

國立台灣大學國際企業學研究所

碩士論文

Graduate Institute of International Business

College of Management

National Taiwan University

Master Thesis

線上新產品推薦系統－以亞馬遜網路書店為例

Internet Recommendation System-

Take Amazon.com as an Example



陳宛伶

Chen Wan-Ling

指導教授：任立中 博士

Thesis Adviser : Jen Li-Chung, Ph.D.

中華民國九十七年六月

June, 2008

謝 詞

終於到了欣然畢業的時刻，這篇論文熬過了酷暑，撐過了寒冬，就這樣誕生了！最感激的當然是對我諄諄教誨不斷的任立中老師，感謝他對我展現了嚴師的專業，又帶著慈父的可親，讓我在碰到挫折窒礙難行時，永遠有最後一道防線。不論是在指導論文，或是陪我規劃人生道路，任老師始終適時的扮演為師的角色，協助我分析是非，評估優缺好壞，也不時的給我信心打氣，讓我能夠將模糊的未來看清，非常感謝任老師！

其次也要感謝我的口試委員陳厚銘老師與謝明慧老師，讓我的論文能有更完整的面貌，陳老師對我們表示的肯定也讓我覺得很欣慰，對於畢業後迎面而來的挑戰更有加倍的勇氣去面對。還有同為任門的好戰友們，非常感謝雅茹、郁茹、亭君、培正，在論文剛起步時與我一起實施高速發展計畫，也很感謝大家的關心與相互扶持，和你們同門真的很幸運。

特別感謝繼堯陪我度過碩班的大部分時間，很高興你也是接受任老師的指導，在整理資料庫的枯燥過程中，有你在身邊與我共同努力，就不再這麼無趣了。當我遇到瓶頸或出現疑惑時，你是第一個我期待能夠給我解答的對象，你也沒讓我失望，總是耐心地與我辯明真理，謝謝你讓我對每件事情都能更確定。

最後要誠摯感恩的就是我的家人，爸爸、媽媽和姐姐，你們給我的真的太多了，不只是在我讀研究所，撰寫論文的這段時間，從我成長至今就不停地在消費你們的愛，你們絕對是我最初、也是永不結束的感謝！

陳宛伶 謹誌
于台大管理學院
民國九十七年六月

摘 要

本研究是以資料庫行銷的技術來剖析每一位顧客的真實消費狀況，利用顧客過去的交易紀錄，來推測其未來的購買行為，同時導入顧客推薦系統的概念，將顧客視為有不同偏好的獨立個體，分別推薦顧客不同的產品，以期能為企業贏得最大的利潤。

全文通篇將資料庫內所有的顧客劃分為兩群組：第一群顧客是交易紀錄超過一筆，顯示為曾經重複購買者，被視為「舊顧客」；第二群顧客是僅有一筆紀錄，則視其為「新顧客」。從舊顧客的資料中個別挑選出最近期的一次紀錄，作為樣本內估計顧客購買偏好，及分析顧客購買行為的依據；剩餘的所有資料則分別用以預測舊顧客與新顧客的購買機率。一旦了解顧客的購買偏好與發生交易的機率後，便可以對顧客進行最直接的產品推薦。

首先由顧客最初的購買行為之需求端切入，將顧客的購買偏好依照相異程度的不同，導入統計方法來分析，並且估計顧客購買偏好，來預測顧客下一次的購買；若預測顧客應該會購買，而資料顯示顧客確實有購買紀錄時，則謂之「擊中」。而後便比較此三種不同的新產品推薦模式，以累積擊中率來判斷其模式的優劣，分別包含平均機率法推薦模式、總合邏吉斯推薦模式，以及在行銷理論模型中，充分展現預測效果最佳的層級貝氏邏吉斯推薦模式。

最終實證後所得出的結論顯示，層級貝氏邏吉斯推薦模式的表現的確最好，但是仍有些許外在因素干擾其預測結果，倘若加以改善則可擁有更好的預測能力。本研究逐一針對各推薦模式的實證結果給予正反向評判，期望能由顧客關係管理中精準的資料庫行銷技術當作指引，一方面提供顧客客製化的推薦服務等，另一方面則帶領企業獲得顧客的終身價值，協助企業帶來更大的商機，創造買賣雙贏的局面。

【關鍵字】：資料庫行銷、顧客推薦系統、層級貝氏模式、亞馬遜網路書店

Abstract

It's getting obvious that customer relationship management could be viewed as a lethal weapon. To give the customers exactly what they want in affordable price can easily enhance the customer satisfaction. And this study is based on database marketing techniques, which is the key point of CRM, to analyze each customer's purchasing behavior. By examining the transaction records, we'll predict each customer's next purchasing behavior, and apply the concept of customer recommendation systems to customize their recommended products. We believe that with one strong and precise recommendation system, we could encourage cross-buying, develop customer loyalty, and finally improve the customer retention, which would lead to great profits.

In this research, we hold two purposes, one is to find out the online books buying preference, the other is to compare the different kinds of recommendation systems. We separate all customers into two groups by their repetitive purchasing in turn representing "Old Customers" and "New Customers". Later we try three types of statistical models, "the Common Average Method", "the Aggregate Logit Recommendation System", and the "Hierarchical Bayesian Logit Recommendation System", and see which one of them can perform the best in the accumulated hit ratio for predicting customers purchasing possibilities.

The study result shows several buying habits on different cluster of customers, for example, people with only high school educational backgrounds prefer buying "Business and Finance" to "Lifestyle" types of books, they would like to buy books that are thicker, and so on. In addition to the purchasing behavior, we also found that Hierarchical Bayesian Logit Recommendation System do the best prediction, just like the Hierarchical Bayes theory proposed, no matter for old customers or new customers. It's is quite evident that more and more customers now they share heterogeneity needs,.

In that, to best serve all individual's need, the company better keep on customizing the recommendation system.

Keywords: Customer Relationship Management; Database Marketing; Internet Recommendation systems; Hierarchical Bayes Model; Amazon.com



目 錄

謝詞	i
中文摘要.....	ii
英文摘要.....	iii
目錄	v
圖次	vii
表次	viii
第一章 緒論.....	1
第一節 研究背景與動機.....	3
第二節 研究目的.....	6
第三節 研究範圍.....	7
第四節 論文架構.....	7
第五節 研究流程.....	8
第二章 文獻探討.....	9
第一節 顧客關係管理.....	9
第二節 顧客推薦系統.....	20
第三節 隨機品牌選擇模式.....	24
第三章 研究方法.....	29
第一節 研究架構.....	29
第二節 研究設計.....	30
第三節 邏吉斯產品選擇模型.....	39
第四節 層級貝氏統計模型.....	42
第五節 層級貝氏邏吉斯模型.....	49

第四章	實證分析.....	57
	第一節 樣本描述.....	57
	第二節 無任何資訊下之推薦系統.....	64
	第三節 總合邏吉斯模型分析.....	67
	第四節 建立線上推薦系統.....	76
第五章	結論與建議.....	89
	第一節 研究發現.....	89
	第二節 策略意涵.....	93
	第三節 研究限制與未來建議.....	95
參考文獻	97



圖次

圖 1-5-1	研究流程圖.....	8
圖 2-1-1	顧客關係管理程序.....	12
圖 2-1-2	顧客關係管理模式.....	15
圖 3-1-1	研究架構.....	29
圖 4-1-1	各商品變數的樣本比例.....	59
圖 4-1-2	人口統計變數的樣本分布比.....	60
圖 4-3-1	舊顧客累積推薦擊中率比較(平均機率法與總合邏吉斯模式).....	74
圖 4.3-2	新顧客累積推薦擊中率比較(平均機率法與總合邏吉斯模式).....	74
圖 4.4-1	舊顧客累積推薦擊中率比較..... (平均機率法、總合邏吉斯模式與層級貝氏邏吉斯回歸模式)	85
圖 4.4-2	新顧客累積推薦擊中率比較..... (平均機率法、總合邏吉斯模式與層級貝氏邏吉斯回歸模式)	85



表 次

表 1-1-1	顧客演進與改變.....	4
表 2-1-1	顧客關係管理定義.....	10
表 2-2-1	常見推薦系統模式.....	21
表 2-3-1	品牌選擇模型之文獻整理.....	27
表 3-2-1	商品屬性與人口變數內容.....	32
表 3-2-2	各商品屬性對應之虛擬變數整理.....	34
表 3-2-3	各人口統計變數對應之虛擬變數表.....	35
表 3-2-4	直交設計所得出之 27 組不同的產品屬性組合.....	36
表 3-2-5	顧客的購書紀錄及各變數之資訊.....	37
表 3-2-6	產品屬性編碼轉換表.....	38
表 4-1-1	本研究使用之資料類型.....	58
表 4-1-2	預測舊顧客購買機率的購買考慮集合.....	62
表 4-1-3	預測新顧客購買機率的購買考慮集合.....	63
表 4-2-1	顧客考慮集合個數.....	65
表 4-2-2	平均機率法的舊顧客推薦累積擊中率.....	65
表 4-2-3	平均機率法的新顧客推薦累積擊中率.....	66
表 4-3-1	利用總合邏吉斯模式估計所得各變數的迴歸係數.....	68
表 4-3-2	考慮集合中的相對效用值(以顧客編號 5662676 為例).....	71
表 4-3-3	總合邏吉斯回歸模式的舊顧客推薦累積擊中率.....	72
表 4-3-4	總合邏吉斯回歸模式的新顧客推薦累積擊中率.....	73
表 4-4-1	利用 Gauss 所估計出的最佳後驗參數值.....	77
表 4-4-2	層級貝氏邏吉斯推薦模式對舊顧客的累積擊中率.....	78
表 4-4-3	利用 Gauss 所求得的 Θ 後驗估計平均值.....	81
表 4-4-4	利用 Gauss 所求得的 Θ 後驗估計平均值(續).....	82
表 4-4-5	新顧客的偏好係數值.....	83
表 4-4-6	層級貝氏邏吉斯推薦模式對新顧客的累積擊中率.....	84
表 4-4-7	顧客交易紀錄筆數對累積推薦擊中率的影響.....	87

第一章 緒論

近來廣為討論的資料庫行銷，本着其能利用資料倉儲（Data Warehousing）及資料探勘（Data Mining）等技術，儲存資料庫，結合行銷學與統計學，將顧客的行為透過模式精準的分析，以其結果作為行銷決策之依據，能讓企業提供客製化（Customization）的商品與服務，也是一對一行銷的利器，更是企業達到顧客關係行銷之重要媒介。於是在若干年後，當眾多行銷及各界學者的研究成果已經累積達一定的程度，便將此一技術推向了實務界，使得企業在擬定行銷策略時，能獲得許多消費者行為的資訊，及更多量化的依據，且其在預測顧客的行為上也變得較為理性，因此資料庫行銷便陸續被廣泛的應用在行銷實務界。

資料庫行銷的技術能夠帶給企業界眾多協助，包含協助企業了解其真正的顧客為何？未來所欲推出的新產品，又該選定何者作為其目標顧客群？除此之外，企業的行銷 4P 策略擬定，更有許多地方可仰賴資料庫行銷的技術，諸如研究消費者購買組合的內容，可以讓企業更加清楚顧客對每一種產品類別的主要與次要需求；研究消費者購買的平均額度，更能幫助企業在替產品定價時，有大概的上下限輪廓；而研究購買行為的間隔時間，能夠讓企業有效的加入某些促銷方案，以干預顧客的消費頻率，例如有些台灣的日用品等賣場業者，會每隔 1~2 個月便透過寄發傳單，來提醒消費者再度購買。

上述所提及的僅只是少部份資料庫行銷的參考價值，仍然有非常大比例的使用，在此考慮篇幅便不提及。總之，資料庫行銷的最終目的是要能夠使企業在行銷其產品或服務時，能夠在正確的時間，以正確的方式，服務正確的顧客，讓顧客的滿意度極大化，企業能夠獲得與其維持長久的顧客關係，獲得顧客的終身價值，提高顧客的再購率及其荷包佔有率(wallet share)。而欲達成這個目的，首要任務便是研究顧客資料庫的內容，有效的針對每一位顧客給予客製化的產品及服

務，利用過去消費的紀錄，針對其購買的產品組合，去思考顧客背後真正的需求，以及對該類型的產品，所屬的偏好為何，並進而將適合的產品銷售給合適的顧客，以達成雙贏的局面。



第一節 研究背景與動機

「我們完全知道未來我們要怎麼走，因為我們的顧客會指引我們方向」全球知名的企業 AT&T 執行長 Jerry Stead 這麼說道。他說得沒錯，科技持續的進步，而人的消費行為也改變著，從前顧客的角色是被動的接受者，而隨著自我意識的高漲，顧客在消費的行為上也逐漸要求自主的權利，於是現在顧客與企業所扮演的角色，是立足在平等的互惠位置，如表 1-1-1 所示。

企業面對顧客態度上的轉變，也發展出因應之道，即將顧客視為朋友，與其經營長久的關係，藉由深入了解每位顧客的習性，提供適合每個人的產品和服務；利用各種工具的幫助，例如 email、或溫馨的生日小卡等，來聯繫與顧客之間的情誼；邀請顧客加入網際網路社群，了解其對企業的意見，並留下更多讓企業深入了解顧客的線索；又或者是透過類似像專人服務的形式，針對顧客過去的消費型態，給予購買上的建議。在網路當道，電子商務發燒之際，瀏覽眾多拍賣網站或網路商店時，不難發現虛擬的商店店員正在提供你類似如下的訊息：少女服飾的衣著賣場—你也可以這樣穿（當游標指向一件白色襯衫時，點進該連結後便出現一個模特兒，身穿該件白色襯衫，搭配圍巾、外套及牛仔褲）；網路書店—有買這本書的人，同時也買了這些書（當點進某一本書籍的頁面時，該頁下方出現 4-5 本書籍可供滑鼠連結）；不僅如此，收發電子郵件時也常發現類似的訊息，過去曾在某網站消費，於是便持續收到該網站所寄出的新產品告知訊息，例如：休閒鞋的網站—勃肯鞋新春天系列（過去曾經在該網站購買過該品牌的一款拖鞋，便會常接收到新鞋上市的訊息）。

表 1-1-1 顧客演進與改變

	顧客是被動的觀眾		顧客主動參與創造價值
顧客角色	顧客是被動的買主，事先決定好的消費者角色		顧客是企業網路的一部份，共同創造和攫取額外的企業價值。 顧客是合夥人、共同開發者，也是競爭者
與顧客的互動	傳統市場研究與調查：產品與服務的創造，沒有太多的顧客回饋	透過服務檯、顧客服務專案，從銷售轉變成協助顧客；經由顧客找出問題，然後根據回饋重新設計產品與服務	觀察使用者，然後提供正確的產品和服務給顧客；從領先使用者找出解決方案，深入了解顧客，以便重新規劃產品與服務
溝通目的	大眾行銷——單向溝通	資料庫行銷——雙向溝通	關係行銷——雙向溝通與存取
			與顧客積極對話、塑造期望並創造風潮——多層次的存取與溝通

資料來源：本研究整理

種種的行為都讓人不難發現，過去顧客傾向自行收集產品相關資訊，直到購買行為發生前都是顧客一個人的事，低涉入的產品更猶是；但今非昔比，現在的顧客普遍擁有網際網路的管道，可以自行收集所有相關資料，但是席捲而來的資訊爆炸，卻可能使顧客無法或不願意在有限的時間條件下，完成所有的資訊收集動作，於是各網路商店便祭出各式推薦系統的虛擬專員，來檢視每位顧客的消費紀錄，以便針對其偏好提供個人化的推薦內容。

本研究是受到來自 Ansari, Essegaiier, and Kohli (2000) 針對電影建立了一套網路推薦系統的啟發，該篇研究結果發現利用模式去推估各部電影的評等，能夠有效的將對各特定顧客而言是優等的電影，推薦給該名顧客，於是本研究便想進一

步探討，使用推薦系統甚多的網路書店，若建立一套同時考量顧客異質性與產品特性的推薦系統，是否也可有效的達成預期的效果，增加購買機率，而使得電子商務得以更加有效率的營運。



第二節 研究目的

本研究擬利用聯合分析法的概念，將顧客發生購買行為之前，會面臨的各種產品屬性組合加以重組配對，列出所有可能的考慮集合，並且結合顧客過去的交易紀錄，進行產品內各屬性的綜合分析，一來可作為提供給實務界在產品的設計上有更多的消費者意見，二來在行銷該產品時，企業能夠事先衡量該產品內各屬性的重要性權重，並且確認顧客認知內最有價值的屬性。上述僅只顧及到產品層面，是早期傳統的行銷手法，將目標顧客定位為彼此享有同質需求的大眾市場，提供一種最合乎大眾口味的產品給所有的顧客；然而，本研究的宗旨便是在了解顧客個別的偏好模式，將每位顧客的異質性融入產品的選擇結果之中，強調目標顧客是由不同區隔的小眾群體齊聚而成，甚至可以是以獨立的個體為單位，而後利用層級貝氏邏輯模式了解不同的顧客針對不同的產品屬性組合，完成購買的機率，這麼一來，不但能夠使企業更深入了解產品各屬性的重要性結構，同時還能藉由顧客關係管理的方式進行一對一行銷，讓企業徹底服務每一位顧客。

承上，本研究主要目的可總結如下：

- 一、結合聯合分析法、邏輯迴歸分析模式、層級貝氏統計理論，建立一套線上新產品的推薦系統，以供企業用來推薦新產品給既有的顧客群，並建立線上新顧客的推薦系統，以供企業用來推薦既有的產品給新的顧客群。
- 二、以合乎邏輯的推演過程，將過去的交易紀錄轉化為具體的預測購買機率資料，提供企業分析、決策，以擬訂對未來營運有幫助的行銷策略。
- 三、利用本研究強調顧客異質性的前提，將個別消費者的偏好作一整合性的分析，以便企業評估其個人化行銷的可能性及獲利性，增益客製化行銷。

第三節 研究範圍

本研究將以全球知名網路商店的顧客為分析的對象，資料來源是取自網路市調公司 comScore，該公司針對 50000 名遍及美國各地的網路使用者，紀錄其上網瀏覽與消費的行為。所提供之資料庫內容包含 2004 年該年度，每一名自願參與的網路使用者，連結至此一線上商店觀賞網頁，或消費任何一項產品的所有時間、產品別、金額等紀錄，當中囊括了各式產品類別的資料（如鞋類、影音光碟、書籍雜誌等），遂從中取出本研究欲討論的產品—書籍—之交易紀錄的資料，利用此一顧客線上購買書籍的交易紀錄資料庫，有助於本研究發展模式，辨別各顧客存在購買行為背後的異質需求，進而更加了解每位顧客的消費黑盒子。最後，藉由研究的成果幫助企業營運電子商務時，能夠擬定更佳的個人化行銷策略，以有效提高企業實質收益及顧客忠誠度。

第四節 論文架構

本研究將由第一章的緒論開始，介紹研究背景及研究動機，並解釋整篇論文的研究目的；透過第二章的文獻回顧，將可推論網路推薦系統實與提升顧客滿意度、達到關係行銷的目的有關，盼獲得顧客關係管理理論上的支持；緊接著由第三章的研究方法，解釋本研究主要的統計模式依據；並遞交至第四章的實證分析，由實際的資料來進行結構上的分析，並且根據真實的交易狀況，將樣本劃分為驗證用及預測用兩部分，一方面提示模式的正確性，另一方面便可用來作為預測購買機率的結果；最後在第五章的結論與建議部分，將整合所有研究發現，將其融入相關的行銷策略，最後解釋研究過程的限制，及提供後續研究方向以作為未來的研究題材。

第五節 研究流程

本論文的研究流程如下圖所示：

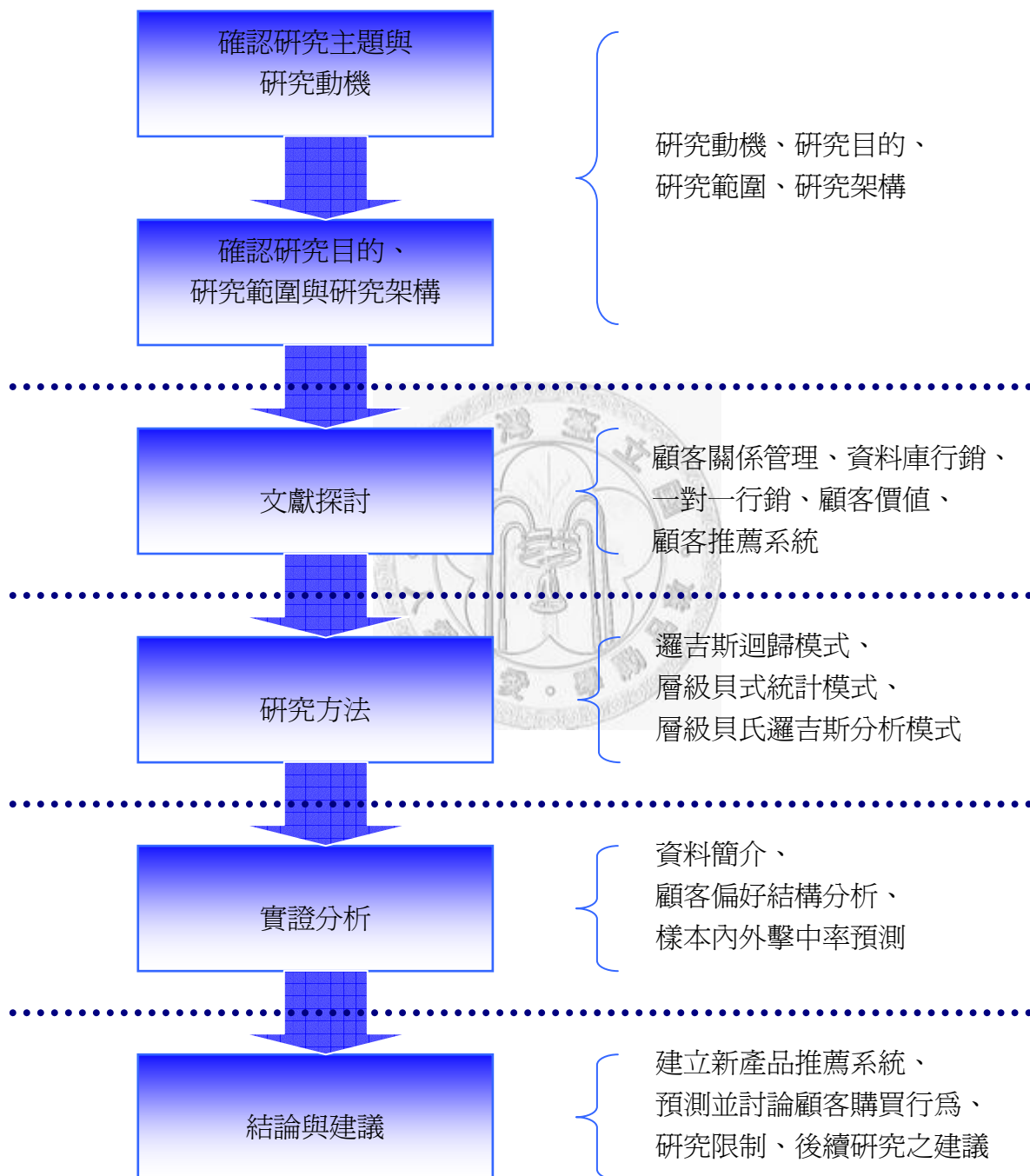


圖 1-5-1 研究流程圖

第二章 文獻探討

第一節 顧客關係管理

綜觀全球各大中小型企業，將顧客關係管理納入策略性思考的公司不在少數。舉例來說：AT&T 便曾根據其內部研究而指出，顧客價值是市場佔有率的領先指標，凡是當其顧客價值增加，四個月後其市場佔有率便會提升。也就是當 AT&T 利用了解顧客的滿意度高低後，執行自我修正，而使顧客的價值明顯提高，便願意再繼續與 AT&T 合作，更甚地還自願幫該企業免費廣告，拉攏其他顧客，使 AT&T 得以增加市佔率。

數十年前，已故的管理學先進 Peter Drucker 就曾給與顧客關係管理的概念高度的關注，並於管理實務一書中提及企業最主要的目的在於創造顧客 (Drucker 1954)； Database Marketing Institute 之副總裁 Arthur Hughes，在策略性資料庫行銷一書中也延續這個想法，如此寫道：要使組織獲利的唯一方式就是首先讓你的顧客獲得好處 (Hughes 2000)。顧客是企業獲利的來源，而有價值的顧客則是企業長期獲利的根本，於是許多學者都曾提出利用顧客關係管理 (Customer Relationship Management, CRM) 的方式，來針對有潛力的顧客，長期維持與其良好的關係，並透過鼓勵其再購行為，創造更多交叉行銷的機會，藉此以獲取顧客的終身價值 (Customer Lifetime Value)，又讓顧客發揮口碑效果引薦更多的顧客，使企業得以正向循環其顧客關係管理的程序 (Morris 1994)。引述管理顧客價值一文之研究結果，證實企業若能讓顧客感受到較高的價值與品質，則可獲得比其他較劣等的企業多出 15% 的投資報酬率 (Gale 1994)。不可否認的，營運在現今這個充滿競爭與資訊透明的世代，企業確實需要將最重要的資源聚焦在顧客身上，才有機會勝出。

豐富的學術研究與實務界成功應用的範例，是過去眾多學者專家的努力所成就，本研究此節先藉由整理文獻，來回顧並說明顧客關係管理的內涵。

一、 何謂顧客關係管理

顧客關係管理的定義從 1997 年首次由 Gartner Group 集團提出至今，已累積有不少學者針對其定義給予小幅度更新修正，整理各方說法如下：

表 2-1-1 顧客關係管理定義

年份	學者	定義
1997	Glazer	顧客關係管理為資訊密集的策略，連結了資訊科技與行銷策略，目的在於建立長久的關係與獲利力。
1999	Berson, Smith & Thearling	顧客關係管理是許多技術與觀念的集合與發展，滿足大多數對企業有價值的顧客之需求。
1999	Spengler	顧客關係管理是結合電腦軟體，並運用資訊科技整合行銷企劃與客戶服務，提供給顧客客製化服務，以提高顧客的忠誠度及企業營運效益的一系列程序。
1999	Davids	顧客關係管理就是關係管理、終身價值行銷、忠誠度行銷、一對一行銷，且可使企業創造長期與顧客間相互獲利的關係，並發展出忠誠度及獲得利潤。
1999	Gordon	顧客關係管理是結合數種資訊科技的綜合應用，目的是要保留對企業有貢獻的客戶，且基於顧客的需求是動態的改變著，則顧客關係管理也必須重複不間斷的改善，於是企業必須從顧客生命週期中了解客戶的行為、偏好，進而提供其所需之商品或勞務。
1999	Kalakota & Robinson	顧客關係管理乃一整合銷售、行銷、售後服務等活動的系統。
1999	Mckinsey & Company. Inc.	顧客關係管理室尋找對企業最有價值的顧客，以微型區隔（Micro-Segmentation）的概念，界定出不同價值的顧客群，且企業以不同的產品、不同的通路滿足不同區隔顧客的個別需求，並在關鍵時刻也持續的與不同層次的顧客溝通，強化顧客價值的貢獻。
2001	Stone&Woodcock	顧客關係管理是公司用以管理顧客關係的一種方法、技術、與電子商務能力。
2001	Sheth&Parvatiyer	顧客關係管理是贏取、維持特定顧客，並與之建立夥伴關係以創造雙方價值的一種過程與全面性策略。

資料來源：邵功新（2003），本研究整理

綜合多位學者的說法，可知顧客關係管理的中心思想不外乎要能了解真正對企業有價值的顧客為何？享有何種消費的偏好？利用資訊科技的技術來分析出解答後，針對這些目標族群給予更多的滿足，包含銷售點發生之前與顧客保持良好關係、開發能滿足顧客潛在需求的商品；銷售時點發生時的通路選擇及優值服務內容等；銷售發生之後的滿意度追蹤、售後服務等，提供超乎目標族群需求的附加價值產品或服務，每項行銷相關活動的內容都以建立顧客的忠誠度為不二目標，期許將顧客的終身價值鎖住（Lock-in）。

實務界敏銳的市場經驗早已看準了顧客關係管理的發展性，基於維持一名舊顧客比開發一位新顧客便宜得多，許多企業紛紛藉由管理顧客存留率以提高其忠誠度，也不乏投入大量硬體設備，建置全天候客服中心、顧客資料庫系統等業者，在在舉動都是因應學理上已證明為可行的目標：藉由延長與顧客之間的關係，來讓利潤持續提升；只要企業能夠辨識其最忠誠的顧客為何，並且明白該如何保留住該顧客，避免轉移至競爭者，則企業便可明顯獲得更高的利潤，儘管只有提高5%的顧客存留率，也可以增加95%的顧客貢獻度之淨現值（Reichheld 1990）。

二、顧客關係管理的程序

在上述對顧客關係管理之定義作一詳細的整理後，不難發現顧客關係管理的本質正是一套系統化的流程，由企業內部建構標準化的顧客關係管理程序，將管理顧客的時間範疇擴大，自顧客第一次與該企業的接觸便開始，直到確認顧客對企業的價值高低排序後，而進行不同程度的關係管理。此套顧客關係管理的流程可以由 Kalakota & Robinson 於 1999 年時所提出的三個階段：獲得、提升、保留來檢視，如下圖所示：



資料來源：Kalakota & Robinson (1999)

圖 2-1-1 顧客關係管理程序

圖中說明顧客關係管理中主要包含三個顧客生命週期的階段：

1. 獲得 (Acquisition)

首先藉由具備便利性與創新性的產品與服務作為促銷的內容，以獲取可能購買的顧客，同時利用較優越的產品與服務來給予顧客更多的附加價值。

2. 提升 (Enhancement)

有效的運用交叉銷售 (cross-selling) 與向上促銷 (up-selling) 等方式，強化企業與顧客之間的關係，創造更多的利潤，增加該顧客的貢獻度；又對顧客而言，其消費便利性隨著搜尋成本等減少而提高，亦是企業帶給顧客的價值提升。

3. 保留 (Retention)

企業透過關係的建立，有效的察覺顧客的需求並主動加以滿足，保留住較具有獲利性的顧客，進而提供高價值的顧客其所偏好的產品，創造服務顧客需求，而非市場需求的適當性。

三、 企業顧客關係管理的應用

企業若能適當運作顧客關係管理，可為企業帶來眾多可見與不可見的好處如下：

1. 幫助企業尋找獨特的市場區隔

在與企業有較良好關係的顧客身上，找出企業獨特的目標市場，繼續滲透現有的客戶群，並從該區隔中開發潛在的顧客，可以減少不必要的資源浪費在價值性不高的區隔之顧客上。

2. 增加企業收益與獲利率

企業可以針對較有價值的顧客進行交叉銷售或向上銷售等，提高收益。

3. 協助決策者擬定最佳策略

藉由與顧客的互動和了解來作為行銷人員決策時的參考。

4. 增加小型目標行銷的次數

基於熟悉各顧客之偏好、生活型態等，而能將顧客更精確的劃分在眾多細小的市場區隔中，企業便可針對不同的市場區隔行銷不一樣的內容，有效提高銷售成功的機率。

5. 提高顧客回應率並增加知識

顧客提供回應的可能性，會隨著與企業之間的關係越好而越高，如此一來，企業便可獲得更多寶貴的顧客消費行為等資訊，而利用企業知識管理的機制，將資料儲存後轉化為企業或行銷人員的知識，以便日後預測消費者態度，或了解新產品的市場大小等時，可派上用場。

而企業該如何有效率的管理與顧客的關係呢？曾任知名管顧公司 Ernst & Young 的資深副總裁 Paul Battista 表示，有五大基本的構面要考慮 (Battista & Verhun 2000)

(一) 顧客關係策略 (Customer Relationship Strategy)

作為策略的主臬，企業應確認其所要瞄準顧客，能為所有在支援顧客關係

管理時所投入的人員、流程、知識與技術等帶來最大的報酬。而欲發展出最佳的關係策略便牽涉到對顧客的分析、研究及預想，必須要擁有正確的顧客組合，傳遞最豐富的價值，了解企業在市場上最應扮演的角色，並決定追求顧客關係發展時，應如何平衡可獲得的獎賞以及所須承擔的風險。

(二) 顧客接觸管理 (Customer Access Management)

建立適當的管道與顧客接觸，確保顧客能夠藉由企業的系統、流程來接觸到其產品、服務與資訊，包含在有人員協助的通路（如實體商店），與沒有人員協助的通路（如網路）。

(三) 顧客程序管理 (Customer Process Management)

整合前端用來攫取及管理必要資訊的系統，與後端系統以提供高價值的顧客服務，重點在於利用 ERP 或工作流等設備，極大化與顧客相關的資訊流。

(四) 顧客體驗管理 (Customer Experience Management)

將服務中心轉化為策略性的顧客關係管理資產，將客服中心調整為擁有符合顧客需求與期望的特色，從中獲得低成本但高價值的顧客關係。

(五) 顧客知識管理 (Customer Knowledge Management)

企業要能運用已知的訊息在顧客的取得、發展與留存上，將關於正確的顧客正確的資訊傳遞給正確的員工，持續循環，將可獲得大幅度的成長。以往在企業中負責管理顧客關係的角色者，多為訓練有素的銷售人員，如仲介商、代理商等職，一但企業能利用現代化的科技，便可將顧客關係管理這門藝術內化成企業獨有的技巧，後提升整體銷售人員的能耐，並擴大關係管理的規模與可及性。

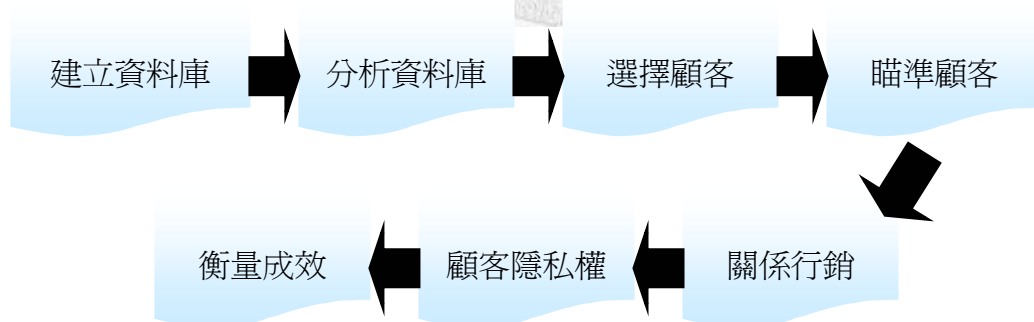
曾有研究指出在 2003 年當年度執行顧客關係管理的成效只有 30% 的比例是成功的。可見得企業要想藉由此技術獲得良好的投資報酬率，與更高的顧客存留率，不但是需要擁有正確的顧客關係管理概念，還需要仔細檢視其所擬定的配套措

施，包含企業的策略是否具有一貫性，以維持顧客關係為首要目標；又在企業本身的營運系統外，增設顧客關係的流程是否有衝突；以及企業所購置的資料庫與顧客關係管理系統使否合宜等，組織都必須多方權衡考量，則顧客關係管理的效益才能全面發揮（Schuster 2005）。

四、 資料庫行銷

Winer（2001）曾提出企業若想有效的管理與顧客之間的關係，便可透過下列模式之步驟逐一實現，如下圖所示：

1. 建立顧客行為的資料庫
2. 分析該資料庫內容
3. 根據該資料分析的結果，選定目標顧客
4. 決定瞄準目標顧客的工具與方式
5. 擬定與目標顧客建立關係的策略
6. 注意顧客的隱私權
7. 執行衡量顧客關係管理方案的準則



資料來源：Russell Winer (2001)

圖 2-1-2 顧客關係管理模式

企業若想有效管理與顧客間的關係，首要任務便是利用資料庫行銷。也就是以資訊科技為基礎，收集有關現有顧客及潛在顧客的資料，真實紀錄每一名消費

者交易與未交易的狀況，並了解其人口統計資料、生活型態、偏好與購買行為等內容，而後再由行銷人員根據統計模式，結合對消費者行為的預測，從中挑選出最能讓企業獲利的顧客，便針對該顧客群給予適當的刺激，達成企業的目標。有研究指出線上推薦系統可間接透過制定更高的價格來影響銷售表現，推薦系統比起顧客反饋的機制（如顧客評價、顧客回應文章數）對銷售表現的影響更有成效（Garfinkel, Gopal, Pathak, Venkatesan & Yin 2007）。不論是成功鼓勵購買，或是售後追蹤以建立忠誠度等，資料庫行銷都是項顧客關係管理的利器

唯近年也有專家學者提出警告，部分的詐欺犯罪行為是來自盜竊企業的資料庫，導致企業與顧客之權益雙雙受損。研究結果也曾指出，若是因為顧客個人資料的揭露，而使顧客察覺到可能出現潛在傷害時，則其對資料持有者的信任，及對該資訊流向之控制權就會覺得大打折扣（Horne 2002）；唯獨執行的前提是企業要能充分知曉顧客對於資訊提供的敏感度，在部分範圍內是可以被接受，一但超出上下界恐怕就會使顧客出現不安的心理。

然而對於資料庫行銷而言，本身就是一個動態資料庫系統的管理，資料庫內包含了有關顧客、詢問者及潛在顧客的廣泛性、即時性、和相關性資料。此技術就是應用上述資料找出最有可能對產品產生回應的顧客和潛在顧客，和顧客發展高品質且長期性的關係。於是整體理論基礎就是來自某種程度的資訊分享，企業藉由與顧客進行偏好的了解、生活資訊的分享，以各式行銷手段來建立彼此的信任，儘管如此但當面對上述所言，企業在導入資料庫行銷的技術後，必須能確實保證資料絕對安全，不論是硬體設備的防盜功能，或是組織內員工的職業道德，都需要相當標準的控管，以避免顧客在與企業接觸時，體驗到的是負向情緒的消費經驗，反而會招致不良的影響。

此外，學者也提出太過於強調要與顧客維持長久的良好關係，當顧客的朋友，在某種程度而言會造成負面的影響（Grayson 2007）。因此友情是根基在具有商業的交易行為之上所發展出來的，所以在該情誼關係的背後，實隱含著工具意義

(Instrumentality concern)，是企業希望能從該段關係中攫取顧客的價值之目的，雖說關係行銷有助於提升企業的利益，但最好與顧客保持適當的平衡。

以下便列舉數名學者對於資料庫行銷之定義：

(一) Shaw & Stone (1990)

資料庫行銷是對行銷溝通的一種互動式的處理方式，它運用某些能夠接觸到特定客戶的傳播媒體，藉由將顧客、潛在購買者的資料以及有關溝通、商業接觸的資訊加以紀錄，並存放在資料庫中，使行銷人員能與目標顧客保持緊密聯繫。

(二) Shani & Chalasani (1993)

資料庫行銷涉及對於過去、現在以及潛在顧客資訊的蒐集。這些資訊包括人口統計資料、顧客的好惡、品味、購買行為和生活型態。

(三) Jackson & Wang (1994)

資料庫行銷是一種以顧客為基礎、資訊密集、長期導向的行銷方式，必須蒐集有關現有與潛在顧客的資料，而這些資料會影響企業決策。

(四) Fletcher, Wright & Desai (1996)

資料庫行銷為在電子化的資料庫中，除了客戶反應的資料外，也儲存了其他顧客資料，以作為長期顧客忠誠的計畫基礎，並促進未來接觸與行銷規劃。

(五) Frederick Newell (1997)

資料庫行銷是一套中央資料庫系統，用來儲存有關企業與顧客所有的資訊，目的不在獲得或是儲存資訊，而是用來規劃個人化的溝通，以創造銷售業績，不但具有整合與業務相關的顧客之資料，並且具備加強顧客終身價值的能力，乃是支持資料庫行銷系統的策略價值，而這種中央資料庫通常稱作資料倉儲 (data warehouse)。

(六) Hughes (2000)

資料庫行銷提供給顧客的並非疲勞轟炸的折扣戰，而是給予注意、認同、友誼以及服務，顧客可以說出其心中的想法，而企業也能藉由變化服務以及產品組合，回應顧客之想法。資料庫行銷的運用使得企業可以鎖定特別的顧客，在適當的時間給適當的客戶提供適當的產品，使企業每花一塊錢在行銷活動上都能增加顧客回應，建立起該企業的顧客基礎以提高企業的獲利能力。

Peppers & Rogers (1995) 認為並非所有的行業都適合採用資料庫行銷，因其所銷售之產品或服務，不見得會讓顧客願意培養長期關係，如下所示即為顧客可能與企業維持長期關係之產業：

1. 線上服務
2. 複雜的產品或服務
3. 大型產品（如汽車、資訊系統等）
4. 奢侈品或專門品（如衣服、香水、化妝品、酒等）
5. 零售服務，特別是顧客欲觸摸、翻閱，或立即消費的產品（如餐廳、酒吧）
6. 需經常購買相同種類的產品（如牙膏、衛生紙等）
7. 可方便遞送到家中的服務（如水管修理等）
8. 顧客本來就重視特定的品牌（如高級洋酒、設計師服飾、名錶等）

一般來說耐久性產業（如銀行）的顧客比較可能產生忠誠度，所以比消費性產業更適合發展資料庫行銷，但若是一般消費品廠商的獲利高（如 Rolex Watch），或短期之內用量大，需經常購買的產品（如紙尿褲）也有發展資料庫行銷的價值。

台灣最早使用資料庫來銷售產品的先鋒應屬出版業和郵購業者，其中又以「讀者文摘」及「台北郵購」最早開始。在美國則可看到資料庫行銷的行業別十分廣泛，從日常用品、電話公司、娛樂業、飲料業都有使用這樣的行銷手法，即使連

一些在大眾媒體運用上受到限制的產業，如香煙、酒類等，也以此種方式來和顧客接觸。現今放眼全球，幾乎所有知名企業都早已應用資料庫行銷的技術，更甚的將之視為引導企業立於不敗之地的重點項目，在經營與顧客的關係管理上扮演關鍵的角色，也因此本研究將根基在此，以圖在擬妥決定性的行銷策略之前，先勾勒出正確的方向。



第二節 顧客推薦系統

一、 何謂顧客推薦系統

Ansari, Essegaier & Kohli (2000) 曾提出在網際網路電子商務的實務界，許多人都嘗試將推薦系統 (Internet Recommendation System) 視為行銷的工具之一，透過系統把其所販售的商品介紹給消費者，像是 Yahoo!、Amazon.com、Movie Critic 等，皆同時採用或擇一使用內容模式 (content filtering method) 與合作模式 (collaborative filtering method) 的推薦系統。

建構在網路之上的推薦系統是企業對顧客進行一對一個人化行銷，兼具低成本與資訊便利性的管道，良好的推薦系統可以使顧客與企業雙贏：前者能夠節省大量搜尋成本，同時可自動獲得即時更新的訊息，以及享有再購的交易便利性；後者得以藉此進行交叉行銷，獲取更多顧客的佔有率，並且能從愈多次的交易，愈發了解顧客真實偏好，推薦更多符合其需求的產品，以提高顧客的忠誠度。且基於電子商務其所銷售之產品，不如實體商店一般，可讓消費者直接觸碰，所以會產生更多的認知風險 (Kim & Kim 2001)，因此一但消費者與特定的網路商店曾有過愉快的交易行為，其移轉成本便能輕易高築。

近來學者們也競相發展各式推薦系統，以期捕捉顧客最真實的偏好，提供企業作為行銷策略的參考依據，以下介紹兩套最基本也最常被應用的推薦系統，並整理各優缺點與應用如下表 2-2-1 所示：

(一) 內容模式的推薦系統

所謂內容模式的推薦系統，即為把顧客針對某一類別的產品，對其曾經購買過的產品之各種屬性與特徵，詳加紀錄後找出影響購買的數個關鍵點，並且利用其過去真實消費的資訊，作為預測其未來是否購買的基礎，也就是假設每項產品的屬性水準是唯一會決定顧客偏好的內容。

(二) 合作模式的推薦系統

所謂合作模式的推薦系統，即為把某一類型的顧客所喜愛的產品，推薦給相同

類型的顧客，消費者透過這樣的系統，將自己對某一特定產品的評價與意見，分享給其他顧客，使其得以較容易作出決定，並且較有可能挑選到其所喜好的產品，這樣的機制假設顧客的偏好是一加權過的線性組合，涵蓋了其他顧客的偏好在其中，正好允許消費者主動進行口碑效果的溝通與傳遞。

表 2-2-1 常見推薦系統模式

	內容模式	合作模式
意涵	系統事先收集顧客的資訊，分析並建立其偏好的檔案，從中找出符合其屬性與水準的產品予以推薦	系統收集眾多顧客的資訊，比對顧客間的相似度，每個顧客都有一組計算得出的同好，便從中推薦各同好喜愛的產品
優點	可有效推薦新產品	不受過去購買的產品所限制，可以推薦其他消費者評價不錯的產品，可以是對顧客而言全新的事物
缺點	<ol style="list-style-type: none"> 1. 只考慮產品屬性，致使當產品的各項特徵皆相同時，便無法判斷各產品的推薦順序 2. 推薦的範圍僅限定在與顧客曾購買過的產品相關性較高的產品 3. 無法推薦適當的產品給新顧客 	<ol style="list-style-type: none"> 1. 當其他顧客的評價相對系統內產品的資訊而言過少，如較為冷門的產品通常不太容易得到其他顧客的評價，便無法有效推薦 2. 當新產品上市時，評價人數無法即時累積，推薦系統便失效 3. 無法推薦適當的產品給新顧客
實際應用	Yahoo!、Alta Vista、Amazon.com、Personal Logic、Active Research	Net Perceptions、Likeminds、Firefly、Amazon.com

資料來源：本研究整理

檢視兩種推薦系統的優缺點後，發現在行銷其他產品替代方案時，應可藉用統計的技術整合顧客的偏好模式，且將可能有用的資訊一併納入分析模型以涵蓋顧客的異質性，便可有效用來預測顧客對各項產品的購買機率，而進行客製化的推薦。過去學者 Ansari, Essegaiier & Kohli (2000) 的做法是收集了顧客個人顯露的偏好、其他顧客的偏好、專家的評估、各項產品的屬性特徵，以及每位顧客的個人特質等，針對電影建立推薦系統的模型，並驗證其準確性，而後發現導入層級

貝氏統計模式，加入了顧客異質性後，整體模式所得出的預測準確度較單純以合作模式的推薦系統還好，爾後網路推薦系統又邁入了新的紀元。

二、 Amazon.com 個案介紹

亞馬遜網路書店著實為一標準的推薦系統應用實例，其網站內所使用到推薦系統機制成功達成了客製化的目標，若一名職業為軟體工程師的顧客進入該網站，則其頁面顯示即為程式設計等書籍為主要標題；若是一名剛當上媽媽的顧客連結進入網站內，頁面即會顯示嬰兒玩具與育兒叢書，此種瞬間能夠巨幅改變以減少顧客時間成本的網路商店，依賴的的是在該網頁背後所隱藏的邏輯，一套會根據顧客不同特質、興趣，及對產品的偏好而推薦不同內容的系統。

就亞馬遜網路書店的推薦系統而言，應是相當成功，其內部資深經理群 Linden、Smith 與 York (2003) 表示，該書店的網路推薦系統明顯有根據特定顧客，顯示不同網站頁面，以及寄發內容皆為個人化的 email，也因此使其廣告點擊率 (click-through) 與轉換率 (conversion rate) 都明顯超越針對一般非顧客的群眾，僅只在正常入口網站頁面提供銷售排行榜與橫幅廣告，所能換得的點擊數與轉換率。而探索其成功原因，可發現在 Amazon.com 光是書籍就擁有超過 500 萬本的資料庫中，能夠讓接近 3000 萬名顧客獲得推薦系統的推薦，且在顧客連上網頁約莫 0.1 秒的等待時間後，便可立即顯示出更新的推薦品，這都必須歸功於亞馬遜書店自行研發的一套功能極強大的推薦系統，稱為項目對項目合作模式 (Item-to-Item Collaborative Filtering)。

此套項目對項目合作模式的推薦系統是結合內容模式與合作模式而發展出來的，其應用離線運算的方式，挑選出所有曾購買或評價相同的產品之顧客後，進一步將挑選出的每一位顧客，所有曾經購買過或評價的項目列出，再進行產品組合的矩陣比對，利用比對結果，把全部有高度相關性的項目搜尋找到，再將相關程度最高或最受歡迎的項目呈現在顧客的頁面，表示為「和您買相同產品的顧客

也買這些 (Customers who bought items in your Shopping Cart also bought)」。乍看之下，此系統似乎相當耗時，但因為亞馬遜網路書店中每位顧客的消費次數都不多，尤其是當購買或評價的項目為暢銷品時，運算的時間更是大幅縮短。此外，為使其推薦系統的正面循環，避免有商品獲得過少評價而不易成功推薦，Amazon.com 也設置了提供顧客一個連結，可以連到推薦書籍的頁面，稱為「推薦給您 (Your Recommendation)」除了能讓顧客了解為何產品會被推薦給自己，並賦予顧客評價該書籍的權利，如此一來系統也可同時得知顧客對該項目的偏好，不論最後購買行為是否有發生，都能增加對顧客更多的了解。

一個大型的網路零售業者如亞馬遜書店，藉由網路推薦系統也能創造顧客個人化的消費經驗，有效的結合內容模式與合作模式的推薦系統，發展出獨特適用的第三種推薦系統，虛擬出書店銷售人員，對每一名顧客實行一對一行銷，讓顧客與亞馬遜書店的關係更加密不可分。



第三節 隨機品牌選擇模式

早在 1950 年時，行銷人員就已經在觀察顧客對品牌的選擇模式，近來隨著資料庫行銷的興起，討論隨機品牌選擇模式 (Stochastic Brand Choice Models, SBC) 的研究也如雨後春筍一般蓬勃發展，本節將先介紹適用於研究品牌選擇的各種計量模式，並探討 Blattberg and Neslin (1990) 在「銷售促銷—概念，方法與策略」一書中提出之批評，佐以其他學者文獻的回顧，並同時為本研究下一章的研究方法選定適當的統計分析模式。

一、 隨機品牌選擇模式的三大特色

(一) 流程順序性 (Order of the Process)

即指前一次購買是否會影響後一次的購買，若前期購買不影響後期則為零階程序 (zero-order process)；反之，若前期購買會影響到後期，則為非零階程序 (nonzero-order process)，而非零階程序中的階層數字則根據各模型之假設有所不同，可以是 1 也可以大於 1。

(二) 購買機率相異性 (Heterogeneity)

即指每一位顧客購買特定品牌的機率是否相同，以及其消費傾向是否一致。早期 SBC 模型多假設顧客購買機率是一樣的，然而現代行銷觀念已明顯進步為微型區隔，相信每一位顧客都是獨立的個體，其購買機率也應不盡相同，於是 SBC 模型也都調整為相信顧客購買機率有異質性存在，顧客的忠誠度以及購買機率都有明顯落差，因此大多數的模型都假設顧客的偏好或其購買頻率，服從某一特定的分配 (如貝塔分配)。

(三) 購買機率恆常性 (Stationarity)

即指對一特定品牌而言，其購買機率是隨著時間變化依然固定不變的。這樣的假設在行銷實務界相當不實際，特別是一但促銷策略執行下，特定品牌的購買機率每週可能都會改變，因此許多學者也都針對此特性，提出統計方法的補救。

二、 各隨機品牌選擇模型介紹

(一) 伯努力模式(Bernoulli Model)

此模式是基於下列假設計算出消費者會購買某種品牌的機率：

1. 前一次購買何種品牌，並不會影響下一次的購買何種品牌的決策，即假設前後期之決策獨立，是為上述所提及之零階程序。
2. 消費者購買某種品牌之機率並不會隨著時間變動而改變，符合上述之購買機率恆常性。
3. 對所有消費者而言其購買機率皆相同，即假設消費者為同質。

在這樣嚴格的假設之下，使得此模式存在著下列問題：

1. 由於假設前一次購買與後一次之決策彼此獨立不相關，則此模式便無法衡量促銷的延續效果 (Carryover Effect)。
2. 假設消費者購買某種品牌之機率不會隨著時間變動而改變，也就意味著廠商的促銷活動，或是顧客每次購買期間的間隔長短，不會影響消費者購買某種品牌的機率。

(二) 馬可夫模式 (Markov Model)

馬可夫模式捨棄前後期購買之決策互相獨立的假設，認為前期的購買行為會影響下一次的購買，但是影響後期購買行為的前期期數僅有一期，即這一次購買行為是上一次購買行為的函數，亦即為上述非零階程序之一階程序的典型。馬可夫模式對於前期的購買行為會影響後期之購買行為的假設，雖然有助於研究促銷策略的延續效果，但是強調會改變後期購買行為的卻僅有前一期的購買，馬可夫模式也說不上具有彈性，儘管有此一大限制，仍可被運用於探討市場佔有率的變化，唯獨相較於伯努力模式，馬可夫模式在計算應用上較為複雜。

(三) 線性學習模式 (Linear Learning Model)

線性學習模式試圖建構一套模式，可看出消費者過去的購買行為對未來購買行為的影響效果。此模式的基本假設就是消費者過去的購買記錄會影響現在的選擇，且離現在愈近的影響效果愈大，但此模式的限制是當市場上只

存在兩種不同品牌時才適用。

(四) 議價價值模式 (Bargain Value Model)

連續購買某一品牌，將導致下一次購買時，轉換購買其他品牌的機率提高。即間隔上一次購買該品牌的時間愈久，而下次會購買此品牌的機會就愈高，這個想法與線性學習模式正好相反。

(五) 邏吉斯模式 (Logit Model)

此模式首次被應用於行銷學上，是出自 Green 學者在 1970 年代後期之手，而後就被大量導入探討品牌選擇模型。由於此模式的根本是從隨機效用理論 (Random Utility Theory) 而來，許多學者在研究消費者選擇品牌時，認為消費者會隨機地給予所考慮的品牌一特定的效用值，並且從所有可能選擇方案中挑出使其產生最大效用的品牌。選擇某一品牌所產生的效用是確定效用值 (Deterministic Utility) 與隨機效用值 (Random Utility) 之和，若假設隨機效用值服從雙指數分配 (Double Exponential Distribution)，則選擇某一品牌的機率模式即為 Logit Model。

確定效用值為可觀察得到的變數之線性函數，因此這樣的假設與理性消費者的購買行為相當吻合；再者，Guadagni and Little (1983) 實證之結論為模式的預測能力相當切合消費者實際購買行為，此結論同時為學術界和實務界提供了一個重要的資訊；最後，此模式一再受到學者的引用是源自於此其基於行為的效用理論，並可以放入行銷變數組合以解釋品牌選擇的狀態，亦可用於衡量新產品各屬性之相對重要性，同時兼具較簡單的計算過程就可得到深遠意涵之效。

三、 品牌選擇模型之實證文獻

回顧過去眾多先進將 SBC 的模型應用在行銷相關議題，本研究整理較具有代表性之內容如下表所示：

表 2-3-1 品牌選擇模型之文獻整理

學者及年份	實證結果
Guadagni & Little (1983)	利用多項式邏吉斯模式 (Multinomial Logit Model) 探討行銷變數對消費者選擇品牌的影響效果，消費者選擇某一個品牌的機率是所有可供選擇品牌的屬性之函數。根據此模式估計消費者購買各品牌之比率，再比對實際購買情況，發現此模式能預測出消費者的購買機率之變化趨勢 (Trend) 與型態 (Pattern)。
Krishnamurthi & Raj (1988)	將選擇品牌與購買數量的決策放入同一個模式中，以多項式邏吉斯模式探討價格在選擇品牌與購買數量決策上所扮演的角色。
Gupta (1988)	將消費者購買的時間點、品牌選擇與購買數量同時考慮，探討促銷活動如何影響消費者的購買時間、品牌與數量之決策。利用負二項分配模式 (Negative Binomial Distribution Model)，探討消費者的購買時機決策；多項式邏吉斯模式，探討消費者品牌選擇的決策；累積邏吉斯模式 (Cumulative Logit Model)，探討消費者購買數量的決策。
Jones and Landwehr (1988)	對邏吉斯模式提出一個新的參數估計之方法，即將異質性的問題納入邏吉斯模式加以考慮，實證的結果發現，此估計參數的方法改善了模式的配適度與解釋力。
Kamakura and Russell (1989)	以多項式邏吉斯模式為基礎，利用消費者過去的購買記錄，探討市場區隔與價格敏感度的議題。所提出之模式能偵測市場結構及提供實務界更多資訊。

資料來源：陳成業 (2002)，本研究整理

經過上述的探討，得知許多關於品牌選擇之研究多是以多項式邏吉斯模式或以此模式為基礎，建立修正的模式作為實證的統計工具，又其結果多符合實際情況，即模式的預測能力相當好，本研究接下來也會以此邏吉斯模式作為主要的研究工具，至於邏吉斯模式又可依據其應變數的結構劃分為多項式邏吉斯模式（Multinomial Logit Model），以及二項式邏吉斯模式（Binary Logit Model），鑑於本研究主張是探討顧客對各產品的購買機率，故將應變數二分為有購買與沒有購買，因此會傾向利用二項式邏吉斯模式作為主要的推論模式，而關於此模式的推導以及如何將顧客異質性加入模型上，將於下一章詳述。



第三章 研究方法

第一節 研究架構

本研究將分成下列各步驟進行，首先為實驗設計、隨後進行資料分析，並將求得之模式用以建立線上新產品的推薦系統。整體流程如圖 3-1-1 所示：

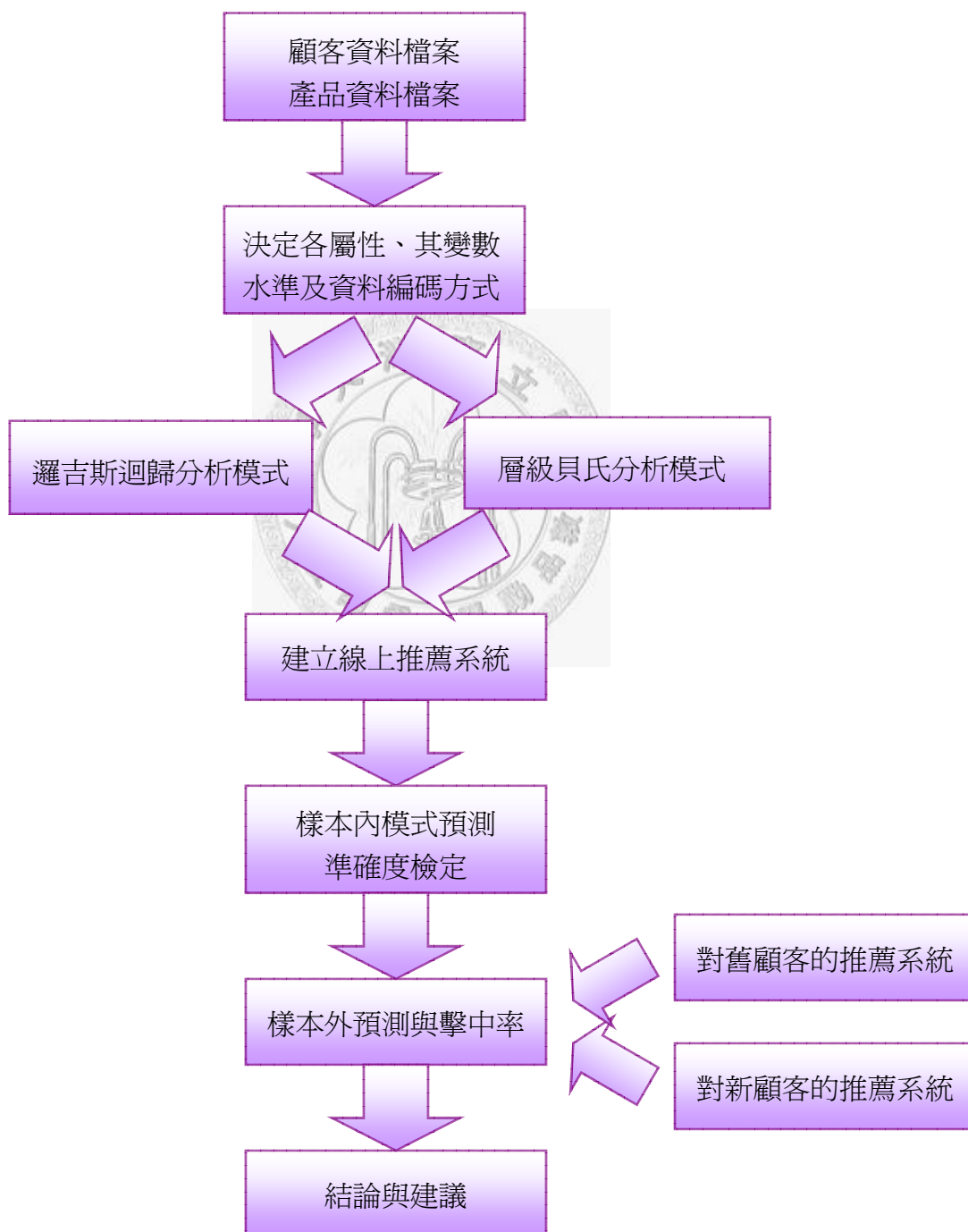


圖 3-1-1 研究架構

第二節 研究設計

一、 決定各屬性及其水準

本研究產品為亞馬遜線上網路書店於 2004 年 1 月 1 日起整年度的書籍交易資料，為了要建立整體輪廓的樣本，在此階段前便已透過專家學者的建議，不斷篩選過濾，最後經由指導教授的修正與整合，而得出最具有代表性的產品屬性，其中包含了書籍標示的價格、書籍的裝訂方式、書籍的頁數、顧客評等、顧客給予回應的文章數目、書籍的銷售排名，以及該書籍所屬的種類。如下為針對各屬性與水準作一定義：

(一) 書籍標示價格

此屬性水準設定為高、中、低三種，將高價格之書籍定義為高於 20 美元的書，中價格之書籍則為 20 至 10 美元之間，低於 10 美元的書籍皆設定為低價。

(二) 書籍的裝訂方式

依次將兩種不同的裝訂方式——紙本與精裝本——定義為此變數的水準。

(三) 書籍頁數

將書籍頁數依照厚度定義為多、中等、少等三種。分別為頁數超過 500 頁者為頁數多；介於 300 頁到 500 頁者為頁數中等；低於 300 頁的書籍頁數則視為頁數少。

(四) 顧客評價

在亞馬遜的網站上所顯示的顧客評價是以顆星數來計算，每本書最高有 5 顆星等，最低則是 0 顆，並且允許給予半顆星。為求資料處理的便利，本研究將顧客的評價乘上 2 倍，每當一書籍獲得 3 顆星的評價則在資料上紀錄為 6 分，3.5 顆星的評價則紀錄為 7 分。而在屬性水準的選擇部分則依照評價分數 0 至 3 分者為低，4-7 分者為中等，8-10 分者為高。

(五) 顧客回應的文章數

將顧客回應的文章數分為多、中、少三個水準，凡獲得超過 50 篇的回應文章

視為多，介於 50 到 10 篇者為中，少於 10 篇者則為少。

(六) 銷售排名

利用亞馬遜網路書店所提供超過數百萬本書籍的資訊作為主要考量基礎，將各書籍的銷售排名水準分別前、中、後三個。當該書籍陳列在網頁上的銷售排名是在 10 萬名之內則將其歸類在排名前面的水準中；介在 10 萬至 50 萬者便將其歸類在排名中間的水準內；而若是銷售排名超過 50 萬者則會將其視為是排名後面的書籍，並歸類之。

(七) 書籍種類

本研究參照亞馬遜的網路書店對各書籍歸類的方式，同時整合過去學者研究與書籍有關的主題時所使用的分類，及台灣主要的網路書店歸類各書籍的方式，決定出本研究所採用的書籍種類水準分別為商業理財、文學小說、人文藝術、普通科學、教科書、心靈養生、生活風格、及親子共享等 8 類。

在該資料庫所提供之交易紀錄中，除了有商品的各項屬性，同時也包含了每位顧客的人口統計資料，如每位顧客的編號、教育程度、家庭平均年所得、年齡、家庭狀況、所在地區、目前有無未成年的小孩等資訊。整理如下頁表 3-2-1 所示：

表 3-2-1 商品屬性與人口變數內容

	屬性	水準	各屬性水準的內容
商品變數	價格	3	低價位 (10 美元以下) 中等價位 (介於 10-20 美元) 高價位 (超過 20 美元)
	裝訂方式	3	紙本 (paperback) 精裝本 (hardcover) 其他 (large print、school & library binding...etc)
	書籍頁數	3	薄 (少於 300 頁) 中等 (介於 301 到 500 頁之間) 厚 (超過 500 頁)
	顧客評價	3	不好 (0-3 星的評等) 普通 (4-7 星的評等) 良好 (8-10 星的評等)
	顧客回應的文章數	3	少 (少於 10 篇的回應文章) 中等 (有 10-50 篇的回應文章) 多 (超過 50 篇的回應文章)
	銷售排名	3	暢銷 (銷售排名前 10 萬名) 普通 (銷售排名介於前 10 萬至前 50 萬名) 不暢銷 (銷售排名在 50 萬名之後)
	書籍種類	8	商業理財、文學小說、人文藝術、普通科學、教科書籍、心靈養生、生活風格、親子共享
人口變數	教育程度	4	高中職以下、大學肄業、大學畢業、研究所以上
	平均收入	4	2 萬 4999 美元以下、2 萬 5000 至 4 萬 9999 美元、5 萬至 7 萬 4999 美元、7 萬 5000 美元以上
	年齡	5	18-29 歲之間、30-39 歲之間、40-49 歲之間、50-59 歲之間、60 歲及以上
	家庭人數	3	2 人或以下同住、3-5 人同住、6 人或以上同住
	所在地區	4	美國東北部、美國中部、美國南部、美國西部
	未成年孩童	2	正在扶養中、家中沒有未成年孩童

二、 決定各變數及其編碼方式

(一) 應變數

此研究所使用的資料庫是由亞馬遜網路書店 2004 年該年整年度的資料中隨機挑選出來的，其中有無購買書籍的結果正是本研究的應變數 Y，所以當 Y=1 時即表示為有購買，Y=0 則無。

(二) 自變數

商品面

本研究主要的商品為書籍，將其分為 7 種屬性與共 26 個屬性水準，編碼方式為書籍標示的價格以 pr1~pr3、書籍的裝訂方式 b1~b3、書籍的頁數 pp1~pp3、顧客評等 rate1~rate3、顧客給予回應的文章數目 rev1~rev3、書籍的銷售排名 sa1~sa3，以及該書籍所屬的種類 cat1~cat8 等變數予以個別編碼。

顧客面

本研究嘗試導入層級貝氏的統計觀念，除了考慮商品的變數之外，顧客的偏好也會是考慮的重點之一，於是我們將在第二層時加入人口統計變量的資料，而第一層的自變數在此則作為第二層的應變數。和商品方面的自變數編碼方式相同，本研究將所有的人口統計變量水準都以虛擬變數的方式個別予以編碼，其中包含教育程度 edu1~edu4、平均所得 inc1~inc4、年齡 age1~age5、家庭狀況 hhs1~hhs3、所在地區 reg1~reg4、有無未成年的孩子、kid1~kid2。

上述所有變數皆為虛擬變數 (Dummy Variable)，茲如下頁表 3-2-2 及 3-2-3 所示：

表 3-2-2 各商品屬性對應之虛擬變數整理

屬性	變數	各屬性水準的內容	編碼	
價格	pr1	低	(pr1,pr2)	(1,0)
	pr2	中等		(0,1)
	pr3	高		(0,0)
裝訂方式	B1	紙本	(b1,b2)	(1,0)
	B2	精裝本		(0,1)
	B3	其他		(0,0)
頁數	pp1	薄	(pp1,pp2)	(1,0)
	pp2	中等		(0,1)
	pp3	厚		(0,0)
顧客評等	rate1	不好	(rate1,rate2)	(1,0)
	rate2	普通		(0,1)
	rate3	良好		(0,0)
顧客回應	rev1	少	(rev1,rev2)	(1,0)
	rev2	中等		(0,1)
	rev3	多		(0,0)
銷售排名	sa1	暢銷	(sa1,sa2)	(1,0)
	sa2	中等		(0,1)
	sa3	不暢銷		(0,0)
書籍類別	cat1	商業理財	(cat1,cat2,cat3,cat4,cat5,cat6,cat7)	(1,0,0,0,0,0,0)
	cat	文學小說		(0,1,0,0,0,0,0)
	cat3	人文藝術		(0,0,1,0,0,0,0)
	cat4	普通科學		(0,0,0,1,0,0,0)
	cat5	教科書		(0,0,0,0,1,0,0)
	cat6	心靈養生		(0,0,0,0,0,1,0)
	cat7	生活風格		(0,0,0,0,0,0,1)
	cat8	親子共享		(0,0,0,0,0,0,0)

表 3-2-3 各人口統計變數對應之虛擬變數表

屬性	變數	各屬性水準的內容	編碼	
教育程度	edu1	高中職以下	(edu1,edu2,edu3)	(1,0,0)
	edu2	大學肄業		(0,1,0)
	edu3	大學畢業		(0,0,1)
	edu4	研究所以上		(0,0,0)
平均收入	inc1	24999 美元以下	(inc1,inc2,inc3)	(1,0,0)
	inc2	25000-49999 美元		(0,1,0)
	inc3	50000-74999 美元		(0,0,1)
	inc4	75000 美元以上		(0,0,0)
年齡	age1	18-29 歲之間	(age1,age2,age3,age4)	(1,0,0,0)
	age2	30-39 歲之間		(0,1,0,0)
	age3	40-49 歲之間		(0,0,1,0)
	age4	50-59 歲之間		(0,0,0,1)
	age5	60 歲及以上		(0,0,0,0)
家庭人口狀況	hhs1	2 人或以下	(hhs1,hhs2)	(1,0)
	hhs2	3-5 人		(0,1)
	hhs3	6 人或以上		(0,0)
所在地區	reg1	美國東北部	(reg1,reg2,reg3)	(1,0,0)
	reg2	美國中部		(0,1,0)
	reg3	美國南部		(0,0,1)
	reg4	美國西部		(0,0,0)
孩童	kid1	有且正在扶養	(kid)	(1)
	kid0	家中沒有未成年孩童		(0)

三、 建立產品組合資料

上述所列的各屬性及其水準若以完全因子設計，則每項商品的產品組合資料將有 $3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 8 = 5832$ 個產品組合，實務上對消費者在購買時而言，不會對每一項商品都進行如此細緻的屬性水準評估，於是為求符合實際狀況，本研究進而利用 SPSS 軟體中的直交設計 (Orthogonal Design) 執行 fractional factorial design 來減少產品特性組合的數目，以便進行下一階段的分析。結果如下表 3-2-4 所示：

表 3-2-4 直交設計所得出之 27 組不同的產品屬性組合

No.	價格	裝訂方式	頁數	顧客評價	顧客回應	銷售排名	書籍種類
1	1	2	2	1	3	2	6
2	2	1	3	2	1	2	3
3	3	1	1	1	3	2	7
4	1	1	1	2	2	1	2
5	1	2	1	3	3	1	3
6	1	1	3	1	2	3	8
7	2	2	1	1	2	3	4
8	3	1	3	3	3	1	4
9	2	2	2	2	2	1	7
10	3	0	1	3	2	2	1
11	3	2	3	1	1	1	5
12	2	1	1	3	1	3	6
13	1	1	2	3	2	2	5
14	3	2	1	2	1	2	8
15	1	0	1	1	1	1	1
16	2	0	3	1	3	2	2
17	1	0	3	3	1	3	7
18	3	0	3	2	2	1	6
19	1	2	3	2	3	3	1
20	2	2	3	3	2	2	1
21	1	0	2	2	1	2	4
22	2	0	1	2	3	3	5
23	3	0	2	1	2	3	3
24	2	1	2	1	1	1	1
25	2	0	2	3	3	1	8
26	3	1	2	2	3	3	1
27	3	2	2	3	1	3	2

不僅只是上述表中所列出的產品組合，本研究也同時考慮從資料庫中擷取出來的交易紀錄。每一次交易的商品屬性水準，若是與表中所提及的任一組合相同，則將該筆商品納入至此正交設計所得出的商品組合之中，並在對應該組合的應變數欄位打上 Y=1。此外，若有重複的情形，則將其屬性的產品組合重複列出；又或者是由交易紀錄內所抽出的商品項目，不符合正交設計得出的產品組合中任何一組，那麼就在這些產品組合下方再列出其特定的屬性與水準組合。以顧客編號 5694553 為例：

表 3-2-5 顧客的購書紀錄及各變數之資訊

購買書籍	1. The add answer : How to help your child now—with questionnaires and family-centered action plans to meet your child' s specific needs. 2. Family first : your step-by-step plan for creating a phenomenal family. 3. Mosaic workshop : A guide to designing & creating mosaics.						
編號	價格	裝訂方式	頁數	顧客評等	顧客回應	銷售排名	書籍種類
1	24.95	1	304	6	26	103381	6,8
2	15	0	304	9	82	20288	6,8
3	17.95	0	144	8	4	786987	3,7

本研究中以虛擬變數將實際購買的資料與考慮組合結合，一併編碼如下頁表 3-2-6 所示，一樣以編號 5694553 的顧客之交易紀錄為例

表 3-2-6 產品屬性編碼轉換表

各產品屬性編碼 (pr1,pr2,b1,b2,pp1,pp2,rate1,rate2,rev1,rev2,sa1,sa2,cat1,cat2,cat3,cat4,cat5,cat6,cat7,y)	
1	The add answer : How to help your child now—with questionnaires and family-centered action plans to meet your child' s specific needs (0 , 1 , 1 , 0 , 1 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 1)
2	Family first : your step-by-step plan for creating a phenomenal family (1 , 0 , 0 , 0 , 0 , 1 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 1)
3	Mosaic workshop : A guide to designing & creating mosaics (0 , 1 , 0 , 1 , 0 , 0 , 0 , 1 , 1 , 0 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 1)
4	(0 , 0 , 0 , 1 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0)
5	(1 , 0 , 0 , 1 , 1 , 0 , 0 , 1 , 0 , 1 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
6	(1 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0)
7	(1 , 0 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
8	(0 , 1 , 0 , 0 , 1 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0)
9	(0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0)
10	(0 , 1 , 0 , 0 , 0 , 1 , 0 , 1 , 0 , 1 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0)
11	(0 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 1 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
12	(0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0)
13	(0 , 1 , 0 , 1 , 1 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0)
14	(1 , 0 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0)
15	(0 , 0 , 0 , 0 , 1 , 0 , 0 , 1 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
16	(0 , 1 , 1 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
17	(1 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0)
18	(0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 1 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0)
19	(1 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
20	(0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 1 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
21	(1 , 0 , 1 , 0 , 0 , 1 , 0 , 1 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0)
22	(0 , 1 , 1 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0)
23	(0 , 0 , 1 , 0 , 0 , 1 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0)
24	(0 , 1 , 0 , 1 , 0 , 1 , 1 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
25	(0 , 0 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
26	(0 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0)
27	(1 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 1 , 0 , 0 , 0 , 0 , 0 , 0 , 0 , 0)

第三節 邏吉斯產品選擇模型

本研究截至目前為止只在建立以書籍為主要商品的產品屬性水準組合，若以亞馬遜網路書店為例，消費者面臨超過 500 萬本以上的書籍，超過上億種的產品組合選擇，個個不盡相同，究竟消費者會選擇哪一種組合的書籍，行銷人員在銷售各書籍時該如何估計消費者購買 K_1 商品組合的機率為何？又 K_2 組合的機率呢？此時就可以應用邏吉斯產品選擇模型了。

本研究的模型其應變數為是否有購買此一商品，若有買則 $Y=1$ ，沒買則 $Y=0$ 。通常就一般線性迴歸模式來分析討論，可能會得出應變數的期望值沒有落在機率值 $(0,1)$ 之間，於是便應用過去生物統計學者常用的 Logit 轉換函數，使應變數的期望值即為所欲預測的購買機率值，介在 $(0,1)$ 之間。

不論是在經濟學或是消費者行為與心理等學科的研究，最初都是以消費者皆為理性的假設出發，也就是說消費者會在有限的條件之下（所得或時間），追求個人行為的效用極大化。在本研究中的前提假設亦如此，也就是說當消費者選擇了某個產品組合時，則此產品組合所能帶來的效用必然是最大的，若是藉由效用函數（Utility Function）來解釋，可知消費者在選擇購買前會基於隨機效用模式

（Random Utility Function），而產生一組效用函數，而此效用函數可劃分為兩部分，由兩種不同的效用加總得到：一為選擇某一產品組合時所確定能得到的效用成分，稱為確定性的效用成分（Deterministic Component of Utility），即由解釋變數與未知參數所構成之非機率性的函數（Nonstochastic Function），例如當商品由平裝本因熱賣而改版推出精裝本時，藉由對可觀測的變數來預測其效用。另一則稱為隨機性的效用成分（Random Component of Utility），即無法觀察到的隨機變數，例如消費者在購買時受到天氣而影響了情緒的變動等。綜合上述，針對不同消費者而有所不同的效用函數，在此對一特定顧客列式如下：

$$U = V + \varepsilon$$

U ：效用函數，

V ：確定性的效用成分 (Deterministic Component of Utility)

ε ：隨機性的效用成分 (Random Component of Utility)

假設 ε 是彼此獨立，且其累積分配函數為「型 I 極值分配」(Type I Extreme

Value Distribution) $F(\varepsilon_i) = \frac{1}{e^{e^{\varepsilon_i}}}$ ，則其機率密度函數為 $f(\varepsilon_i) = \left(\frac{1}{e^{\varepsilon_i}}\right) \left[\frac{1}{e^{e^{\varepsilon_i}}}\right]$ 。

基於消費者會在有限的條件下進行購買，又會盡可能使其效用達到最大，則可用下列式子表示之，

$$U_{ik} > U_{jk} \dots i, j \in C_k, i \neq j$$

U_{ik} ：對顧客 k 而言，購買產品 i 帶來的效用

U_{jk} ：對顧客 k 而言，購買產品 j 帶來的效用

C_k ：顧客 k 所能購買的所有產品組合

若進一步將顧客 k 購買第 i 種產品的機率設為 P_{ik} ，則

$$\begin{aligned} P_{ik} &= \Pr(U_{ik} > U_{jk}) \\ &= \Pr[V_{ik} + \varepsilon_{ik} \geq \max(V_{jk} + \varepsilon_{jk})] \\ &= \Pr[V_{ik} + \varepsilon_{ik} \geq V_k^* + \varepsilon_k^*] \\ &= \Pr[(V_k^* + \varepsilon_k^*) - (V_{ik} + \varepsilon_{jk}) \leq 0] \\ &= \Pr(\varepsilon_k^* \leq V_{ik} - V_k^* + \varepsilon_{jk})，又 \varepsilon_k^* 之累積分配函數為型 I 極值分配， \end{aligned}$$

$$\text{則, } P_{ik} = \frac{1}{1 + e^{(V_k^* - V_{ik})}} = \frac{e^{V_{ik}}}{e^{V_{ik}} + e^{V_k^*}} = \frac{e^{V_{ik}}}{\sum_{j \in C_k} e^{V_{jk}}}$$

假設確定性的效用成分值為可觀測的變數之線性函數，則可表示為

$$V_i = \sum_{f=1}^F \beta_{fi} X_{fi}$$

其中 X_{fi} 是應用於購買產品 i 時的解釋變數，如書籍的價格、裝訂

方式等； β_{fi} 則為解釋變數的參數，因此當顧客 k 在面對所有的產品組合 C 時，會

選擇購買具有產品屬性 X 的產品 i 之機率為 $\Pr_k(i|X, C, \beta) = \frac{e^{X_i\beta}}{\sum_{j \in C} e^{X_j\beta}}$ ，左式則為多項

式邏吉斯迴歸模式(multinomial logistic regression model)，是研究品牌的選擇型為最常使用的統計模式，在本研究中將延伸為對書籍的選擇。若顧客 k 選擇產品 i 的機率 P_{ik} 除以選擇基礎產品 m 的機率 P_{im} 之成敗比取對數 (Logit) 則可轉換成線性模式，如下所示：

$$\ln\left(\frac{P_{ik}}{P_{im}}\right) = y_k = X_i\beta' \quad \dots, i=1, 2, \dots, m-1, \dots, \dots, \dots$$

式中，參數 $\beta' = \beta - \beta_m$ ，即對顧客 k 而言，所有產品屬性 X_i 對產品機率的成敗比(產品 i 相對於基礎產品 m 而言)的影響效果。



第四節 層級貝氏統計模型

過去傳統統計模式常以抽樣的理論為基礎，考慮樣本的資料以得出參數的估計值，並且假設消費者擁有同質的需求；而在使用聯合分析法時，常會需要大量受測體以估計多個參數，卻使得受測體與參數個數的比例越小，其預期得到的期望均方差 (Expected Mean Square Error) 就越大，而降低估計參數的期望值之信效度 (Green & Srinivasan 1978)。為此將導入 Cattin、Gelfand，與 Danes 在 1983 年提出的簡單貝氏迴歸程序，此模式是建立在貝氏定理 (Bayes' Theorem) 之上，其特色在於合併事前的資訊與整體輪廓資料，以產生參數的事後機率分配，估計多元屬性的偏好模式，且所求得的參數貝氏估計值比利用最小平方法 (Ordinary Least Square Regression, OLS) 所得到的預測結果還要好。因此本研究藉由實際的購買紀錄，建立一套個人化的新產品推薦系統，基於偏好結構是建立產品推薦系統的基礎，於是採用層級貝氏的統計模式來估計出每位顧客的個人偏好結構，並得出消費者彼此具有異質性的推薦系統 (Lenk 2001)。

在本研究所使用之資料庫中出現一常見的問題，即指產品及顧客數量眾多，而每位顧客的交易紀錄相對而言較少，使得在估計其個人化的偏好結構上較為困難。但層級貝氏統計模式 (Hierarchical Bayes Methodology) 結合了參數的先驗分配與樣本分配，產生未知參數 (unobserved parameters) 的條件後驗分配後，再藉由系統模擬的方式，如馬可夫鏈蒙地卡羅 (MCMC, Markov Chain Monte Carlo) 透過十萬次的疊代過程產出近似的參數後驗分配，既可有效的產生個體層級 (individual level) 的參數估計，同時也能給出總體層級 (aggregate level) 的結果 (Rossi & Allenby 1998)；此外，層級貝氏統計模式對先驗知識有兩層假設：假設第一層的先驗分配是在描述各產品的異質性；而第二層則是用來描述第一層的先驗分配無法觀察到的異質性，在此則引申為每位顧客的偏好結構，以下則為進一步說明。

一、一般貝氏統計模型

$$P(\text{parameters}|\text{data}) = \frac{P(\text{data}|\text{parameters})P(\text{parameters})}{P(\text{data})}$$

$$\propto P(\text{data}|\text{parameters})P(\text{parameters})$$

1. $P(\text{data}|\text{parameters})$ ：概似函數 (Likelihood Function)

即指樣本的聯合機率密度函數

2. $P(\text{parameters})$ ：先驗分配 (Prior Distribution)

在進行抽樣分配之前，由研究者主觀設定其分配與參數的事前知識

3. $P(\text{parameters}|\text{data})$ ：後驗分配 (Posterior Distribution)

透過先驗知識及概似函數所得到的事後機率分配

二、各層級的模式介紹

(一) 第一層：個體層次 (Individual Level)

個體層次乃為顧客的購買模式 (Within-Subjects Model)，是用來探討顧客的偏好與產品屬性之間的關係，模式如下：

$$Y_i = X_i\beta_i + \varepsilon_i$$

其中

$$Y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix},$$

(n_i×1)

表示顧客 i 對第 n_i 個產品組合的購買行為，當 Y_i=1 即表示該顧客有購買；Y_i=0 則為無購買

$$X_i = \begin{pmatrix} x_{i11} & x_{i12} & \cdots & x_{i1K} \\ x_{i21} & x_{i22} & \cdots & x_{i2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{in_i1} & x_{in_i2} & \cdots & x_{in_iK} \end{pmatrix}$$

(n_i×K)

表示顧客 i 面對的第 n_i 個產品組合，在 K 個屬性水準上的表現，由虛擬變數 (dummy variable) 構成，如書籍價格及其裝訂方式等。

$$\beta_i = \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{iK} \end{pmatrix}$$

(K×1)

表示顧客 i 對 K 個屬性水準的成分效用值 (part-worth)，即為該顧客的偏好結構 (preference structure)

另外， ε_i 即為誤差項，假設其遵循多變量的常態分配 $[\varepsilon_i | \sigma_i] = N_{n_i}(\varepsilon_i | 0, \sigma^2 I_{n_i})$

(二) 第二層：總體層次 (Aggregate Level)

在第一層的模式中， β_i 即為每位顧客不同的偏好結構，在此層次可再進一步針對參數的估計給予修正，以更充分的反應異質性。此一總體層次則是跨顧客的模式 (Between-Subjects Model)，探討每個顧客的偏好模式與其個人特質的關係，模式如下：

$$B = Z\Theta + \Delta$$

其中

$$B_{(N \times K)} = \begin{pmatrix} \beta_{11} & \beta_{12} \cdots & \beta_{1K} \\ \beta_{21} & \beta_{22} \cdots & \beta_{2K} \\ \vdots & \vdots & \vdots \\ \beta_{N1} & \beta_{N2} \cdots & \beta_{NK} \end{pmatrix} = \begin{pmatrix} \beta'_1 \\ \beta'_2 \\ \vdots \\ \beta'_N \end{pmatrix}$$

表示在個體層次得出的偏好結果

$$Z_{(N \times P)} = \begin{pmatrix} z_{11} & z_{12} \cdots & z_{1P} \\ z_{21} & z_{22} \cdots & z_{2P} \\ \vdots & \vdots & \vdots \\ z_{N1} & z_{N2} \cdots & z_{NP} \end{pmatrix} = \begin{pmatrix} z'_1 \\ z'_2 \\ \vdots \\ z'_N \end{pmatrix}$$

表示顧客特質的矩陣，假設共有 P 個變數

$$\Theta_{(P \times K)} = \begin{pmatrix} \theta_{11} & \theta_{12} \cdots & \theta_{1K} \\ \theta_{21} & \theta_{22} \cdots & \theta_{2K} \\ \vdots & \vdots & \vdots \\ \theta_{P1} & \theta_{P2} \cdots & \theta_{PK} \end{pmatrix} \quad \text{表示在總體層次的模式迴歸係數矩陣}$$

$$\Delta_{(N \times K)} = \begin{pmatrix} \delta'_1 \\ \delta'_2 \\ \vdots \\ \delta'_N \end{pmatrix}$$

表示總體層次的誤差項，是 $(N \times K)$ 的維度，並假設其亦遵循多變量常態分配，也就是 $[\Delta] = N_{N \times K}(\Delta | 0, I_N, \Lambda)$

在上式中， z_i 則表示顧客的個人特質，如本研究中有眾多人口統計的變數； Θ 則是用來衡量顧客的特質與偏好結構的關係，能修正顧客在估計偏好結構時得出的參數。

三、 後驗分配

後驗分配 (posterior distribution) 是貝氏統計理論推論的依據，一般簡化後都是以參數的概似函數乘上其參數的先驗分配，以推倒出完整的條件後驗分配，最後在使用馬可夫蒙地卡羅模擬術對完整條件後驗分配進行參數後驗分配的推導。

(一) 概似函數 (Likelihood Function)

在層級貝氏模型的假設下可知，參數的估計是在不同層次進行，不同層次的後驗分配之推倒與其概似函數皆有所不同。在本研究中，第一層個體層次的概似函數是來自顧客對不同產品組合的購買行為 $(y_{i1}, y_{i2}, \dots, y_{in_i})$ ，所組合而成的聯合機率密度函數，而每位顧客的購買行為向量 (Y_i) 則服從維度是 $(n_i \times 1)$ 的多變量常態分配，即為第一層模式的概似函數，如下式：

$$L(\beta_i, \sigma^2 | Y_i, X_i) = [Y_i | X_i, \beta_i, \sigma^2] = N_{n_i}(Y_i | X_i \beta_i, \sigma^2 I_{n_i})$$

另外，第二層總體層次的概似函數是由第一層的迴歸係數值 β_i 所構成，亦即上式裡對 β_i 設定的先驗分配，在此將之設定為服從 $(K \times 1)$ 維度的多變量常態分配，如下式：

$$L(\Theta, \Lambda | \beta_i) = [\beta_i | z_i, \Theta, \Lambda] = N_{K \times 1}(\beta_i | \Theta' z_i, \Lambda)$$

(二) 先驗分配

概似函數必須與先驗分配相乘才能推導出參數的後驗分配，因此在設定參數的先驗分配之前，也先做分層：第一層的參數包含迴歸係數 β_i 及齊質變異數 σ^2 ；第二層的參數則是包含迴歸係數 Θ 與共變異數矩陣 Λ 。Lenk (2001) 對上述兩兩參數所設定的先驗分配分別為多變量常態分配與 Inverse Gamma 分配，以及多變量常態分配與 Inverse Wishart 分配，各列式如下：

$$[\beta_i | z_i, \Theta, \Lambda] = N_{K \times 1}(\beta_i | \Theta' z_i, \Lambda)$$

$$[\sigma^2 | r_0, s_0] = IG\left(\sigma^2 \left| \frac{r_0}{2}, \frac{s_0}{2} \right.\right)$$

$$[\Theta^* | u_0, V_0] = N_{K \times P}(\Theta^* | u_0, V_0) \quad \text{又} \quad \Theta^*_{(K \times P)} = \text{vec}(\Theta)_{(K \times P)}$$

$$[\Lambda | f_0, G_0^{-1}] = IW_K(f_0, G_0^{-1})$$

第二層的先驗分配又稱為純粹先驗分配 (pure prior)，因為這些分配的參數都來自研究者主觀的設定，如 $u_0, V_0 \dots$ 等，則不需要推導其後驗分配。

(三) 完整的條件後驗分配

如前所述，若將第一、二層模式的概似函數及其先驗分配各自相乘，就得出參數的完整條件後驗分配，最後再利用馬可夫蒙地卡羅抽樣模擬過程，便可產生後驗分配的參數估計值。

第一層模式的參數包括 β_i 和 σ^2 ，完整的條件後驗分配如下推導：

$$\begin{aligned} [\beta_i | Rest] &\propto L(\beta_i, \sigma^2 | Y_i, X_i) \cdot [\beta_i | z_i, \Theta, \Lambda] \\ &\propto N_{n_i}(Y_i | X_i \beta_i, \sigma^2 I_{n_i}) \cdot N_{K \times 1}(\beta_i | \Theta' z_i, \Lambda) \\ &\propto N_{K \times 1}(\beta_i | u_i, V_i) \end{aligned}$$

$$V_i = \left(\frac{1}{\sigma^2} X_i' X_i + \Lambda^{-1} \right)^{-1}, \quad u_i = V_i \left(\frac{1}{\sigma^2} X_i' Y_i + \Lambda^{-1} \Theta' z_i \right)^{-1}$$

$$\begin{aligned} [\sigma^2 | Rest] &\propto \prod_{i=1}^N L(\beta_i, \sigma^2 | Y_i, X_i) \cdot (\sigma^2 | r_0, s_0) \\ &\propto \prod_{i=1}^N N_{n_i}(Y_i | X_i \beta_i, \sigma^2 I_{n_i}) \cdot IG\left(\sigma^2 \left| \frac{r_0}{2}, \frac{s_0}{2} \right.\right) \\ &\propto IG\left(\sigma^2 \left| \frac{r_N}{2}, \frac{s_N}{2} \right.\right) \end{aligned}$$

$$r_N = r_0 + \sum_{i=1}^N n_i, \quad s_N = s_0 + \sum_{i=1}^N (Y_i - X_i \beta_i)' (Y_i - X_i \beta_i)$$

由上式可知， β_i 的條件後驗分配是因人而異的，會受到參數 (u_i, V_i) 的影響，也就是因為結合了樣本資料 (X_i, Y_i) 與先驗分配的結果，使得接下來在第二層中能夠有效的修正個人參數估計值，又第二層模式的參數包括 Θ 和 Λ ，其完整的條件後驗分配推導如下：

$$\begin{aligned} [\Theta | Rest] &\propto \prod_{i=1}^N L(\Theta, \Lambda | \beta_i) \cdot [\Theta] \\ &\propto N_{NK \times 1}(B^* | (Z \otimes I_K) \Theta^*, I_K \otimes \Lambda) \cdot N_{KP \times 1}(\Theta^* | u_0, V_0) \\ &\propto N_{KP \times 1}(\Theta^* | u_N, V_N) \end{aligned}$$

$$V_N = \left[(Z' Z \otimes \Lambda^{-1}) + V_0^{-1} \right]^{-1}, \quad u_N = V_N \left[(Z' \otimes \Lambda^{-1}) B^* + V_0^{-1} u_0 \right]^{-1}$$

$$\begin{aligned} [\Lambda | Rest] &\propto \prod_{i=1}^N L(\Theta, \Lambda | \beta_i) \cdot [\Lambda | f_0, G_0^{-1}] \\ &\propto N_{N \times K}(B | Z \Theta, I_N, \Lambda) IW_K(\Lambda | f_0, G_0^{-1}) \\ &\propto IW_K(\Lambda | f_N, G_N^{-1}) \end{aligned}$$

$$f_N = f_0 + N, \quad G_N^{-1} = G_0^{-1} + (B - Z \Theta)' (B - Z \Theta)$$

上述所列示的條件後驗分配還必須進一步利用前面提到過的MCMC來模擬以逼近出邊際後驗分配的觀察值，而其中最常用的方法即為吉布斯抽樣法（Gibbs Sampling）。在模擬的過程符合馬可夫鏈的性質，抽樣結果僅受上一次抽樣結果影響，與其他次抽樣結果無關，這種疊代過程（iterative process）是將上一次的抽樣結果視為這一次抽樣的已知條件，再代入這一次的條件分配，重複抽樣多次之後，參數訊息的條件後驗分配會持續更新，而從中抽出的觀察值就會逼近真正的後驗分配。而根據這些模擬過程所產生的觀察值，就可以被用來計算後驗分配的所有特徵函數，如平均值與標準差等，進而描繪出後驗分配的性質，且其後驗分配的平均數（posterior mean）正是貝氏統計理論所欲推導的參數估計值。



第五節 層級貝氏邏吉斯模型

前文已指出在探討品牌或產品的選擇模式時，當其選擇行為是屬於分類型的變數，便適合以隨機效用模式來解釋，於是當消費者是理性購買的前提下，其選中的產品 m ，就代表其效用 (U_{ijm}) 必定大於其餘的產品 (U_{ijk})，即為：

$$U_{ijm} > U_{ijk}, \text{ for } k=1,2,\dots,M+1, K \neq m$$

又第 ($M+1$) 項產品一如往常在本研究中也是效用比較的基準，並將之設定為 0，稱為基礎產品 (base alternative)，於是便可將上式改寫為：

$$u_{ijm} = U_{ijm} - U_{ij(M+1)} = U_{ijm} = V_{ijm} + \varepsilon_{ijm}$$

$$= (w_{ijm} - w_{ij(M+1)})' \beta_{im} + \varepsilon_{ijm}$$

$$= x'_{ijm} \beta_{im} + \varepsilon_{ijm}$$

本式中各變數的意涵說明如下：

u_{ijm} ，表示顧客 i 在第 j 個購買時點，覺得第 m 項產品相對基礎產品在整體效用上的差異

x'_{ijm} ，表示顧客 i 在第 j 個購買時點，覺得第 m 項產品相對基礎產品在其他環境變數上的差異

β_{im} ，表示顧客 i 對產品 m 之偏好結構

ε_{ijm} ，表示研究者無法觀測到的隨機效用成分，假設其遵循極值分配 (Extreme Distribution) 正是邏吉斯迴歸模式。

接下來就是針對每位顧客都不同的偏好結構來給予參數的估計，學者們在推導時有兩種方法可用，一為資料擴充 (Data Augmentation)，另一則為梅托普里斯—海斯丁運算 (Metropolis—Hasting Algorithm)。前者適合在 Probit

模式，基於其需要透過資料擴充的方式，建立一個潛伏變數以及完整的條件後驗分配，再利用吉布思抽樣結果推論參數的後驗分配；而後者則適用在 Logit 模式，只需藉由將參數的樣本分配及先驗分配相乘，並在沒有封閉型式 (close form) 的條件後驗分配中抽樣，抽出適當的參數數值進而推論後驗分配為何。

一、 邏吉斯模式

(一) 第一層：個別顧客層次

由傳統的 Logit 模式假設隨機效用的誤差項服從極值分配可知，產品選擇的機率可以簡化為

$$P(C_{i,j} = k | \beta_i) = P(C_{i,j,k} \geq Y_{i,j,m}) \text{ for all } m = \frac{e^{x'_{i,j,k} \beta_i}}{1 + \sum_{m=1}^M e^{x'_{i,j,m} \beta_i}}, \text{ 若產品 } k \text{ 非基礎}$$

產品，即 $k \leq M = \frac{1}{1 + \sum_{m=1}^M e^{x'_{i,j,m} \beta_i}}$ ，若產品 k 即是基礎產品，即 $k = M + 1$

上式中

$$X_{ijm} = \begin{pmatrix} X_{ijm1} \\ X_{ijm2} \\ \vdots \\ X_{ijmK} \end{pmatrix}$$

表示顧客 i 在購買時點 j 時認為產品 m 在第 K 個屬性水準上的表現

$$\beta_i = \begin{pmatrix} \beta_{i1} \\ \beta_{i2} \\ \vdots \\ \beta_{iK} \end{pmatrix}$$

表示顧客 i 對第 K 個屬性水準的成分效用值，即為其偏好結構

根據產品選擇模式便可推導第一層模式的概似函數 $L(\beta_1 | x_{ijm}, y_{ijm})$ ，亦即

$$L(\beta_1 | x_{ijm}, y_{ijm}) = \left(\frac{e^{x'_{i,j,1}\beta_1}}{1 + \sum_{m=1}^M e^{x'_{i,j,m}\beta_1}} \right)^{a_1} \cdots \left(\frac{e^{x'_{i,j,M}\beta_1}}{1 + \sum_{m=1}^M e^{x'_{i,j,m}\beta_1}} \right)^{a_M} \left(\frac{1}{1 + \sum_{m=1}^M e^{x'_{i,j,m}\beta_1}} \right)^{n_i - \sum_{m=1}^M a_m}$$

上式等號右邊是顧客 i 購買產品 1 共 a_1 次的聯合機率，以及依此類推購買產品 m 共 a_M 次的聯合機率，選擇基礎產品共 $\left(n_i - \sum_{m=1}^M a_m \right)$ 次。

將此一概似函數搭配 β_i 的先驗分配，便可使用梅托普里斯－海斯丁運算產生 β_i 之後驗模擬值。

(二) 第二層：跨顧客層次

層級貝式 Logit 模式的顧客偏好結構假設為相異，遂得出第二層模式，加入個人特質以進一步求得精確的個人喜好，達成客製化行銷的目的。

$$B = Z\Theta + \Delta$$

其中

$$B_{(N \times K)} = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1K} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2K} \\ \vdots & \vdots & & \vdots \\ \beta_{N1} & \beta_{N2} & \cdots & \beta_{NK} \end{pmatrix} = \begin{pmatrix} \beta'_1 \\ \beta'_2 \\ \vdots \\ \beta'_N \end{pmatrix} \text{ 表示個人層次的偏好結構}$$

$$Z_{(N \times P)} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1P} \\ z_{21} & z_{22} & \cdots & z_{2P} \\ \vdots & \vdots & & \vdots \\ z_{N1} & z_{N2} & \cdots & z_{NP} \end{pmatrix} = \begin{pmatrix} z'_1 \\ z'_2 \\ \vdots \\ z'_N \end{pmatrix} \text{ 表示顧客特質的矩陣，設共有 P 個變數}$$

$$\Theta_{(P \times K)} = \begin{pmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1K} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2K} \\ \vdots & \vdots & & \vdots \\ \theta_{P1} & \theta_{P2} & \cdots & \theta_{PK} \end{pmatrix} \text{ 表示第二層模式的迴歸係數矩陣}$$

$$\Delta_{(N \times K)} = \begin{pmatrix} \delta'_1 \\ \delta'_2 \\ \vdots \\ \delta'_N \end{pmatrix}$$
 表示第二層模式的誤差項，遵循多變量常態分配，即

$$[\Delta] = N(\Delta | 0, I_n, \Lambda)$$
 ，式中的 z_i 表示個人特質，如年齡、教育程度，與所得收入等； Θ 則表示用來衡量顧客的個人特質與偏好結構之間的關係，有助於預測顧客的產品偏好和對行銷變數的反應。

二、梅托普里斯—海斯丁運算法

當層級貝氏模式結合參數的先驗分配、樣本分配而產生出未知參數的條件後驗分配時，通常都是利用馬可夫鏈蒙地卡羅的方法，產生近似的參數後驗分配。但是當參數的條件後驗分配不易推導時，可使用梅托普里斯—海斯丁運算法 (Metropolis—Hasting Alogrithm)，直接根據概似函數與先驗分配之乘積密度值，模擬產生參數的可能值。這套梅托普里斯—海斯丁運算法是由 Metropolis 在 1953 年與 Hasting 在 1970 年先後發展，再予以一般化，也是 MCMC (Markov Chain Monte Carlo) 的方法之一。先假設隨機變數與參數 $[x]$ 所遵循機率分配是正比 (proportional to) 於某一函數 f ，即 $[x] \propto f(x)$ ，左式中的 $[x]$ 式原始的機率分配，也許不具有封閉型式 (close form) 的機率密度函數，但是依然可以推導 $[x]$ 是正比於另一個具有封閉型式的函數 $f(x)$ ，若能先找到一個起始值 x_i 後，再根據 $f(x)$ 選取適合的模擬值，建立一串馬可夫鏈，而此串馬可夫鏈的機率分配最終會收斂為 $[x]$ ，即可根據模擬值的性質 (如平均值或變異數)，反推回原始分配 $[x]$ 的參數 (Lenk, 2001)。梅托普里斯—海斯丁運算法的模擬過程如下：

- (一) 產生初始值 x_i
- (二) 選擇跳躍分配，給定起始值為 x_i 後，隨機產生下一個可能值 $X_{i+1} = Y$ ，以令跳躍分配 (jump distribution) 的函數為 $g_i(y|x_i)$ ，是為產生新模擬值的依據。跳躍分配的設定除了決定該如何抽取可能值，也可能會影響到其所抽出的可能值，是否符合馬可夫鏈的資格。基本上跳躍分配可以分為三種型式：

獨立 (Independence) : $g(y|x) = g(y)$

對稱 (Symmetric) : $g(y|x) = g(x|y)$

條件常態 (Conditional, Normal) : $g(y|x) = N(y|x, Y)$

- (三) 判斷抽取出的可能值是否符合馬可夫鏈資格

假設下一個被抽取出來的可能值 $X_{i+1} = Y$ 之機率為

$$p(x_i, y) = \min \left\{ \frac{f(y)g_i(x_i|y)}{f(x_i)g_i(y|x_i)}, 1 \right\},$$

左式為假設跳躍分配是對稱型，則上

述機率可簡化為 $p(x_i, y) = \min \left\{ \frac{f(y)}{f(x_i)}, 1 \right\}$ ，也就是說若 $f(y) > f(x_i)$ ，

則可知 $p(x_i, y) = 1$ ，也就是說 y 即符合馬可夫鏈資格，可作為下一個模擬值；反之，若 $f(x_i) > f(y)$ ，則其 $p(x_i, y) < 1$ ，這時 y 是否有符合馬可夫鏈資格，就必須和另外一個機率比較。

- (四) 若 $p(x_i, y) < 1$ 時，必須從均勻分配 $(0,1)$ 間隨機抽出一值，令其為 u ，若 $p(x_i, y) > u$ 則仍然可以將 y 視為符合馬可夫鏈資格；否則便將此值捨棄，並使該輪的 y 存在缺失值。

在層級貝氏 Logit 模式中，梅托普利斯—海斯丁運算法應用在產生第一層迴歸係數 β_i 的模擬值，而上述所提及的機率函數 $f(x_i)$ ，相當於層級貝氏模型中的概似函數與先驗分配之乘積。一旦得出 β_i 的模擬值，就可繼續其他的參

數推導，不需要像 Probit 一般產生潛伏變數。

三、條件後驗分配—各層級之迴歸係數模擬

(一) 第一層：個人層次

根據層級貝氏統計理論，在第一層的個人層次中其迴歸係數的條件後驗分配，和其概似函數及先驗分配的乘積成正比，在前文已開列了第一層模式之概似函數，並且將第二層模式的 β_i 之先驗分配設定為服從多變量常態分配，所以 β_i 的條件後驗分配應正比於 $f(\beta_i)$ ，如下說明：

$$f(\beta_i) = L(\beta_i | x_{ijm}, y_{ijm}) \pi[\beta_i | z_i, \Theta, \Lambda]$$

$$= \left(\frac{e^{x'_{i,j,1} \beta_i}}{1 + \sum_{m=1}^M e^{x'_{i,j,m} \beta_i}} \right)^{a_1} \cdots \left(\frac{e^{x'_{i,j,M} \beta_i}}{1 + \sum_{m=1}^M e^{x'_{i,j,m} \beta_i}} \right)^{a_M} \left(\frac{1}{1 + \sum_{m=1}^M e^{x'_{i,j,m} \beta_i}} \right)^{n_i - \sum_{m=1}^M a_m} \pi[\beta_i | z_i, \Theta, \Lambda]$$

梅托普里斯-海斯丁運算法之模擬過程如下(Lenk,2001)：

1. 產生初始值 $\beta_i(0)$ ，一般設為 0 向量或是上一輪 β_i 的抽樣結果。
2. 選擇跳躍分配，給定起始值 $\beta_i(0)$ ，隨機產生下一個可能值 $\beta_i(1)$

假設跳躍分配是對稱型式的常態分配，即：

$$g[\beta_i(1) | \beta_i(0)] = N_K(\beta_i(0), c^2 \Lambda)$$

上式中， c^2 是由研究者自行控制的參數，倘若 c^2 值太大，則抽出的 $\beta_i(1)$ 將很可能與 $\beta_i(0)$ 相差太遠，使得造成資格不符而需要重抽，這不僅將大幅增加程式重覆執行的次數，甚至有可能根本跑不出符合資格的 β_i 值；然而，若 c^2 值太小，則抽出的 $\beta_i(1)$ 會與 $\beta_i(0)$ 太過相近，雖然說這樣必然可以產生 β_i 值，但其所形成的「馬可夫鏈」可能存在高度的自我相關，與原先隨機產生模擬值的精神不符。因此 c^2 值的設定要非常小心，必須經過多次試誤。

3. 判斷 $\beta_i(1)$ 是否符合馬可夫鏈資格

先假設抽到下一個值是 $\beta_i(1)$ 的機率是：

$$p[\beta_i(0), \beta_i(1)] = \min \left\{ \frac{f[\beta_i(1)]g[\beta_i(0)|\beta_i(1)]}{f[\beta_i(0)]g[\beta_i(1)|\beta_i(0)]}, 1 \right\},$$

本研究繼續對上式假設，若跳躍分配是對稱型式，則上述機率可簡化為：

$$p(x_i, y) = \min \left\{ \frac{f[\beta_i(1)]}{f[\beta_i(0)]}, 1 \right\}$$

隨後檢視上式中比值，

$$\frac{f[\beta_i(1)]}{f[\beta_i(0)]} = \frac{L[\beta_i(1)|x_{ijm}, y_{ijm}] \prod_{k=1}^K N_k(\beta_i(1)|\Theta' z_i, \Lambda)}{L[\beta_i(0)|x_{ijm}, y_{ijm}] \prod_{k=1}^K N_k(\beta_i(0)|\Theta' z_i, \Lambda)}$$

若上式的比值大於 1，則 $\beta_i(1)$ 就是個合格的模擬值，可投入後續的統計推論；反之，若比值小於 1，則必須進行至下一個步驟，也就是要再與另一個指標值 (u) 做比較。這個指標值 (u) 是隨機選自均勻分配 (0, 1)，若上式中的比值確實大於 u ，則 $\beta_i(1)$ 就依然算是合格的模擬值；但若比值仍舊小於 u ，則將予以捨棄，而該輪的 $\beta_i(1)$ 就會存在缺失值。

(二) 第二層：跨顧客層次

上述已經使用梅托普里斯—海斯丁運算法，求出個人層次之迴歸係數的模擬值，接下來便針對第二層模式的參數之條件後驗分配進行推導，在此層次中包含了 Θ 的迴歸係數矩陣與共變異數矩陣 Λ ，說明如下：

Θ 矩陣

$$\text{概似函數：} \prod_{i=1}^N L(\Theta, \Lambda | \beta_i, z_i) = \prod_{i=1}^N L(\beta_i | z_i, \Theta, \Lambda) = \prod_{i=1}^N N_{(M+K)}(\beta_i | \Theta' z_i, \Lambda)$$

先驗分配： $[\text{vec}(\Theta')] = N_{(M+K)P}[\text{vec}(\Theta') | u_0, V_0]$ ，將 Θ 向量化後，遵循多變量常

態分配；令 $\Theta^* = \text{vec}(\Theta')$ 。

條件後驗分配：

$$[\Theta^* | \text{Rest}] \propto \prod_{i=1}^N L(\Theta^*, \Lambda | \beta_i, z_i) [\Theta^*]$$

$$\begin{aligned} &\propto N_{N(M+K)} \left[\mathbf{B}^* \left(\mathbf{Z} \otimes \mathbf{I}_{(M+K)} \right) \boldsymbol{\Theta}^*, \mathbf{I}_{(M+K)} \otimes \boldsymbol{\Lambda} \right] N_{(M+K)P} \left(\boldsymbol{\Theta}^* \mid u_0, V_0 \right) \\ &\propto N_{(M+K)P} \left(\boldsymbol{\Theta}^* \mid u_N, V_N \right) \\ V_N &= \left[\left(\mathbf{Z}' \mathbf{Z} \otimes \boldsymbol{\Lambda}^{-1} \right) + V_0^{-1} \right]^{-1}, \quad u_N = \left[\left(\mathbf{Z}' \otimes \boldsymbol{\Lambda}^{-1} \right) \mathbf{B}^* + V_0^{-1} u_0 \right]^{-1} \end{aligned}$$

$\boldsymbol{\Lambda}$ 矩陣

$$\text{概似函數：} \prod_{i=1}^N L(\boldsymbol{\Theta}, \boldsymbol{\Lambda} \mid \beta_i, z_i) = \prod_{i=1}^N L(\beta_i \mid z_i, \boldsymbol{\Theta}, \boldsymbol{\Lambda}) = \prod_{i=1}^N N_{(M+K)}(\beta_i \mid \boldsymbol{\Theta}' z_i, \boldsymbol{\Lambda}),$$

$$\text{先驗分配：} \left[\boldsymbol{\Lambda} \mid f_0, G_0^{-1} \right] = IW_{(M+K)}(f_0, G_0^{-1})$$

條件後驗分配：

$$\begin{aligned} \left[\boldsymbol{\Lambda} \mid Rest \right] &\propto \prod_{i=1}^N L(\boldsymbol{\Theta}, \boldsymbol{\Lambda} \mid \beta_i, z_i) \left[\boldsymbol{\Lambda} \right] \\ &\propto N_{N(M+K)} \left[\mathbf{B} \mid \mathbf{Z} \boldsymbol{\Theta}, \mathbf{I}_N, \boldsymbol{\Lambda} \right] IW_{(M+K)} \left[\boldsymbol{\Lambda} \mid f_0, G_0^{-1} \right] \\ &\propto IW_{(M+K)} \left[\boldsymbol{\Lambda} \mid f_N, G_N^{-1} \right] \\ f_N &= f_0 + N, \quad G_N^{-1} = G_0^{-1} + (\mathbf{B} - \mathbf{Z} \boldsymbol{\Theta})' (\mathbf{B} - \mathbf{Z} \boldsymbol{\Theta}), \end{aligned}$$

以上所開列之第二層次的條件後驗分配，只要以 MCMC 的方法模擬產生來自邊際後驗分配的抽樣值，透過疊代過程(iterative process)，可將上一次抽樣的結果視為本次抽樣的已知條件，再代入本次的條件分配；重覆抽樣多次之後，由不斷更新參數訊息的條件後驗分配所抽出來的觀察值，將逼近真正的後驗分配。根據上述這些模擬產生的抽樣值，便可計算後驗分配的所有特徵函數，如平均數與變異數等，進而描述出後驗分配的性質；後驗分配的平均數(posterior mean)即為貝氏統計 Logit 模式之參數估計值。

第四章 實證研究

延續上一章研究方法，本研究接下來會利用實際的資料來做實證的探討。本章內容總共可分為四節：第一節為樣本的介紹；第二節為模擬企業在無任何顧客或產品的資訊時，利用平均機率法進行的推薦系統；第三節則為引入總合邏吉斯模型的推薦系統；第四節則是利用層級貝氏邏吉斯模式來建立顧客推薦系統，在這個階段便將推薦系統的概念由總體層次縮小為個體層次，除了考量產品屬性面外，也衡量顧客的異質偏好會產生的改變，使推薦系統更加細緻。

第一節 樣本描述

一、樣本描述

本研究是選定全球知名線上零售商亞馬遜網路商店（Amazon.com）之顧客交易資料為研究工具。資料時間點為選取自該業者於2004年1月1日至2004年12月31日整一年的書籍交易資料，並由其中篩選出1556位消費者，共包含5861筆實際成交的資料。在每位顧客的交易紀錄中，各分別涵蓋了以下幾個項目：

（一）顧客基本資料檔案：包含教育程度、所在地區、家庭人口狀況、年齡、平均收入、是否正扶養未成年的孩童、種族背景、網路連線速度、國籍等。

（二）書籍的敘述資料：包含書籍名稱、標示價格、實際售價、出版年月份、裝訂方式、頁數、顧客評價、顧客回應文章數、銷售排名、書籍種類等。

（三）交易時點的資料：包含顧客編號、顧客停留在網頁的時間、顧客瀏覽的網頁頁數、交易的日期與時間等。

考慮該資料庫中的豐富資訊，與本研究主題相關之資料皆予以留下，如顧客基本資料檔案中便將顧客編號、教育程度、所在地區、家庭人口狀況、年齡、平均收入、是否扶養未成年的孩童等資訊，納入本研究的人口統計變項之中，其他如種族背景及國籍等資訊，一來因為該資料庫內的交易紀錄多以美國籍的白人為主，其餘國籍與人種皆不成比例的低，二者是由於此變項非本研究主要討論的內

容，因此為保持樣本的結構平衡，與研究意涵的連貫性則不予計入，剩餘如網路連線速度等不予採計的資訊，皆源自相似的理由；另外，本研究面對資料不齊全的狀況，會依據其變數的相對重要性，考量遺失值的處理方式，例如本研究中將教育程度視為購書的重要影響變數之一，於是便將資料內教育程度為不詳者，利用多變量統計的處理方式，以該變數的平均值來取代遺失值；而其他相對影響較小的變數，則視情況嚴重性決定是否要直接捨去。上述是為顧客基本資料的變數採用標準，與處理方式簡介，其他類型的資料皆依此類推，評估後整理本研究會用到的資料如下表 4-1-1 所示：

表 4-1-1 本研究使用之資料類型

顧客基本資料	教育程度、所在地區、家庭人口狀況、年齡、平均收入、是否扶養未成年的孩童
書籍的敘述資料	書籍名稱、標示價格、裝訂方式、頁數、顧客評價、顧客回應文章數、銷售排名、書籍種類
交易時點的資料	顧客編號、交易的日期與時間

在第三章研究方法的內容當中，已經揭示了各變數的內容及其水準，以下便將實際的交易紀錄資料作為樣本內容，套入各變數之中，得出各變數水準的樣本比例如下各圖 4-1-1 與 4-1-2 所示：

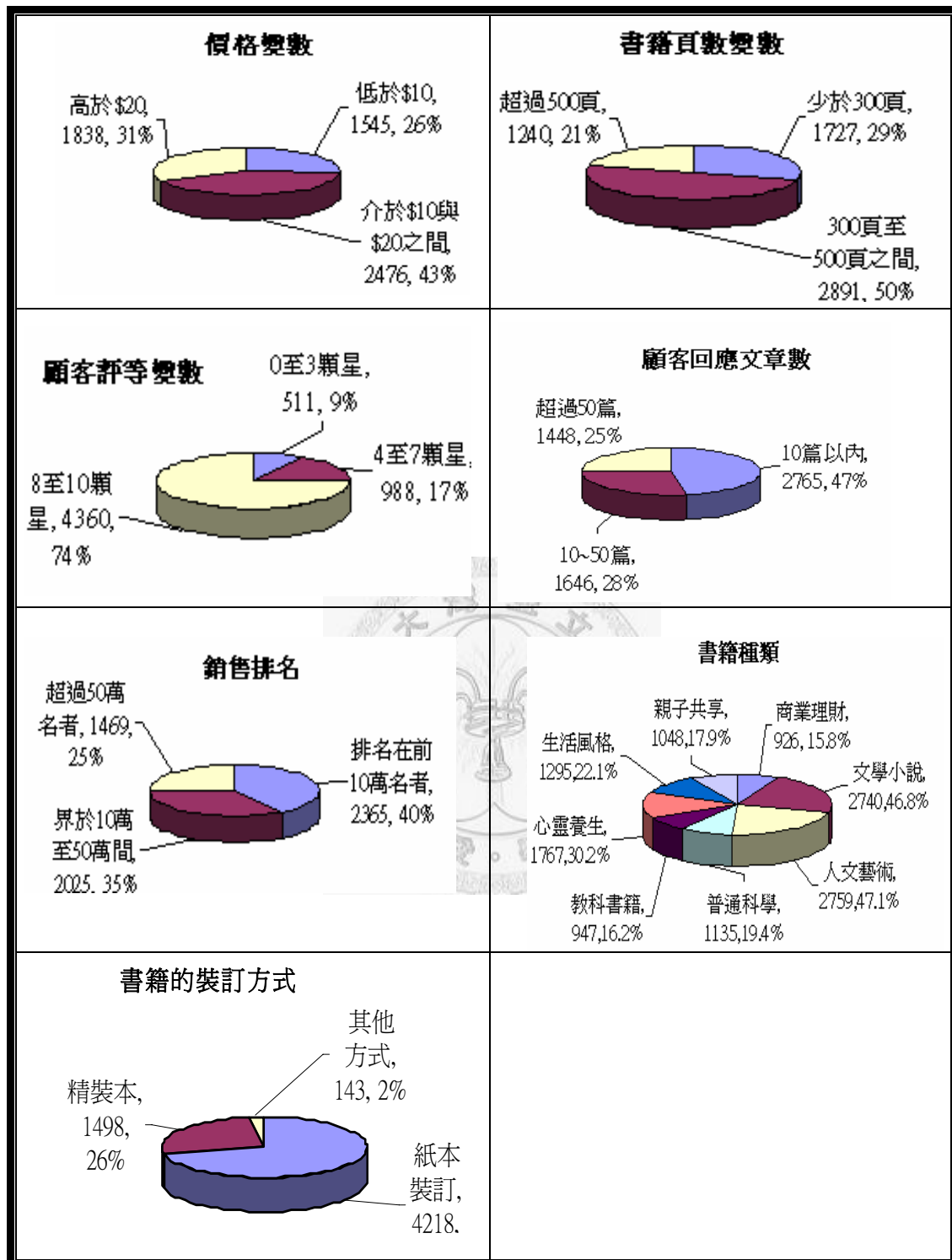


圖 4-1-1 各商品變數的樣本比例

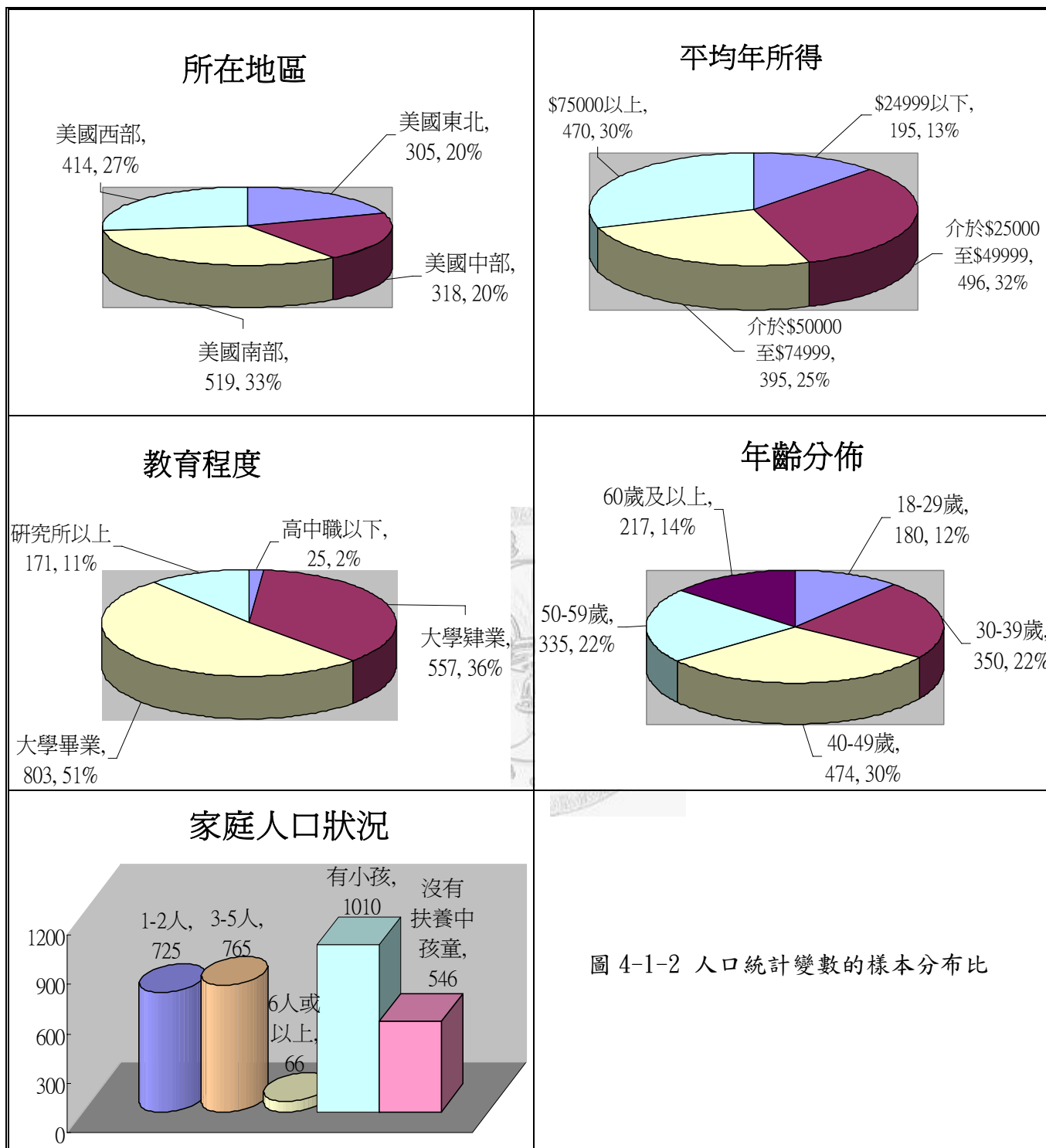


圖 4-1-2 人口統計變數的樣本分布比

該資料庫中篩選出來的 5861 筆交易紀錄，共包含 1556 位顧客的資料，為建立本研究的顧客推薦系統，遂將該資料劃分為樣本內與樣本外的保留資料。樣本內資料即指在這 5861 筆資料中，抽選出曾重複購買的顧客共 987 人，並在這 987 人之中分別去掉每位顧客最近期交易的一筆資料，剩餘的部分則為樣本內資料，用來建立顧客推薦系統以及檢驗推薦系統的有效性；另外，從中被選取出來的顧客最近期之交易紀錄，遂形成樣本外的資料，可作為之後建立完成顧客推薦系統後，能夠被用來預測舊顧客的購買機率；此外，還有一部分樣本外的資料不在這 987 位曾經重複購買過的顧客中，也就是剩餘僅只有一次購買紀錄的 569 位顧客，這些顧客的交易資料也是在建立完成顧客推薦系統之後，用來預測新顧客的購買機率。

二、顧客購買的考慮集合

由舊顧客與新顧客的樣本中，個別挑選出在 2004 年度最常被購買之書籍，將其前 30 名各種屬性的組合，分別列表如下頁表 4-1-2 與 4-1-3，本研究稍後需以此購買考慮的集合作為比較各種模型對新、舊顧客的預測擊中率。

表 4-1-2 預測舊顧客購買機率的購買考慮集合

No.	價格	裝訂 方式	頁數	顧客 評價	顧客 回應	銷售 排名	書籍 種類
1	0	1	1	0	1	0	0
2	0	1	1	0	1	0	0
3	0	1	1	0	1	0	0
4	1	0	1	0	1	0	1
5	1	0	1	0	1	0	0
6	0	1	1	0	1	0	0
7	0	1	0	1	1	0	0
8	0	1	1	0	1	0	0
9	0	0	0	1	1	0	0
10	0	1	0	1	1	0	0
11	0	1	1	0	1	0	1
12	0	1	1	0	1	0	0
13	0	0	1	0	1	0	1
14	0	1	1	0	1	0	0
15	1	0	1	0	1	0	0
16	1	0	1	0	0	1	0
17	0	1	0	1	1	0	0
18	0	0	1	0	1	0	1
19	0	1	1	0	0	1	0
20	0	0	0	1	0	1	0
21	0	0	0	1	0	1	0
22	0	0	0	1	1	0	0
23	0	1	1	0	1	0	1
24	0	0	0	1	1	0	0
25	1	0	1	0	1	0	1
26	1	0	0	0	0	1	1
27	1	0	0	1	0	0	1
28	0	1	0	0	1	0	1
29	1	0	0	0	0	1	0
30	1	0	0	1	0	0	0

表 4-1-3 預測新顧客購買機率的購買考慮集合

No.	價格	裝訂方式	頁數	顧客評價	顧客回應	銷售排名	書籍種類
1	0	1	1	0	0	1	0
2	0	1	1	0	1	0	0
3	1	0	1	0	1	0	0
4	0	1	1	0	1	0	0
5	0	1	1	0	1	0	0
6	0	1	1	0	1	0	0
7	0	1	1	0	0	1	0
8	0	1	1	0	0	1	0
9	0	1	1	0	1	0	0
10	1	0	1	0	0	1	0
11	1	0	1	0	1	0	0
12	0	1	1	0	1	0	0
13	1	0	1	0	0	1	0
14	0	1	1	0	0	1	0
15	1	0	1	0	0	1	0
16	1	0	1	0	0	1	0
17	0	1	0	1	1	0	0
18	0	0	1	0	1	0	1
19	0	1	1	0	0	1	0
20	0	0	0	1	0	1	0
21	0	0	0	1	0	1	0
22	0	0	0	1	1	0	0
23	0	1	1	0	1	0	1
24	0	0	0	1	1	0	0
25	1	0	1	0	1	0	1
26	0	1	0	1	0	1	1
27	0	1	1	0	0	1	0
28	0	0	0	1	0	1	0
29	0	0	0	0	0	1	0
30	1	0	1	0	1	0	1

第二節 無任何資訊下之推薦系統

顧客客製化的推薦系統可說是在資料庫行銷興起之後才被開發出來的一種技術，以過去實體商店的銷售情形為例，除非是對顧客相當熟稔的店員，能夠猜到特定顧客的消費習性與偏好，否則一般而言，商店的店員在推銷商品時，多是隨機的推薦，許多通常都必須等到顧客已明確的表示所要尋找的商品類型，或某種樣式的產品，才能夠更加確定推薦的產品是符合顧客需求的。因此，有鑒於並非所有的企業都能掌握顧客的資訊，本研究便欲先了解在這種缺乏資訊的情況下，隨機推薦書籍所能夠增加的購買機率為何，並可與之後的其他情境比較，以作為未來線上書店欲交叉銷售其他產品時的參考。

一、基本概念

當行銷人員對消費者一無所知的時候，只能隨機推薦產品給消費者，所以當企業或廠商要猜測消費者是否會購買某一推薦的商品時，可以使用平均機率的觀念來了解，在眾多的書籍種類可選時，推薦的產品可能會有多少機會受到消費者的青睞。

二、模式準確性

(一) 資料型態

由於不需要了解顧客的偏好、過去曾購買過的書籍是什麼樣的屬性，故只需要知道每一位顧客面對的產品組合選擇，便可知道其購買特定書籍的平均機率為多少？在前一節便提及，本研究利用挑選出的 30 種常購買書籍之屬性組合，當作顧客的購買考慮集合 (Consideration Set)，其每位顧客面臨的購買集合個數多為 30 組，但比較特殊的狀況是有的顧客實際購買的資料型態，正好都不符合 30 組購買考慮集合中任一組，因此其真實面對的選擇集合則定義為 $30+1=31$ 組，以下便列出顧客的考慮集合個數

表 4-2-1 顧客考慮集合個數

顧客編號	購買集合的個數
5662676	31
5669738	31
5689812	30
⋮	⋮
8812747	31
8815390	31

(二) 推薦書籍給舊顧客

計算的方式相當簡易，即當每推薦 10 本書籍給甲顧客，甲顧客便購買了其中一本，則可得到購買機率為 10%，而本研究的考慮集合個數可能多達 30 個或 31 個，其中只有一個被消費者購買，於是我們便開始計算到底需要隨機推薦到第幾本書籍時，才能刺激顧客購買。在此共有 987 個樣本外的保留資料，其平均機率的求算方式如下：

1. 推薦一本書顧客即購買之機率：

$$P_1 = \frac{1}{987} \left(\frac{C_1^1}{C_1^{31}} + \frac{C_1^1}{C_1^{31}} + \dots + \frac{C_1^1}{C_1^{30}} + \frac{C_1^1}{C_1^{30}} + \dots + \frac{C_1^1}{C_1^{31}} \right) = 0.03249$$

2. 推薦兩本書時至少有一本被購買之機率：

$$P_2 = \frac{1}{987} \left(\frac{C_1^2}{C_1^{31}} + \frac{C_1^2}{C_1^{31}} + \dots + \frac{C_1^2}{C_1^{30}} + \frac{C_1^2}{C_1^{30}} + \dots + \frac{C_1^2}{C_1^{31}} \right) = 0.19497$$

3. 推薦 10 本書時至少有一本被購買之機率：

$$P_{10} = \frac{1}{987} \left(\frac{C_1^{10}}{C_1^{31}} + \frac{C_1^{10}}{C_1^{31}} + \dots + \frac{C_1^{10}}{C_1^{30}} + \frac{C_1^{10}}{C_1^{30}} + \dots + \frac{C_1^{10}}{C_1^{31}} \right) = 0.32495$$

整理推薦前 10 本書籍便推薦成功（亦即顧客有購買）之平均機率如下：

表 4-2-2 平均機率法的舊顧客推薦累積擊中率

推薦個數	1	2	3	4	5	6	7	8	9	10
累積擊中率 (%)	3.25	6.50	9.75	13.00	16.25	19.50	22.75	26.00	29.25	32.50

(三) 推薦書籍給新顧客

用相同的做法來預測隨機推薦書籍給新顧客時，在此共有 569 位樣本外的保留資料，其平均購買機率如下：

表 4-2-3 平均機率法的新顧客推薦累積擊中率

推薦個數	1	2	3	4	5	6	7	8	9	10
累積擊中率(%)	3.24	6.48	9.72	12.96	16.20	19.44	22.68	25.92	29.16	32.40

比較兩種類型的顧客之平均推薦成功機率，可發現擊中率純粹受到顧客考慮集合內數目多寡的影響，且隨著推薦個數的增加而提高成功率，是相當直觀的結果，因此對科技日臻成熟，書籍的印製和出版都加速的環境而言，其各式屬性組合是有增無減，將使得顧客面對的考慮集合愈加多元，導致隨機推薦的成效下降。

在下一節中，本研究試圖將書籍的各種屬性帶進模式裡，以了解不同屬性水準對購買機率的影響，亦即為總合邏吉斯分析。推測由於本節推薦時未包含書籍的屬性以及消費者的偏好，故推薦成功機率應遠低於其他有牽涉到產品屬性與顧客偏好的推薦模式。



第三節 總合邏吉斯模型分析

一、模型

$$y_j = \sum_{p=1}^N \beta_p X_{pj}$$

其中， y_j = 個別受測者對第 j 個產品所作之決定，若 $y_j = 1$ 代表有產生實際購買行為， $y_j = 0$ 代表無產生購買行為。

β_p = 第 p 個產品屬性水準之成分效用值， $p=1,2,\dots,N$ 。

X_{pj} = 第 j 個產品、第 p 個產品屬性。

二、參數估計

在此節的總合邏吉斯模式主要是假設所有顧客皆視為同質，彼此同屬在單一的市場區隔內，因此模式會將所有的顧客，不論是曾經消費過的舊顧客，抑或是未曾光顧的新顧客，其偏好結構都看作相同，且僅只估計出一組迴歸預測式，所有顧客都根據這組迴歸式的係數進行購買預測，本節便依照這組迴歸式所預測之購買結果，來為顧客做產品的推薦。下頁表 4-3-1 即為該總合邏吉斯模式估計出之各變數的迴歸係數

表 4-3-1 利用總合邏輯模式估計所得各變數的迴歸係數

變數		B	S.E.	Sig.
Constant		-23.0008	0.5492	0.000
標示價格	低於\$10 元	-3.4401	0.1772	0.000
	\$10-\$20 元	-1.0598	0.1520	0.000
裝訂方式	紙本	10.8793	0.3237	0.000
	精裝本	6.7752	0.2769	0.000
頁數	少於 300 頁	4.0947	0.1612	0.000
	300-500 頁	4.4115	0.1656	0.000
顧客評價	0-3 顆星	-5.2938	0.1743	0.000
	4-7 顆星	-3.5530	0.1457	0.000
顧客回應文章	少於 10 篇	2.5499	0.1884	0.000
	10-50 篇	-1.5092	0.1673	0.000
銷售排名	前十萬名	5.1888	0.2184	0.000
	十萬至五十萬名	2.7650	0.1553	0.000
書籍種類	商業理財	3.9844	0.2821	0.000
	文學小說	12.6980	0.3275	0.000
	人文藝術	11.3328	0.3389	0.000
	普通科學	6.3348	0.2778	0.000
	教科書籍	8.8499	0.3448	0.000
	心靈養生	8.3443	0.2788	0.000
	生活風格	9.7364	0.3548	0.000

從上表我們可以看到所有的變數，對於影響購買的程度都具有一定的顯著重要性，可歸結出以下幾點：

(一) 標示價格

在價格的部分是利用虛擬變數分為三類，由係數可以發現，當書籍的價格低於\$10，或是介於\$10至\$20時都會對購買機率產生負向的影響，又當書籍標示的價格是在\$10元以下時，得到反向影響的幅度竟是最大，超過書籍價格為\$10至\$20元，甚至超過標價為\$20元以上的書籍，可推論網路書店的消費群眾中具有一定比例的消費能力，儘管面對稍高價格的書籍，依舊不減購書的機率。

(二) 裝訂方式

書籍的裝訂方式明顯以紙本最受到網路購書的群眾喜愛，其次是精裝本，二者都是正向影響購書可能性，前者更是可以有效提高購買機率。

(三) 頁數

在頁數的部分，無法從係數直接判斷少於 300 頁和介於 300 頁到 500 頁之間的書籍，何者較能有效提高購書機率，但可確定的是二者皆為正向影響，且其影響力也都不差，惟頁數超過 500 頁的書籍，似乎對吸引顧客購買的能力較弱。

(四) 顧客評價

在顧客評價的部分可明顯看到當顧客給予該書籍的評價是較差，只獲得 0-3 顆星等，將強烈降低顧客的購買機率，對獲得 4-7 顆星等的書籍亦是，只不過負向影響購買機率的程度稍弱一些。這也可從樣本的分佈情形窺知一二，在本研究的資料中就有近 74% 的書籍獲得 8-10 顆星等的評價，由此可見得網路書店在經營口碑效果時，萬萬馬虎不得。

(五) 顧客回應的文章數

顧客回應的文章數目常會發生數量相對懸殊的情況，但在維持樣本的比例平衡且合乎邏輯的條件下，將回應的文章數目區分為三種水準，由其係數發現文章回應數目界在 10-50 篇之間的書籍，竟然會出現負向影響購買機率的結果，而文章數目儘管少於 10 篇，顧客購書的機率非但不因此降低，甚至還能正面提高購買機率，推斷可能原因來自當回應的文章數目增加時，對該書籍的負面評論文章會隨著正面說法不等比例的出現。

(六) 銷售排名

在銷售排名的部分，發現顧客在購買書籍時會考量到該書的銷售狀況，當該書籍在擁有超過 500 萬本書籍的亞馬遜網路書店裡被視為暢銷品，排名在前十萬名，則顧客購買該書籍的可能性就會提高，若是排名在前十萬至前五十萬之間，也會正向影響購買機率。

(七) 書籍種類

書籍種類的部分應是購書人決定購買的重要因素之一，由本研究的樣本資料可推論，會在線上購書的顧客較偏重購買的書籍種類，依序應為人文藝術類、文學小說類，以及心靈養生類別；若從迴歸式中得到的係數結果可發現，最能提高顧客線上購書機率的書籍種類，是由文學小說類、人文藝術類，及生活風格類別分居前三，估計得出的結論雖稍有不同，但文學小說類及人文藝術類別二者著實在影響購買機率的差異上不大，相較之下商業理財類別與親子共享類別，能提高購買機率的可能性就遜色些。

三、模式準確性

在上個步驟得出了各個變數的參數估計值後，便開始進行樣本外的顧客預測。不論是對新顧客或舊顧客都一樣，先將每位顧客所購買的書籍之屬性水準，帶入此估計的迴歸式中，所得到的值即可代表該名顧客對此書籍的屬性偏好。以下便分別描述舊顧客的推薦準確性及新顧客的推薦準確性。

(一) 對舊顧客的推薦

首先將樣本外保留的舊顧客挑選出來，並將每位顧客面對的產品購買考慮集合 (Consideration Set) 列出，將各個屬性水準的參數估計值代入，在每位顧客的購買考慮集合內的每一筆資料，都能求得一個相對效用值，當此相對效用值越大，即表示購買的機率也越大，於是將所有顧客的相對效用值由大至小排列，以顧客編號 5662676 為例，將其考慮集合中所有產品屬性的迴歸係數帶入，並按照其偏好結構排序，可得結果如下頁表 4-3-2：

表 4-3-2 考慮集中的相對效用值(以顧客編號 5662676 為例)

考慮集中的產品	購買決策	預測購買機率 (相對效用值)
1	0	32.315
2	0	28.360
3	0	27.429
4	0	26.373
5	0	22.777
6	0	21.297
7	0	21.019
8	0	17.583
9	0	17.583
10	1	17.145
11	0	17.145
12	0	16.633
13	0	16.029
14	0	15.964
15	0	15.516
16	0	14.723
17	0	14.532
18	0	14.100
19	0	9.082
20	0	2.608
21	0	1.055
22	0	0.562
23	0	0.096
24	0	-0.749
25	0	-2.878
26	0	-5.196
27	0	-5.866
28	0	-6.413
29	0	-6.891
30	0	-6.909

由上表可知，當顧客編號 5662676 面對考慮集中共 30 項產品，我們可以

按照每樣產品對顧客而言的相對效用值，由大至小排列後逐一進行產品的推薦。於是便先從第一樣相對效用值最高的產品開始推薦，但是顧客並沒有購買，第二樣相對效用值次高的產品也沒有命中顧客的需要，如此照順序推薦下去，直到推薦至第 10 樣產品時，顧客即發生購買行為，則謂之命中。

相同地，運用考慮集合中產品的各個變數與迴歸參數的估計值，把每一位顧客的相對效用值排序後，便可得到預測的購買機率，而進行產品的推薦。藉由這種方式，我們可以看出對每一位顧客的推薦成效，也可了解對不同的顧客，應該要推薦幾項產品才會真正令其購買。將所有顧客之推薦順位列出後，可以計算總合邏吉斯模式累積擊中的個數與擊中率，遂將其推薦的成功機率高低列表如下表所示：

表 4-3-3 總合邏吉斯回歸模式的舊顧客推薦累積擊中率

推薦個數	累積擊中個數	顧客總數	累積擊中率
1	24	987	2.432%
2	24	987	2.432%
3	41	987	4.154%
4	62	987	6.282%
5	118	987	11.955%
6	148	987	14.995%
7	168	987	17.021%
8	245	987	24.823%
9	252	987	25.532%
10	279	987	27.964%

解讀上表可發現，在對舊顧客的推薦時，將推薦相對效用值排序後第一順位的產品，僅只能擊中約 2.4% 的購買，且當推薦的產品繼續增加，第二個、第三個直到第五個，能提高擊中顧客購買產品的成功率相當有限，但是在稍後推薦的產品，也就是當推薦個數若累積至第六、第七和第八項產品，擊中顧客欲購買的書籍比例遂逐步提高，若是增加推薦個數至第十個，則累積下來能夠命中顧客購買之機率可提升至 28%。

(二) 對新顧客的推薦

如前所述，利用總合邏吉斯模式對新顧客的推薦方式，也是先將樣本外所保留的新顧客挑選出來，並開列其產品考慮集合，將迴歸式中所估計得到的參數係數值帶入，可求得每位顧客對每項產品的相對效用值，而後加以排序以進行推薦，整理推薦的擊中個數與擊中率如下表：

表 4-3-4 總合邏吉斯回歸模式的新顧客推薦累積擊中率

推薦個數	累積擊中個數	顧客總數	累積擊中率
1	21	569	3.691%
2	34	569	5.975%
3	38	569	6.678%
4	54	569	9.490%
5	55	569	9.666%
6	58	569	10.193%
7	62	569	10.896%
8	64	569	11.248%
9	69	569	12.127%
10	73	569	12.830%

解讀上表對新顧客的推薦擊中率，得到的結果與對舊顧客推薦時相當不一樣。首先是在推薦個數為一個時，模式命中新顧客購買的機率為 3.7%，這是較好的擊中率。在針對新顧客推薦其相對效用最高的產品時，較推薦相對效用最高的產品給予舊顧客的成功機率為高；其次，進一步再推薦相對效用值次高的產品，累積命中率可達到 6%，若再順勢往下推薦第三個、第四個至第十個產品時，其購買機率的預測準確度提升都不是太高，也就是說這數個推薦順位的產品幾乎無法猜中顧客的需求，直到前十個產品被推薦給顧客時，累積擊中率才能向上推進至 12.8%。

(三) 小結

比較運用平均機率模式時可得到的累積擊中率，與運用總合邏吉斯模式的擊中率，如下頁圖 4-3-1 與圖 4-3-2 所示：

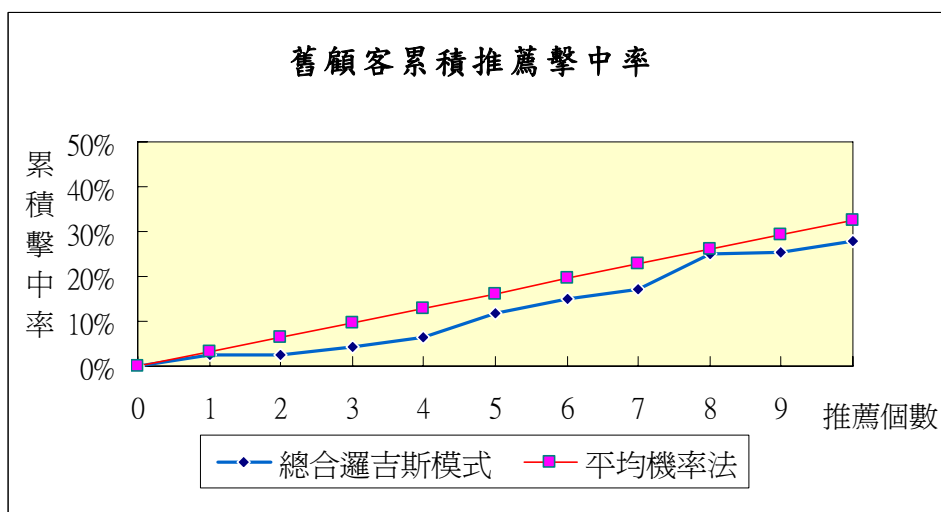


圖 4-3-1 舊顧客累積推薦擊中率比較 (平均機率法與綜合邏吉斯模式)

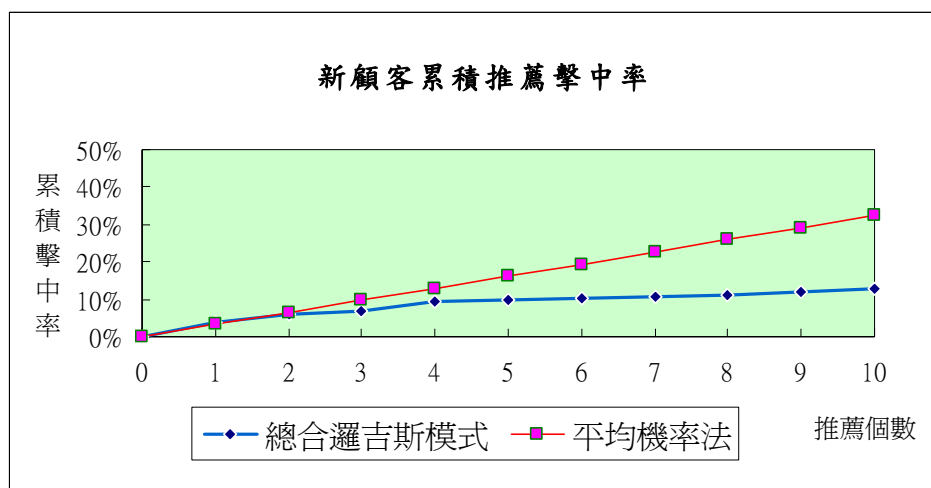


圖 4-3-2 新顧客累積推薦擊中率比較 (平均機率法與綜合邏吉斯模式)

由上圖可知，不論是在推薦新、舊顧客時，綜合邏吉斯模式的預測效果和平均機率模式的預測結果相比，都略顯要差，尤其是在針對新顧客作推薦時，其結果更是不如一般平均模式，這與原先本研究所預定的假設有所出入。本研究第三章的研究方法模式推導以及過去眾多學者的文獻皆相信，採用綜合邏吉斯模式來衡量顧客購買產品的機率，是加入了產品的屬性對購買機率的影響，換句話說也就是在進行產品推銷時，不再是只有隨機推薦，還會加入對各種顧客的基本知識，了解可能會提高顧客購買的產品屬性為何，而後推薦之。

比較應用綜合邏吉斯推薦模式，在針對舊顧客及新顧客的推薦時，發現兩者

的結果有些微差異，即便舊顧客所透露出來的資訊較新顧客為多（包含其過去的購買紀錄與偏好），但二者在此模式中，預測購買機率的主要變數來源，都僅只是以現階段每位顧客其對不同書籍的屬性所反應出來的行為，與其他個人特質或過去消費行為較無著墨。也就是說在總合邏吉斯推薦模式中，依舊是在視所有顧客為同質的前提下發展，而這就極有可能是導致新顧客與舊顧客的保留樣本，在接受相同的迴歸模式計算後，一樣順著相對效用值推薦產品，擊中率卻有落差的原因——忽略了顧客的異質性。

現在顧客多擁有一個別獨立的需求，資料庫行銷的觀念也強調要重視顧客單一個體的偏好，特別是在電子商務、網路商店的經營，越來越偏向個人化服務時，推薦機制不應只考量同質需求，更應加入考慮個體異質的部分，影響顧客對書籍各種屬性的偏好結構，會更貼近真實情況。因此在下一節的顧客推薦系統便導入層級貝氏統計理論的意涵，利用層級貝氏邏吉斯模式去針對每一位獨一無二的顧客進行新產品的推薦。



第四節 建立線上推薦系統

一、 基本概念與模型

在本節所利用的模式是結合層級貝氏之統計理論，以及邏吉斯迴歸分析法而成的層級貝氏邏吉斯推薦系統。主要的觀念在於重述每位顧客的異質需求，會表現在其購買行為上，因此在研究推論各式購買機率高低的現象時，必須儘可能將顧客的個體層次與跨群體的總合層次合併解釋，以期針對每一獨立個體特有的偏好，加以預測其絕對個人的購買機率，進而決定產品的推薦項目及次序。模式的設定及推導詳見第三章研究方法。

二、 參數估計

在層級貝氏統計理論當中都已將每位顧客視為異質，因此在進行迴歸分析時，每一位獨立的個體也都會得到一組自己的迴歸係數，是與他人都不相同的。延續 Peter Lenk 在估計層級貝氏邏吉斯模式 (HB Logit Model) 的參數時，其所撰寫的程式碼，稍作修改後應用在本研究，並利用 Gauss 軟體分別進行 50000 次、75000 次、100000 次、120000 次及 150000 次等模擬與比較後，所得到的最佳後驗參數估計值，茲將此參數組合整理如下頁表 4-4-1 所示。

三、 模式準確性

和前一節的做法相同，在得出了該迴歸係數的組合之後，便開始將每位顧客的資料帶入保留樣本，讓每位顧客的考慮集中所有產品都能得到一組相對效用值，進而將此相對效用值由大至小排列，排序完成後便可以決定產品推薦的順序，且了解其推薦成功的機率。以下便分別整理此模式對舊顧客及新顧客的推薦累積擊中率。

表 4-4-1 利用 Gauss 所估計出的最佳後驗參數值

B 屬性水準		顧客編號					
		5662676	5669738	5689812	...	8812747	8815390
Constant		-5.854	-3.588	-8.056	...	-6.665	-7.789
標示 價格	低於\$10 元	-1.596	-3.043	-3.104	...	-0.642	-2.319
	\$10-\$20 元	-0.500	-1.747	-0.567	...	-0.391	-0.733
裝訂 方式	紙本	2.891	2.750	4.013	...	3.247	3.409
	精裝本	0.530	0.827	0.372	...	1.964	1.376
頁數	少於 300 頁	0.532	-0.839	3.102	...	0.737	2.762
	300-500 頁	-0.049	-0.375	0.272	...	1.733	1.143
顧客 評價	0-3 顆星	-2.953	-4.520	-2.714	...	-4.094	-2.413
	4-7 顆星	-1.874	-2.883	-3.236	...	-1.381	-1.545
回應 文章	少於 10 篇	-0.408	-0.153	2.951	...	-1.298	1.195
	10-50 篇	-0.656	-0.361	0.575	...	-1.495	-1.004
銷售 排名	前十萬名	0.032	0.447	2.402	...	-0.187	0.864
	十萬至五十萬名	-0.157	-0.544	-0.649	...	1.035	1.300
書籍 種類	商業理財	0.975	-0.158	-1.456	...	1.138	-0.392
	文學小說	3.392	2.397	3.972	...	4.525	3.747
	人文藝術	4.344	3.674	3.379	...	3.741	3.730
	普通科學	1.877	0.648	1.904	...	0.741	1.563
	教科書籍	1.080	2.277	2.535	...	1.912	2.079
	心靈養生	3.074	0.816	0.018	...	2.882	2.069
	生活風格	1.697	0.511	2.312	...	1.783	2.8450

(一) 對舊顧客的推薦

首先必須將保留樣本中所有的舊顧客，其個人針對各書籍屬性的係數值，帶

入所屬的產品考慮集合中，包含一筆真實有購買之書籍的資訊，並可得出每項產品的相對效用值，經排序後可得到結果如上節表 4-3-2，而後即計算所有舊顧客之預測購買機率，及此推薦模式的累積擊中率。整理如下表所示：

表 4-4-2 層級貝氏邏吉斯推薦模式對舊顧客的累積擊中率

推薦個數	累積擊中個數	顧客總數	累積擊中率
1	33	987	3.344%
2	63	987	6.383%
3	104	987	10.537%
4	129	987	13.070%
5	166	987	16.819%
6	194	987	19.656%
7	231	987	23.404%
8	268	987	27.153%
9	298	987	30.193%
10	334	987	33.840%

若以層級貝氏邏吉斯的模式帶入每位顧客過去的購買紀錄等資料，可以更個人化地計算出每位顧客自己的一組偏好值，如上列作法所敘述，此時便可以針對每位異質需求的顧客，提供其獨有的推薦書籍順序。由上表可發現在推薦產品給舊顧客時，推薦第一項產品雖僅能命中約 3.3% 的顧客，但比較起平均機率法則及總合邏吉斯模式而言，確實是有稍高一些，再往下繼續推薦產品時，可發現擊中率是逐步在攀升中，直到推薦至第十項產品時，已能累積約 34% 的產品擊中率了，雖然本研究預期在推薦產品給舊顧客的擊中率部分是應該要有更好的表現，不過這樣的表現確實也差強人意，能在邏輯上映證加入顧客異質性的考量，是有助於廠商在向舊顧客推薦新產品時，提高其購買的機率。

(二) 對新顧客的推薦

由於保留樣本中對新顧客樣本的設定為沒有過去其他交易資料，因此在推斷新顧客購買機率時，就沒有辦法得到絕對個人的參數估計值，此時就必須利用模

式當初在模擬的過程，所得到的 Θ 後驗估計平均值，與新顧客的個人特質統計資料相乘，以便得到新顧客的迴歸係數值，進而才能計算其購買機率，並給予適當的推薦品。

下二頁的表 4-4-3 與 4-4-4 即為利用 gauss 軟體模擬上萬次後得到的 Θ 後驗估計平均值。

從這兩張表中我們可以看到幾個現象說明如下：

1. 書籍的標示價格變數

家庭人口數在 2 人以下，喜愛購買低價書籍勝過中價位（購買低於\$10 元的機率較大），又家庭人數為 3-5 人者亦是，且購買偏好更明顯（ $1.4379 > 0.8971$ ），居住在美國中部的顧客也透露出偏好低價位書籍的訊息；相反的，教育水平為高中職以下者，最不會在線上購買價格低於\$10 元的書籍，換句話說，其所偏好的書籍價格是中價位（介於\$10-\$20 元間）勝過低價位。

2. 書籍裝訂方式

高中職以下學歷的顧客顯著傾向購買紙本的書籍勝過精裝本（ $1.7239 > -0.6103$ ）。此外，若由顧客年齡來觀察其對此變數的反應，可發現消費者介於 18 至 39 歲者，購買精裝本的書籍意願較購買紙本為低。

3. 書籍頁數變數

在書籍頁數的多寡上，大部分顧客表現出的態度皆無明顯的偏好，僅只有教育背景是高中職以下的顧客表示，頁數在 300-500 頁間的書籍，其購買的意願頗低（平均不購買的機率比購買的機率還大），頁數少於 300 頁者購買意願也不高，相較之下頁數在 500 頁以上似乎較受歡迎。又家庭人口數在 2 人以下者，對較薄的書籍之好感度稍低些。

4. 顧客評等變數

顧客的學歷為高中職以下者，明顯由購買機率的高低可看出所偏好的書籍評價，對只有 0-3 顆星評等的書籍表示購買機率最小，其次為 4-7 顆星，也就是唯有 7 顆星等以上的書籍，購買的可能性才存在；大學畢業者，對書籍的評等變數

也有些相似的考量，獲得顧客 0-3 顆星等的書籍，購買的機率比較低，但是在選購評價為 4-7 顆星等的書籍部份，則消費行為便剛好相反；反之，對家庭人口數是 3-5 人的顧客而言，購買獲得顧客評價低於 3 顆星的書籍與購買評價介於 4-7 顆星的書籍，二者成交的機率皆大於不買的機率（前者的購買機率為 0.6229, 後者為 0.692），這是相當不一樣且有趣的消費偏好。

5. 顧客回應的文章數量

由該數據顯示，教育背景為高中職以下的顧客，強烈傾向購買回應文章數量可能完全沒有或是 10 則以內的書籍 ($0.9437 > 0.248$)；正好與家庭人數在 2 人以下的顧客持相反的意見，其對顧客回應文章數量較少的書籍較不感興趣，特別是擁有不到 10 篇回文的書籍，傾向購買有多篇討論文章的書籍。

6. 該書籍的銷售排名

在書籍銷售排名的部分，顧客的教育背景若是在高中職以下者，較不喜歡非暢銷類的書籍，也就是購買銷售排名為第 10 萬名至第 50 萬名之間的書籍之意願明顯較低；但是相反的，家庭人數在 2 人以下及 3-5 人之間者，則呈現一致的偏好反應，傾向購買銷售排名為中等的書籍，而較不喜愛銷售排名在前十萬名的暢銷書籍。

7. 書籍種類變數

在書籍種類變數的部分，每位具有不同特質的顧客所偏好的書籍種類可說是相當分歧，在此僅列舉數個較明顯的偏好：如學歷背景為高中職以下的顧客，明顯喜愛購買商業理財與人文藝術等類別，卻較不偏好購買心靈養生或生活風格類別的書籍；而介於 18-29 歲的年輕顧客，除了文學小說外，幾乎各類別書籍都有表示購買偏好，包含高度對商業理財型態與教科書籍等類有興趣。

表 4-4-3 利用 Gauss 所求得的 Θ 後驗估計平均值

	常數項	常數	標示價格		裝訂方式		頁數		顧客評價		回應文章	
			<\$10	\$10-\$20	紙本	精裝本	<300	300-500	0-3	4-7	<10	10-50
	常數項	-6.2035	-2.9925	-1.1419	2.8654	1.2167	1.1648	1.0031	-3.1692	-2.7735	0.7101	-0.4680
教育程度	高中職以下	-0.7609	-0.7113	0.1417	1.7239	-0.6103	-0.8000	-2.1077	-0.8339	-0.0335	0.9437	0.2480
	大學肄業	0.0753	0.2274	0.0872	0.5712	0.1229	-0.2108	-0.5369	-0.5386	-0.0405	-0.1875	0.0708
	大學畢業	0.1046	-0.0625	0.1092	0.2028	0.0230	-0.0129	-0.2620	-0.7145	0.3478	-0.2218	-0.063
平均收入	\$24999 以下	0.3127	-0.4636	-0.3422	-0.5785	-0.428	0.3992	0.1787	-0.2886	-0.3616	0.0344	0.3305
	\$25000-\$49999	0.4988	-0.3211	-0.4252	-0.2713	-0.1501	0.1722	0.2172	-0.2904	0.3059	-0.2573	0.0652
	\$50000-\$74999	0.1560	0.0323	-0.2051	-0.3658	-0.3179	0.3911	0.4952	0.0926	0.0163	-0.1175	0.1871
年齡	18-29 歲	-0.0926	-0.1910	0.1244	-0.1285	-0.7896	0.2042	0.1916	0.3732	-0.0154	-0.0706	-0.1210
	30-39 歲	-0.3926	-0.2063	0.3735	-0.1893	-0.7085	0.5200	0.3342	0.1162	0.2295	0.3978	-0.1671
	40-49 歲	-0.1752	0.0028	0.0931	0.1591	-0.1607	0.1474	0.3722	-0.1129	-0.1640	-0.0093	-0.1524
	50-59 歲	-0.7871	-0.2699	-0.1896	0.4328	-0.0620	0.3771	0.2544	0.4421	-0.0792	0.6863	-0.0916
家庭人數	2 人或以下	0.2170	1.2496	0.8635	-0.5322	-0.0872	-0.8922	-0.2785	0.2000	0.6902	-0.8182	-0.0288
	3-5 人	-0.6133	1.4379	0.8971	-0.1171	0.2771	-0.3447	-0.2016	0.6229	0.6920	-0.6400	-0.6043
所在地區	美國東北部	0.1879	0.4087	0.2377	0.1650	0.0652	-0.1773	-0.1114	-0.1132	-0.2361	-0.5557	-0.0456
	美國中部	-0.3115	0.7040	0.6586	0.0245	-0.3746	0.1314	-0.4449	0.0625	-0.2685	0.2890	0.3553
	美國南部	0.0436	0.2234	0.1570	0.1753	0.3448	-0.2438	-0.3272	0.0002	-0.0900	0.1769	0.0651
孩童	正在扶養中	0.7667	-0.2628	-0.1439	-0.3479	-0.6310	-0.2625	-0.6229	-0.0291	-0.2509	0.0194	0.5366

表 4-4-4 利用 Gauss 所求得的 Θ 後驗估計平均值 (續)

		銷售排名		書籍種類						
		前十萬名	第十萬至 第五十萬名	商業理財	文學小說	人文藝術	普通科學	教科書籍	心靈養生	生活風格
教育程度	常數項	1.2965	0.0594	0.4910	2.4876	3.5696	0.7815	2.8348	1.8457	1.9210
	高中職以下	-0.1202	-1.5238	1.1715	0.5307	0.9965	-0.0785	-0.2503	-1.0847	-2.0379
	大學肄業	-0.2423	-0.2233	-0.3735	0.2446	-0.3555	0.2812	-0.6415	-0.6448	-0.3111
	大學畢業	-0.2062	-0.0995	-0.3183	0.3555	0.0308	0.0559	-0.1415	-0.6589	0.0122
平均收入	\$24999 以下	0.0644	0.6393	-1.0053	-0.1692	-0.1646	-0.1818	0.1973	0.1172	-0.3693
	\$25000-\$49999	-0.1481	0.5673	-0.8578	-0.1482	-0.3475	-0.2020	0.0412	-0.0288	-0.1341
	\$50000-\$74999	0.1258	0.6240	-0.8041	-0.0645	-0.8651	0.3009	0.1795	0.0117	-0.4454
年齡	18-29 歲	-0.0865	-0.5300	1.0141	-0.3265	0.4188	0.0420	0.7373	0.6324	0.2053
	30-39 歲	0.4506	-0.1987	0.7368	-0.4274	-0.213	0.5454	0.1588	-0.2466	0.4864
	40-49 歲	-0.0450	-0.0233	0.3045	-0.1033	0.1575	-0.0348	0.2044	0.3853	0.4922
	50-59 歲	0.3240	0.0273	0.2412	0.0910	0.3592	-0.0745	0.2169	0.0842	0.2319
家庭人數	2 人或以下	-0.8353	0.6736	-0.0940	0.6610	0.6401	0.6594	-1.0876	1.4071	0.5000
	3-5 人	-0.8829	0.6900	0.7079	0.6615	0.7681	0.4540	-1.1947	1.3322	0.5444
所在地區	美國東北部	-0.1458	-0.2923	-0.4016	0.2517	-0.2956	0.0956	-0.5743	0.2080	0.3349
	美國中部	0.1697	-0.5525	0.2763	0.0463	-0.5493	0.4677	-0.3844	0.0449	0.2144
	美國南部	0.0361	-0.5354	0.3865	0.0089	0.0002	-0.1498	-0.3668	-0.2487	-0.0732
孩童	正在扶養中	0.0606	-0.1112	-0.9774	0.2400	0.2520	0.5855	-0.2005	-0.1993	-0.4594

其次，需求明顯較為不同的顧客，如家庭人數僅只有 2 人或以下的顧客，對心靈養生類別的書籍顯示強烈的偏好，又如家庭人數介於 3-5 人之間者，對於教科書以外的其他類別書籍，都表示較高的購買意願：對於心靈養生類便透露濃厚的購買興趣，對於人文藝術和商業理財等類別也表示不錯的購買意願。有趣的是，當顧客家中有正在扶養的孩童，其對商業理財類別的書籍就沒有太大的購買意願，且當考慮到平均年收入對於購書時的影響，便可發現平均年收入越低的族群，越排斥購買商業理財類型的書籍。

接著便討論新顧客的偏好結構，只需將新顧客的個人特質之統計資料，在本研究中設定為年齡、年平均所得與家庭狀況等，和上表中的後驗參數估計平均值相乘後加總，即可得出每位顧客所獨有的一組迴歸係數值，然後便可用相同的預測方式，將新顧客的購買機率與推薦品的累積擊中率算出，以下列表說明：

表 4-4-5 新顧客的偏好係數值

5661417	(0,1,0,0,0,1,0,1,0,0,1,0,0,1,0,0)
	(-6.4593,-0.9855,0.6358,2.3738,-0.1486,1.1043,0.5722,-3.2365,-2.1465,0.2737,-0.0507,0.965,0.3825,0.2325,2.9476,2.2268,3.0361,1.0596,2.418,2.3653)
5672296	(0,0,1,1,0,0,0,0,0,0,1,0,1,0,0)
	(-6.711,-1.3767,0.1808,2.3971,0.7142,1.3378,0.2733,-3.4869,-2.3638,0.1717,-0.4495,0.4415,0.7367,0.1516,3.3817,3.6546,1.5773,1.3115,2.6811,2.3227)
...	...
8812939	(0,0,1,0,0,1,0,0,0,0,0,1,0,1,0,0)
	(-6.8677,-0.8808,0.3179,2.6098,0.8243,1.3297,0.5898,-3.1057,-1.9859,0.0198,-0.5929,0.5029,0.7214,0.3528,3.4864,2.9541,2.06,1.2937,2.5756,2.2466)

上表即為每位新顧客獨有的偏好迴歸係數值，本研究根據其人口特質的不同而合理推算出其線上購買書籍的屬性偏好，依據此偏好進行新產品的推薦過程，成功使顧客購買的機率表示如下表：

表 4-4-6 層級貝氏邏吉斯推薦模式對新顧客的累積擊中率

推薦個數	累積擊中個數	顧客總數	累積擊中率
1	17	569	2.988%
2	36	569	6.327%
3	58	569	10.193%
4	77	569	13.533%
5	92	569	16.169%
6	115	569	20.211%
7	137	569	24.077%
8	152	569	26.714%
9	167	569	29.350%
10	188	569	33.040%

由上表可知，藉由層級貝氏邏吉斯模式來推薦新產品給新顧客的時候，會依據不同顧客的人口特質，而給予其對不同產品屬性喜好的權重，並且從中挑選出能滿足其需求的書籍，依據購買機率的高低排序來作為推薦的順序。首先看到推薦第一項產品，便能使新顧客完成購買的機率約為 3%，略低於一般平均機率法能達到的擊中率，不過當推薦品持續累積至第第二個、第三個時，應用此推薦模式的擊中率便逐步優於平均機率法以及總合邏吉斯迴歸模式，直到推薦的書籍多達十個時，其擊中率明顯較其他二種模式來得高。

一、 小結

綜合上述層級貝氏邏吉斯推薦模式的結果，可以發現以此模式進行推薦時，對新顧客的累積擊中率比對舊顧客的推薦成功機率來得稍低，這就理論而言是相當合理的，因為擁有顧客過去的交易紀錄時，則較容易掌握顧客的偏好與購買習性，於是進行預測時也可以較為精準；而對新顧客而言，過去既然沒有消費的紀錄，僅只能從會員註冊的資料得知其個人特質的部分訊息，並且需借助模式內總體層次的後驗估計值，以跨群體的方式得到每個人口統計變數對產品屬性的係數，再求出新顧客的購買機率，換言之也就是利用總體而非個人的偏好結果來進行預測，理當得到較差的推薦擊中率。接著比較三種不同推薦模式的累積擊中

率如下圖 4-4-1 與 4-4-2 所示：

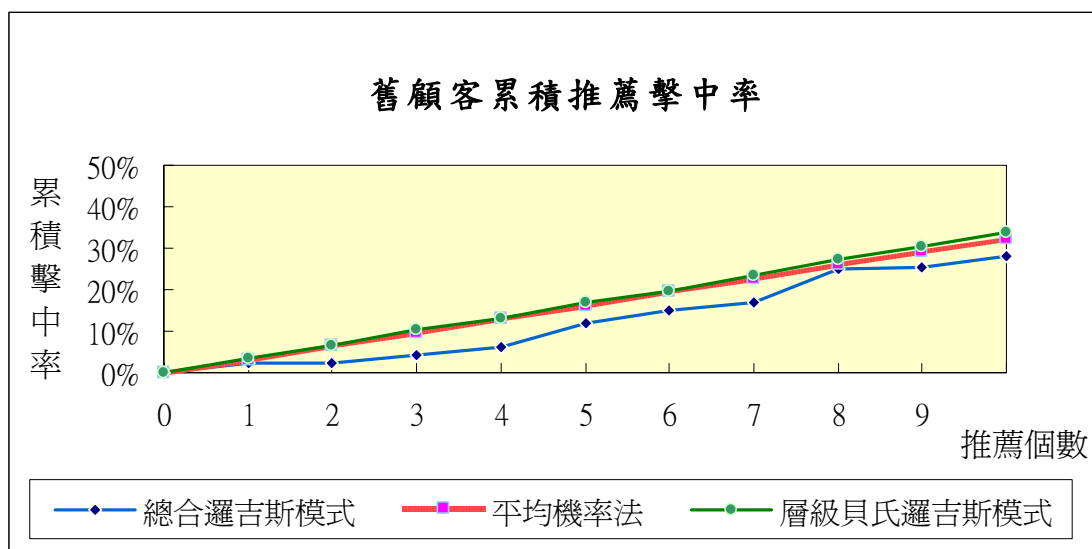


圖 4-4-1 舊顧客累積推薦擊中率比較
(平均機率法、總合邏吉斯模式與層級貝氏邏吉斯回歸模式)

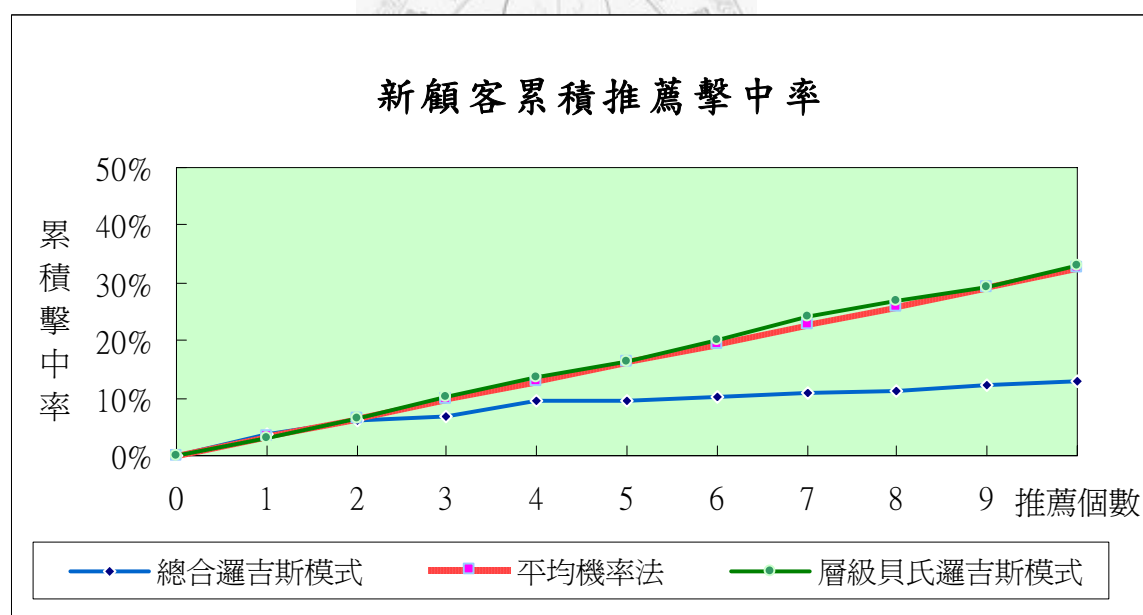


圖 4-4-2 新顧客累積推薦擊中率比較
(平均機率法、總合邏吉斯模式與層級貝氏邏吉斯回歸模式)

本研究的理論結果應為層級貝氏邏吉斯迴歸模式的推薦擊中率，明顯優於總合邏吉斯迴歸模式；而總合邏吉斯迴歸模式則要較一般平均機率法的擊中度高。但是在前述的模式驗證結果中，實證結果發現與第三章研究方法所推斷的結

論有些許出入。在這三個模式當中，確實是以層級貝氏邏吉斯推薦模式的累積擊中率表現較佳，不論是針對舊顧客的推薦或新顧客皆是如此，但遺憾的是在本次的研究中，並無法得到非常顯著的差異，來印證層級貝氏推薦模式以個體層次的偏好，可更加有效率的解釋總體層次未觀察到的異質性，不過若就此研究背後的貢獻而言，實證結果也的確有效的彰顯出顧客在線上購書時，其行為表徵以及消費型態，這些觀察可以幫助企業更加了解其顧客的形貌，可作為日後其目標篩選顧客，或擬定行銷策略時的參考。

此外，針對以層級貝氏邏吉斯模式來推薦新產品給新、舊顧客的擊中率不如預期理想一事，合理的解釋應該是當預估新顧客時所用的資料為總體層次，只能以總體顧客的偏好來作為推測其個別偏好的基準，難免會遺失掉個體獨立的需求，因此在推論其購買的可能性就容易產生偏差，不過就概念上而言，層級貝氏的模式是額外加入考慮顧客在人口特質變項，與教育背景等的差異，這也的確能夠帶出更多顧客異質需求的表徵，而使得進行新產品的推薦時，獲得比僅只考慮產品屬性的總合邏吉斯模式還要好的結果。

針對舊顧客的部分，在這個 2004 年整年度的交易紀錄資料庫中，我們所持有的舊顧客交易的資訊中，每位顧客個別的真实交易資料量有限，個體層次過去曾購買的次數每人平均僅有約 4.3 次，記錄稍嫌過少，以至於在估計個別舊顧客的偏好時，容易出現誤差；本研究為提高前述解釋之合理性，於是檢視各個顧客過去的交易紀錄量，並觀察是否會影響到本研究在估計個體偏好，以及預測舊顧客個別的購買傾向，而導致推薦新產品時的累積擊中率有所不同，結果如下表 4-4-7 所示：

表 4-4-7 顧客交易紀錄筆數對累積推薦擊中率的影響

個別顧客當年的 交易紀錄筆數	顧客編號	推薦品擊 中的順位	推薦前十項產品 便擊中的總個數	平均前十項推薦 品可擊中的比例
1~4	5662676	6	162	0.2275 (162/712)
	5669738	17		
		
5~6	8785298	11	66	0.6346 (66/104)
	5749487	3		
	5780231	13		
7~8	36	0.6207 (36/58)
	8799331	16		
	5880540	15		
9~10 及以上	6450668	3	70	0.6195 (70/113)
		
	8783670	5		
	5879080	4		
	6421590	9		
		
	8802565	2		

從上頁表可看出，當每位顧客的資料平均能增加至 5 筆時，在累積前十項推薦品的擊中率就有顯著的提升，利用新產品推薦模式來預測其購買的準度可由 21%、28% 增加至 63%、62% 並維持在這個機率的水準。經由合理的推論致使本研究結果稍差之原因後，發現層級貝氏推薦模式的預測能力，是可以藉由資料的累積來逐步提高成功命中購買的可能性，因此，企業不妨朝著推薦系統的方向去努力，在資料庫建立的較完整之後，推薦效果的強度就會更明顯。



第五章 結論與建議

第一節 研究發現

一、 總合邏輯迴歸模式結論

由總合邏輯迴歸模式所估計出的迴歸係數，可發現顧客對各書籍的屬性水準，會影響其線上購買的程度如下所示：

(一) 當顧客在考慮\$10元以下的書籍與介於\$10-\$20元之間者時，研究結果發現顧客購買偏高價位的可能性較大，且結果亦顯示顧客較喜歡於線上購買紙本的書籍，勝過精裝版本或其他裝訂的方式。此外，顧客在購書時，書籍的厚度確實是會影響購買的決定，惟顧客對於書籍的頁數多寡，其各自的偏好影響購買的機率甚大，不論頁數是少於300頁或是介於300-500頁之間者，都有顧客會被吸引，而頁數偏多者又更容易引起購買。

(二) 由研究發現顧客在線上購書時，多半會受到其他顧客對該書籍的評價而影響購買的決定，當書籍被評為0-3顆星的等級時，顧客的購買意願明顯較低，而當書籍被評為4-7顆星時，顧客的購買意願便提升，由本研究之資料庫內購買的書籍比例顯示，約7成4的顧客所選擇之書籍是具有高度評價的。

另一方面關於顧客回應文章數目，顧客在選擇時傾向購買少於10篇回應文章的書籍；過去曾有文獻提及較多的回應文章數量，真的能夠提高銷售，但是從係數來看好像不盡然是這樣的結果。所以推論應該要檢視討論文章的內容來做銷售判斷，會較為合理，畢竟關於回應文章的討論內容究竟是美言還是批評，實在是需視每一本書籍而定，而本研究因在這方面的資料整理不易，所以著墨較少。

(三) 書籍本身在亞馬遜網站上的銷售排名是顧客購買時考量的依據之一，排名為前十萬名的書籍能夠正面影響顧客購買，其次為排名介於10萬至50萬名者。另在顧客購買的書籍種類部分，文學小說類別最能刺激網路購書的客群，其次人文藝術類別也是顧客產生線上購書行為的大宗，生活風格類型的書籍，在吸

引顧客購買的第三位；至於教科書籍也是顧客再選擇購買時的重點項目之一。

二、以層級貝氏邏吉斯模式分析顧客需求偏好之結論

(一) 教育水準為高中職以下者其所透露出線上購書的偏好，明顯較其他來得強烈且迥異，例如其所偏好的書籍價格是中價位（介於\$10-\$20 元間）為主，且顯著傾向購買紙本的書籍勝過精裝本；在大多數顧客表達出對書籍厚度沒有明顯偏好時，此類區隔的顧客則表示，對中薄厚度的書籍之購買意願頗低，偏好選購 500 頁以上的書籍。

此外，若由購買機率的高低也可判斷出，其對只有 0-3 顆星評等的書籍表示購買機率最小，其次為 4-7 顆星，唯有獲得高度好評的書籍，才有購買的可能性。區隔內的顧客傾向購買可能完全沒有任何討論文章，或是只有 10 則以內的回應文章之書籍，且比較喜歡暢銷類的書籍，也就是購買銷售排名為中等（第 10 萬名至第 50 萬名之間）的書籍之意願明顯較低，偏好購買排名在前十萬名的暢銷書籍。

此類型的顧客在選擇書籍的種類時，明顯喜愛購買商業理財與人文藝術等類別，卻較不偏好購買心靈養生或生活風格類別的書籍，種種特殊的線上購書偏好模式，為此區隔內的顧客添加更鮮明的個性色彩。

(二) 家庭人口數在 2 人以下的顧客也有相當顯著的偏好，例如喜愛購買低價書籍勝過中價位（購買低於\$10 元的機率較大），這也和居住在美國中部的顧客透露出相同消費習性的訊息；此外，家庭人口數在 2 人以下的區隔群，對較薄的書籍好感度稍低些。其對顧客回應文章數量較少的書籍也較不感興趣，特別是擁有不到 10 篇回文的書籍，最無法引起此類型顧客的共鳴，他們傾向購買有多篇討論文章的書籍，並且喜愛銷售排名在中等的書籍，又對心靈養生類別的書籍顯示強烈的偏好，這與教育背景為高中職以下的顧客明顯有所不同。

(三) 消費者介於 18 至 39 歲者，購買精裝本的書籍意願較購買紙本為低。

(四) 大學畢業者，對獲得評價 0-3 顆星等的書籍，購買的機率比較低；反之，

對家庭人口數是 3-5 人的顧客而言，購買獲得顧客評價低於 3 顆星的書籍與購買評價介於 4-7 顆星的書籍，成交機率皆相仿。

(五) 家庭人數為 3-5 人之間者，呈現的偏好是傾向購買銷售排名為中等的書籍，而較不喜愛銷售排名在前十萬名的暢銷書籍。

(六) 年齡介於 18-29 歲的年輕顧客，除了文學小說外，幾乎各類別書籍都有表示購買偏好，包含高度對商業理財型態與教科書籍等類有興趣；而家庭人數介於 3-5 人之間則剛好相反，對於教科書以外的其他類別書籍，都表示有高度興趣：對於心靈養生類及人文藝術和商業理財等類別都有不錯的購買意願。有趣的是，當顧客家中有正在扶養的孩童，其對商業理財類別的書籍就沒有太大的購買意願，且當考慮到平均年收入對於購書時的影響，便可發現平均年收入越低的族群，越排斥購買商業理財類型的書籍。

三、 層級貝氏邏輯迴歸模式推薦結果

(一) 層級貝氏邏輯迴歸模式的推薦結果，雖在整體擊中率的表現上，僅明顯勝出總合邏輯迴歸推薦模式，和平均機率法的推薦結果相比，反而不如預期般顯著，不過本研究仍然有其貢獻之處，也就是幫助解讀消費者購書時的習性，利用資料庫內的交易紀錄去分析顧客的購買行為如上述。

(二) 在利用此模式去推薦新產品給新顧客時，擊中率不如預期理想，合理的解釋應該是當預估新顧客時所用的資料為總體層次，只能以總體顧客的偏好來作為推測其個別偏好的基準，難免會遺失掉個體獨立的需求，因此在推論其購買的可能性就容易產生偏差；而在推薦新產品給舊顧客時，我們檢視其擊中率較低的原因，發現了資料庫內交易紀錄數量會影響其結果的優劣，也就是在我們所持有的舊顧客交易的資訊中，每位顧客個別的真实交易資料量有限，每人平均僅有約 4.3 次的購買紀錄，以至於在估計個別舊顧客的偏好時出現誤差；若當每位顧客的資料平均能增加至 5 筆，在累積前十項推薦品的擊中率就有顯著的提升，預測

其購買的準度可由 21%、28% 增加至 63%、62%，並維持在這個機率的水準。合理的推論後發現層級貝氏推薦模式的預測能力，是可以藉由資料的累積來逐步提高成功命中購買的可能性。



第二節 策略意涵

一、 利用平均機率模式推薦商品時，得到的擊中率是成等差級數遞增，可解釋為當完全沒有顧客任何資訊時，隨機推薦的產品越多，顧客購買的可能性就越高，因此當企業沒有導入資料庫行銷技術時，隨著推薦品累積至第十樣時，命中顧客的購買約可達三成的機率；但若是利用層級貝氏推薦模式來預測購買行為，可以藉由每位顧客自身資料的累積愈多，逐步提高成功命中購買與否的機率，因此，企業不妨朝著推薦系統的方向去努力，在資料庫建立較完整之後，推薦效果的強度就會更明顯，而在深耕顧客的忠誠度方面可著墨更多，並提高交叉銷售與整體盈餘。

二、 考慮推薦產品給顧客時，可視企業策略的方向而定，若欲應用層級貝氏推薦模式則能夠使推薦商品更加顧客個人化，且可提供針對每位顧客不同需求更多樣化的選擇。此系統再建立完成後可大量紀錄顧客的交易偏好，使預測命中率提高，惟初期導入此推薦系統時可先利用過去既有的會員資料之紀錄來做銷售模擬，而後可隨著企業業務量的增大而擴充顧客資料庫規模，此系統僅需作些小設計，門檻不高但長遠財務利潤等效益皆頗被看好。

三、 新產品推薦系統已經越來越被線上零售商當作特殊的武器了，基於其可以減少顧客的搜尋成本，又可以降低顧客對購買新產品時可能會有的不確定性。再者，本研究結果也顯示，利用各種線上推薦系統的模式，最成功的應屬層級貝氏邏吉斯迴歸模式，不論是推薦產品給舊顧客或新顧客，對亞馬遜網路商店而言，只要能夠擊中一次便可以再多增加一筆書籍的銷售，確實是相當值得一用，也無怪乎亞馬遜網路商店會如此致力於研發各式推薦系統，正是因為其所能帶來的銷售增加著實不容小覷，因此當電子商務已經逐漸成為主要銷售管道，或至少是許多實體商店的輔助通路（click-and-brick）時，尚未採行新產品推薦系統機制的企業不妨可以思考一下，往後或許可嘗試應用顧客推薦系統而提高顧客的忠

誠，加速交叉銷售的可能，達到最終使獲利成長的目的。

四、對於資料庫行銷而言，其理論基礎就是來自某種程度的資訊分享，企業藉由與顧客進行偏好的了解、生活資訊的分享，以各式行銷手段來建立彼此的信任；然而，執行的前提是企業要能充分知曉顧客對於資訊提供的敏感度，在部分範圍內是可以被接受，一但超出上下界恐怕就會使顧客出現不安的心理。因此企業在導入資料庫行銷的技術後，必須能確實保證資料絕對安全，不論是硬體設備的防盜功能，或是組織內員工的職業道德，都需要相當標準的控管，以避免顧客在與企業接觸時，體驗到的是負向情緒的消費經驗，反而會招致不良的影響。



第三節 研究限制與未來建議

一、 此研究所用之資料庫是在 2007 年 6 月時取得，而該樣本的時間點是 2004 年，所以這當中已經有許多的變化。對本研究而言，在資料收集與整理上多少會有些窒礙，例如像是書籍的產品屬性資料，這些需要一一由亞馬遜網路書店的網頁中尋獲填入，就會衍生出時間差的問題：常在尋找某一本書籍的時候，發現該書籍的網頁已經被亞馬遜書店移除，於是無法找到最符合當時消費者在購買時的書籍資訊；其次，也曾發生過某一書籍的印刷次數多，而導致無法單從書籍名稱判斷究竟該書籍的購買是第幾版，或是該書的各種屬性資料。

索性藉由顧客消費的日期、及其他相關訊息，可以試著合理推論最符合其購買的書籍，況且書籍的所有屬性中，改變幅度最大的不外乎書籍價格，於是本研究在針對標示價格的處理上，便將書籍價格以區間尺度的方式，劃分出不同水準，可令其網頁隨著時間而改變的衝擊減少。

二、 未來針對這個誤差的改善方式，除了儘可能在該資料記錄的年度內，完成所有訊息收集的動作外，也可建議亞馬遜資料庫在收錄其顧客的交易資訊時，能夠擴充記載的訊息，將所有值得探討的產品屬性變數，皆納入顧客購買資料庫之中，以便未來進行各式研究時，得有最精確的資訊。

三、 單一年度的資料量雖多，但相對亞馬遜網路商店資料庫中所登記的會員人數而言，比例不算高，每位顧客的交易資料筆數平均只有 4 筆，也有不少舊顧客只有一筆資料可供作為樣本內分析用，這對於在檢視顧客之間異質性的需求而言，確實算是有點少，欲審視樣本內的顧客發生過的交易紀錄，以推論其偏好作為未來預測購買機率時，就容易出現偏差，又若每位顧客購買的書籍，其資料結構相當分散時，更會拿不準顧客真正的需求及消費習性，因而導致層級貝氏邏吉斯推薦模式的累積擊中率較弱。

未來可藉由縱斷面的研究，橫跨一段較長的時間再進行分析，讓樣本的資料

數目累積，每位顧客的平均交易紀錄增加，結果應該可以更好。

四、 在本次的研究中，僅只以人口統計變項如年齡、教育程度、收入、家庭狀況及居住地等，作為討論顧客異質性的主要來源。但是在探討顧客線上購書的需求時，若能有其他行為變數納入考量，例如顧客的生活型態、閱讀書籍的頻率，或是可以加入其性別與職業的變項，一同作為偏好異質性的根據，可能更貼近事實，並且可以利用層級貝氏邏輯推薦模式中第二層總合層次，來消弭第一層個人層次的誤差，有效提高模式的預測與解釋能力。

五、 此次研究的對象是由 comScore 市調公司提供的 scanner panel data，其樣本主要範圍是侷限在美國地區，有鑒於亞馬遜網路書店是全球知名，且經營無時空界限的線上零售產業，因此未來若有後續延伸的研究，實可考慮將樣本的來源擴大，不必僅只限制在美國本土，還有歐洲、澳洲及亞洲等地的顧客，可以增加樣本的多元性全貌。



參考資料

一、 中文部分

1. Brian Spengler 著(姜怡如譯)(1999)，1999年度台灣業者之顧客關係管理應用現狀調查，電子化企業經理人報告，11月份，第三期，第9-15頁
2. Peppers & Rogers 著(謝晶瑩譯)(1995)，1：1行銷，台北：時報出版
3. Peter Drucker 著(周文祥，詹文明，江政達譯)(1954)，管理的實務，第一版，台北：中天出版
4. Philip Kotler 著(方世榮譯)(2000)，行銷學原理，第十版，東華書局。
5. 王治平，客觀行為與主觀認知在新產品推薦系統之比較，國立台灣大學國際企業研究所碩士論文，民國92年6月
6. 邵功新，資料庫行銷之客製化新產品推薦系統，國立台灣大學國際企業研究所碩士論文，民國92年6月
7. 陳成業，資料庫行銷之品牌選擇模式之研究，國立台灣大學國際企業研究所碩士論文，民國91年6月
8. 楊昌憲，資料庫行銷之新產品推薦系統：以3C家電業為例，國立台灣大學國際企業研究所碩士論文，民國91年6月

二、 英文部分

1. Allenby, G. M. and Rossi, P. E. (1998), "Marketing Models of Consumer Heterogeneity", *Journal of Econometrics*, 89,57-78
2. Ansari; Essegaiier, S. and Kohli, R. (2000), "Internet Recommendation Systems", *Journal of Marketing Research*, 37(3),363-375
3. Battista, P. and Verhun, D. (2000), "Customer Relationship Management: The Promise and the Reality", *CMA Management*, 74(4), 34-37
4. Berson, A.; Smith, S. and Thearling, K. (1999), *Building data mining applications for CRM*, New York, NY, McGraw-Hill
5. Blattberg, R. and Neslin S.A. (1990), "Sales Promotion-Concepts, Methods, and Strategies," Prentice- Hall Inc. NJ
6. Chevalier, J.A. and Mayzlin, D. (2006), "The Effect of Word of Mouth on Sale: Online Book Reviews", *Journal of Marketing Research*,43(3),345-354
7. Cramer, J.S. (2003). *Logit Models from Economics and Other Fields*, Cambridge University Press, Cambridge .
8. Danaher P.J. ; Mullarkey G.W. and Essegaiier S. (2006), "Factors Affecting Website Visit Duration: A Cross-Domain Analysis", *Journal of Marketing Research*, 43(2),182-194
9. Davids, M. (1999), "How to avoid the 10 biggest mistakes in CRM", The

- Journal of Business Strategy; Nov/Dec 20, 6, 22
10. DeMaris, A. (1992), "*Logit modeling: Practical applications*. Newbury Park," CA: Sage Publications.
 11. Fletcher, K.; Wright, G. and Desai, C. (1996), "The Role of Organizational Factors in the Adoption and Sophistication of Database Marketing in the UK Financial Services Industry", *Journal of Direct Marketing*, 10(1), 10-21.
 12. Gale, Bradley T. (1994), *Managing customer value : creating quality and service that customers can see*, New York : Maxwell Macmillan International
 13. Garfinkel, R.; Gopal, R.; Pathak, B.; Venkatesan, R. and Fang, Y. (2007) , Empirical Analysis of the Business Value of Recommender Systems, working paper
 14. Glazer, R. (1997), "Strategy and Structure in Information-Intensive Markets: The Relationship Between Marketing and IT," *Journal of Market Focused Management*, 2 (1), 65–81.
 15. Gordon, I. (1999). *Relationship Marketing: New Strategies, Techniques and Technologies to Win the Customers You Want and Keep Them Forever*. John Wiley and Sons Publishers, 336.
 16. Grayson, K. (2007), "Friendship Versus Business in Marketing Relationships", *Journal of Marketing*, 71(4), 121-139
 17. Green, P.E.; Srinivasan, V. (1978), "Conjoint Analysis in Consumer Research: Issue and Outlook," *Journal of Consumer Marketing*, 5,103-123
 18. Gupta, S. (1988), "Impact Of Sales Promotions On When, What, and How Much To Buy," *Journal of Marketing Research*, 25 (4), 342-355
 19. Horne, D.A.; Horne, D.R. (2002) , "Database marketing: When does good practice become an invasion of privacy?" *American Marketing Association. Conference Proceedings*; 13, 480-486
 20. Hughes, Arthur M. (2000), "*Strategic Database Marketing-The Master plan for Starting and Managing a Profitable, Customer-Based Marketing Program*." McGraw-Hill Professional Publishing; 2nd Ed.
 21. Jackson, R. and Wang, P. (1994), *Strategic Database Marketing* , NTC Business Book. 11, Lincolnwood
 22. Kalakota, R. (1999), *e-Business: Roadmap for Success* , Addison-Wesley Longman, USA
 23. Kim, B. D. & Kim, S.O. (2001), "A new recommender system to combine content-based and collaborative filtering systems", *Journal of Database Marketing*, 8(3), 244-252
 24. Lenk, P. (2001), *Bayesian Inference and Markov Chain Monte Carlo*, Bayesian Applications and Methods in Marketing Conference and Tutorial

25. Linden, G.; Smith, B. and York J. (2003), “Amazon.com Recommendations: Item-to-Item Collaborative Filtering”, *IEEE Internet Computing*, Jan/Feb, 76-80
26. Montgomery, A. L.; Li, S.; Srinivasan, K.; Liechty, J.C. (2004), “Modeling Online Browsing and Path Analysis Using Clickstream Data”, *Marketing Science*, 23(4), 579-595
27. Morris, T. (1994), “Customer Relationship Management”, *CMA magazine*, 68(7), 22-25
28. Newell, F. (1997), *The New Rules of Marketing: how to use one –on –one relationship marketing to be the leader in your industry*, New York: McGraw-Hill, Inc.
29. Parvatiyar, A. and Sheth, J.N. (2001), “Conceptual Framework of Customer Relationship Management,” in *Customer Relationship Management—Emerging Concepts, Tools and Applications*, eds. New Delhi, India: Tata/McGraw-Hill, 3–25.
30. Payne, A. and Frow, P. (2005), “A Strategic Framework for Customer Relationship Management”, *Journal of Marketing*, 69(4), 167-176.
31. Reichheld, F.F. and Sasser, W. E. Jr. (1990), “Zero Defection: Quality Comes to Services”, *Harvard Business Review*, 68(3), 105-111
32. Schuster, Camille (2005), “Customer Relationship Management Can Work for You, But Is It?”, *Business Credit*, 107(4), 65-66
33. Shani, D. and Chalasani, S. (1993), ”Exploiting Niches Using Relationship Marketing,” *Journal of Business & Industrial Marketing*, 8 (4),58-66
34. Shaw, R. and Stone, M. (1990), *Database Marketing : Strategy and Implement*, New York :John Wiley & Sons
35. Srinivasan, R. and Moorman, C. (2005), “ Strategic Firm Commitments and Rewards for Customer Relationship Management in Online Retailing”, *Journal of Marketing*, 69(4), 193-200.
36. Stone, M. and Woodcock, N. (2001), “Defining CRM and Assessing its Quality,” in *Successful Customer Relationship Marketing*, Brian Foss and Merlin Stone, eds. London: Kogan, 3–20.
37. Winer, R. (2001), “A Framework for Customer Relationship Management”, *California Management Review*, 43(4), 89-105