

利用類神經模糊理論預測壽險保單早期失效之研究

陳建勝
朝陽科技大學保險金融管
理系教授
jschen@cyut.edu.tw

陳美菁
朝陽科技大學企
業管理系副教授
mcchen@cyut.edu.tw

顏鳳妮
朝陽科技大學保險金融管
理研究所碩士

摘要

壽險保單早期失效率是業務品質的重要評估指標之一，但隨著壽險市場的開放及我國加入世界貿易組織的影響，壽險公司如何有效地降低保單早期失效率，以提升公司的經營績效，便成為壽險公司的重要課題之一。本研究係利用資料採礦技術中的類神經模糊理論來預測壽險保單早期失效之情況，其研究變數的選取方式採因素分析及文獻回顧法兩種，並採用整體正確率、型 I 誤差、型 II 誤差及誤判成本來衡量預測模式的優劣性。本研究實證結果發現利用因素分析所篩選的變數會比根據文獻回顧所篩選出的變數對保單是否失效的影響大，且類神經模糊在整體上有較穩定之預測能力；而利用類神經模糊規則庫，可更具體地顯示出變數之間的相互關連性。其中，以被保人年齡及婚姻、要保人婚姻、要保人和被保人關係、是否為員工自保件、年繳化附約保費、附約佔總保費比例對保單是否失效的影響最為顯著。

關鍵字：類神經模糊、型 I 誤差、型 II 誤差、誤判成本、規則庫。

1. 前言

壽險保單早期的失效率(early lapse rate)是評價壽險業務品質的重要指標，其對壽險公司的影響相當深遠。我國壽險市場在民國 82 年全面開放新壽險公司的設立後，因為新公司對業務人才與知名度的需求而引爆的挖角風暴，也造成許多連人帶件的跳槽事件不斷地上演，進而促使我國壽險保單的早期失效率持續地惡化。為改善此一問題，財政部保險司於民國 85 年實施「業務經營自律準則暨招攬體制改善計劃」，要求國內壽險公司分階段達成法定的第十三個月與第二十五個月保單繼續率水準，換句話說，保單早期的失效率(1-繼續率)必須控制在合理的範圍內，否則監理機關將不會核准該公司新保險商品的發行。實施至民國 88 年底止的統計數據顯示，以件數計算之壽險保單第十三個月與第二十五個月的失效率，已從民國 85 年的 26%與 33%下降至 20%與 31%，壽險業致力於改善壽險保單早期失效率已有初步的成效。但隨著我國加入世界貿易組織與投資型保單的銷售，在市場競爭與外商保險公司的大舉入侵下，未來是否會再次造成

另一股解約的風潮，值得業者與監理機關的重視。

根據壽險公會統計資料發現，保單生效後兩年內的失效率最高，過高的保單早期失效率對於壽險公司衝擊最大，當然也是保險監理機關的關注焦點。主要的原因在於長年期壽險保單的新契約成本大都集中在承保初期所致；包括業務人員的佣金、發單費用、體檢費用及銷售管理費用等，因此長年期壽險保單在未達損益平衡點前即失效，將使得保險公司無法攤回新契約成本而造成損失，進而減少壽險公司的利源。對被保險人而言，除喪失原繳保費之經濟損失外，亦將喪失原有的壽險保障。因此，保單早期失效不但造成保單持有人與保險公司雙方的經濟損失，隨之喪失的保險保障更加重社會福利成本。因此如何及早發現即將失效的保單，避免原有保戶因業務人員服務品質不佳或不當的話術之影響而輕易解約，將是壽險公司有效地維持高保單繼續率或降低保單失效率的積極方法，這也是本研究的主要目的。

壽險經營的基本原理為大數法則(law of large number)之運用，其特點在於如何大量地維持長期而

穩定的有效契約，使得因收取的保費所聚集之資金能夠充分運用於投資，以獲得最大的利潤。因此如何從大量的保單資料庫中擷取出早期失效可能性較高的壽險保單，將有助於業務品質與保單繼續率之提升。過去在探討保單失效率的研究中，常應用的分析方法包括區別分析、邏輯特迴歸分析及類神經網路。區別分析在使用上的限制較多；而邏輯特迴歸分析雖較區別分析的限制少，但其仍有臨界值設定的問題；類神經網路具有「捕捉非線性關係之能力」、「事先毋須設定變數間的函數型式」及「自我學習調整的特性」等優點之資訊處理系統，但無法將結構性的知識納入其中；而模糊邏輯則利用歸屬函數描述知識的特質，把傳統數學的二元邏輯推展到連續的數學模式，但模糊邏輯之學習能力不如類神經網路有效率（林金賢等，2002）。類神經模糊技術的基本精神即在利用模糊理論來描述結構性的知識，然後藉由類神經網路的學習能力來調整決定歸屬函數的形狀與參數，計算每一條規則之信賴度，進而取捨是否保留某些規則來做推論。基於上述理由，為能及早因應保單早期失效所帶來的顧客流失問題，本研究根據影響壽險公司的保單早期失效之保戶因素、保單因素、核保因素與業務人員因素，利用模糊邏輯的規則庫描述影響保單失效因素的結構性知識，以及模糊邏輯的歸屬函數描述參數的大小，並利用類神經網路的學習能力對各規則之重要性與參數之大小作調整，將可以得到一更接近真實狀況之分辨準則，進而更有效地預測保單失效的情況，以建立完整的評估模式並比較其預測效果，期能有效地預測保單早期失效的潛在保戶，以使保險經營者能夠針對具有高度保單失效風險之保戶，儘早採取積極的契約保全工作，以避免業務品質的繼續惡化而衝擊到公司的營運。

2. 研究設計

2.1 樣本資料及變數說明

本研究的樣本資料是以國內某一壽險公司 2000 年度新生效之長年期壽險契約，共計 11,993 筆。本研究係以保單件數為基礎，並根據財政部保險司所頒布之業務經營自律準則暨招攬體制改善

計畫中的定義，早期保單失效係指保單生效後第 25 個月內解約、停效或失效之保單，但不包括效力終止(如死亡或全殘給付)與契約撤銷保單，並排除躉繳與一年期定期壽險保單。為能充分瞭解壽險公司早期保單失效率過高之問題，本研究以第 25 個月內是否失效為依變數，並將影響保單失效率的可能因素分為保戶特性、保單特性、核保特性及業務員特性等四類，其所選取之研究變數的定義及說明如表 1。

表 1 輸入變數的說明

類別	變數名稱	說明
保戶特性	被保人性別	1 表男性，0 表女性。
	被保人保險年齡	計算投保時的保險年齡。
	被保人婚姻	未婚、已婚、離婚、喪偶。
	被保人職業	6 類職業類別。
	要保人性別	1 表男性，0 表女性。
	要保人保險年齡	投保時的保險年齡。
	要保人婚姻	未婚、已婚、離婚、喪偶。
	要保人與被保人關係	公司、本人、配偶、子女、父母、其他。
	是否為原保戶	1 表是，0 表否。
	要保人收入	投保當時的收入。
保單特性	是否員工自保件	1 表是，0 表否。
	保險種類	終身險、定期壽險、終身健康險、新防癌險、養老險。
	繳費別	月繳、季繳、半年繳、年繳。
	繳費方式	派員收費、銀行轉帳及信用卡、業代代收、郵政劃撥。
	繳費年期	5、6、10、15、20、25、30。
	主約保險金額	投保當時的主約保險金額。
	年繳化總保費	每年各繳別按係數計算的總保費。
	有無附約	1 表有，0 表無。

	年繳化附約保費	每年各繳別按係數計算的附約保費。
	附約佔總保費比	附約保費/總保費。
	是否為孤兒保單	保單失效時，業務員是否已離職。1 表是，0 表否。
	保單預定利率	計算保費時的預定利率。
核保特性	有無體檢	1 表有，0 表無。
	是否為標準體	1 表是，0 表否。
	有無既往病症	1 表有，0 表無。
業務員特性	業務員年齡	保單生效時業務員的年齡。
	業務員服務年資	業務員到職至保單生效之期間。
	業務員教育程度	國小、國中、高中、專科、大學、研究所(含)以上。
	是否為再登錄	1 表是，0 表否。

業務員職位	保單生效時職位。壽險顧問、主任、襄理、經理。
業務員 88 年 FYP	業務員 88 年的首年度保費。
業務員 88 年 FYC	業務員 88 年的首年度佣金。

由於本研究所收集到的自變數多達 32 個，為避免不重要的變數影響模式之執行結果，故分別利用因素分析及文獻回顧的方式來選取變數。在因素分析方面，經取特徵值大於 1 的主成分後，共萃取出 4 個主成分，其組成變數如表 2 所示，累積解釋變異量高達 70.17%。由於經因素分析粹取出的成分所包括的變數和本研究所定義的特性一致，因此沿用特性名稱，而不另對因素命名。而另一種變數選取方法則是採用過去文獻所提出的變數，經整理後共選出 14 個輸入變數來建構模式，如表 3。

表 2 因素分析萃取變數

主成分	說明	解釋變異量	累積解釋變異量
保戶特性	被保人保險年齡	23.46%	23.464%
	被保人婚姻		
	要保人婚姻		
	要保被保人關係		
保單特性	是否員工自保件	17.74%	41.208%
	有無附約		
	年繳化附約保費		
	附約佔總保費比		
核保特性	是否為標準體	15.80%	57.009%
	有無既往病症		
業務員特性	業務員 88 年 FYP	13.16%	70.166%
	業務員 88 年 FYC		

表 3 文獻回顧所採用變數

變數選取方式	所使用的變數	
文獻變數	被保人性別	繳費年期
	被保人保險年齡	主約保險金額
	被保人職業類別	年繳化總保費

	要保人收入	有無體檢
	保險種類	是否為標準體
	繳費別	有無既往病症
	繳費方式	是否為再登錄

2.2 邏輯特迴歸和類神經網路模式

陳建勝與林明宏(2003)以國內某家壽險公司 1998 年生效之長年期壽險保單作為預測模式建立的研究樣本，針對保戶、保單、核保及業務員等四個構面，利用類神經網路來預測壽險保單早期失效之機率。研究結果發現，業務員是否跳槽或離職及其服務品質、繳費方與保費負擔能力等因素對早期保單是否失效有很顯著的影響。其研究指出，邏輯特迴歸與類神經網路模式在整體的預測能力都非常具有說服力且模式配適度良好，適合運用於壽險保單失效之預測工具。因此，本研究嘗試利用此二法來建構預測模式，並利用因素分析和文獻回顧所萃之變數分別建立模式。但其實證結果顯示，不論從整體正確率、型 I 和型 II 誤差及誤判成本來評估，類神經網路和邏輯特迴歸的準確度都不高，表示兩種模式的預測能力皆不理想。

由上述結果可知，在採樣公司及樣本期間不同的情況下，雖然皆是利用邏輯特迴歸和類神經網路來建構模式，但本研究建構的模式其區別效果卻不如預期，其顯示出不同年度的保單和不同的公司可能也是影響保單失效的因素之一(Chenghsien et al.,2002、沈小琪, 1998)，總體經濟環境的變化和公司間的差異使得原本具有失效特徵的保單仍維持有效的情況，或是原本具備有效特徵的保單反而失效，以致模式的區別能力降低。本研究將針對邏輯特迴歸和類神經網路區別效果不佳的原因說明如下：

1. 樣本資料夾雜強烈矛盾案例：由於 Berardi and Zhang(1999)曾提及樣本資料的結構對模式預測能力有相當的影響力，當使用的樣本不具代表性、或存在太多矛盾案例時，都會導致模式的預測效果不佳。一般而言，使用一些差異不大的矛盾樣本資料，將有助於模式的建立；但是，

當所使用的訓練資料夾雜太多強烈矛盾案例，則其訓練結果就會不太理想。由於本研究樣本資料，發現資料中夾雜許多強烈矛盾案例。例如，30 歲以下已婚者以自己為被保人，健康情況良好、無附約且業務員 88 年績效為 0 的樣本共有 83 筆，其中失效件為 27 筆，有效件為 56 筆。自變數相似但保單失效與否卻呈現相反的情況，導致模式的預測效果不佳。

2. 邏輯特迴歸臨界值設定的問題：由於邏輯特迴歸可以直接計算事件發生的機率，故可依據預測值找出模式最佳的臨界值，以判別保單是否失效。圖 1a 為標準邏輯特迴歸曲線圖，當機率值為 0.5 時，其曲線最為陡峭，因此取 0.5 為臨界值，若樣本的預測值大於 0.5 則判為失效；反之，則判為有效。圖 1b 為本研究運用邏輯特迴歸所得之預測值分佈圖，因為不易從預測值的分佈狀況判斷最適之臨界值，進而影響預測模式的分類結果。因此，在處理類似問題時，邏輯特迴歸的分類效果會因臨界值之不同而有極大的差異。

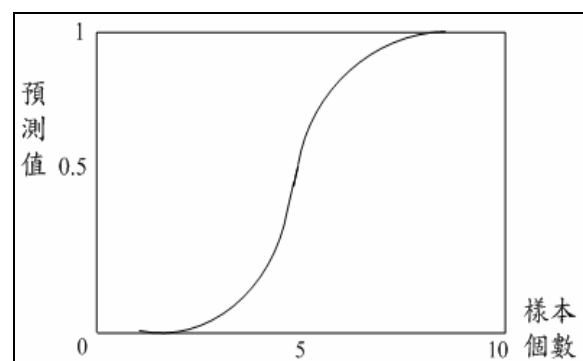


圖 1a 標準邏輯特迴歸曲線圖

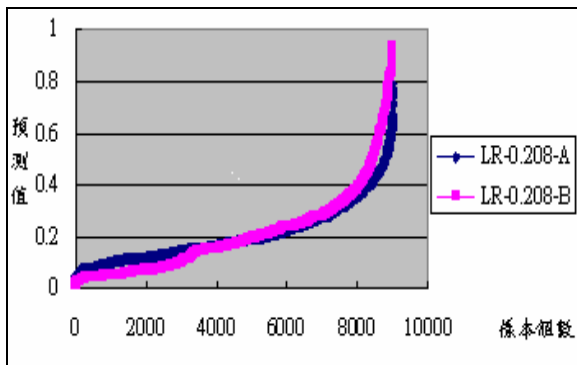


圖 1b 利用邏輯特迴歸所得到的預測值

2.3 類神經模糊模式之建構

鑒於邏輯特迴歸在臨界值的設定不易決定，必須逐一嘗試找出最佳值，以及樣本資料存在強烈矛盾案例等情況，本研究將以類神經模糊理論建構預測模式，以提升模式之預測能力。

2.3.1 類神經網路

類神經網路是由許多的人工神經細胞(artificial neuron)所組成，人工神經細胞又稱類神經元、人工神經元、處理單元(processing element)。每個處理單元的輸出以扇狀送出，成為其他許多處理單元的輸入。其基本的運作精神如圖 2 所示。輸出值與輸入值的關係式，一般可用輸入值的加權乘積和的函數(如方程式所示)來表示，其中： Y_j =輸出值。 f =轉換函數。 W_{ij} =連結加權值。 X_i =輸入層資料。 θ_j =閾值，又稱偏權值。而每一個連結上有一個數值的加權值 W_{ij} ，用以表示第 i 個處理單元對第 j 個處理單元的影響強度(葉怡成，2002)。

$$Y_j = f\left(\sum_i W_{ij} X_i - \theta_j\right)$$

輸入層資料(X)經由計算以及函數轉換，從隱藏層輸出作為輸出層之輸入，經由最後之計算與轉換由輸出層產生輸出值(Y)，然後計算其與真實值的誤差，並將此誤差從後往前推，找出應該調整的神經元及其權重(W_{ij})。整個學習的過程一直重複著

上述的動作，直到整體的誤差值低於事先所設定的門檻值為止。換言之，類神經網路可以從資料集中尋找輸入層變數與輸出層變數間的數學關係，是一種不需事先給定模式，而可以捕捉非線性關係的動態系統工具。

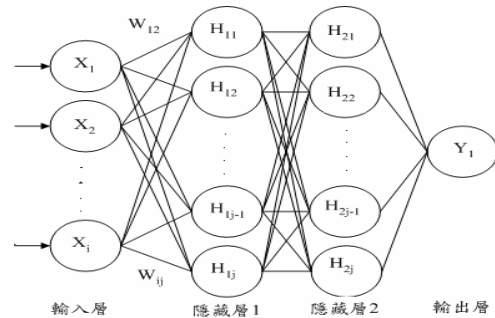


圖 2 類神經網路架構圖

2.3.2 模糊邏輯

模糊集合自從 Zadeh 於 1965 年提出發展至今，已被廣泛應用到各領域。基本上模糊理論是在處理一個物件歸屬於某一模糊集合之程度，通常用 $\mu_A(x)$ 來描述物件 x 歸屬於模糊集合 A (或稱之術語)之程度。而模糊邏輯模式之建構即在利用“IF-THEN”規則建立輸入變數與輸出變數間之對應關係，但其輸出入變數並非數值變數而是利用術語來描述的語言變數。模糊邏輯是將人們決策思考的自然語言描述，轉換成數學模式演算法的應用科學(Von Altrock, 1996)，其包括三個主要的處理程序分別為：模糊化(fuzzification)、模糊邏輯推論(inference)和反模糊化(defuzzification)。

(1) 語言變數的模糊化

模糊化為將數值變數轉換為語言變數的過程。若將某變數以三種程度來描述，則以三種程度描述的變數稱為語言變數。例如主約保險金額介於 10 萬到 1,500 萬之間，可以用 low、medium、high 等三個術語加以描述，每個術語皆有相對應的歸屬函數(membership function)。常用的標準歸屬函數分別為 Λ 型、 Π 型、 Z 型及 S 型等四種模式，本研究則採用 Λ 型、 Z 型及 S 型歸屬函數作為實證分析之用。假設某一筆保戶資料在各個變項為 {被保人保險年齡，年繳化附約保費，附約保費佔總保費比

例，是否為標準體} = { 23, 11,838, 58.6, 1 }，則各變項在各個術語上的歸屬程度可從圖 3 中對應得到如下的數值：

- 被保人保險年齡： $\mu_{low}(23) = 1$ ，
 $\mu_{medium}(23) = 0$ ， $\mu_{high}(23) = 0$
- 年繳化附約保費： $\mu_{low}(11,838) = 1$ ，
 $\mu_{medium}(11,838) = 0$ ， $\mu_{high}(11,838) = 0$
- 附約保費佔總保費比例： $\mu_{low}(58.6) = 0$ ，
 $\mu_{medium}(58.6) = 0.87$ ， $\mu_{high}(58.6) = 0.13$
- 是否為標準體： $\mu_{yes}(1) = 1$ ， $\mu_{no}(1) = 0$

換言之，該保戶在各個術語的對應值分別為：

- 被保人保險年齡： $\{ low, medium, high \} = \{ 1, 0, 0 \}$
- 年繳化總保費： $\{ low, medium, high \} = \{ 1, 0, 0 \}$
- 附約保費佔總保費比例： $\{ low, medium, high \} = \{ 0, 0.87, 0.13 \}$
- 是否為標準體： $\{ yes, no \} = \{ 1, 0 \}$

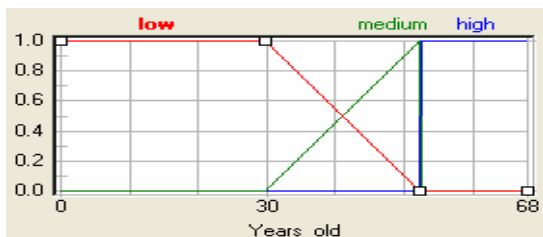


圖 3a 被保人保險年齡

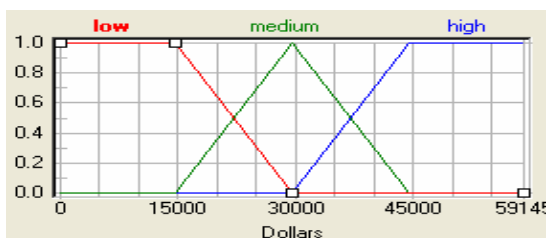


圖 3b 年繳化附約保費

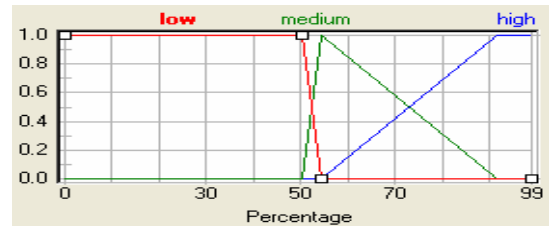


圖 3c 附約保費佔總保費比例

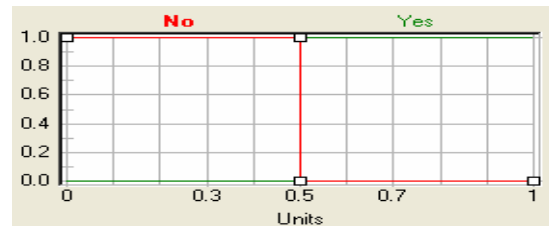


圖 3d 是否為標準體

表 4 各變數及其術語彙整表

變數名稱	變數型態	術語名稱	模糊化與否
保單第25個月內是否失效	輸出	是、否	
被保人性別	輸入	男、女	
被保人保險年齡	輸入	高、中、低	V
被保人婚姻	輸入	單身、已婚、離婚、喪偶	
被保人職業	輸入	1~6 類	
要保人性別	輸入	男、女	
要保人保險年齡	輸入	高、中、低	V
要保人婚姻	輸入	單身、已婚、離婚、喪偶	
要保被保人關係	輸入	雇主、本人、配偶、子女、父母、其他	
是否為原保戶	輸入	是、否	
要保人收入	輸入	高、中、低	V
是否為員工自保件	輸入	是、否	
保險種類	輸入	終身險、定期險、終身健康險、新防癌險、養老險	
繳費別	輸入	月繳、季繳、半年繳、年繳	
繳費方式	輸入	派員收費、轉帳、信用卡繳費、劃撥	
繳費年期	輸入	5~30	V
主約保險金額	輸入	高、中、低	V
年繳化總保費	輸入	高、中、低	V
有無附約	輸入	是、否	
年繳化附約保費	輸入	高、中、低	V
附約佔總保費比例	輸入	高、中、低	V
是否為孤兒保單	輸入	是、否	
保單預定利率	輸入	高、中、低	V
是否為再登錄	輸入	是、否	
業務員職位	輸入	壽險顧問、主任、襄理、經理	

由於同一物件可以在不同的術語上有不同的歸屬程度，故突破了過去二元邏輯中不是對即是錯之限制，而此過程即稱為模糊化。本研究所使用的變數類型包括類別變數和數值變數，由於類別變數本身已是明確的情況，因此，僅針對數值變數予以模糊化。本研究所使用的語言變數、術語及變數型態如表 4 所示。

(2) 推論機制的建構

推論機制是由一連串描述輸出入語言變數的 IF-THEN 規則所構成。在推論機制的建構過程中，為完整捕捉輸出入變數間的關係，故利用類神經網路誤差倒傳遞(error back propagation)的處理模式進行學習，逐一地調整每一條規則的權重，如此循環訓練直到停止準則被滿足而得到模糊邏輯之規則庫。其中“IF”部分衡量現狀滿足條件式的程度，“THEN”則描述系統的反應。因此模糊邏輯推論的運算基本上包括對“IF”與“THEN”等兩部分的推論計算。首先需建立的「IF-THEN」規則，如一模式有 3 個輸入變數、一個輸出變數。每個輸入變數有 5 個術語，而輸出變數有 2 個術語，則共有 $5 \times 5 \times 5 \times 2 = 250$ 條規則的組合。在類神經模糊的訓練過程中，會自動賦予每一規則一個權重(degree of support, DoS)，代表此規則相對於其他規則之重要性，權重值愈接近 1 表示規則觸動愈頻繁，即相對於其他規則之重要性愈高。

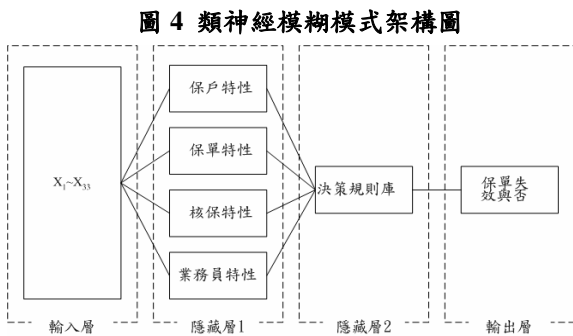
(3) 語言變數反模糊化

輸入變數在經過模糊化以及規則庫之推論後，輸出變數的每一個術語(term)會有其相對應的歸屬程度，而將規則推論的語言結論轉換到數值變數的過程則稱為「反模糊化(defuzzification)」。反模糊化的方法有三種，分別為最大中心法(Center of Maximum, CoM)、面積中心法(Center of Area, CoA)或稱重心法(Center of Gravity, CoG)及最大平均法(Mean of Maximum, MoM)。

反模糊化的目標是產生一種明確且能代表語言變數的非模糊數值。然而在選擇反模糊方法時，必須瞭解「決定最佳的折衷值」與「決定最合理的決策」等兩種口語上的意義，其在實際應用時是很重要的(Von Altrock, 1997)。在決策支援系統中，反

模糊化方法的選擇將視決策內容而定。定量的決策通常使用最大中心法，如預算分配、信用額度評估或是計畫執行優先次序；定性的決策則使用最大平均法，如信用卡真偽檢測或是客戶分類。由於本研究的輸出變數為保單失效與否，係屬於定性的決策，故選擇最大平均法做為反模糊化的方法。

由於規則庫中的規則除了其歸屬函數之決定較為主觀外，各個規則之權重亦令人難以決定，故需藉助其他方法的協助。類神經網路的學習方法即有助於對此模式的進一步修正。結合此類神經網路之訓練方法與模糊邏輯之結構化知識即構成所謂的類神經模糊。本研究所提出之類神經模糊的架構如圖 4。以類神經模糊模式的架構來看，此模式有二個隱藏層，隱藏層中的神經元個數即等於模糊規則數。



在建構完整之資料庫時，先將所有模糊規則之起始權重設為 0，利用類神經網路的學習能力來調整各規則之權重。如果某一規則的特性在資料集中確實存在，則該規則將會被觸動，其權重將被強化；若某一規則之特性並不存在於資料集中，則該規則不會被觸動，其權重會維持為 0。當預測值與實際值之誤差低於事先設定的某一門檻值時，訓練過程即結束，將所有權重低於另一門檻值的規則刪除掉，剩下的規則就代表資料集所存在的關係。將修正訓練完後的模式用來對測試集的資料作預測，如果準確度高，則表示模式建構完成。

2.4 模式預測能力之衡量

一般在評估分類模式的預測能力時，常以模式

的個別與整體正確率或誤判率高低衡量預測的效果，但二分類問題對臨界值的選定相當敏感，為克服臨界值的影響，Zweig and Campbell(1993)以接受者操作特徵曲線所包含的面積(area under receiver operating characteristic curve；簡稱 AUROC)的大小來比較模式在不同臨界值下之分類結果，並評估分類模式的區別效果。根據定義，明確度(specificity)表示正確分類有效保單之比率，敏感度(sensitivity)表示正確分類失效保單之比率，通常 AUROC 值介於 0 到 1 之間，愈大表示模式的區別效果愈佳，典型的 ROC 曲線如圖 5 所示。

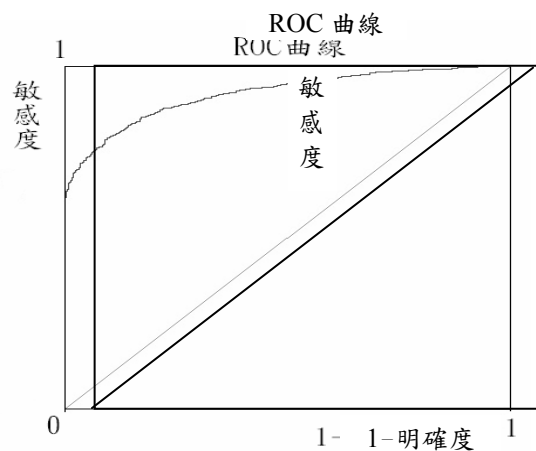


圖 5 ROC 曲線圖

利用 AUROC 的大小來衡量模式區別效果時，必須假設分類錯誤時，其型 I 錯誤(type I error)與型 II 錯誤(type II error)所造成的成本率 (cost of ratio) 等於 1，但如此的假設並不切實際(Adams and Hand, 2000)，因為就保單失效預測而言，將會失效的保單歸類為不會失效保單(即犯下型 I 錯誤)對保險公司的資金與業務品質之影響，遠較將會有效的保單歸類為會失效的保單(即犯下型 II 錯誤)之影響來的嚴重許多。因此 Boritz and Kennedy(1995)、Lin(1996)與 Boonyanunta and Zeepongsekul(2000)考慮在不同的成本率下，以誤判成本(misclassification cost)來衡量模式因誤判所造成的損失程度，進而比較預測模式的優劣性。理論上，當誤判成本最小時，二分類模式可根據特定的成本率來決定最佳臨界值，但成本率卻常因環境或管理等因素的變化而改變，因此並無固定的數據可供參考。本研究將根據

所建立的模式，以模擬的成本率來探討其變化所造成的影響。

根據 Boritz and Kennedy 的定義，型 I 錯誤是指實際為失效保單但預測為有效保單之誤判率，型 II 錯誤是指實際為有效保單但預測為失效保單之誤判率，其誤判成本的定義如下：

$$\text{誤判成本}=(A+B)/C$$

其中，A=型 I 錯誤的百分比*失效保單的比率*成本率；

$$B=\text{型 II 錯誤的百分比}*\text{有效保單的比率}；$$

C=失效保單的比率*成本率+有效保單的比率。

整體而言，本研究先以模式在訓練與測試樣本的誤判率來比較，並檢定模式之預測能力是否存在顯著的差異，再以 AUROC 排除臨界值差異對模式區別效果的影響。另外，根據模式的誤判成本來比較邏輯特迴歸與類神經網路的優劣性，並檢定兩者是否存在顯著的差異，最後以最適的模式說明影響壽險保單早期失效之重要因素。

3. 研究結果

3.1 敘述性統計

本研究針對原始樣本資料所作敘述統計分析如表 5 所示。由表 5 要保人收入、主約保險金額、年繳化總保費及附約保費的標準差都相當大，且業

務員 88 年 FYP 及 FYC 也有相同的情況，顯示保戶的經濟情況、保單的保障額度及業務員的績效差異性頗大。而要保人保險年齡和業務員年齡的平均值分別為 32.02 和 34.67，顯示業務員的銷售對象是以年齡相近的客戶為主。另外，在業務員 88 年 FYP 及 FYC 的部份，其最小值有負值的情況，這是因為契約撤銷件在業績的計算上屬於減項⁽¹⁾。若業務員上一期之契約撤銷件的 FYP 和 FYC 高於本期，那麼就會有負值的情況發生。而從平均值來看，該公司業務員 88 年 FYP 及 FYC 分別僅為 24 萬及 7 萬，推測造成業務員績效平均都很低的原因在於該公司業務員服務年資太短(平均再 2 年左右)，脫退率太高，業務員工作期間過短。假設業務員僅工作 3 個月，但業績計算卻是以一年為單位，因此將整體的平均降低了。而以類別變數的分佈情況來看，如表 6 所示。從要保人性別、婚姻狀況及要被保人關係中可看出該公司的銷售對象以已婚男女為自己及子女購買保險者居多。至於保戶繳費方式則以公司派員收費為主，佔 42.82%；繳費年期以 20 年期居多，年繳保費為主，商品銷售以終身壽險為主佔 85.47%；另外，從孤兒保單所佔的比率高達 23.78% 及業務員平均服務年資為 2 年的情況來看，顯示公司在業務員的定著率上有待加強。在核保方面，該公司的保件以無體檢件及標準體居多，這可能與被保人的保險年齡在 21 歲左右有關。

表 5 原始樣本資料數值變數的敘述統計

變數名稱	平均數	標準差	最大值	最小值	中位數
被保人保險年齡	21.33	14.86	68	0	22
要保人保險年齡	32.02	9.30	68	0	31
要保人收入	945,282.20	2,699,490.00	99,990,000	0	600,000
繳費年期	18.66	3.38	30	5	20
主約保險金額	478,024.70	532,988.10	15,000,000	100,000	26.50
年繳化總保費	15,547.43	26,261.82	540,300	211	8,180
年繳化附約保費	5,956.58	6,832.87	310,701	0	4,950
附約佔總保費比例	55.35	37.27	99	0	71
保單預定利率	6.56	0.26	8	4	7
業務員年齡	34.67	7.17	59	20	34
業務員服務年資	2.31	2.10	6	0	2
業務員 88 年 FYP	244,628.80	399,604.30	4,066,899	-190,263	93,248
業務員 88 年 FYC	73,722.66	103,696.80	625,224	-55,269	29,450

表 6 原始樣本資料類別變數的分佈情況

變數名稱	筆數	比例(%)	變數名稱	筆數	比例(%)
被保人性別			繳費別		
—男	5,786	48.24	—月繳	874	7.29
—女	6,207	51.76	—季繳	677	5.64
被保人婚姻			—半年繳	635	5.29
—未婚	7,829	65.28	—年繳	9,807	81.77
—已婚	3,967	33.08	繳費方式		
—離婚	125	1.04	—派員收費	5,135	42.82
—喪偶	72	0.60	—轉帳	2,528	21.08
被保人職業類別			—信用卡繳費	2,781	23.19
—1 類	9,705	80.92	—劃撥	1,549	12.92
—2 類	1,307	10.90	有無附約		
—3 類	504	4.20	—有	9,088	75.78
—4 類	426	3.55	—無	2,905	24.22
—5 類	25	0.21	要保人婚姻		
—6 類	26	0.22	—未婚	3,288	27.42
要保人性別			—已婚	8,384	69.91
—男	4,729	39.43	—離婚	198	1.65
—女	7,264	60.57	—喪偶	123	1.03
是否為孤兒保單			是否為標準體		
是	2,852	23.78	是	11,924	99.42
否	9,141	76.22	否	69	0.58
要被保人關係			有無既往病症		
—僱傭關係	2	0.02	—有	120	1.00
—本人	6,568	54.77	—無	11,873	99.00
—配偶	426	3.55	業務員教育程度		
—子女	214	1.78	—國小	53	0.44

—父母	4,730	39.44	—國中	1,093	9.11
—其他	53	0.44	—高中	2,975	24.81
是否為原保戶			—專科	6,738	56.18
—是	5,795	48.32	—大學	1,127	9.40
—否	6,198	51.68	—研究所以上	7	0.06
是否為員工自保件			是否為再登錄		
—是	43	0.36	—是	806	6.72
—否	11,950	99.64	—否	11,187	93.28
保險種類			業務員職位		
—終身險	10,250	85.47	—壽險顧問	6,982	58.22
—定期壽險	125	1.04	—主任	1,800	15.01
—終身健康險	1,187	9.90	—襄理	1,102	9.19
—新防癌險	201	1.68	—經理	2,109	17.59
—養老險	230	1.92			

3.2 類神經模糊預測模式

本研究先利用類神經模糊的軟體 fuzzyTECH 5.5®所提供的聚集(Clustering)功能，對樣本資料作事先篩選的工作，排除強烈矛盾的樣本，以提升訓練效果，並可縮短模式訓練的時間。fuzzyTECH 5.5®所提供的聚集功能是先計算所有資料點間的幾何距離，從距離最近的一對資料點開始，所有近似點都會被結合，這將產生一個新資料點和同時清除兩個原始點。每次一個新點產生時，其他點與這一新點的距離都將被重新計算。透過聚集功能將會解決資料內的衝突關係，以及移走多餘的資料，以便產生典型代表性的取樣資料。本研究即利用這些代表性的取樣資料來訓練及測試類神經模糊模式。

現以 NF 表示類神經模糊模式，使用訓練集資料以建構模式，並使用測試集資料驗證所建構之模式，A 為運用因素分析所萃取的變數，B 為運用文獻回顧所採用的變數。兩個類神經模糊模式之分類結果如表 7 所示，就整體正確率來而言，在排除強烈矛盾樣本後，模式正確率高達 80%，其中 NF-A 模式的整體正確率較 NF-B 模式的正確率高。而模式的型 I、型 II 誤差亦較低，顯示對樣本資料作事先篩選的工作以排除強烈矛盾樣本，確實可以改善模式的預測結果。另就模式的誤判成本和成本比值的關係圖(如圖 6)發現，NF-A 模式的誤判成本變動隨著成本比值的增加而較為平穩。

另 Berardi and Zhang(1999)亦提及，若一群體的觀察

值太少，則會因為不充分的資訊而使得分類結果不值得信賴。由於本研究之失效樣本數佔全部樣本數的比例不高，故在選擇訓練及測試樣本時，採用分層隨機抽樣的方式，以避免訓練集(in-sample data)和測試集(out-sample data)的失效率和母體的失效率相差過於懸殊，進而影響模式的判斷。

表 7 類神經模糊模式分類結果 (單位：%)

模式 簡稱	訓練結果			測試結果		
	型 I	型 II	整體	型 I	型 II	整體
	誤差	誤差	正確率	誤差	誤差	正確率
NF-A	22.67	18.64	80.90	22.67	19.69	79.97
NF-B	32.56	27.41	78.95	31.03	24.89	71.18

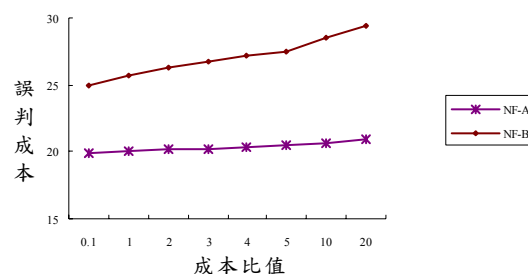


圖 6 類神經模糊模式誤判成本和成本比值的關係

3.3 規則庫之分析

為了進一步了解各個變數對保單早期失效的影響，根據模式 NF-A 在整體的分類預測能力皆較 NF-B 模式為佳，故以 NF-A 模式來作規則庫分析（表 8）。由於本研究使用的變數很多，為避免規則陳述過於繁雜，因此在模式建構過程中是以一個特性來表達多個自變數間的關係，而此特性在類神經模糊中即所謂的中介變數。從表 8 的規則 1-8 條來看，可發現當保戶特性、保單特性都為 low，核保特性為 low 或 medium 的情況下，不論業務員特性為 low、medium、high，保單都不會失效。但當保戶特性、保單特性為 low，核保特性、業務員特性為 high 時，保單卻是失效（規則 19）。如果以變數來看的話，則是表示如以本人為被保人、年繳化總保費不高且非同時為標準體、無既往症的情況下，則不論被保人年齡、要保人或被保人婚姻狀況、自保件、業務員 88 年 FYP 及 FYC 如何變化，保單都是有效的。但是如果其他條件一樣，而被保人同時為標準體、無既往症，且業務員 88 年 FYP 及 FYC 很高的情況下，保單就會失效。而從 9-18 條來看，

當保戶特性、保單特性為 high，核保特性為 medium 或 low，業務員特性為 medium 或 high 時，保單就會失效。但是當保戶特性及保單特性的條件不變下，當核保特性及業務員特性都為 low 或者核保特性為 medium 而業務員特性為 low 時，保單都不會失效。

從上述規則來看，可見保戶特性和保單特性的重要性。也就是說當保戶特性及保單特性為 low 的情況下，不論核保特性及業務員特性如何，保單都不會失效；然而，若保戶特性為 high，保單特性為 high 或 medium 時，則保單較易失效。以變數來看的話，則是指被保人年齡及婚姻、要保人婚姻、要保人和被保人關係、是否為自保件、年繳化附約保費、附約佔總保費比例對保單是否失效的影響最為顯著。因保單失效可能同時受到多個變數的影響，故單某一個變數無法準確地推論保單失效與否。換言之，在判斷保單是否失效時，必須同時考量多個變數和保單是否會失效間的關係，而非單就某一變數來做探討。

表 8 決策規則庫(部分規則)

	IF				THEN	
	保戶特性	保單特性	核保特性	業務員特性	權重	保單是否失效
1.	low	low	medium	medium	1	non_lapse
2.	low	low	medium	low	1	non_lapse
3.	low	low	low	high	1	non_lapse
4.	low	low	low	low	0.98	non_lapse
5.	low	low	medium	high	0.96	non_lapse
6.	low	low	low	medium	0.82	non_lapse
7.	low	low	high	medium	1	non_lapse
8.	low	low	high	low	0.94	non_lapse
9.	high	high	low	medium	1	lapse
10.	high	high	low	low	0.97	non_lapse
11.	high	high	medium	medium	0.96	lapse
12.	high	high	low	high	0.88	lapse
13.	high	high	medium	high	0.71	lapse
14.	high	high	medium	low	0.97	non_lapse
15.	high	medium	medium	high	1	lapse
16.	high	medium	high	medium	0.95	lapse
17.	high	medium	low	high	0.95	lapse
18.	high	medium	medium	medium	0.73	lapse
19.	low	low	high	high	1	lapse

4. 結論與建議

本研究運用類神經模糊理論的資料萃取技術辨認失效保單，可用來輔助壽險公司核保作業之執行，以期降低保單失效率及人事成本，以穩健公司的經營。過去大多數的相關研究，其結果均指出類神經網路的分類預測能力有較佳成效。然本研究實證結果顯示，就整體正確率、型 I 型 II 誤差及誤判成本而言，類神經網路和邏輯特迴歸的差距不大，其分類預測結果呈現一致的情況。再者，Boritz and Kennedy(1995)比較類神經網路、區別分析和邏輯特迴歸，發現線性模式的預測能力優於類神經網路，由此可見，類神經網路並不一定適用於所有的資料屬性。

另類神經模糊規則庫對變數間的關聯有較詳盡且具體的解釋功能，本研究發現在同時考量多個變數的情況下，被保人年齡及婚姻、要保人婚姻、要保人和被保人關係、是否為自保件、年繳化附約保費、附約佔總保費比例對保單是否失效的影響最為顯著。此外，本研究亦發現單從某一個變數並無法準確推論保單是否失效，因為保單失效可能是同時受到多個變數影響所造成的。因此，在判斷保單是否失效時，必須同時考量多個變數和保單是否會失效間的關係，而非單就某一變數來做探討。

根據過去研究指出，保單早期失效受市場利率及保單貸款利率之差的大小規模所影響。然由於本研究是針對 2000 年新生效保單在 25 個月內失效的情況作分析，因此忽略了總體經濟環境變動對保單失效的影響，且此時大部分的保單都尚未累積有保單價值準備金，而無法提供保單貸款。因此，建議未來研究可將研究期間延長，並將這二個變數納入分析，以使模式更符合實際環境。不同的公司也是影響保單失效的因素之一，公司間的差異使得原本具有失效特徵的保單仍維持有效的情况，或是原本具備有效特徵的保單反而失效，以致模式的區別能力降低。因此，建議未來研究可將公司特性加入考慮，同時比較多家公司，以找出公司間差異和保單失效的關係。

本研究所建立的規則可作為壽險公司之參考依據，但多個變數間與保單是否會失效的關係錯綜

複雜，單憑人工的判斷不僅費時且耗費人力，壽險公司可透過本研究所建立的模式，事先對投保人作篩選，且承保後亦能對可能失效的保單提早做契約保全的工作。如此，不僅壽險公司有更科學化的判斷依據，同時亦可降低保單失效率，以使企業更具競爭力。

參考文獻

- [1] 沈小琪, 1998, 「市場利率與保單失效解約率關係之研究」, 逢甲大學保險學研究所碩士論文
- [2] 林金賢、許碧芬、鄭妃君, 2002, 「利用類神經-模糊理論評定契合程度—以管理人員徵選為例」, 管理學報, 第十九卷, 77-108 頁。
- [3] 陳建勝、林明宏, 2004, 「壽險保單早期失效之預測」, 風險管理學報, 第五卷, 第三期, 341-361 頁。
- [4] Adams, N. M. and Hand D. J., "Improving the practice of classifier performance assessment," *Neural Computation*, Vol. 12, pp. 305-311, 2000.
- [5] Berardi V. L. and Zhang G. P., "The effect of misclassification costs on neural network classifiers," *Decision Sciences*, Vol. 30, No. 3, pp. 659-682, 1999.
- [6] Boonyanunta, N. and Zeephongsekul, P., "State of the Art Credit Risk Analysis Model: Comparative Analysis between Statistical Approaches and Neural Network Approaches," *The Fifth Conference of the Association of Asian-Pacific Operations Research Societies within IFORS*, Singapore, 2000.
- [7] Boritz, J. E. and Kennedy D. B., "Effectiveness of Neural Network Types for Prediction of Business Failure," *Expert System with Applications*, Vol. 9, No. 4, pp. 503-512, 1995.
- [8] Chenghsien T., Weiyu K. and Chen W. K., "Early surrender and the distribution of policy reserves," *Insurance: Mathematics and Economics*, Vol. 31, pp. 429-445, 2002.
- [9] Lin, S. L., "Financial Distress Classification in

- the Life Insurance Industry,” *Journal of Insurance Regulation*, Vol. 14, No. 3, pp.314-342,1996.
- [10] Von Altrock, C., *Fuzzy Logic & Neuro Fuzzy Applications in Business and Finance*, Prentice-Hall, Upper Saddle River, New Jersey, 1997.
- [11] Zadeh, L. A., “Fuzzy sets,” *Information Science*, Vol. 8, No. 3, pp. 338-353,1965.
- [12] Zweig, M. H. and Cambell, G., “Reciever-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine,” *Clinical Chemistry*, Vol. 39, No. 4, pp. 551-577,1993.