

以 RNC 基因表達規畫法建構顧客流失預警模型

林文修

天主教輔仁大學資訊管理系

wslin@im.fju.edu.tw

黃柏鈞

天主教輔仁大學資訊管理系

400346058@mail.fju.edu.tw

摘要

隨著個人化服務與顧客流失議題的重視，以及對企業獲利的影響越趨重要，本研究旨在提出一套有效的顧客流失預警模型，提供企業顧客服務決策支援。此外，本研究認為基礎基因表達規畫法(Gene Expression Programming, GEP)具有不錯的演算能力，但是在模組設計與運作上尚有很大改善空間。因此，本研究巧妙地運用 GEP，納入 RFM 指標建構顧客流失預警模組，並透過 RNC 機制改良模型，建立一套新的顧客流失預警模型，希望能夠協助企業擬定動態的顧客服務與行銷策略。研究結果顯示，(1)本研究模型之演化機制有效地增加解答多樣性，並達到流失預警的效果。(2)在穩定度測試之標準差為 0.039，表示本研究模型具備良好的穩定性。(3)在模型適用性方面，在不同樣本大小之下，模型的績效穩定，代表本模型能夠處理不同數量大小之樣本資料。

關鍵詞：顧客流失、RFM 模型、基因表達規畫法、RNC。

Using RNC Gene Expression Programming on the Prediction Model for Customer Churn

Wen-shiu, Lin

Department of Information Management
Fu Jen Catholic University
wslin@im.fju.edu.tw

Bu-Jyun, Huang

Department of Information Management
Fu Jen Catholic University
400346058@mail.fju.edu.tw

Abstract

With personalized service and customer churn issues more and more attention, the propose of this study is to build a customer churn warning model which helps business on decision-making. The churn model based on gene expression programming combines with the RFM customer analysis technology. Also using RNC mechanism to improve the model performance. The results show that (1) the evolution mechanism of this model increase the diversity of answers (2) in the stability test show that the standard deviation is 0.039, indicating that the model of this study get a good stability scores. (3) the applicability test show that the customer churn model can handle the different number of sample data stably.

Keywords: Customer Churn, Gene Expression Programming, RFM Model 、 RNC

壹、緒論(Introduction)

一、 研究背景與動機

近年來，隨著市場逐漸重視個人化服務以及差異行銷，企業策略的擬定也與從前的產品導向大不相同，因此企業紛紛進化成為顧客為中心的思維。然而在有關企業策略競爭研究領域中「微笑理論」(Smile Curve)，常被學者們提及的重要概念之一(施振榮，1992)，主要概念是指引企業如果企圖增加附加價值，企業應該致力往微笑曲線的左端或右端努力，如圖 1-1 所示。附加價值的高點在於技術專利的發展，或者是注重品牌及服務。然而，服務的概念在顧客關係管理領域中，一直是研究的核心之一。

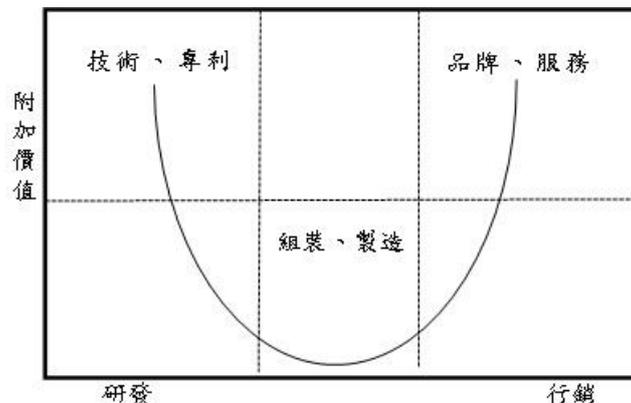


圖 1-1 微笑曲線

在顧客關係管理的研究領域中，Peppers and Rogers (1997) 認為開發一個新客戶所花費的成本，要比保有一個現有的客戶之成本高出五倍以上，如能將客戶流失率減少 5%，利潤將有 100% 的成長。Zeithaml, et al. (1996) 則認為顧客的流失將對公司所造成各種財務損失，並指出新爭取的顧客並不會馬上對企業的營收有所助益。Kotler (2001) 更清楚的指出，吸引新顧客的成本約為保留現有顧客的五倍，維持舊顧客成功的機率高達 60%，而爭取新顧客成功機率卻低於 30%。

再者，在 Google 上搜尋關鍵字「顧客流失」一詞，約有 1,210,000 筆相關搜尋結果，而以英文「customer churn」搜尋也有 1,320,000 項結果 (Google, 2011.12)。上述顯示，近年來企業實務界與學術界對顧客流失 (Churn) 議題的重視，體認其對企業與顧客價值的重要性，終至影響企業的獲利。此外，亦指出舊有客戶為企業重要的獲利來源，如果能夠有效的解決顧客流失問題，也意味著能夠替企業創造更高的價值與利潤。

若從學術觀點切入，可發現近年研究顧客流失議題越來越豐富。Kaymak (2001) 應用 RFM 顧客分析技術以最近購買日 (Recency)、購買頻率 (Frequency)、購買金額 (Monetary) 三種構面分析顧客價值。廖儷雪 (2009) 更進一步地將 RFM 因子融入遺傳規劃法 (Genetic Programming, GP)，據以建構顧客流失模型。該研究證實，利用 GP 的樹狀結構在編碼上較具彈性，擅長解決複雜的問題，應用的領域相當廣泛。但是，GP 的問題在於解決複雜問題時，其計算量龐大，導致演化效率不佳。並且 GP 為了改善演化效率的問題，將犧牲一些計算的複雜度，導致演化效果受到限制。因此，Ferreira (2001)

結合遺傳演算法(GA)與 GP 的特性，發展了基因表達規劃法 (Genetic Expression Programming, GEP)，解決 GP 的侷限性。因此，吳明輝 (2010) 在其研究中融入 RFM 分析因子來增進基因表達規劃法 (GEP) 之效果，用於建構電信業者之顧客流失模型，證實了 GEP 在建構顧客流失模型之可行性，惟須留意樣本之特性以及樣本結構對模型穩定度之影響。

二、 研究問題與目的

根據上述，本研究將提出之模型原理為，透過分析顧客基本資料、顧客交易紀錄與相關衍生變數等資訊，用以解析顧客之特徵，進一步判斷出可能流失之客戶。事實上，欲建構一個有效的顧客流失預警模型，必須讓模型具備高度的彈性，解答空間也不能太過於狹隘，並且需要良好的的演化機制藉以訓練模型。近年來的研究當中，已有不少學者使用柔性計算之演算法，例如遺傳演算法 (Genetic Algorithm, GA)、遺傳規劃法 (Genetic Programming, GP)、類神經網路 (Neural Network, NN)、基因表達規劃法 (GEP) 等，其中基因表達規劃法結合簡單編碼與樹狀結構的優點，對於產生公式或規則發現皆有良好的成效。因此，本研究將以基因表達規劃法建構顧客流失預警模型。本研究將改良基因表達規劃法，並且結合 RFM 顧客價值分析技術，將顧客購買頻率、顧客購買金額、顧客購買日等重要因子融入 GEP 模組設計中，並使用 RNC 機制，增強了解答的多樣性。因此，，本研究之研究目的如下：

- (一) 支援企業解決顧客流失預警議題，建構顧客流失預警模型，以供企業擬定服務與行銷策略之參考。
- (二) 建構兼具效果及穩定度之模型，嘗試解析模型之顧客行為模式規則。

貳、文獻探討

一、 顧客流失問題

顧客流失分析為，企業利用資料分析、統計分析或其他智慧型演算等方法，在顧客基本資料、顧客交易記錄資料、或顧客消費行為資料中，找出可能流失的顧客及可能流失的原因，進而讓企業能夠進行相關管銷計畫之擬定或顧客挽留策略..等，以達到事先預警的效果。

本研究彙整顧客流失預測相關研究如下表 2-1，可以發現顧客流失問題是一個廣泛存在於各領域中的重要議題。因此，本研究認為顧客流失模型在不同樣本大小之下的表現以及穩定度是非常重要的，亦是本研究的探討重點之一。

表 2-1 顧客流失預測相關研究

年分	研究者	研究對象	預測方法
2003	王正雄	直銷化妝品業-顧客流失模式	遺傳規劃法
2006	Ahn, et. al.	電話通訊業顧客流失分析	Mediation Analysis
2007	許瀞方	金融業-顧客資料與銀行往來資料	Logistic
2009	廖儷雪	洗衣業顧客流失模型	遺傳規劃法
2010	吳明輝	電信公司顧客流失預測模型	基因表達規劃法

二、RFM 顧客模型

(一) RFM 模型定義

Hughes (1996) 的研究指出，客戶資料庫中，最近購買時間 (Recency)、購買頻率 (Frequency)、購買金額 (Monetary)，為三個關鍵的要素，其各別之基礎定義如下：

1. 最近購買時間 (Recency)：

最近購買時間是指顧客最近一次購買距離分析時點的天數。當顧客最近一次購買某產品或服務的時間距離現在愈近，企業認為該名顧客會重複購買的機率也愈高。

2. 購買頻率 (Frequency)：

購買頻率是指顧客在一定期間內購買產品或服務的次數，一般而言，購買頻率愈高代表了顧客與公司之互動程度愈高。

3. 購買金額 (Monetary)：

購買金額是指顧客在一定期間內購買該企業產品的總金額。一般而言，顧客購買總金額也代表著對此產品或服務之認同。

(二) RFM 指標應用

近年研究中，CRM 相關模型設計常會融入 RFM 指標，本研究整理為三大類，其一，將 RFM 運用於調整權重。以 RFM 指標代表顧客個別差異，計算絕對距離來做為顧客價值指標(楊清潭 2003)。其二，依照行業的特性，自行增減更適合的評分指標。廖麗雪 (2009) 的研究中，適其應用領域之特性，除基本 RFM 因子之外還加上了總送洗次數因子，形成 RFNM 顧客權重因子，再將顧客分為五個等級給予不同權重值，反映該客戶貢獻度。其三，將 RFM 因子視為重要變數之一，交由模型進行演化。吳明輝 (2010) 的研究中，將 RFM 模型結合基因表達規劃法，建構顧客流失預測模型，從電信業客戶之話務資料、帳務資料中取出 RFM 因子，作為研究模型的分析變數基礎。

綜合本節研究之論述，RFM 模型具備良好的彈性，能夠依照模型以及不同行業特性做更精細的調整。因此，本研究提出之顧客流失預警模型將融入 RFM 指標於演算法中，藉以提高模型之成效。

三、基因表達規劃法

(一) 基因表達規劃法簡介

基因表達規劃法 (GEP) 是由葡萄牙學者 Ferreira 在 2000 提出，透過模仿自然生態演化機制「適者生存，不適者淘汰」的概念，來尋找欲解決之問題之解答 (Ferreira, 2000)。遺傳演算法 (GA) 使用簡單固定長度的編碼方式，操作簡易，但不適合處理複雜度高的問題，而遺傳規劃法 (GP) 是採用非線性、不等長的資料結構來編碼，較能解決複雜問題，但在操作上也較為複雜。基因表達規劃法結合了 GA 與 GP 的優點。透過類似 GA 的固定編碼方式之基因型，以及類似 GP 樹狀結構的方式做為表現型，實現了以簡易的編碼方式來解決複雜之問題。

基因表達規劃法 (GEP) 之演化流程為：先隨機產生初始族群，之後再將這些個體表達出來，進行適應值之評估。個體適應值評估完後，會去檢查是否滿足終止條件，

若達到條件則停止演化，若無，則將適應值最佳的個體保留到下一代，並選擇適應值較佳的個體進行再製，以產生新的子代，再次進行以上演化步驟，重新對新的個體進行表示、執行、計算適應值、判斷是否達到終止條件、選擇、再製等動作，直到找到最佳解或是達到最大演化代數，才終止演化的動作。

(二) 隨機數值常數(Random numerical constants, RNCs)

透過隨機數值常數的機制，使得 GEP 模組能夠透過類似對應的方式將參數替為數值代入模型中，透過常數的擾動 GEP 模組產生的解答可以更具多樣性。隨機數值常數在 GEP 表現型中以問號(?)表示，並且只會發生在終端節點上，當編碼中出現問號(?)代表該位置之參數須對應至另一組隨機常數矩陣。而 RNCs 在 GEP 染色體編碼方式為，在尾部基因後再加上一段與尾部等長的隨機常數基因片段，該片段上隨機產生的數值則表示對應至矩陣 Dc 之位置，對應方式為依序由左到右，最後在取出該位置上之數值做為模型的計算參數。如圖 2-1(a) 所示，色體在第 2、5、11 個 bit 需要做常數轉換，依照隨機常數基因片段之順序，可得到矩陣位置依序為位置 6、位置 8、位置 0，再對應矩陣 Dc 之位置 6、8、0 之數值，代入模型相對位置中進行運算，如圖 2-1(b)。

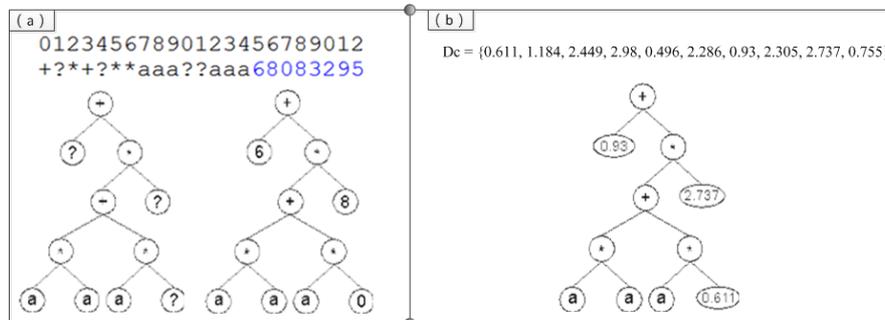


圖 2-1 GEP-RNC 示意圖

(三) 基因表達規劃法相關研究

GEP 相關研究近年來應用越來越廣闊，其中有學者將 GEP 運用在分類問題上，並且與常用的決策樹 C4.5 做比較，實驗結果顯示，在該研究議題中，GEP 模型的準確度較高 (Chi, et al., 2003)。再者，也有不少學者致力於改良 GEP 模型，透過加入多基因體編碼或者 ADF (Automatically Defined Function, ADF) 機制、RNC 機制…等方式，甚至是結合其他方法構成一個新的模組，使得模型更具彈性，改善基礎 GEP 之效能。本研究整理了近年 GEP 應用之相關文獻如表 2-2。

表 2-2 GEP 相關應用文獻

年分	研究者	研究對象
2003	Chi, et. al.	將 GEP 運用於分類規則之研究，其準確度相較於 C4.5 之模組，正確率提升約 20%，並且產生之規則比 GP 更加簡短，模型效率也較高。
2004	Li, et. al.	提出 GEP 應用於預測多變量的問題上，其精確與穩定性上皆優於遺傳規劃法 (Genetic Programming) 與線性迴歸 (Linear

		Regression) 等方法。
2008	陳玲穎	以 GEP 為基礎並使用 ADF，以改進 GEPCLASS (Weinert, 2006) 在染色體操作上受限且複雜性高的問題。
2010	吳明輝	以 GEP 為基礎並結合了 RFM 指標，替電信公司建構顧客流失預警模型。
2011	黃怡婷	以兩階段的 GA 找出共同基金之最適投資組合與資產配置，並利用 GA 與 GEP 的共演化，來找出最適交易策略。

總之，從上述文獻研究成果可知，基礎 GEP 確實具有不錯的演算能力，但是在模組設計與運作上皆還有發展空間。再者，目前罕見使用 GEP 來解決顧客流失議題。透過演算法挖掘顧客特徵，捕捉顧客流失的訊號，這正是本研究欲解決的問題。

四、 文獻彙總

綜合文獻探討可以得知，顧客流失一直是個重要的研究議題，也有許多學者提出各種方法來進行相關研究，其中 RFM 模型為一個廣為使用的顧客活躍性分析技巧，可以有效地劃分顧客的群集，便於分析顧客或者增進模型成效。然而基因表達規劃法 (GEP) 結合了遺傳演算法與遺傳規劃法之優點，模型編碼頗富彈性，並且對於訊號捕捉、預測預警等議題也有不錯的成效，近年來已成了熱門的演化式計算之一。

本研究認為基礎 GEP 演化能力強大，但同時存在著穩定性不足之問題，因需要視問題調整模型，才能發揮更好的效果。再者，目前罕見使用 GEP 研究顧客流失預警的文獻。綜合以上論述，本研究以傳統的 GEP 為基礎，巧妙地融入 RFM 指標以及 RNC 機制，建構新的顧客流失預警模型，期望模型能提升預警系統的效能及穩定性。

參、研究方法

一、 研究架構

本研究提出改良式基因表達規劃法來建構顧客流失預警模型，透過分析顧客的基本資料、交易資料、RFM 顧客指標來預測顧客的流失可能性。模型中，採用 RNC 機制改善 GEP 產出規則不易解釋之問題，以及強化演化的多樣化。本節將依照研究架構、變數選擇與定義、顧客流失預警模組設計、實驗設計之順序逐步介紹。

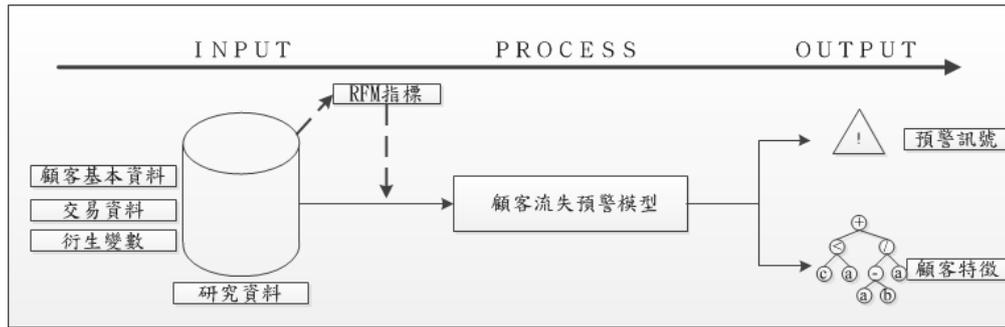


圖 3-1 研究模型架構圖

二、變數選擇與定義

本節主要針對模組中所會用到之解釋變數進行說明。而模型所需之解釋變數可分成三大類，分別為（一）顧客基本資料、（二）RFM 指標、（三）交易資料或衍生變數，而依據資料集特性不同，可以彈性調整變數。

- （一）顧客基本資料：顧客基本資料包括了：性別、年齡、居住地、是否為會員…等，將顧客基本資料處轉換為類別化的基本資訊，進一步建立模型所需之資料倉儲。
- （二）RFM 指標：RFM 指標之來源為，客戶資料庫中：最近購買時間（Recency）、購買頻率（Frequency）、購買金額（Monetary），三種關鍵的要素。依據 Hughes（1994）之顧客五等分法，將顧客之 RFM 因子由大到小、多到少的順序予以排列，並將每一構面分成五等分並給予得分，排名在最前面 20%的顧客其得分為 5 分，次之的 20%為 4 分，再次之的 20%為 3 分，依此類推。
- （三）交易資料或衍生變數：使用者交易行為之資料庫，或從中所萃取、計算後的衍生變數，例如：購買行為之資料庫、金融卡交易紀錄、電信交易資料庫…等，並將變數做類別化處理。

三、顧客流失預警模型設計

（一）編碼規則：

GEP 染色體之編碼可以分為頭部（head）及尾部（tail），頭部可為函數節點或終端節點，而尾部只能為終端節點。本研究將頭部長度設為 12，而尾部長度可由公式(1)算得，其中 n_{max} 為函數節點集合中所需最多運算元各數，而本研究使用的函數節點為加（+）、減（-）、乘（*）、除（/）外，並將納入小於（<）、等於（=）、大於（>）等函數，因此 $n_{max}=2$ ，可得尾部長度為 13，再加上隨機數值基因片段長度與同源染色體長度，每一代 GEP 染色體總長度為 $(12+13+13)=38$ 。每個 GEP 長度為 38，前 25 個 bits 做為顧客基本資料、RFM 因子、顧客交易資料、隨機常數符號之編碼，後 13 個 bits 為隨機數值常數之對應編碼。本研究將根節點限制為比較性質之函數，如 >、<、=，其目的是要讓每個 GEP 都產出一個二元的解答。

隨機數值常數（RNCs）在基因型中編碼為符號「？」，代表需替換為隨機常數，此機制能夠改善 GEP 樹狀結構可能不具解釋性之問題。參照圖 3-2 GEP 編碼基因型示意圖，編碼位置第 16、18、20 出現了問號，代表這三個位置需對應替換成常數。



圖 3-2 GEP 編碼基因型示意圖

RNC 對應的方法為，將第 i 個問號則對應至 RNC 編碼中順序第 i 個 bit 之數值，再將此數值對應至隨機常數矩陣，將值加以替換之。如下圖 3-3 編碼表現型示意圖所示。

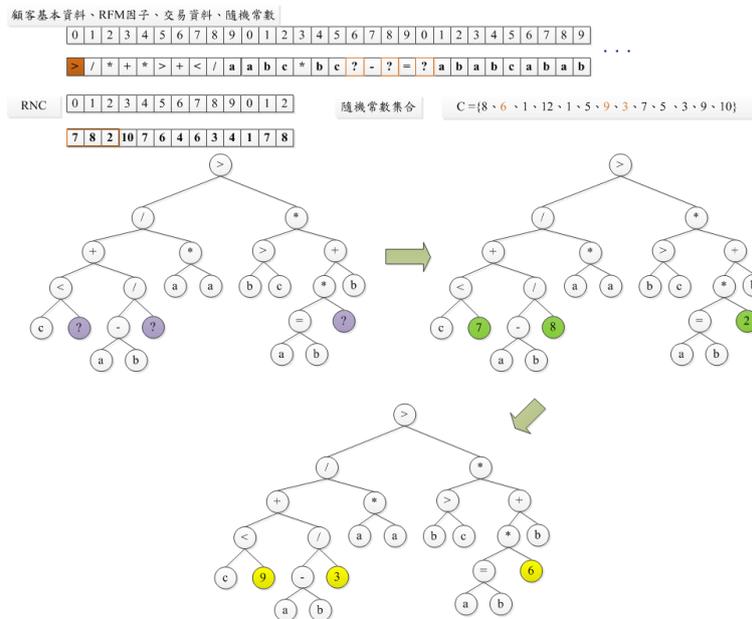


圖 3-3 GEP 編碼表現型示意圖

(二) 適應函數：

本研究以混亂矩陣 (Confusion matrix) 的方式來衡量預測模型之優劣，依據表 3-1，其可能預測組合結果有四種 (TP/FN/FP/TN)。矩陣透過計算之後可得出模型之敏感度 (Sensitivity)，精確率 (Precision Rate) 及 F-measure 等指標，各指標計算方式列於表 3-2。

表 3-1 混亂矩陣

	預測顧客會流失	預測顧客不會流失
實際顧客會流失	TP	FN
實際顧客未流失	FP	TN

表 3-2 適應值指標

指標	計算方法
正確度(Accuracy)	$R_a = (TP + FN) / (TP + FP + FN + TN)$
敏感度 (Sensitivity)	$R_s = TP / (TP + FN)$
精確率 (Precision)	$R_p = TP / (TP + FP)$
F-measure	$F_\beta = (1 + \beta^2) \frac{R_p * R_s}{\beta^2 * R_p + R_s}$

就顧客流失問題而言，模型若將會流失之顧客判定為不會流失是較嚴重的，因此本研究採用 F-measure 指標為適應值，並設定 $\beta = 0.5$ ，代表提高敏感度之重要性，相對的精確度之重要性會下降，可得到適應值公式如下。

$$\text{適應函數: } [F_{\beta=0.5} = (1 + 0.5^2) \frac{R_p * R_s}{0.5^2 * R_p + R_s}] \quad (2)$$

(三) 參數設定

本研究之模型參數可分為函數節點、演化參數，其詳細參數設定請參考表 3-3。

表 3-3 顧客流失預警模組參數設定

類型	控制項目	GEP 參數		
函數節點	算術運算元	+、-、×、÷、<、>、=、AND、OR		
演化參數	族群大小	100	終止條件	200 代
	選擇方式	菁英法、輪盤法		
	突變率	0.2		
	單點重組率	0.15	雙點重組率	0.15
	基因重組率	0.05	基因轉換率	0.05
	IS 轉換率	0.05	RIS 轉換率	0.03

四、實驗設計

(一) 研究資料

本實驗之研究資料為某洗衣業者之實際資料，內容包含洗衣業之顧客基本資料、交易資料，資料範圍為 2004 年 5 月起至 2008 年 8 月止，共計約 1,700 筆，參數皆為正規化後之類別變數，欄位說明如表 3-4 所示。對於洗衣業者來說，掌握重點顧客群，降低支出成本，是非常重要的，建構洗衣業客戶流失之預測模型，以供洗衣業能儘早進行客戶挽留之行為，以避免顧客流失之狀況。

表 3-4 顧客流失預測相關研究

資料屬性	變數說明
基本資料	是否會員 (a)、性別 (b)、年齡 (c)、所在區域 (d)
歷史交易資料	總送洗金額 (e)、總送洗件數 (f)、總送洗次數 (g)、平均每件金額 (h)、送洗頻率 (i)、最近送洗期間 (j)

(二) 實驗設計

1. 實驗一：探討顧客流失模型之適應值演化過程以及模組績效。
2. 實驗二：將模型執行五次，探討五次的演化情形，以分析模組的穩定度。
3. 實驗三：以不同的樣本數量測試模組之演化情況。

肆、實驗結果

根據本研究之目的，希望建立一個顧客流失預警模型，讓企業更加了解自身客戶，協助企業擬定行銷策。本研究巧妙地運用基因表達規劃法(GEP)，納入 RFM 指標建構顧客流失預警模組，並透過 RNC 機制改良模型，據以達成演化多樣化的效果並且同時具備良好的效能。

為了檢驗模組之績效進行了三項實驗，實驗一評估預警模組的績效表現；實驗二評估模組之穩定度；實驗三探討模組在不同樣本大小之下是否有顯著差異。

一、顧客流失預警模組績效

(一) 演化狀況

依據圖 4-1 演化收斂圖所示，第 1 至 20 代演化的速度較為劇烈，然而在第 20 代至 65 代，適應值開始呈現停滯的情況，直到 65 代之後才繼續往上演化，演化終止條件為到達第 100 代。適應值停滯的情況代表模型可能陷入了區域最佳解，需要等到下一次擾動發生，並且產生更優秀的個體才有機會跳離，而數據顯示在 65 代之後適應值繼續成長，代表本研究之演化機制能夠有效地增加解答的多樣性，讓模型能夠跳脫區域最佳解，找到更佳的答案。表 4-1 為模型預測之結果，可以算得如表 3-2 之適應值指標，Accuracy = 84.8 %、Precision = 73.4 %、Sensitivity = 92.7 %，數據結果符合本研究之問題核心，重視敏感度的績效表現，代表模型能夠有效的找出可能流失之客戶，達到預警的功效。

表 4-1 模型預測結果

	預測顧客會流失	預測顧客不會流失	
實際流失顧客	590	46	636
實際未流失顧客	213	852	1065
	803	898	1701

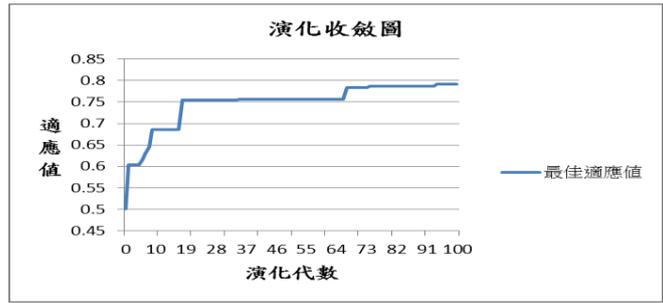


圖 4-1 演化收斂圖

(二) 最適模型演化規則實例說明

本節將探討模型之演化規則，並對應資料庫欄位，顯示演化後的規則，如表 4-2 所示，模型產出演化後之染色體(final chromosome)，解析後可以得到樹狀結構圖，再將節點與資料欄位加以對應出如圖 4-2 之規則。

表 4-2 最終染色體範例

final chromosome	>j-c*-/fjb-jdfhaif?jchfe
Best tree	$(j > (c - (((d - f) / j) - f) * (h / b)))$

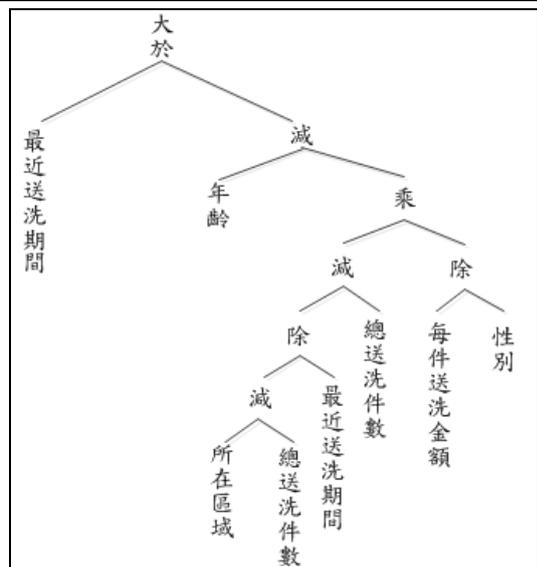


圖 4-2 中文樹狀結構解

二、 模型穩定度測試

為了驗證模型的穩定度，本實驗將模組重複執行五次，並綜合比較之。依據圖 4-3 顯示，本實驗 5 個 run 之最佳適應值呈現一致性的收斂，平均適應值為 0.72，實驗數據如表 4-3 所示，平均適應值的標準差為 0.039，代表本模型具有良好的穩定性。

表 4-3 模型穩定分析表

	Run_1	Run_2	Run_3	Run_4	Run_5	平均適應值	標準差
平均適應值	0.72	0.70	0.72	0.68	0.80	0.724	0.039

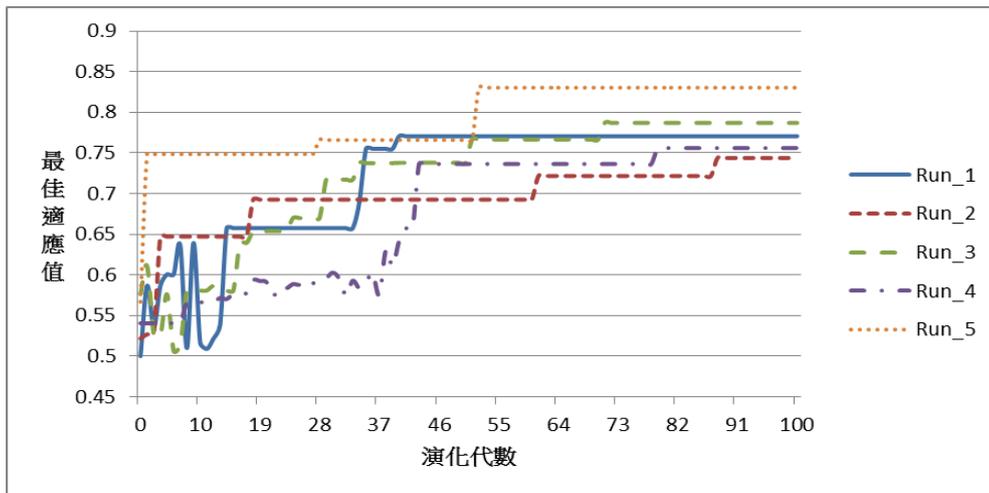


圖 4-3 模型 5 次執行適應值

三、 模型資料大小測試

為了測試模型在各種資料大小績效是否有顯著差異，本實驗使用三種不同的樣本數量做測試，分別為(1)原始資料 1700 筆、(2)三倍的原始資料以及(3)隨機抽樣 400 筆來進行實驗，實驗結果如表 4-4 所示。平均適應值之標準差為 0.027，表示模型在三種不同的資料大小之下仍然具有穩定的表現。而從圖 4-4 也能看出，在三種情況下適應值收斂狀況良好，雖然大樣本在最後適應值較低於其他兩者，但仍然有達到 0.775 的水準，表示本模型在樣本數較大的形況下一樣能夠有不錯的表現。

表 4-4 資料數量穩定分析表

	1700 筆	5100 筆	400 筆
最佳適應值	0.82	0.77	0.81
平均適應值	0.67	0.73	0.75
標準差	0.027		

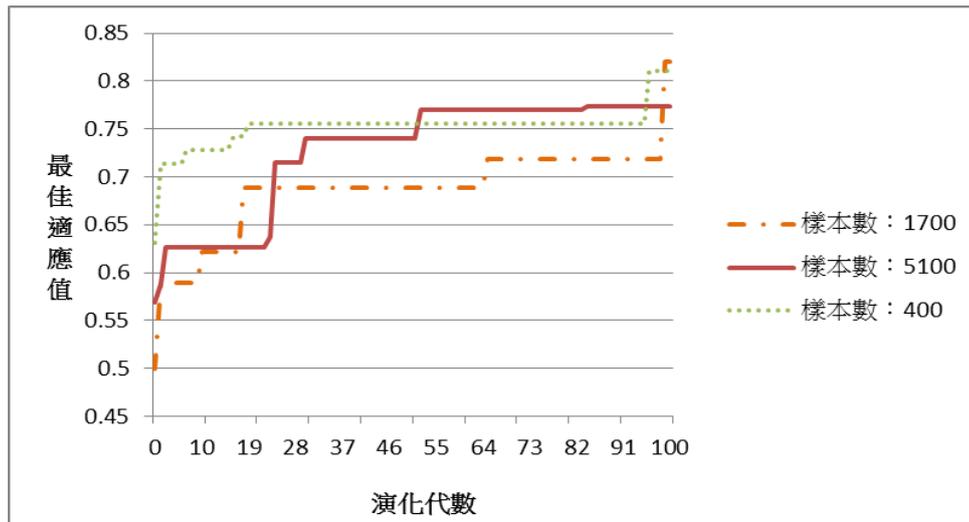


圖 4-4 資料數量測試結果

伍、結論與未來發展

一、 結論

本研究認為基礎 GEP 確實具有不錯的演算能力，但是在模組設計與運作上皆還有發展空間。因此，本研究巧妙地運用基因表達規劃法(GEP)，納入 RFM 指標建構顧客流失預警模組，並透過 RNC 機制改良模型，建立一個顧客流失預警模型，希望能夠協助企業擬定行銷策略。經由實驗結果分析，可得到以下結論。

首先，根據模型績效評估實驗之實驗數據，模型之敏感度績效良好，代表模型能夠達到預警的效果，替企業找出可能流失之顧客。其次，在穩定度測試的實驗中，模型呈現一致的收斂狀況，最佳適應值的標準差僅為 0.039，代表新模型具有良好的穩定性。最後，在模型適用性實驗中，發現大樣本在最終適應值表現會略低於小樣本的資料，但是整體來看模型具有一定的穩定性，能夠接受不同樣本大小之資料。

總之，目前罕見利用基因表達規劃法解決顧客流失議題，並且 GEP 是一個較新的方法，雖然具有不錯的演算能力，但需要更精細的改良調整，才能夠讓系統成效表現穩定，然而本研究所做之改良，使得模型能夠更加穩定且達到不錯的績效。據此，新的預警模型將能幫助企業提早發現可能流失之顧客群，預先擬定挽回或補救服務策略加以因應。換言之，藉由挽留更多可能流失的客戶以提升顧客價值，為企業帶來更大的獲利機會。

二、 未來研究建議

建議有興趣的研究者，可以個案方式導入不同業種的資料集，探討模型之外部效度。此外，可將 GEP 進一步改良為多 GEP 之架構，讓每個 GEP 代表一個決策的專家，一同評估顧客是否可能流失，以增加解答之多樣性。最後，亦可進一步解析 GEP 產生之規則中是否有重要的模式或規則，藉此更加了解顧客行為模式。

參考文獻

1. 王正雄，應用遺傳規劃法在顧客流失預警之研究-以直銷化妝品為例，輔仁大學管理學院資訊管理學系碩士在職專班碩士論文，2003。
2. 吳明輝，應用基因表示規劃法於顧客流失預測模型之研究-以某電信公司為例，天主教輔仁大學資訊管理學系在職專班碩士論文，2010。
3. 施振榮，再造宏碁：開創、成長與挑戰，台北：天下文化，2004。
4. 陳伊伶，演化式計算於證券投資組合與擇時規則建構之研究，輔仁大學資訊管理系碩士論文，2011。
5. 陳玲蘋，以 GEP 為基礎的分類規則產生器，元智大學資訊管理學系碩士論文，2008。
6. 許瀟方，現金卡客戶流失預警模型之建立，世新大學財務金融學研究所碩士論文，2006。
7. 黃怡婷，演化式計算於共同基金投資組合與交易策略推薦模型建構之研究，輔仁大學資訊管理系碩士論文，2011。
8. 廖儷雪，應用遺傳程式規劃在顧客流失模型的研究-以百麗洗衣為例，天主教輔仁大學資訊管理學系在職專班碩士論文，2009。
9. Chi Zhou; Weimin Xiao; Tirpak, T.M.; Nelson, P.C., Evolving Accurate and Compact Classification Rules With Gene Expression Programming, *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 6, 2003, pp. 519-531.
10. Ferreira, C. *Gene Expression Programming : Mathematical Modeling by an Artificial Intelligence*, 2nd, Springer-Verlag, Germany, 2006.
11. Ferreira, C., *Gene Expression Programming : A New Adaptive Algorithm for Solving Problem*, Complex System, Vol.13, 2001, pp.87-129.
12. Holland, J. H., *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.
13. Kotler, P., 2001. *Analysis, planning, implementation, and control. Marketing management, 11th ed.*, New Jersey : Prentice Hall, Inc.
14. Koza, J.R. *Genetic Programming : On the Programming of Computers by Means of Natural Selection*, Cambridge, MA : MIT Press, 1992.
15. Li, Q., Cai, Z.H., Jiang, S.W., & Zhu, L., *Gene Expression Programming in Prediction, Proceedings of the 5th World Congress on Intelligent Control and Automation*, June 15-19, 2004, pp.2171-2175
16. Stone, Bob. *Successful Direct Marketing Methods*, NTC Business Books, 1994.
17. Hughes, A.M., *Boosting response with RFM*, Marketing Tools, Vol.5, 1996, pp.4-10.
18. Kaymak, U., *Fuzzy target selection using RFM variables; IFSA World Congress and 20th NAFIPS International Conference*, 2001. ; Joint 9th, Vol2 : pp. 1038-1043, 2001.
19. Peppers, D. and Rogers, M., *Enterprise One to One : Tools for Competing in the Interactive Age*, New York : Currency Doubleday, 1997.
20. Zeithaml V. A., Berry L. L., & Parasuraman, A., The behavioral consequences of service quality. *Journal of marketing*, 1996, 60 : 31-4.