

行政院國家科學委員會專題研究計畫 成果報告

應用知識探勘於顧客回應模式之研究

計畫類別：個別型計畫

計畫編號：NSC92-2416-H-168-003-

執行期間：92年08月01日至93年07月31日

執行單位：崑山科技大學資訊管理系

計畫主持人：高淑珍

共同主持人：吳建興

計畫參與人員：陳侑瑜, 麥靜怡

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 9 月 1 日

行政院國家科學委員會專題研究計畫成果報告

應用知識探勘於顧客回應模式之研究

A customer response model using data mining approach

計畫編號：NSC 92-2416-H-168 -003

執行期限：92年8月1日至93年7月31日

主持人：高淑珍 崑山科技大學資訊管理系

一、中文摘要

近年來智慧型計算(Intelligent Computing)和顧客回應模式(Customer Response Model)由於能夠協助企業挖掘出顧客隱性知識並預測具回應傾向的顧客因此逐漸受到重視。本研究希望能夠利用在智慧型計算中頗受重視的資料探勘技術，將交易資料所挖掘所得的購買行為型樣藉以建立顧客回應模式，協助企業更正確地預測出行銷推廣的目標顧客以及後續之行銷策略。研究中以國內 A 壽險公司為例，蒐集該公司 188464 筆顧客交易資料，透過決策樹歸納演算法加以探勘，模式包括 943 項具代表性的顧客回應之預測性規則，並經過驗證達到 80.7% 的預測準確率。此外，研究中將所得到的規則加以歸納，進一步探究屬性與熱賣壽險類別之間的關係以供擬定行銷策略之參考。

關鍵詞：顧客回應模式，資料探勘，目標行銷

Abstract

With the increased use of intelligent computing, target marketing that utilizes data mining in databases to seek for the potential customers as well as to derive marketing strategies is gradually capturing attentions of management. The Customer Response Model (CResM) with the capability of potential customers prediction is a useful tool that can help in targeting customers who are most likely to reply marketing promotion. In this paper, a data mining based CResM is proposed to help elicit potential customers who is likely to response for marketing promotions. The size of the transaction database collected from a leading insurance company in Taiwan was 188464. The mining mechanism used was the induction-based algorithm. An 80.7% test accuracy was obtained, indicating that the mining mechanism used was adequate. The final results obtained were 943 decision rules. Knowledge interpretation as well as managerial implications with respect to the application case were also provided in this research also.

Keywords: Customer Response Model, Data Mining, Target marketing

二、緣由與目的

隨著全球性經濟的興起以及市場環境 科技的改變，傳統的行銷組合 4P 在 e 世代應該有個全新的競爭策略上的改變，其中在促銷(Promotion)方面應強調雙向互動及有效的廣告回應[1]。因此企業必須著重目標行銷(target marketing)，亦即針對顧客的個人化需求以從事行銷活動才能有效地找到目標顧客(target)、對其行銷適當的產品，同時有效地節省推廣(promotion)的成本，因此尋找和預測目標顧客這兩項已被視為企業促銷推廣的核心決策[2]。近年來智慧型計算(Intelligent Computing)在行銷領域的應用已逐漸受到重視即是一例，其主要的應用在於挖掘出隱藏於大量交易資料的顧客隱性知識，以進行各種顧客區隔以及行銷策略之制定[3,4,5,6]。另一方面，顧客回應模式(Customer Response Model)根據某顧客的獨立之預測資訊(例如：年齡、居住區域、收入、教育程度)而得到一個分數用以預測該顧客將對於某特定的促銷活動有所回應的可能性[7]，因此可以協助行銷人員進行市場區隔(segmentation)和鎖定目標(targeting)這兩項攸關行銷資源配置的主要活動。

近年來已有一些應用在行銷方面的顧客回應模式被發展出來，其中所使用的方法大致上以統計工具為主，近年來則有逐漸應用學習演算法(learning algorithm)的趨勢[8]。在目前較常使用的資料探勘技術中，統計、決策樹和類神經網路常被用來解決分類(classification)的問題。根據文獻上針對資料探勘技術效能的相關研究，將幾種解決分類問題的技術之預測效能加以比較後發現：在分類預測準確率方面，ID3 與類神經網路的預測準確率相當，而優於 C4.5, CHAID 以及 CART 等其他決策樹演算法[10]；就解釋能力來說，ID3 易於將決策樹展開成為決策規則，因此優於類神經網路；而就規則的精簡度來看，ID3 演算法所得到的規則數相較於 C4.5 則更為精簡[9]，因此本研究決定採用 ID3 為整個顧客回應模式的工具。

另一方面，國內保險產業因為競爭激烈而使彼此差異化程度逐漸降低，因此無不希望藉由資

訊科技的輔助創造本身的競爭優勢。對保險業來說，雖資料探勘並非萬靈丹，但對未來銀行保險與新保單的開發將扮演重要的角色。整體而言，資料探勘在保險業的應用主要可以協助解決以下問題：(1) 保單制定；(2) 保戶區隔及增進顧客關係；(3) 找出保險詐騙的可能性並加以事前預防 [3, 11, 12]。黃品豪(2002)在採訪國內某外商保險公司將資料探勘技術應用在電話行銷的結果時發現，原本為 13%的成交率，在依過去成交紀錄設計模型與分析資料後，篩選出來「最容易成交」的前 10%客戶名單進行電話行銷，成交率便高達 50% [13]。由此可見資料探勘在保險業的行銷上將可省下不少的時間與人力。然而如何利用資料探勘的技術將保險業之大量交易資料作深入的探探，並萃取出有價值的知識，以進一步藉此來建構顧客回應模式，對於國內的保險業者而言仍是一項尚未被深入研究的課題。

綜合上述之文獻探討，本研究以國內 A 壽險公司為例，藉由壽險交易資料的探勘發掘顧客回應知識，並進而建構出顧客回應模式。目的除了協助壽險業者挖掘出系統性的知識和規則之外，也藉此探討整體過程對於 A 公司之管理啟示。

三、研究方法

(一) 研究對象

A 壽險公司創立於民國五十二年七月三十日，目前是我國第二大保險公司，總資產為 5,980 億 2,200 萬，總營業額達 1,979 億 7500 萬，保戶總數為 513.4 萬人。公司的組織包括總公司的 30 個部室及 75 課、1 個教育中心及 4 個地方研修處，並在北京及香港各設立辦事處，員工總數達 17,918 人。在產品方面，公司提供完整的壽險產品，而在公司電腦化的進程方面，公司持續積極地推行電腦化。此外，公司內部已設置常態性的市場行銷部門，利用人口變項來作集群分析以針對市場開發以及新產品設計加以分析研究。其所使用的行銷工具及通路仍停留在傳統工具的階段，尚未利用公司建置的資料庫進行任何資料庫行銷、目標行銷等行銷活動。

(二) 研究架構

本研究架構包括模式產生、模式驗證以及結果三大部份：在模式產生的部份主要是將蒐集所得三分之二的交易資料加以前置處理以利於資料探勘的挖掘，所產生的顧客回應規則，將用以建構顧客回應模式；而模式驗證的部份則在於驗證所產生之顧客回應模式在預測準確性方面是否良好，將以蒐集所得三分之一的交易資料加以驗證；至於結果的部份則著重在此顧客回應模式所衍生之行銷推廣之策略建議以及管理意涵的探討。

(1) 內部資料

本研究蒐集個案公司保戶投保資料自民國 62 年 7 月至民國 90 年 11 月止共 188464 筆，欄位包括 14 個屬性。

(2) 資料之前置處理

主要任務是將不完整資料加以去除，並針對連續性資料加以間斷化。

(3) 行銷推廣之預測性規則

If 條件集合 Then 險別分類結果(支持筆數, 支持度)

(4) 規則篩選

a. 篩選標準 1

$$\text{支持筆數} > \frac{N}{\prod_{i=1}^n x_i}$$

N : 所有交易記錄的總筆數

x_i : 屬性之分級數目, $i=1, 2, \dots, n$

n : 規則屬性之屬性總數

b. 篩選標準 2

$$\text{支持度} > \frac{1}{n}, n > 1$$

n : 相同條件下所產生之結論數

(5) 預測準確率之驗證

驗證步驟 1 : 以規則來預測驗證資料之結論，並將結果與實際交易的結論加以比較。

驗證步驟 2 : 統計所有預測結論與實際結論相同的記錄筆數，並計算其佔所有驗證資料的比例。

驗證步驟 3 : 以 60%為準確率驗證之門檻，並將前一步驟所得到的模式預測準確率與之作比較。

(三) 研究結果

(1) 顧客回應模式的預測性知識

本研究經 ID3 決策樹歸納演算法，得到各屬性之訊息量以及屬性訊息獲取量，從中選擇最具分類能力之七項屬性：依序為保費分級、財力、繳費年期、投保始期、投保年齡、繳費方式以及保額分級。共獲得 2689 項規則，經篩選後總共得到 943 項具代表性之規則，附件一為所產生之部份規則。

(2) 預測準確率的驗證結果

利用所獲得的 943 項規則來預測剩餘的三分之一投保資料(68370 筆)，發現其能夠正確預測結果的記錄為 55172 筆，佔所有驗證資料筆數之 81%。

(3) 顧客回應之描述性知識

a. 財力越低者越傾向購買終身健康醫療型之產品，但隨著財力的增加，則投保終身年金型產品的比例也逐漸提高。另一方面，中低等財力的保戶最不喜歡的是定期型或還本型產品，而隨著財力越高，保戶越不喜歡功能有限或只有死殘理賠功能的險種。

b. 舊投保人過去所購買的險種幾乎都集中於終身年金型，只有少部份的保戶購買終身健康醫療型；中期保戶所購買的終身健康醫療型的數量最多，其次為終身年金型產品、終身保障型產品；新投保人所購買的產品則以終身健康醫療型或終身健康醫療還本型為大宗。

c. 青少年與壯年為目前壽險市場的主力年齡層，其次為幼童族群，老年壽險市場預期將有許多空間。而幼童與青少年之投保產品均以保障兼保本的類型居多，壯年與老年這兩個年齡層則以醫療保障和死殘保障為主。此外，最受到幼童及青少年青睞的終身年金型產品隨著被保險人年齡的增長，銷售比例隨之快速遞減。

四、結論

本研究以擁有大量顧客交易資料且產業競爭激烈的壽險業為例，試著初探將知識探勘技術推廣到壽險公司之行銷決策上的實務價值。整個研究在蒐集了 A 壽險公司十八萬多筆交易資料，並經過資料清潔、資料轉換等前置處理步驟之後，將其中三分之二的交易資料以廣為被使用的 ID3 歸納演算法加以萃取出 943 項較具代表性的規則，用以預測保戶對於某類壽險產品的回應。所得之顧客回應規則也經準確率之驗證以確認此模式的預測能力頗佳。隨後本研究更將所萃取出來的 943 項規則施以深入的歸納分析，就屬性不同的程度與險別分類之間的關係加以探究推論其背後之原因，藉以累積更為精煉的顧客回應知識。

五、參考文獻

[1] Pine, J. (1999), "Mass customization: the new frontier in business competition", *Harvard Business School Press*, Boston

[2] Zahavi, J. & Levin, N. (1995), "Issues and problems in applying neural computing to target marketing", *Journal of Direct Marketing*, 9(3), pp. 33-45

[3] Adriaans, P. & Zantinge, D. (1996), *Data mining*, Reading Mass.: Addison-Wesley

[4] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R. (1996), *Advances in knowledge discovery and data mining*, AAAI Press

[5] Joseph P. B. (1996), *Data mining with Neural Network*, McGraw-Hill

[6] Berson, A. & Smith, S. J. (1997), *Data warehousing, data mining, and OLAP*, McGraw-Hill

[7] Frothinger, C. (1993), "Results oriented direct marketing", *Bank Marketing*, 6, pp.24-28

[8] Levin, N., Zahavi, J & Olitsky, M. (1995), "AMOS - A probability-driven, customer-oriented decision support system for target marketing of solo mailing", *European Journal of Operational Research*, 87(3), pp.708-721

[9] Ohmann, C., Moustakis, V., Yang, Q. & Lang, K. (1996), "Evaluation of automatic knowledge acquisition techniques in the diagnosis of acute abdominal pain", *Artificial Intelligence in Medicine*, 8(1), pp.23-36

[10] Katharina, D. C. Stark & Dirk, U. Pfeiffer (1999), "The application of non-parametric techniques to solve classification problems in complex data sets in veterinary epidemiology- an example", *Intelligent Data Analysis*, 3(1), pp.23-35

[11] Hoffman, T. (1999), "Finding a rich niche", *Computerworld*, February, pp. 44

[12] Chae, Y. M., Ho, A. H., Cho, K. W., Lee, D. H. & Ji, S. H. (2001), "Data mining approach to policy analysis in a health insurance domain", *Intern. J. of Medical Informatics*, 62(2-3), pp. 103-111

[13] 黃品豪, 2002/7/29, 『資料採礦在保險行銷的應用』, 現代保險雜誌

附表一：顧客回應模式之部份結果

規則編號	保費分級	財力	繳費年期	投保始期	投保年齡	繳費方式	主約保額	險別分類	支持筆數	支持度
1	中	中	中	新	幼童	月繳	中	1	72	72/85
2	中	中	中	新	幼童	月繳	少	2	35	35/167
3	中	中	中	新	幼童	年繳	中	12	76	76/378
4	中	中	中	新	幼童	年繳	少	1	650	650/913
5	中	中	中	新	老年	月繳	中	10	10	10/11
6	中	中	中	新	老年	月繳	少	10	20	20/27
7	中	中	中	新	老年	半年繳	中	10	10	1
8	中	中	中	新	老年	年繳	少	10	100	100/185
9	中	中	中	新	老年	年繳	少	9	43	43/185
10	中	中	中	新	壯年	月繳	中	10	90	90/123