出顧客價值階層模型(Customer Value Hierarchy Model)及 Parasuraman [8]提出顧客價值觀測架構。

#### 2. 顧客貢獻價值

Mulhern [7]提出企業為了提供產品或服務顧客，針對個人、家庭或其他企業所產生的成本與從顧客所獲得的利益相較，利益超出成本的部份，即為顧客獲利性，也就是以顧客對公司的貢獻狀況評量顧客價值。最為業界普遍使用的 RFM 分析模型，依據 Hughes [4]研究中描述，RFM 分析模型在直效行銷(Direct Marketing)領域中已使用超過 30 年，但由於電腦科技的進步及資料庫系統的成熟應用，才使 RFM 分析模型自 1990 年以來廣泛被使用。RFM 分析模型是利用顧客過去的歷史交易記錄，包括最近一次的購買日期(Recency)、一段期間內的購買頻率(Frequency)及一段期間內的購買金額(Monetary)來進行顧客價值的衡量。

RFM 分析模型具有可適用於各個行業的

特性，且計算邏輯簡單，許多行銷人員在不需專屬資訊系統輔助下也可進行顧客分析。但 RFM 分析模型存在二項問題：一是個別 RFM 屬性針對不同產業有不同的重要性。例如某些產業對 R 屬性有很好的區隔能力，但其他產業可能對 FM 屬性有較佳的區隔能力，而 RFM 分析模型無法針對 RFM 屬性重要性的不同，整合成單一的區隔指標。二是 RFM 分析模型不具有預測能力，僅就顧客過去的歷史交易資料區隔顧客。

### 三、研究方法

關於資料探勘流程多位學者已提出各種版本，例如[3,6]，其內容大同小異各有其特點。本研究歸納整合及加入顧客價值計算後的研究流程如圖 2 所示。資料前置處理是資料探勘中非常重要的步驟，前置處理直接影響後續資料探勘是否能夠得到預期的結果。圖 3 是完整資料前置處理流程圖。

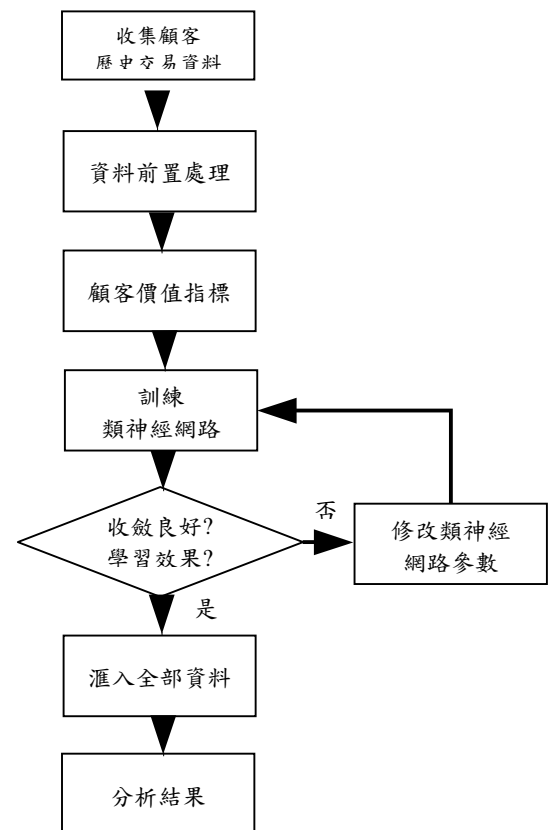


圖 2 研究流程圖

#### (一) 收集顧客歷史交易資料

資料收集是資料探勘的首要工作，足夠、正確的資料內容及品質才能確保後續的研究得到預期的效果。以下針對健康檢查業的資料內容加以說明：

1. 營業內容：包括個人身體健康檢查及其他醫療單位(診所或檢驗所)轉送代檢的血液檢查。

- 營業對象：個人身體健康檢查約佔 50% 營業比重，團體簽約體檢(如學校新生入學體檢、公司行號員工體檢)約佔 40% 營業比重，其他醫療單位轉送代檢的血液檢查約佔 10% 營業比重。
- 資料收集範圍：收集最近六年的個人顧客歷史檢查記錄。以 1997 年至 1999 年為類神經網路輸入來源資料，以 2000 年至 2002 年為訓練的目標資料。同時為避免團體公司行號資料影響個人健康檢查的消費特性，因此只分析個人健康檢查部份。

## (二) 資料前置處理

資料前置處理主要包括「區隔期間資料」、「統計 RFM 資料」及「計算 RFM 百分位數」等步驟，如圖 3 所示。

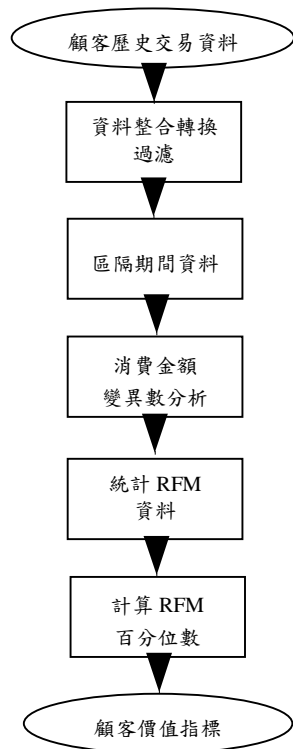


圖 3 資料前置處理流程圖

### 1. 區隔期間資料

健康檢查業的區隔期間設定為三年，是因為依健保局提供免費成人健診規定「40 歲至 64 歲每三年可檢查一次，65 歲以上每年皆可檢查」，因此設定區隔期間為三年。有些顧客每年至少做一次綜合性身體檢查，有些則數年才一次。設定期間為三年相信大部份的顧客至少已做一次身體檢查。

### 2. 統計 RFM 資料

將過去期間及目前期間的交易資料分別計算 RFM 數值。例如，某健康檢查顧客在過去期間有三次體檢明細如下：

順序	體檢日期	金額
1	1999/05/01	5000
2	2000/02/01	3000
3	2000/12/15	7000

最近購買日期(R):2000/12/15,相距 2000/12/31  
天數為 16 日

2000/12/31-2000/12/15=16(日)

購買頻率(F):3(次)

購買金額(M):5000+3000+7000=15000(元)

分別就個別期間內的交易資料統計 RFM 資料。

### 3. 計算 RFM 的百分位數(Percentile Rank, PR)

傳統的 RFM 模型將顧客 RFM 的屬性標示為 1~5 的數字，數字愈大代表價值愈高。這樣清楚的界線區隔顧客等級可能會產生二個顧客 RFM 屬性非常接近只是前後排列的順序差異，造成不同的顧客價值等級差距。為解決以上明顯界線造成誤差問題，本研究改採以百分位數的表示法，將個別 RFM 屬性的數值排序，以 0~1 之間的數值表示。舉例說明，某顧客 A 的 R 屬性的 PR 為 0.2、F 屬性的 PR 為 0.8、M 屬性的 PR 為 0.7，則以 A(0.2, 0.8, 0.7) 代表該顧客 A 的顧客價值向量，如圖 4 所示。

### (三) RFM 屬性與消費金額變異數分析

本研究以顧客貢獻價值為顧客價值的分析基礎，因此對於 RFM 屬性針對消費金額進行交互作用之多因子變異數分析及 V(顧客價值指標)針對消費金額進行單因子變異數分析，以確定是否與消費金額有明顯差異。最後將有明顯差異的屬性做為類神經網路的輸入源，訓練類神經網路模型。

### (四) 顧客價值指標

傳統 RFM 分析模型依據 Hughes [4]的研究，應用的行業別不同對 RFM 屬性反應的顧客消費程度有不同的影響。直效行銷行業中，R 的效果最好，F 其次，而 M 最差。但不是所有行業 RFM 的重要性都是 R>F>M。因此本研究提出以獲利性(顧客消費金額)為評量重點，同時加入個別差異的標準差做為權重值，將 RFM 屬性當成三度空間向量，再計算至原點的歐幾里德距離(Euclidian Distance)成為顧客價值，距離越長表示顧客價值越高。這顧客價值是相對的，所以最後再轉換為百分位數，成為顧客價值指標(Customer Value Index)，顧客價值指標越接近“1”表示顧客價值越高，反之越接近“0”表示顧客價值越低。圖 4 為顧客價值指標概念圖。

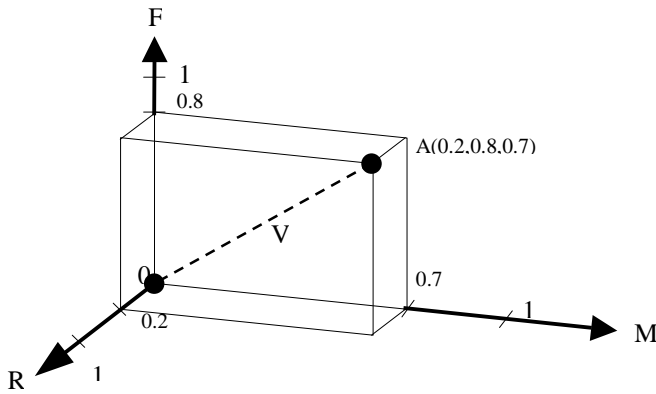


圖 4 顧客價值指標概念

歐幾里德距離：

$$V = \sqrt{(W_r R)^2 + (W_f F)^2 + (W_m M)^2}$$

$W_r$ ：R 屬性的標準差

$W_f$ ：F 屬性的標準差

$W_m$ ：M 屬性的標準差

V：顧客價值

### (五) 訓練類神經網路

#### 1. 倒傳遞類神經網路運算邏輯

倒傳遞類神經網路由輸入資料經過網路計算得到計算輸出值，將計算值與目標值比較不斷修改人工神經元權重值參數，直到獲得滿意數值或達到最大循環次數為止，詳細運算邏輯如下：

(1) 假設輸入資料向量陣列

$$X(P,1) = (X_{p,1}, X_{p,2}, X_{p,3}, \dots, X_{p,i})$$

P：資料的順序號。

i：輸入點的號碼。

(2) 每一個隱藏層或輸出層都由上一層接收完全相似的向量 X。每一個神經單元依下列公式計算：

$$Y(P,L,J) = X(P,L-1) * W(J,L) + B(J,L)$$

$Y(P,L,J)$ ：計算 P 資料在 L 層 J 神經元的結果

$X(P,L-1)$ ：L-1 層的輸入資料

$W(J,L)$ ：L 層 J 神經元外部的權重值(weight)

$B(J,L)$ ：L 層 J 神經元的門檻值

(3)  $Y(P,L,J)$ 的結果再經過轉換函數(Transfer Function)的運算做為下一層神經元的輸入，本研究以最常用的 S 型(Sigmoid Function)為轉換函數。

(4) 經過轉換函數後的結果做為一下層神經元的輸入。

$$X(P,L) = (f(Y(P,L,1)), f(Y(P,L,2)), \dots, f(Y(P,L,K)))$$

K：L 層中所有的神經元數量。

(5) 最後的輸出向量結果為  $X(P,O)$ 。

(6) 接著計算與學習目標向量 T(P)的誤差值。

$$E(P,O) = (T(P) - X(P,O)) * X'(P,O)$$

$$X'(P,O) =$$

$$(f'(Y(P,O,1)), f'(Y(P,O,2)), \dots, f'(Y(P,O,K)))$$

$f'$ 是轉換函數 f 的導函數(Derivative)

針對 L 層 J 神經元的誤差向量計算公式如下：

$$E(P,L,J) =$$

$$X'(P,L,J) * \text{SUM}(E(P,L+1,K) * W(K,L+1,J))$$

K：在 L-1 層的第 K 個神經元

(7) 依據上一步驟所得到的誤差向量，修改權重值。

$$W(J,L)_{T+1} =$$

$$W(J,L)_T + (\text{eta}) * \text{SUM}(E(P,L,J) * X(P,L-1)) +$$

$$(\text{alpha}) * (W(J,L)_T - W(J,L)_{T-1})$$

eta：學習率(Learning Rate)

alpha：動量因子(Momentum Factor)

T：重複循環(Iteration Cycle)

(8) 將所有輸入向量，透過類神經網路並調整權重值，重複所設定次數使誤差向量縮到最小。

(9) 均方根誤差(Root Mean Square Error, RMSE)是計算自訓練輸出值與目標值之間的誤差程度，數值越小表示模式的預測成效越佳。

$$\text{RMSE} = \sqrt{\text{SUM}(K((T(P,K) - X(P,O,K))^2) / (PT * KT))}$$

P：第 P 個輸入資料

K：第 K 個輸出點(node)

PT：所有 P 的總數量

KT：所有 K 的總數量

重複執行步驟 1 至 9 重複計算 RMSE，並調整權重值，直到達到設定的訓練次數。

#### 2. 學習效果評估

倒傳遞類神經網路學習效果的評估，可以從 RMSE 的數值變化，監控學習曲線收斂(Convergence)的情況。而學習率(Learning Rate)的大小直接影響收斂的速度及收斂的品質。同時過多的隱藏層及人工神經元數量也會影響學習的結果，以下針對不良學習結果加以說明：

##### (1) 無法有效收斂

無法有效收斂的原因有下列三種情況：

- 類神經網路過於複雜：過多的隱藏層或過多的人工神經元數量，使 RMSE 無法下降到預期範圍。
- 學習率數值太小：太小的學習率影響每次循環修改的權重值變量，在設定的最大學習循環次數內無法達到收斂效果。

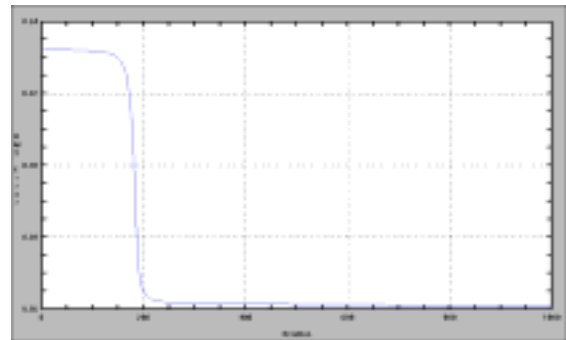


圖 5 200 次以上有效收斂圖

- 學習循環次數太小：最大學習循環次數太小，造成未達有效收斂時已到學習循環次數。當類神經網路無法有效收斂，解決方法是「降低類神經網路複雜度」、「加大學習率數值」及「提高最大學習循環次數」。如圖 5 所示，學習次數在 200 以上才得到良好收斂。

(2) 產生波動(Fluctuate)

RMSE 發生波動的原因有下列二種：

- 學習率數值太大：太大的學習率數值，使每次修改的權重值過大而造成 RMSE 忽大忽小無法成為平滑曲線。
- 學習循環次數過多：過多的學習循環次數，使學習系統在降低訓練資料的 RMSE 數值時卻發生不穩定現象。

當類神經網路產生波動，解決方法是「減小學習率數值」及「降低學習循環次數」。如圖 6 所示，學習次數在 1500 以上產生波動的情況。

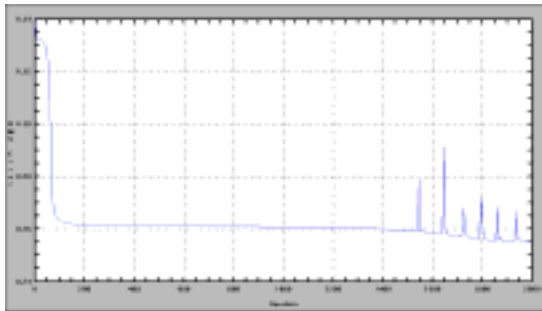


圖 6 產生波動圖

(3) 過度訓練

過度訓練是由於學習循環次數設定過大，造成類神經網路持續降低訓練資料的 RMSE，但測試資料的 RMSE 不降反增。這表示類神經網路對訓練資料有很好的學習效果，但測試資料卻效果不佳。這不是我們所樂於見到的，因為將來的真實資料也會產生效果不佳的情況。如圖 7 所示，學習次數在 5500 以上，RMSE 不降反增。

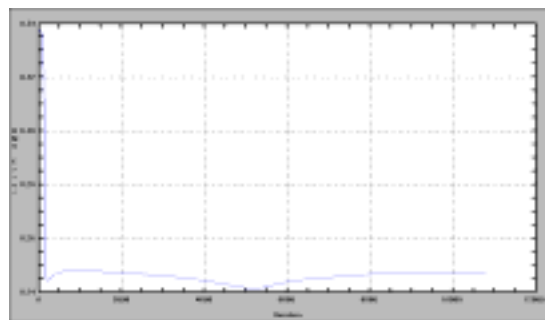


圖 7 過度訓練圖

四、實驗建置及結果分析

本研究以顧客貢獻度為顧客價值衡量基

礎，因此針對 RFM 屬性，首要步驟是統計 RFM 屬性的分佈情況及檢定 RFM 屬性與顧客消費金額是否存在顯著性影響。然後計算 RFM 屬性個別標準差做為權重值，計算顧客價值指標。接著將顧客歷史交易資料隨機取樣 10%~30% 為訓練資料及 5%~10% 為測試資料，做為訓練類神經網路之用。最後將所有資料輸入類神經網路評估模型的正確率。

本研究之實驗對象為國內知名專業健康檢查機構，應用之開發工具包括系統開發語言 Microsoft Visual Basic 6.0、資料庫系統 Microsoft SQL Server 2000、類神經網路開發工具 Vesta Qnet V2000 及統計分析系統 SPSS 10.0.7C。

(一) 健康檢查業資料內容

健康檢查業顧客歷史交易資料中過去期間的顧客數量為 17034 人，交易總金額 139,317,932 元。目前期間顧客數為 18317 人，交易總金額 194,252,270 元。過去期間及目前期間同時有消費的顧客 2919 人，占目前期間的 15.94%，交易總金額為 40,688,620 元，占目前期間交易總金額的 20.95%。

健康檢查業顧客類別屬於個人顧客，所以顧客數量多，因此訓練資料取 20% 的數量，測試資料取 5% 的數量，做為訓練類神經網路的資料。輸入屬性共有 4 個，為過去期間的 R、F、M 及 V，而學習目標值為目前期間的顧客價值指標。

(二) RFM 屬性與消費金額的分佈

由表 1 可得知過去期間(1997~1999)顧客歷史交易資料中，M 的標準差(9.91)大於 F 的標準差(4.53)大於 R 的標準差(2.51)，因此健康檢查業過去期間的 RFM 屬性的重要性  $M > F > R$ 。

表 1 健康檢查業 RFM 分佈表(過去期間)

	R		F		M	
	消費金額	比率	消費金額	比率	消費金額	比率
>0.0-<=0.1	10863522	7.8	10365440	7.4	503150	0.4
>0.1-<=0.2	11107830	8.0	10562950	7.6	1788000	1.3
>0.2-<=0.3	10679570	7.7	17165230	12.3	3097170	2.2
>0.3-<=0.4	9875320	7.1	9948310	7.1	4063340	2.9
>0.4-<=0.5	11457400	8.2	8159140	5.9	7050462	5.1
>0.5-<=0.6	15482270	11.1	9563240	6.9	12316610	8.8
>0.6-<=0.7	19250960	13.8	10243410	7.4	117356100	12.5
>0.7-<=0.9	17008950	12.2	14917830	10.7	22766980	16.3
>0.8-<=0.9	15420830	11.1	20280890	14.6	26990770	19.4
>0.9-<=1.0	18153140	13.0	28076182	20.2	43305000	31.1
標準差(百分比)	2.51		4.53		9.91	

由表 2 可得知目前期間(2000~2002)顧客歷史交易資料中，M 的標準差(8.18)大於 F 的標準差(3.51)大於 R 的標準差(2.30)，健康檢查業目前期間的 RFM 屬性的重要性  $M > F > R$ ，同時可以發現「過去期間」與「目前期間」RFM 屬性沒有很大的改變。

表 2 健康檢查業 RFM 分佈表(目前期間)

	R		F		M	
	消費金額	比率	消費金額	比率	消費金額	比率
>0.0~<=0.1	18720010	9.6	10106550	5.2	1064060	0.5
>0.1~<=0.2	17149610	8.8	19326920	9.9	3033170	1.6
>0.2~<=0.3	12300510	6.3	14632560	7.5	5061890	2.6
>0.3~<=0.4	19665180	10.1	20701560	10.7	9510920	4.9
>0.4~<=0.5	23153520	11.9	11177390	5.8	16737000	8.6
>0.5~<=0.6	17269840	8.9	19766750	10.2	23619560	12.2
>0.6~<=0.7	25356130	13.1	17369730	8.9	25106920	12.9
>0.7~<=0.9	26853450	13.8	26389220	13.6	26424360	13.6
>0.8~<=0.9	16562870	8.5	22028600	11.3	30481450	15.7
>0.9~<=1.0	17213100	8.9	32748150	16.9	53115340	27.3
標準差	2.30		3.51		8.18	

(三) RFM 屬性對消費金額變異數分析結果

對於 RFM 屬性進行變異數分析，以確定是否與消費金額有明顯影響。採用有交互影響之多因子變異數分析，檢定最近購買時間、購買次數及購買金額的強度愈高是否對消費金額貢獻愈大有顯著性影響。

由表3及表4結果顯示，無論過去期間及目前期間，p值小於 $\alpha$ 值=0.05、R值、F值、M值三因子交互效果檢定顯著，因此可說明隨著最近購買時間、購買次數及購買金額的強度愈高對消費金額貢獻愈大有顯著性影響。

表 3 RFM(過去期間)多因子變異數分析檢定

來源	平方和	自由度	平均平方和	F 檢定	p值
校正後的模式	1.328E+12	548	2422653418	357.814	.000
截距	2.910E+11	1	2.910E+11	42981.805	.000
R值 * F值 * M值	1.328E+12	548	2422653418	357.814	.000
誤差	1.203E+11	17768	6770699.160		
總和	3.508E+12	18317			
校正後的總數	1.448E+12	18316			

表 4 RFM(目前期間)多因子變異數分析檢定

來源	平方和	自由度	平均平方和	F 檢定	p值
校正後的模式	1.068E+12	539	1981148054	361.919	.000
截距	1.909E+11	1	1.909E+11	34871.523	.000
R值 * F值 * M值	1.068E+12	539	1981148054	361.919	.000
誤差	9.029E+10	16494	5474008.464		
總和	2.298E+12	17034			
校正後的總數	1.158E+12	17033			

(四) 顧客價值指標與消費金額的分佈

經過以標準差為權重值，將 RFM 屬性轉換為三度空間向量，計算所得到的顧客價值指標與消費金額分佈如表 5 所示。其中可得知在過去期間顧客價值最高的 20% 的顧客，交易金額佔總金額的 50.4%；而目前期間顧客價值最高的 20% 的顧客，交易金額佔總金額的 43.3%。健康檢查業體檢價格以人為單位，每位顧客除非增加消費次數否則消費金額差異性不大，但身體健康檢查不是需要常常做，通常是 1~3 年做一次。所以顧客價值最高的 20% 顧客，只佔約一半的總消費金額。

分別針對過去期間及目前期間的顧客價值指標與消費金額進行變異數分析，以確定是否與消費金額有明顯影響。採用單因子變異數分析，檢定顧客價值指標的強度愈高是否對消費金額貢獻愈大有顯著性影響。由表 6 及表 7

結果顯示，p 值為 0 小於  $\alpha$  值=0.05，故可以說明顧客價值指標的強度愈高對消費金額貢獻愈大。

表 5 顧客價值指標與消費金額的分佈表

	V(過去期間)		V(目前期間)	
	消費金額	比率	消費金額	比率
>0.0~<=0.1	785920	0.6	1408230	0.7
>0.1~<=0.2	2028020	1.5	3079360	1.6
>0.2~<=0.3	2886000	2.1	5305020	2.7
>0.3~<=0.4	4094890	2.9	10896850	5.6
>0.4~<=0.5	7599260	5.5	16050540	8.3
>0.5~<=0.6	12031942	8.6	21878600	11.3
>0.6~<=0.7	17468620	12.5	24856260	12.8
>0.7~<=0.9	22110510	15.9	26697860	13.7
>0.8~<=0.9	27512360	19.7	31444090	16.2
>0.9~<=1.0	42744730	30.7	52590780	27.1

表 6 V(過去期間)單因子變異數分析檢定

V值	過去期間健診交易資料					
	總和	總和的百分比	平均數	標準差	變異數	個數
0->0.1	785920	.6%	461.49	340.44	115901.534	1703
0.1->0.2	2028020	1.5%	1190.85	572.59	327858.675	1703
0.2->0.3	2886000	2.1%	1693.66	497.74	247743.951	1704
0.3->0.4	4094890	2.9%	2404.52	767.31	588770.843	1703
0.4->0.5	7599260	5.5%	4462.28	1467.28	2152923.7	1703
0.5->0.6	12031942	8.6%	7061.00	1811.83	3282745.3	1704
0.6->0.7	17468620	12.5%	10257.56	2009.74	4039048.7	1703
0.7->0.8	22110510	15.9%	12975.65	1215.31	1476969.5	1704
0.8->0.9	27512360	19.7%	16155.23	1923.28	3699005.7	1703
0.9->1.0	42800410	30.7%	25117.61	9337.65	87191686	1704
總和	139317932	100.0%	8178.81	8245.80	67993137	17034

ANOVA 摘要表

V值	平方和	自由度	平均平方和	F檢定	p值
組間	9.825E+11	9	1.092E+11	10583.245	.000
組內	1.756E+11	17024	10315258.2		
總和	1.158E+12	17033			

表 7 V(目前期間)單因子變異數分析檢定

V值	目前期間健診交易資料					
	總和	總和的百分比	平均數	標準差	變異數	個數
0->0.1	1408230	.7%	769.10	540.92	292599.744	1831
0.1->0.2	3079360	1.6%	1680.87	694.58	482445.660	1832
0.2->0.3	5305020	2.7%	2895.75	1322.04	1747801.4	1832
0.3->0.4	10896850	5.6%	5951.31	2666.21	7108682.3	1831
0.4->0.5	16050540	8.3%	8761.21	3491.29	12189138	1832
0.5->0.6	21878600	11.3%	11942.47	2060.73	4246625.7	1832
0.6->0.7	24856260	12.8%	13575.24	969.80	940514.028	1831
0.7->0.8	26697860	13.7%	14573.07	401.30	161044.653	1832
0.8->0.9	31444090	16.2%	17163.80	2229.68	4971456.9	1832
0.9->1.0	52635460	27.1%	28731.15	9977.33	99547123	1832
總和	194252270	100.0%	10605.03	8891.12	79051969	18317

ANOVA 摘要表

V值	平方和	自由度	平均平方和	F檢定	p值
組間	1.207E+12	9	1.341E+11	10181.082	.000
組內	2.411E+11	18307	13170445.5		
總和	1.448E+12	18316			

(五) 類神經網路學習結果

類神經網路相關參數設定，隱藏層分別由 1 至 3 層，人工神經元數量由 8 至 24 個，學習率分別由 0.002 至 0.01，以 0.002 為間隔。訓練停止設定 RMSE 值小於或等於 0.01，最多訓練次數為 10000 次。

表 8 為不同網路結構及不同學習率的訓

練結果，其中在隱藏層數量為 2、人工神經元數量為 18、學習率為 0.006 時有最低的 RMSE 值及最佳的關聯係數。在部份學習率加大時產生過度訓練或波動的現象，雖然有更低的 RMSE 值，還是捨棄不列入考慮。限於篇幅表 8 僅列隱藏層 2 至 3 層、人工神經元數量 15 至 24 個的學習結果。

表 8 類神經網路學習結果

隱藏層	神經元數量	學習率	Training RMSE	Testing RMSE	Training Correlatiol	Remark
2	10,5	0.002	0.265997	0.247656	0.564	
		<b>0.004</b>	<b>0.255186</b>	<b>0.250006</b>	<b>0.610282</b>	
		0.006	0.247834	0.256267	0.63899	4500 以上過度訓練
		0.008	0.247675	0.261548	0.639424	5500 以上過度訓練
		0.01	0.246485	0.267493	0.646526	2200 以上過度訓練
	12,6	0.002	0.267036	0.24796	0.559254	
		0.004	0.257263	0.244945	0.60193	
		<b>0.006</b>	<b>0.250677</b>	<b>0.250309</b>	<b>0.628172</b>	※
		0.008	0.247181	0.258711	0.641405	6000 以上過度訓練
		0.01	0.245512	0.26373	0.647394	3000 以上過度訓練
3	10,6,4	0.002	0.266992	0.248489	0.559468	
		0.004	0.258904	0.250041	0.594956	
		<b>0.006</b>	<b>0.257223</b>	<b>0.254549</b>	<b>0.601945</b>	
		0.008	0.253457	0.25684	0.617179	7000 以上過度訓練
		0.01	0.250335	0.254098	0.629394	5500 以上過度訓練
	12,8,4	0.002	0.266391	0.248091	0.562209	
		<b>0.004</b>	<b>0.26277</b>	<b>0.246288</b>	<b>0.57839</b>	
		0.006	0.254163	0.24924	0.614368	7000 以上過度訓練
		0.008	0.253939	0.252474	0.615294	5500 以上過度訓練
		0.01	0.252426	0.256719	0.621283	4000 以上過度訓練

表 9 是依據類神經網路參數隱藏層數量為 2、人工神經元數量為 18、學習率為 0.006 時，輸入過去期間全部資料，將運算所得輸出值比對目標值。誤差在 0.05 範圍內正確率為 54.33%，誤差在 0.1 範圍內正確率為 76.91%。

表 9 類神經網路測試結果

誤差範圍	±0.05	±0.1
正確率	54.33%	76.91%

由以上顧客價值分析得知，過去期間的顧客價值指標大於 0.6 以上的顧客佔消費總額的 78.8%，目前期間的顧客價值指標大於 0.6 以上的顧客佔消費總額的 69.8%。因此建議健康檢查業將舊顧客維繫資源，著重在顧客價值指標大於 0.6 以上的顧客。而其餘顧客價值指標 0.3~0.6 之間的顧客投入一般的維繫資源，最後顧客價值指標 0.0~0.3 之間的顧客可考慮列為放棄顧客，以降低顧客維繫支出。

## 五、結論及未來研究

### (一) 結論

依據顧客價值的高低，提供顧客不同的服務及行銷策略，在企業界已行之多年。如何識別出有價值的顧客及預測未來高價值的顧客，是所有企業希望獲得的資訊。

本研究從顧客對公司的貢獻價值，提出以 RFM 分析模型為基礎。首先以變異性分析檢定 RFM 屬性與消費總金額是否呈現顯著差異。然後增加以 RFM 屬性個別標準差為權重值，將 RFM 屬性轉換成三度空間向量，計算絕對距離做為顧客價值。接著應用類神經網路技術，將過去期間的顧客交易資料做為輸入源，以目前期間的顧客價值為學習目標。調整類神經網路模型的相關參數，監控學習過程曲線，達到最佳的模型。應用所學習的類神經網路模型，做為未來顧客價值的預測。最後以實際產業的顧客交易資料，導入顧客價值指標的計算及顧客價值的預測，並針對分析結果提出顧客價值的區隔建議。

### (二) 未來研究

影響顧客重複購買的因素很多，包括產品屬性、目的及使用結果，都會是影響因素。另外，屬於顧客個人屬性的尚有居住地、年齡、性別等，本研究經 ANOVA 檢定與消費總金額

不顯著。而模糊理論可處理數值型及類別型資料型態的優點，將可提高學習及預測的準確率。以下為未來我們幾個研究方向：

1. 研究顧客心理認知層面，對產品屬性、目的及使用結果的滿意度，增加可能影響消費金額的屬性。
2. 利用模糊理論可處理數值型及類別型資料型態的優點，與類神經網路結合，提高學習及預測的準確率。

## 六、參考文獻

- [1] C. H. Chu and D. Widjaja, "Neural Network System for Forecasting Method Selection," *Decision Support System*, vol. 12, pp. 13-24, 1994.
- [2] M. W. Craven and J.W. Shavlik, "Using Neural Networks for Data Mining," *Future Generation Computer System*, vol. 13, pp. 221-229, 1997.
- [3] U. Fayyad, "Data Mining and Knowledge Discovery in Databases," *Communications of the ACM*, vol. 39, no. 11, pp. 22-25, 1996.
- [4] A. M. Hughes, "Boosting response with RFM," *Marketing Tools*, vol. 5, pp. 4-10, 1996.
- [5] A. R. L. Keeney, *Value-Focused Thinking*, London: Harvard University Press, 1992.
- [6] C. Kleissner, "Data Mining for the Enterprise," in *Proceedings of the 31<sup>st</sup> Annual Hawaii International Conference On System Sciences*, pp. 295-304, July 1998.
- [7] F. J. Mulhern, "Customer Profitability Analysis: Measurement, Concentration, and Research Directions," *Journal of Interactive Marketing*, vol. 13, pp. 25-40, 1999.
- [8] A. Parasuraman, "Reflections on Gaining Competitive Advantage Through Customer Value," *Journal of the Academy of Marketing Science*, pp. 154-161, February 1997.
- [9] R. B. Woodruff, "Customer Value: The Next Source for Competitive Advantage," *Journal of The Academy of Marketing Science*, vol. 25, pp. 139-153, 1997.
- [10] V. A. Zeithaml, "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence," *Journal of Marketing*, vol. 52, pp. 2-22, 1988.