

應用旅遊業資料建立推薦系統-以日本北海道為例

The Application of Data Mining for Building Personalized Recommender Systems-Base on Hokkaido, Japan

葉惠忠¹

國立高雄應用科技大學 企業管理系 副教授

hcyeh@cc.kuas.edu.tw

蕭竣謙²

國立高雄應用科技大學 企業管理系碩士班 研究生

howlincastle@gmail.com

摘要

現今大部分的人都有旅遊經驗，可是隨著到國外觀光的人變多後，旅行社也越來越多，相對之下要與同業競爭也越來越難，所以只有多加了解顧客需求才能贏得顧客信任。顧客通常在旅遊上最想知道的就是如何安排行程去遊玩，所以要做好推薦顧客的旅遊景點行程就相對重要，旅行社要有效運用先佔優勢來增加客源及利潤，並可更加快速的尋找到更多顧客、消費族群的哪些特徵，提升整體績效是本次的研究重點。

本研究以日本北海道為例下去設計相關行程及問卷，同時也跟專家討論後定出問卷題項的評分分數，並透過資料探勘中的推薦系統的技術來做應用，過程可分為四個階段：第一階段為蒐集資料，第二階段中將針對第一階段蒐集的資料做整理，在第三階段中利用前面所整理出來的資料來做資料探勘應用，來做推薦系統分析，最後第四階段根據問卷及基本資料兩種推薦系統來做比較，期望能幫助旅行社能有效快速的推薦給顧客最適合的行程。

關鍵詞：旅遊業、資料探勘、推薦系統、專家評分

Keywords : Tourism, Data Mining, Recommended System, Expertise Rating,

1. 緒論

1.1 研究背景

近年來鑑於國民所得有所提高，國民對於休閒旅遊的需求也漸漸增加。所以自民國六十八年政府開放出國觀光政策以來國民出國人數持續的成長，在加上周休二日制的實施，國人對於旅遊休閒更加重視，根據交通部觀光局統計年報資料顯示民國九十四年國民出國人次為 8,208,125 人次，到了民國一百零三年時國人出國人次已達到 11,844,635 人次。

而隨著出國人數的增加，提供旅遊服務的旅遊業者也相對的快速成長。由於旅行社數目成長速度極快，同業之間的競爭也越來越激烈，業者必須持續針對不同國家的文化特性、風俗民情與觀光景點安排等大小不同的旅遊路線來做為他們的產品銷售。

旅行業的產品就是透過各項觀光資源的整合與勞務服務所組成之無形商品，利用業者本身優質服務與專業來樹立消費者對於這些無形商品選購的信心，是以「人」為中心的產業類別，而該行業的興盛也可以反映我國人民的生活水準與我國推動觀光產業的成果。

1.2 研究動機

由於大部分旅行社的業務範圍同質性很高，所以旅行業競爭也是相當激烈的，要如何在眾多旅行社中脫穎而出，除了業者本身整合各種觀光資源來組合成一套商品來進行銷售外，更需仰賴人員的服務、銷售與推廣。所以旅行社要先想到顧客的需求是甚麼，往往顧客在旅遊上最想知道的就是哪裡適合安排行程去遊玩，所以做好推薦

顧客的景點行程就相對重要，旅行社要如何利用先佔優勢來增加利潤，將是一個重要的問題，並有效地找尋到更多潛在顧客，消費族群有哪些特徵，如何提升整體績效是本次研究的重點。

所以本研究將利用資料探勘(Data Mining)的相關技術來讓大量的資料能夠轉化成有用的資訊，並藉由設計問卷的方式來蒐集資料並應用於推薦系統(Recommender System)，再依據顧客的偏好、興趣、行為或需求，推薦給旅遊業者所需的潛在資訊、服務或產品 Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., & Riedl, J (2002)。藉由推薦系統的幫助，可得到個人化的資訊服務，減少搜尋資訊的成本，進而增加效率。

1.3 研究目的

一、利用問卷回收後並做資料探勘的分析，挖掘出不同的旅客群體，之後使用推薦系統的功能來找出有用的資訊，以作為旅遊業者在進行組合行程銷售上的參考，同時增加被推薦之行程之能見度。

二、利用整理過後的行程資料庫分數來做資料探勘的分析，找出它們與顧客的關聯性與組合，這裡將使用推薦系統的功能來找出有用的資訊，以作為旅遊業者在進行組合行程銷售上的參考，同時增加被推薦之行程之能見度。

三、利用分群過後的顧客資料來進行推薦系統的設計，並找出最適合顧客的旅遊行程，以提供給旅遊業者有用的參考資料。

四、利用上述之資料探勘方法以及使用推薦系統所得到的資料來應用在對旅遊業，讓業者可以多加了解顧客以及如何做好推薦行銷手法來留住顧客。

1.4 研究流程

整體研究流程如下圖 1-2 所示：

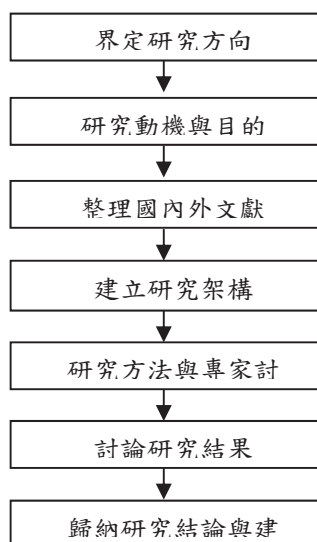


圖 1-2 研究流程圖

2.文獻探討

本章將相關的文獻整理為三節，第一節將介紹旅遊業，說明其旅遊業的背景與發展過程，以及觀光景點的概述；第二節將介紹資料探勘，包含其種類及應用領域；第三節將介紹推薦系統，闡述其種類及應用的範圍；最後再以本研究的觀點來做結尾。

2.1 旅遊業

Ryan (1991) 認為觀光旅遊是對遊客離家期間的住宿與其他服務設施之需求與供給所造成支出、所得產生與就業型態之研究；同時它是人們在假日離家庭停留一個夜晚以上，從事除了國外就學或半永久性就業的其他目的性活動；此外觀光旅遊是人們藉由對新地點的體驗進而追求心理效益的方法。

2.1.1 旅遊業的定義

根據美國旅遊協會(American Society of Travel Agents, 簡稱 ASTA)對旅遊業所下的定義為：「個人或公司使用一些方法以銷售旅遊及相關服務。」Poon (1993)認為旅遊市場是由大眾旅遊轉變而成的，主要是經濟因素，帶起新興旅遊，其特點有個性化設計，細分及彈性。「發展觀光條例」第二條第八項：旅遊業旨為旅客代辦出國及簽證手續或安排觀光旅客旅遊、食宿及提供有關服務而收取報酬之事業。

交通部觀光事業委員會(1996)對旅遊業之定義如下：旅行社為觀光事業之中間媒介，實際推動旅遊事業之主要橋樑。交通部觀光局旅遊業從業人員基礎訓練教材之定義如下：旅遊業係一種為旅遊大眾提供有關旅遊方面服務與便利的行業，其主要業務為憑其所具有旅遊方面之專業知識、經驗所蒐得之旅遊資訊(travel information)，為一般旅遊大眾提供旅遊方面的協助與服務，其範圍包括接受諮詢、提供旅遊資料與建議、代客安排行程與食宿、交通與遊覽活動及提供其他有關之服務。

2.2 資料探勘

2.2.1 資料探勘的定義

資料探勘這個名詞在 Frawley, Piatetsky and Matheus (1991)所提出，他認為「資料探勘 (Data Mining) 是從資料庫挖掘出不明確、前所未知以及潛在有用的資訊過程。」，而另一位學者 Fayyad, Piatetsky-Shapiro (1996)亦對資料探勘的定義提出類似看法，他認為「資料探勘是指由以存在的資料中挖掘出新的事實及發現尚且不知的新關係」，Berry and Linoff (1997)資料探勘是從大量的資料中，利用自動或半自動的方式進行分析，進而挖掘最有意義的關係或法則。

2.2.2 資料探勘的類型

學者 Berry and Linoff (1997)將資料探勘之功能分為下列六項：分類、估計、預測、關聯分組、集群化以及順序，以下將分別針對這六類功能加以定義：

(一) 分類 (Classification)：

所謂分類就是檢視、分析新物件的所有特性，然後將其分配到預先定義好的類別集群中，之後再更新資料以及標上類別編號。而分類的工作就是將每一個群集的特徵清楚定義，並且透過資料分類，建立起模型，將未歸類的原始資料分門別類。較常使用的分類技巧有決策數或類神經網路等。

(二) 推估 (Estimation)：

推估能夠針對連續行數值進行預測，同時將之輸出為連續型的數值顯得較為分析者使用。使用的方法包含統計方法中之相關分析、迴歸分析及類神經網路。

(三) 預測 (Prediction)：

其實預測與分類和推估是相當接近的，依據對象屬性之過去歷史資料建立模型或，檢視新增觀察值之資料，來預測「未來」的數值以及趨勢。其概念為將目前新的數值輸入到此模型中，運算結果就是未來狀態的預測；可再根據某些未來行為的預測來分類，或推估某變數未來可能的值。所需的資料探勘技術包含類神經網路、決策數及統計方法中的時間序列、迴歸分析等。

(四) 關聯分組 (Affinity grouping)：

關聯分組的功能是去發掘哪些事情總是同時發生，用以辨識資料之間的關聯性，而這些關聯通常用來判定哪些事物會一起出現且描述出彼此的關聯性，最典型的案例就是分析超級市場的購物籃，也因此有時候會稱其為購物籃分析(Basket purchases analysis)。

(五) 群集化 (Clustering)：

也稱為同質分組或群集。將許多異質的群體區隔，分成一些同質性較高、更相似的子群組或群集，主要目的要找出組與組之間的差別；與分類不同的是，集群化並沒有根據事先定義明確的類別來進行分類。

(六) 順序/因果 (Sequential Modeling) :

根據既有連續性數值之相關屬性資料，以獲致某一屬性未知之值，這項技術會辨識過去的樣式，如分析客戶過去數次的購物行為。

2.3 推薦系統

推薦系統是一種進行資料訊息過濾之系統，以協助解決消費者在選取眾多的商品或新商品中，精準地且快速地購買到最滿意的商品。推薦系統主要的目的是根據消費者購買偏好來過濾移除無關聯的商品訊息及減少資訊超載和產生不必要的雜訊之問題，並且提供相關性的商品資訊給予消費者。

2.3.1 推薦系統的定義

推薦系統的概念起源於 Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992) 學者等提出合作 (協同) 過濾 (Collaborative Filtering) 後，到 Resnick and Varian (1997) 正式提出推薦系統 (Recommender Systems)，來過濾龐大資訊。推薦系統主要是以合作推薦和內容根據 Rashid et al (2002) 等學者定義推薦系統為用來在複雜及混亂的資訊超載量中協助消費者做決策及決定，根據消費者的輪廓以及了解，針對可能的需要及需求作分析，推薦其可能感興趣的產品或項目。

小結：

過去大部分有關推薦系統的研究都只有使用單一一個推薦系統來做資料探勘的分析，本篇研究將會利用推薦系統加上有專家討論過後的評分標準與問卷來做分析與應用，讓本研究對旅遊業的顧客探勘這塊過程更趨完善。

3. 研究方法

本章節將依據研究動機與目的，做為本研究之規劃、設計與實施的基礎。本研究的方法，總共分成四個部分，以下內容依序為：研究架構、資料蒐集、輔助軟體的介紹，最後是資料探勘的方法與應用。

3.1 研究架構

本研究所要探討之重點在於如何以資料探勘的其中一個應用方法，推薦系統來做分析，這種方法屬於非監督式學習的探勘方法。研究的主要方法是利用問卷設計及專家意見的評分所得到的資料來進行分析，首先把專家評分的資料整理成資料庫後，在將顧客填寫完的問卷與專家討論後所訂出的旅遊行程分數來做關聯性，最後產生出推薦的結果。

根據本研究的研究動機及目的整理出研究架構如圖 3-1 所示，總共分成四個階段：第一階段為蒐集資料，資料的類型有專家意見所整所整的「資料庫資料」、及填寫完後回收的「問卷資料」，第二階段中將針對第一階段蒐集的資料做整理，將這些資料淨化、整合、轉換及換算後，在第三階段中利用前面所整理出來的資料來做資料探勘應用，來做推薦系統分析，第四階段將對分析完的結果進行討論與建議。本研究的整體研究架構如下圖 3-1 所示。

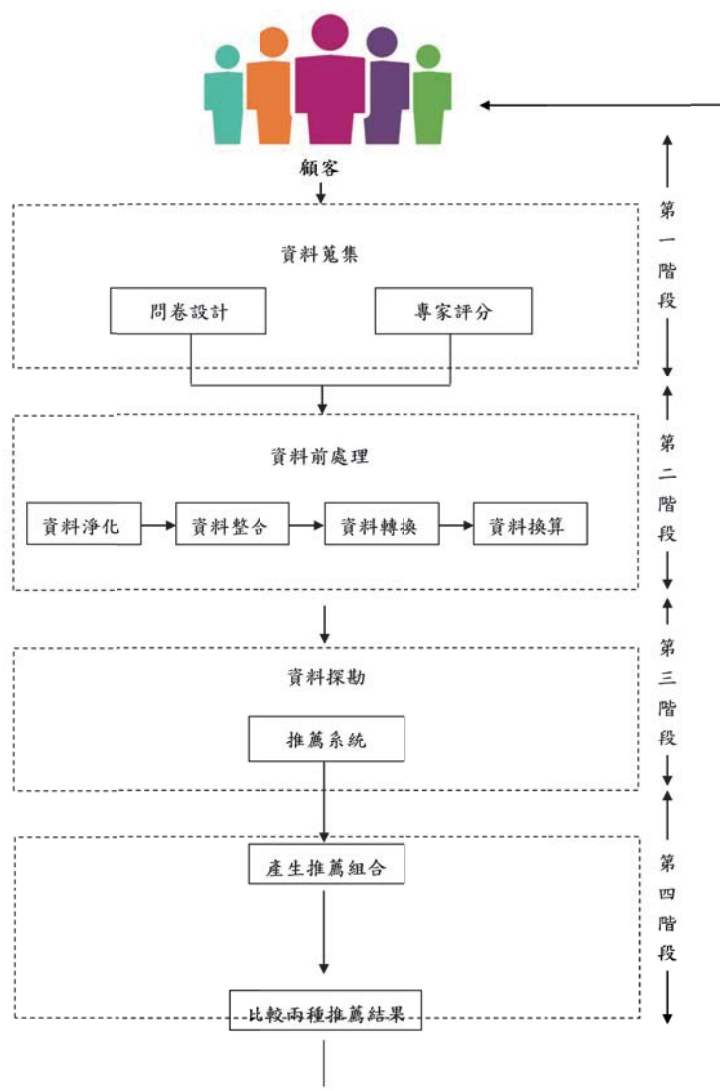


圖 3-1 本研究架構圖

3.2 資料蒐集

3.2.1 問卷設計與發放

(一)問卷設計

本研究主要目的是希望透過問卷調查來幫助旅遊業在對於了解顧客的喜好上更加完善，並參考了一些學者與知名旅行社對於在旅遊上的分類，因此在問卷設計方面會依「顧客期望」與「人口統計變數」來做設計。

1.顧客期望

這部分將以日本北海道為例，根據顧客對於旅遊行程的期望來進行設計，總共有 9 個題項，透過這部分的題目來了解顧客並加以分類。題項包含經驗、時間、假期、天數、預算、旅遊方式、目的、地區等等。題項如下表 3-1 所示。

表 3-1 問卷題項

題項	選項
1. 請問您在過去曾有出國遊玩的經驗嗎?	有(接續填寫)
	沒有(跳至第三題)
2. 請問您在過去曾有出國去日本北海道的經驗嗎?	有
	沒有

表 3-1 問卷題項

題項	選項
3. 請問您希望去日本北海道旅遊的時間是在幾月份?	一~三月
	四~六月
	七~九月
	十~一二月
4. 請問您希望去日本北海道旅遊主要是利用哪種假期?	國定假日
	週末、星期日
	平常日
	寒暑假
5. 請問您希望去日本北海道旅遊的天數是幾天幾夜?	四天(含)以下
	四~六天
	六天(含)以上
6. 請問您希望去日本北海道旅遊每個人的預算金額大約是?	三萬元(含)以下
	三~五萬元
	五~七萬元
	七萬(含)以上
7. 請問您希望去日本北海道旅遊的方式是?	家族成團旅遊
	朋友成團旅遊
	自行規劃行程旅遊
8. 請問您希望去日本北海道旅遊的主要目的是?	生態旅遊
	休閒度假
	心靈放鬆
	探訪親友
	購物
9. 請問您希望去日本北海道的哪些地區遊玩?	自然風景區
	觀光勝地區
	美食地區
	溫泉休憩區
	冰(雪)上活動區
	其它

2. 人口統計變數

在人口統計變數方面總共有 6 個題項。內容包含性別、年齡、工做別、教育程度、婚姻狀況、個人每月平均所得。題項如下表 3-2 所示。

表 3-2 人口統計變項

題項	選項
1. 性別	男
	女
2. 年齡	20 歲(含)以下
	21 歲~30 歲
	31 歲~40 歲
	41 歲~50 歲
	51 歲~60 歲
	60 歲(含)以上
3. 工作別	軍公教
	服務業
	自由業
	金融業
	科技業
	製造業
	學生
其他	
4. 教育程度	國小(含)以下
	國中
	高中/職
	大學/科技大學
	研究所(含)以上
5. 婚姻狀況	未婚
	已婚
6. 個人每月平均所得	20,000 元(含)以下
	20,001~40,000 元
	40,001~60,000 元
	60,001~80,000 元
	80,001~100,000 元
	100,000 元(含)以上

(二)抽樣方法

本研究採用的方法為立意抽樣，所以問卷將根據與旅行社有接觸的顧客族群為主，將會在特定三家旅行社來做發放。

(三)問卷發放

本研究主要的發放對象為旅行社的顧客及對旅遊有興趣的族群，分別採人工問卷發放及網路問卷，人工問卷發放的部分會去旅行社對顧客進行發放，而網路問卷的部分會找有旅遊相關的網站、論壇進行發放。本問卷發放前會請教專業人士來進行指導與建議，並適時的修改問卷後再去做發放。

3.2.2 專家評分建立資料庫

1. 配套行程設計

關於日本北海道的相關配套行程，本研究參考了一些國內的知名旅行社及旅遊網站設計出來。

2. 專家評分資料庫

這部分將找有經驗的旅行社老闆與資深導遊來參與討論，並對問卷的題項之選項與配套行程評分。

3.3 輔助軟體

本研究的輔助軟體將使用 SQL Sever 2014 來進行專家評分資料庫建立及推薦系統資料分析與比對，先與專家討論後訂定各題項之選項對各行程的分數，並使用 EXCEL 來整理成資料庫，以便將來顧客填寫完問卷後能方便分析。之後將蒐集回來的問卷依照推薦系統的相關模式並加入專家評分做進一步的分析，之後再使用 Matlab 進行聚類分群。

3.4 資料探勘方法

協同過濾是目前使用較為成功的推薦系統技術，它解決了內容過濾式的一些限制問題，此方法不需要分析項目內容，系統根據與使用者有相同興趣的社群成員曾購買的產品或瀏覽過的資訊做為推薦參考依據，意即依據其他喜好相似的使用者歷程來做為目標使用者的推薦，故所推薦之項目可能與使用者從前的喜好大不相同，但是卻能挖掘出使用者的潛在需求。

其主要目的是在眾多使用者當中尋找出興趣偏好相似的使用者做為相互推薦之依據。主要工作可分為三個步驟如下圖 3-2 所示：(一)分群(二)挖掘關聯規則(三)形成推薦

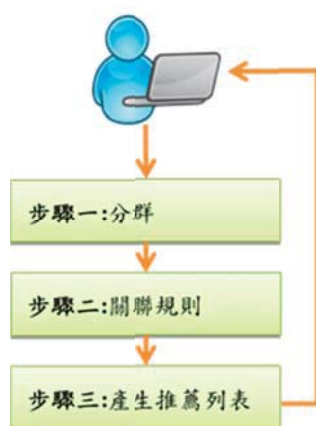


圖 3-2 推薦系統步驟

3.4.1 專家討論

(一)訂出日本北海道之配套行程

以往要作推薦的第一步驟就分群，這部分以專家討論的方法取代，不但可以加強對顧客的分類，在分析的速度的也會變快，本研究參考知名旅行社及旅遊網站來設計北海道的配套行程，如雄獅、東南、燦星……等旅遊網對於去北海道遊玩所做的行程，進而設計出 7 種配套行程，以下為各行程之簡述。

- (1)行程 1-景色賞玩
- (2)行程 2-冰雪活動
- (3)行程 3-花海絢麗
- (4)行程 4-名產藥妝
- (5)行程 5-美食探索
- (6)行程 6-知性文物
- (7)行程 7-自由搭配

(二)建立評分資料庫

經由專家的討論與評估後，可以建立起完整的資料庫，可以做為之後問卷分析時推薦系統的使用資料，這部分總共找了六位旅行社的老闆及資深導遊協助建立評分標準，其中各題項以 abc 為代稱。

請六位專家填完評分表後，把六位專家對問卷各題項的評分去頭尾取平均，經過整理，得到以下表 3-3 的評分標準。

表 3-3 專家評分標準資料表

分數 題項	行程一	行程二	行程三	行程四	行程五	行程六	行程七
a1	6	6	6	4	5	9	6
a2	4	5	6	4	5	6	6
b1	6	6	7	5	6	6	7
b2	3	4	6	4	4	7	6
c1	7	8	7	4	6	7	7
c2	6	7	8	2	5	6	8
c3	8	5	6	4	5	6	7
c4	7	8	7	5	6	7	7
d1	5	6	8	5	6	6	7
d2	7	6	8	5	7	6	6
d3	6	8	7	4	5	7	8
d4	6	7	8	6	6	5	7
e1	6	8	8	4	7	7	7
e2	6	8	7	5	6	6	7
e3	6	7	7	4	6	7	7
e4	5	7	7	6	6	6	7
f1	8	7	7	4	5	7	8
f2	7	7	7	4	6	7	8
f3	6	5	7	6	6	6	8
f4	6	6	6	6	7	8	7
f5	6	5	8	7	6	5	7
g1	8	6	6	6	6	7	5
g2	6	8	8	4	6	7	6
g3	5	6	6	5	7	6	7
g4	4	7	7	4	6	7	7
g5	6	8	7	4	7	7	8
g6	5	7	8	4	6	6	7

3.4.2 適應共振理論ART分群

(一)適應共振理論

ART 的學習系統要有足夠的穩定性來抵抗環境中不相干之干擾或事件，但同時也要有足夠的可塑性才能夠快速的改變與學習，為了因應環境的變化，ART 能夠快速學習，但又不會洗去舊有的記憶，ART 均符合這些特點。

介於輸入層與輸出層之間的網路連結加權。由 ART 所發展出來的，有處理二元值輸入之 ART1 及處理連續信號之 ART2，也有運用模糊之 Fuzzy ART 及 ARTMAP 等相關的網路。在本研究中使用 ART1 將顧客的特徵樣本資料轉為二元值來分群。

(二)ART 編碼

學者陳致傑於 2001 年時提出以三種不同屬性之特徵編碼方式，分別有二進位、位移及堆疊設計法，如表 3-4 所示。二進位設計法是把不同的特徵屬性以二進位的方式來編碼；而位移設計法是以不同的特徵屬性下用編碼位移的方式來表示每一個特徵屬性，來達到區別效果；最後的堆疊設計法運用編碼堆疊的方式來區別每一個不同之特徵屬性。

依照學者將這幾種編碼方式的優缺點比較出來後，可以明顯看到使用堆疊設計法來編碼的效果為最好的，所以在本研究 ART 資料編碼的部分使用推疊設計法。

表 3-4 自適應共振理論特徵屬性設計法比較

設計法	屬性設計方式					比較
	很少	略少	普通	略多	很多	
二進位	001	010	011	100	101	1.無法明顯聚類 2.聚類效果不彰 3.耗用較少節點數
位移	10000	01000	00100	00010	00001	1.無法明顯聚類 2.聚類效果不彰 3.需使用最多之節點
堆疊	0000	0001	0011	0111	1111	1.可明顯聚類,且屬性相似時,亦會歸為同類 2.可表現屬性程度大小分佈 3.需要較多的節點數

資料來源：陳致傑(2001)

3.4.3 推薦系統

(一)推薦流程

一般而言，推薦系統的推薦流程可分為三大步驟：1.蒐集使用者的資料；2.根據使用者的資料進行分析以產生推薦；3.針對使用者對推薦項目的喜好與否來進行適當的修正。

根據這些步驟，我們利用消費者的基本資料及專家評分所得到的分數進行個人化推薦，之後再依相似度計算後找出同好名單並記錄在消費者資訊檔案中，當消費者上線後即可利用其同好們最近的評分記錄來共同進行配套行程的推薦，形成協同過濾推薦。

4. 資料分析與研究結果

4.1 回收樣本結構分析

本研究總共發放了 185 份，回收 145 份，剔除並篩選掉填答不完整的問卷 10 份後，有效問卷為 135 份，有效回收率為 73%。下表 4-1 為有效問卷 135 份樣本之樣本特性統計。

表4-1 有效問卷135份樣本之樣本特性統計

1.性別	男	52	38.5 %
	女	83	61.5 %
2.年齡	20歲(含)以下	11	8.1 %
	21歲~30歲	78	57.8 %
	31歲~40歲	22	16.3 %
	41歲~50歲	14	10.4 %
	51歲~60歲	10	7.4 %
	60歲(含)以上	0	0 %
3.工作別	軍公教	10	4.4 %
	服務業	31	23.0 %
	自由業	0	0 %
	金融業	6	4.4 %
	科技業	14	10.4 %
	製造業	15	11.1 %
	製造業	59	43.7 %
	學生 其他	0	0 %
4.教育程度	國小(含)以下	0	0 %
	國中	5	3.7 %
	高中/職	7	5.2 %
	大學/科技大學	85	63.0 %
	研究所(含)以上	38	28.1 %
5.婚姻狀況	未婚	96	71.1 %
	已婚	39	28.9 %
6.個人每月平均所得	20,000元(含)以下	59	43.7 %
	20,001~40,000元	33	24.4 %
	40,001~60,000元	12	8.9 %
	60,001~80,000元	28	20.7 %
	80,001~100,000元	3	2.2 %
	100,000元(含)以上	0	0 %

4.2 建立顧客各行程分數之資料庫

The screenshot shows an Excel spreadsheet with the following structure:

- Columns: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U.
- Rows: 1 to 21, with a total row at the bottom.
- Columns 2-8 are grouped under '行程1' (Trip 1), columns 9-15 under '行程2' (Trip 2), columns 16-22 under '行程3' (Trip 3), and columns 23-29 under '行程4' (Trip 4).
- Column 30 is labeled 'SUM'.
- Columns 31-34 are labeled '行程5' (Trip 5), columns 35-38 under '行程6' (Trip 6), and columns 39-42 under '行程7' (Trip 7).
- Columns 43-46 are labeled 'SUM'.

圖 4-1 顧客各行程分數資料庫

依照上章整理過的專家評分資料庫(表 3-3 所示)來對回收之有效樣本進行評分，如上圖 4-1 所示。之後再隨機選取 5 筆顧客資料，用來呈現單一樣本所得到的各行程分數，如下表 4-2 所示。

根據上表的分數可以針對這些各別顧客推薦最適合他們的前幾名行程，如 A3 這位顧客的前三名依序為行程 3，行程 6，行程 2；而 A78 前三名為行程 6，行程 2 及 3；A111 前三名依序為行程 3，行程 1 及 2；A115 前三名依序為行程 3，行程 2 及 5；A135 前三名依序為行程 3，行程 2，行程 6。

表 4-2 單一顧客之各行程分數

序	顧客 行程	行程 1	行程 2	行程 3	行程 4	行程 5	行程 6	行程 7
1	A3	39	43	48	30	40	47	41
2	A78	42	46	46	29	38	49	40
3	A111	39	39	42	26	34	38	35
4	A115	35	39	41	28	39	40	34
5	A135	36	39	42	26	34	38	36

4.3 資料聚類

4.3.1 聚類方法

這裡的聚類是使用 Matlab 的 ART 外掛程式來開發，其輸入的值一定要是二元值的檔案才能輸入 ART 內，才能得由類別變數的聚類分群，先使用顧客的基本資料來做相似性的分類，在輸入 ART 網路前要把基本資料編碼，本研究使用堆疊設計法來對個人特徵屬性做分類並轉成二元值。

本研究使用的個人特徵屬性，分別是性別(gender)、年齡(age)、工作別(job)、教育程度(education)、婚姻狀況(marriage)和個人每月平均所得(salary)共 6 個屬性做 ART 的輸入向量，編碼的內容如下表 4-3 所示，gender 為 X1、age 為 X2、job 為 X3、education 為 X4、marriage 為 X5 和 salary 為 X6。

表 4-3 個人特徵編碼表

(gender)	男	女
B ₁ ~B ₂	01	10

(age)	20歲(含) 以下	21歲~30 歲	31歲~40 歲	41歲~50 歲	51歲~60 歲
B ₃ ~B ₆	0001	0011	0111	1111	1101

(job)	軍公教	服務業	金融業	科技業	製造業	學生
B ₇ ~B ₁₀	0001	0011	0111	1111	1101	1011

(education)	國中	高中/職	大學/科技大學	研究所(含)以上
B ₁₁ ~B ₁₄	0001	0011	0111	1111

(salary)	20,000元 (含)以下	20,001 ~40,000元	40,001 ~60,000元	60,001 ~80,000元	80,001 ~100,000元
B ₁₇ ~B ₂₀	0001	0011	0111	1111	1101

下一步，把資料庫中的顧客特徵屬性取出，依上表二十所示將編碼表依序編碼，並將每位顧客之個人特徵值對應其值，gender 為 B1~ B2、age 為 B3~ B6、job 為 B7~ B10、education 為 B11~ B14、marriage 為 B15~B16 和 salary 為 B17~ B20，共 20 個輸入節點，其編碼後的內容如下圖 4-2 所示，該圖即為輸入 ART 網路的輸入向量。

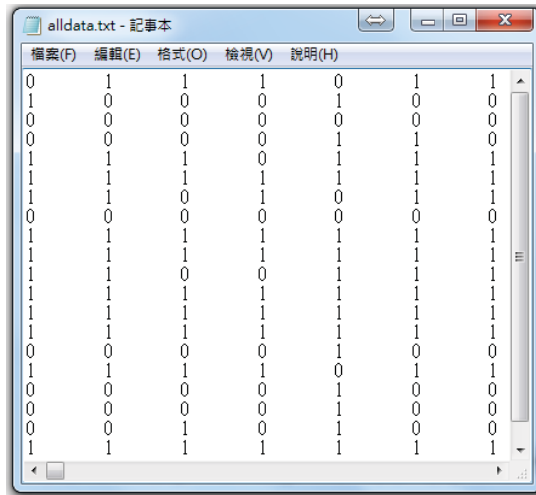


圖 4-2 ART 輸入檔

以下圖 4-3 做說明，將經過編碼後的輸入檔匯入 Matlab 軟體中的外掛軟體 ART 後，接下來 ART 網路會依照輸入向量的每位顧客之特徵屬性(B1~B20)，進行聚類，輸出向量就會為 Y1, Y2, ..., Yk。之後就可得到 4-4 畫面中依代表輸入向量所產生的聚類結果。

圖 4-3 ART 聚類結果

```
>>art1(Artdata,0.3)
Category      Pattern
-----
1      11, 12, 21, 22, 27, 30, 36, 37, 42, 45, 55, 56, 68, 69, 78, 79, 84, 87, 93, 94, 99, 102, 1
-----
2      1, 2, 4, 6, 7, 10, 13, 15, 16, 20, 24, 25, 28, 29, 31, 35, 38, 39, 40, 41, 43, 47, 51, 52,
-----
3      3, 5, 8, 9, 14, 17, 18, 19, 26, 33, 44, 48, 49, 51, 60, 62, 65, 66, 71, 74, 75, 76, 83, 90
-----
4      23, 32, 34, 46, 80, 89, 91, 103, 121
-----
```

經過統整後，可得到如下表 4-4 所示，4 群顧客的 ART 群集結果，也可以清楚看到每個群集內與其他顧客特徵屬性相似的成員。

表 4-4 ART 群集結果

群集編號	群內顧客
1	A11、A12、A21、A22、A27、A30、A36、A37、A42、A45、A55、A56、A68、A69、A78、A79、A84、A87、A93、A94、A99、A102、A112、A113、A119、A120、A125、A131、A132

表 4-4 ART 群集結果(續)

群集編號	群內顧客
2	A1、A2、A4、A6、A7、A10、A13、A15、A16、A20、A24、A25、A28、A29、A31、A35、A38、A39、A40、A41、A43、A47、A50、A52、A53、A54、A57、A58、A59、A60、A63、A64、A67、A70、A72、A73、A77、A81、A82、A85、A86、A88、A92、A95、A96、A97、A98、A100、A104、A107、A109、A110、A111、A114、A118、A122、A123、A126、A128、A129、A130、A133
3	A3、A5、A8、A9、A14、A17、A18、A19、A26、A33、A44、A48、A49、A51、A60、A62、A65、A66、A71、A74、A75、A76、A83、A90、A101、A105、A106、A108、A115、A116、A117、A124、A127、A134、A135
4	A23、A32、A34、A46、A80、A89、A91、A103、A121

4.3.2 分群後的顧客類別命名

經聚類後得到了 4 個群集，如下表 4-5 所示，再根據各群集的基本資料特徵來對各群集命名，根據下表的統整後將群集 1 命名為事業有成型、群集 2 命名為年輕學生型、群集 3 命名為小資生活型、群集 4 命名為壯年樂活型。

表 4-5 ART 群集命名

演算法	ART 聚類			
群集	群集1	群集2	群集3	群集4
命名	事業有成型	年輕學生型	小資生活型	壯年樂活型

4.4 關聯性分析

在這裡以第四群壯年樂活型為例，把這些相似性高的顧客資料的各行程分數都整理好如下表 4-6 後，接著要將形成分數轉換成 2 元進制，並以行程分數 45 分以上為門檻標準，45 分以上為 1，表示分數較高的，其餘為 0 表示分數較低，以是和非兩種型態代表來做轉換。其轉換後的結果如表 4-7 所示。

表 4-6 群集四之顧客行程分數

序	顧客 行程	行程 1	行程 2	行程 3	行程 4	行程 5	行程 6	行程 7
1	A23	32	37	43	26	36	37	37
2	A32	39	45	48	30	41	49	40
3	A34	41	45	50	31	41	49	40
4	A46	39	37	41	27	34	39	33
5	A80	32	37	43	26	36	37	37
6	A89	39	45	48	30	41	49	40
7	A91	41	45	50	31	41	49	40
8	A103	39	37	41	27	34	39	33
9	A121	32	37	43	26	36	37	37

4.5 形成推薦

最後的推薦清單是由同好的分數所形成的。這裡以表 4-8 來說明，在此例中，當新會員 A_i 使用此系統後，假設新加入的會員基本資料與表 4-7 這群差不多，系統就可以從會員紀錄檔中得知，於是由表 4-7 就可以得知行程 3、行程 1 及行程 6 合計後的分數較高，因此行程 3、行程 1 及行程 6 就會優先被列入這位新會員的推薦清單內。

表 4-7 群集四之顧客行程分數推薦清單

序	顧客 行程	行程 1	行程 2	行程 3	行程 4	行程 5	行程 6	行程 7
1	A23	0	0	1	0	0	0	0
2	A32	1	1	1	0	1	1	1
3	A34	1	1	1	0	1	1	1
4	A46	1	0	1	0	0	1	0
5	A80	0	0	1	0	0	0	0
6	A89	1	1	1	0	1	1	1
7	A91	1	1	1	0	1	1	1
8	A103	1	0	1	0	0	1	0
9	A121	0	0	1	0	0	0	0
	合計	6	4	9	0	4	6	4

旅遊行程的推薦不同於一般購物的推薦，顧客在不同的時間點對於購物的需求也會因人而異，因為對於旅遊行程推薦，顧客不可能每次旅遊都會希望玩相同的地方。因此，在介紹遊玩行程時也就更加重要，而推薦的群體也是相當重要的一環，藉由這樣的資料探勘方法，可以使推薦的過程更加完善。

4.6. 系統建置與操作

按照本研究所提出來的日本北海道推薦系統建置及操作，進而設計出推薦系統與網頁畫面，對本研究提出之理論基礎及架構來進行接下來的驗證與評估，以下是此系統的說明。

(一)系統架構

在系統架構的部分使用 ASP.net 作為主要的開發程式來撰寫，至於資料庫是使用 SQL，並依照先前蒐集專家評分外再加上顧客分群後的推薦系統流程資料表來做運算。

(二)系統流程

首先先開啟網頁畫面，即可看到選單中日本北海道的推薦系統，點選後可看到兩種推薦方式，分別是填寫問卷及填寫個人資料所產生的推薦，接受推薦後可得到 TOP3 的行程，並可查看各行程的說明，如下圖 4-5 所示。

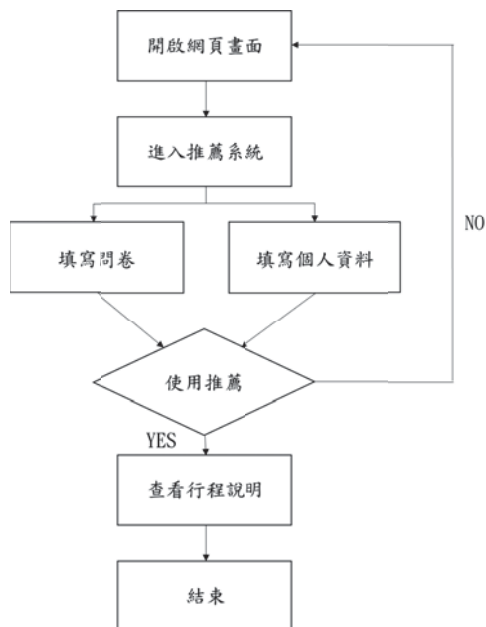


圖 4-5 系統執行流程

(三)系統操作畫面

首先使用 google chrome 瀏覽器打開操作的網頁介面，其畫面顯示如下圖 4-6 所示，此頁面為本研究系統的首頁畫面，其內容有首頁畫面、日本北海道的介紹、推薦系統及其他連結。

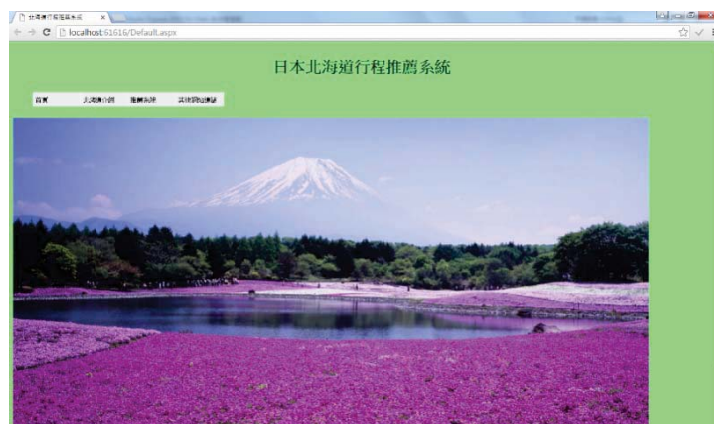


圖 4-6 日本北海道推薦系統首頁

使用者接下來可以點選推薦系統來進行兩種不同方式的推薦方法，首先是第一種推薦方法，使用問卷的方式來進行推薦，這裡所使用到的資料庫為專家評分資料表來進行計算，如圖 4-7 所示，以下是操作頁面。



圖 4-7 第一種推薦系統

點選進入第一種問卷推薦系統，點選進去後可看到問卷題目，如下圖 4-8 所示，依畫面所示，填答完後按送出即可查看推薦行程。

說明

問卷

- 請問您在過去曾有出國遊玩的經驗嗎?
 有 * 沒有(跳至第三題)
- 請問您在過去曾有出國去日本北海道旅遊的經驗嗎?
 有 沒有
- 請問您希望去日本北海道旅遊的時間是在幾月份?
 一~三月 四~六月 七~九月 * 十~十二月
- 請問您希望去日本北海道旅遊每個人的預算金額大約是?
 三萬元(含)以下 * 三~五萬元 五~七萬元 七萬元(含)以上
- 請問您希望去日本北海道旅遊的方式是?
 家族或團體旅遊 * 朋友或團體旅遊 自行規劃行程旅遊(自由行) 旅行社團體旅遊
- 請問您希望去日本北海道旅遊的主要目的是?
 生態旅遊 * 休閒度假 心靈放鬆 探訪親友 購物
- 請問您希望去日本北海道的哪些地區遊玩?
 自然風景區 觀光勝地區 美食地區 溫泉休憩區 * 冰(雪)上活動區
 其他:

送出並查看推薦

圖 4-8 問卷推薦系統

問卷推薦結果就如下圖 4-9 所示，按照排名依序為行程 2、3、6。



圖 4-9 問卷推薦系統結果

接著點選行程說明可察看詳細行程的安排，如下圖 4-10 所示。



圖 4-10 北海道推薦行程說明

而第二種推薦方法，則是使用資本資料的方式來進行推薦，這裡所使用到的資料庫為分群後的資料表，並且經過轉換後得到群體間的 TOP3 行程推薦，而這裡填寫問卷後，判斷填答資料最接近哪個群體，即推薦此群體的 TOP3 行程，如下圖 4-11 所示。



圖 4-11 第二種推薦系統

點選進入第二種基本資料推薦系統，點選進去後可看到基本資料選項，如下圖 4-12 所示，依畫面所示，填答完後按送出即可查看推薦行程。



圖 4-12 基本資料推薦系統

經過運算資料後可發現最接近第一個群體，事業有成型，所以推薦此一群體之 TOP3 行程，圖下圖 4-13 所示，一樣可以按下面之行程說明來查看行程。

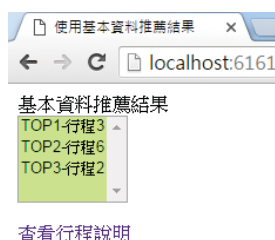


圖 4-13 基本資料推薦系統結果

5. 結論與建議

5.1 研究結論

(一)問卷推薦系統

本研究的第一個推薦系統是填寫問卷來讓顧客找到最適合的個人化推薦系統，其中在設計問卷一開始的時候分別去請教了六位旅行社的專家們來幫忙設計與修改，之後進一步參考相關知名旅遊網站與專家意見來設計出本研究要推薦給顧客的七個行程，最後再請他們訂出各行程對題項之選項分數，最後再將六位專家的分數平均後整理成專家評分標準資料表。

所以當顧客填寫完問卷題目後就可以根據專家的評分資料庫計算出分數最高的前三名並推薦給他們作為個人化的行程。

(二)基本資料推薦系統

而關於第二個推薦系統是使用基本資料來進行推薦，把蒐集回來的問卷使用基本資料來作顧客分群，使用 Matlab 的類神經外掛軟體 ART1 進行分群，最後共分成四群，等顧客填寫基本資料後可以從資料庫中比對與哪一群最為相近就將該群體的前三名行程推薦給該名顧客。

5.2 管理意涵

本研究重點在幫助旅行社如何鎖定特定族群、替顧客規劃出最符合需求之旅遊行程，所以為顧客量身打造的人性化的服務就相當重要了，首先旅行社需要了解顧客的需求，本研究使用推薦系統的方式來幫助旅行社再幫顧客規畫行程能更加快速且方便管理，同時可了解顧客族群有哪些特徵，最後提升整體旅行社績效是本研究的重點。

(一)建立網頁資料庫

如同上一章的研究系統，若業者把後臺要做推薦的行程分數都建好後，以後就能快速收到並整合顧客資料，並對他們發送相關的旅遊行程規劃的廣告，同時也可以減少業者本身的一些基本開銷。

(二)個人化與群體的推薦

本研究採取的推薦方式與以往推薦系統設計有所差異，以往多是使用歷史資料庫來分析已經存在的顧客的購買行為，這裡用問卷題目蒐集顧客對旅遊的相關期望與基本資料，再使用與專家討論而產生出來的相關行程評分來做推薦系統，更能有效客觀的來對顧客做推薦。

5.3 研究限制與後續研究

一、專家數量的問題，本研究雖然使用專家評分標準來判定行程分數與以前相關推薦系統研究有所不同，不過因為全台旅行社目前將近快三千家，而本研究所找的專家是從三家旅行社中給六位專家做評分而已，所以來可以再加入多一點專家討論可以讓研究更加客觀。

二、再來是樣本數的問題，因為本研究所鎖定的研究對象主要為三家旅行社的顧客，還有在一些旅遊網站回收樣本，不過未來可以將研究的樣本數來擴大建立更多樣性的資料庫，也能對後續想要研究這方面的學者有所幫助。

三、地區限制的問題，本研究因為只針對日本北海道為範例來做相關的問卷與行程設計，比較沒辦法顧及到各個客群的顧客，所以未來可以把研究的地區範圍在做擴大，也可以讓旅行社業者在做推薦上更加完善。

四、顧客後續追蹤，雖然本研究的結果對顧客有產生推薦，不過卻無法得知該顧客接受這樣的推薦行程後是否滿意，所以未來可以針對推薦行程後的滿意度來做調查，這樣才能適時的調整推薦行程的內容。

6. 參考文獻

中文文獻

1. 于學謙(1981)。旅行業入門。旅遊教育出版社：北京。
2. 朱玉槐(1988)。旅遊學辭典。人民出版社：陝西。
3. 吳緯閔(2008)。網際網路服務推薦系統。國立中央大學資訊工程研究所碩士論文。桃園縣。

英文文獻

1. Adomavicius, G., and Tuzhilin A., (2003), "Recommendation Technologies: Survey of Current Methods and Possible Extensions", Working paper, Stern School of Business, New York University.
2. Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. Paper presented at the ACM SIGMOD Record, 22(2),207-216.
3. Boucher-Ryan, P.D., and Bridge, D., 2006, "Collaborative Recommending using Formal Concept Analysis", Knowledge-Based Systems, 19(5),309-315.
4. Breese, J.S., Heckerman, D., and Kadie, C., (1998), "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence, 43-52.
5. Balabanovic Y. and Y. Shoham(1997), "Fab: content-based, collaborative recommendation", Communications of the ACM, Vol. 40, Issue 3, 66-72.
6. Breiman, L.(1984), Friedman, J.H., Olshen, R.A. and Stone,C.J., Classification and Regression Trees, Wadsworth Belmont.
7. Berry, M.J.A.,Linoff, G. S.(1997). "Data Mining techniques: For marketing, sales, and customer support."New York:Wiley Computer.
8. Charles J. Metelka (1990) "The Dictionary of Hospitality", Travel and Tourism, N.Y.: Delmar Publishers Inc.
9. D. Billsus and M. J. Pazzani, "Learning collaborative information filters," 1998,48.
10. Fayyad, Piatetsky-Shapiro, et al. – (1996), "From Data Mining to knowledge Discovery: An Overview".

11. Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70.
12. GOOD, N., SCHAFFER, J. B., KONSTAN, J. A., BORCHERS, A., SARWAR, B. M., HERLOCKER, J. L., AND RIEDL, J. (1999). Combining collaborative filtering with personal agents for better recommendations. In *Proceedings of the 16th National Conference on Artificial Intelligence (AAAI-99)*, J. Hendler, and D. Subramanian, Eds. AAAI Press, Menlo Park, Calif., 439-446.
13. Grossberg, S. (1976a), "Adaptive pattern classification and universal recoding, I: Parallel development and coding of neural feature detectors", *Biological Cybernetics* 23, 121-134.
14. Haubl, G. & Murray, K. B. (2005). Personalized product recommendations and consumer purchase decisions.
15. Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5-53.