


南 華 大 學

資訊管理學系

碩士論文

資料挖掘技術在旅遊行銷之應用

The Study of Applying Data Mining Technology into Tourist Marketing



研 究 生：鄭丞君

指導教授：謝昆霖 博士

中華民國九十三年六月

南 華 大 學
資 訊 管 理 學 系
碩 士 學 位 論 文

資料挖掘技術在旅遊行銷之應用

研究生：鄭丞君

經考試合格特此證明

口試委員：
陳 彥 巨
沈 耀 成

指導教授：謝品輝

所 長：


口試日期：中華民國 93 年 6 月 10 日

南華大學碩士班研究生
論文指導教授推薦函

資訊管理學系碩士班 鄭丞君 君所提之論文
資料挖掘技術在旅遊行銷之應用
係由本人指導撰述，同意提付審查。

指導教授 謝品霖

93年5月10日

誌 謝

二年的時光很快就過去了，在南華大學的日子中除了在學業的精進努力外，也認識很多學弟妹，在學習及生活上給我很多的幫助，在這裡，首先要感謝蔣志堅老師在學業上的指導，其次要感謝謝昆霖老師，兩位指導老師在課業上的諄諄教誨才能如期完成論文，也要謝謝所上的老師，無私地在課堂上把課業上及畢業後工作上該有的態度不厭其煩的耳提，最後要謝謝同學及學弟妹在平常生活上的幫助，使我能順利在南華學習，完成學位。

鄭丞君 謹識
于 南華大學
中華民國九十三年七月六日

資料挖掘技術在旅遊行銷之應用

學生：鄭丞君

指導教授：謝昆霖

南華大學 資訊管理學系碩士班

摘要

隨著社會經濟的發展，民眾的生活型態有著逐漸的改善，這也使得休閒旅遊已成為民眾重視的生活議題。顧客關係管理(Customer Relationship Management, CRM)是一項近幾年最受重視的研究領域，要如何挖掘潛在顧客與如何快速地開發新顧客通常是飯店業者最積極投入資源進行探討，飯店業者可以根據過去顧客的交易記錄進行有意義的規則或是模型等資料探勘(Data Mining, DM)，再根據所挖掘的結果進行最適行銷建議。人工智慧技術(Artificial Intelligence, AI)在近幾年在許多領域(例如：預測、辨識、控制、分類等)都有不錯的應用成效，其中的類神經網路(Artificial Neural Networks, ANNs)方法更是受到多數研究者的青睞，尤其是資料挖掘中的聚類分析(Clustering Analysis)，類神經網路也因為它的便利性和簡易性而有不錯的成效，在本論文中亦將引用類神經網路中非監督式的自適應性共振理論模式(Adaptive Resonance Theory, ART)來進行必要的顧客屬性群的聚類分析。此外，我們也設計一個指標與評比程序來找出到底哪些舊顧客屬性群對飯店業者而言是最適合進行資源投資以及行銷開發、並利用舊顧客的屬性區隔資訊來對新顧客提供最適行銷建議。

此外，飯店服務人員可能會被詢問到顧客想去玩的旅遊景點應如何搭車較方便？怎麼進行是最省錢？這些顧客服務就要考慮到前往各個景點移動時間或者移動費用(成本)的最適性問題，本研究

將應用基因遺傳演算法(Genetic Algorithm, GA)來規劃出旅遊行程最適解，求出顧客欲前往的各個景點總移動時間最少的最佳路徑、總移動成本最少的最佳路徑、以及總移動時間與總移動成本同時為最少的最佳組合路徑，提供顧客旅遊行程規劃參考，以提升飯店業者在市場的競爭力。

關鍵字：休閒旅遊、類神經網路、自適應性共振理論模式、基因遺傳演算法、顧客關係管理

The Study of Applying Data Mining Technology into Tourist Marketing

Student: Chen-Chun Cheng

Advisor: Dr. Kun-Lin Hsieh

Department of Information Management
The M.B.A Program
Nan-Hua University

ABSTRACT

As for the social economic rapidly developing, the living style of people in Taiwan had been improved. It will let most people pay more attention to the leisure tour. Recently, customer relationship management (CRM) is a new issue to be addresses for many Hotels or Tourism centers. How to efficiently mine the potential customers or attract the new customers are two primary goals to invest in studying for them. The marketing suggestions can be made by mining the meaningful rules or models from the historical transaction records. Up to now, the performance of using the techniques of artificial intelligence (AI) had presented well for many applications, e.g. prediction, recognition, control, classification and clustering. Among them, the issue of clustering analysis can be done well by using the artificial neural networks (ANNs). Especially, the adaptive resonance theory (ART) in the unsupervised model is the popular one with the characteristics of simplification to be applied into clustering analysis. Hence, we will employ it to implement the necessary clustering analysis in this study. Besides, an evaluated index and a complete evaluation procedure are developed to obtain the related conclusion from the results of DM. That is, the information about the clusters of the potential customers being

suitable to invest and that about the corresponding cluster being suitable to develop can be obtained.

Besides, the attendant will be frequently asked about how to travel conveniently by public transits? or how to travel with the less cost? Such cases will focus on the problems of the adaptive tour scheduling with moving time or moving cost limitations. Genetic algorithm (GA) is an approach to address the optimization problem well in most industries. In this thesis, we will apply GA to address such problem. The suggestions of adaptive tour scheduling with the minimum tour's moving time, minimum tour's moving cost or simultaneously minimize the tour's moving time and moving cost can be derived. Not only the customer's requirement can be achieved, but the necessary market strategies can also be created in the subsequent customer relationship management (CRM). The service competence for hotel can be then promoted.

Keywords: Leisure Tour, Artificial Neural Networks (ANNs), Adaptive Resonance Theory (ART), Genetic Algorithm (GA), Customer Relationship Management (CRM)

目 錄

書名頁	I
博碩士論文授權書	II
論文指導教授推薦函	III
論文口試合格證明	IV
誌謝	V
中文摘要	VI
英文摘要	VIII
目錄	X
表目錄	XII
圖目錄	XIII
第一章 緒論	1
第一節 研究背景	1
第二節 研究動機	1
第三節 研究目的與方法	2
第四節 論文架構	3
第二章 相關研究文獻	5
第一節 相關研究	5
第二節 資料挖掘	5
第三節 旅遊行銷	8
第三章 解決的方法及使用的技術	10
第一節 類神經網路	10
第二節 形心	15
第三節 歐氏距離	15

第四節 顧客重住率	16
第五節 基因演算法	16
第四章 問題解析程序	22
第一節 新、舊顧客問題解析程序	22
第二節 最適行程問題解析程序	26
第五章 應用案例展示	29
第一節 新、舊顧客問題範例展示	29
第二節 最適性行程範例說明	35
第六章 結論	51
參考文獻	52

表 目 錄

表 5-1 顧客屬性對應之屬性值及編碼所需向量數	30
表 5-2 年齡屬性編碼設定範例	30
表 5-3 顧客屬性群型態解析	31
表 5-4 整體及各聚類重住率	32
表 5-5 新顧客消費型態及說明	34
表 5-6 第二群聚中各住宿樣式型態所佔比例	35
表 5-7 飯店附近旅遊景點	35
表 5-8 飯店預估兩旅遊景點之間的移動時間 (分鐘).	36
表 5-9 兩旅遊景點移動費用 (元).	36
表 5-10 基本型最適性行程規劃分析結果	40
表 5-11 對 (d) 問題解析結果	44
表 5-12 針對 (e) 問題解析結果	47

圖 目 錄

圖 1-1 論文架構圖	4
圖 3-1 ART 網路架構	13
圖 3-2 基因演算法應用流程	21
圖 4-1 舊顧客保留解析程序架構圖	23
圖 4-2 新顧客推薦架構圖	24
圖 4-3 最適性行程規化架構圖	27
圖 4-4 最適性行程問題解析程序步驟	28
圖 5-1 ART 網路分群結果	31
圖 5-2 歐氏距離計算結果與推薦住宿樣式型態	34
圖 5-3 設定族群大小及終止條件	39
圖 5-4 設定交配機率及突變機率	39
圖 5-5 設定總移動時間最小化	40
圖 5-6 套裝軟體針對問題 (a) 執行的結果	42
圖 5-7 針對問題 (a) 在最終世代族群中各物種的適應度	42
圖 5-8 針對 (d) 問題設定染色體	44
圖 5-9 針對 (e) 問題設定總移動時間之限制條件	46
圖 5-10 針對 (e) 問題之總移動時間限制 (240 分鐘) 解析結果其 中之一	46
圖 5-11 飯店網站最適性行程規劃應用	48
圖 5-12 線上最適性行程規劃網頁	48
圖 5-13 使用者選擇所需要的旅遊景點	49
圖 5-14 最適性行程規劃中使用者以移動時間考量結果之一	49
圖 5-15 最適性行程規劃中使用者以移動成本考量結果之一	50

圖 5-16 最適性行程規劃中使用者以移動時間及移動成本考量結果
之一 50

第一章 緒論

第一節 研究背景

近年來政府積極發展休閒觀光，尤其是週休二日的推行，更讓國人愈來愈重視休閒活動；加上旅遊資訊均可以透過相關的雜誌、網路、電視、報紙等媒體獲得，這也使得休閒旅遊 (Leisure Tour) 變成社會大眾提升生活品質的重要活動之一，相對地這也可說是休旅飯店業者的一大利多。休旅飯店業者在這樣競爭的市場環境中要如何強化自己的競爭優勢來獲得利潤，已成為大多數業者的重要研究議題。在眾多的研究議題中，如何有效地利用各種顧客關係管理 (Customer Relationship Management, CRM) 是學者研究的重點[8]。

第二節 研究動機

要挖掘潛在顧客與如何快速地開發新顧客通常是業者最積極投入資源進行探討，然在實際的狀況下，多數的業者發現到投資在舊客戶上所獲得的利潤回報遠超過投資在新顧客上所獲得的，這也使得業者在資源分配上會較傾向於如何有效地挖掘潛在的顧客。除了對於舊有顧客的掌握之外，新顧客的行銷策略也是業者所面對的課題之一，但因新顧客缺乏其相關的交易消費資訊，如何利用既有顧客資料所挖掘的資訊來對新顧客提供最適切的行銷建議，也是一項值得投入的研究議題，而當顧客登入飯店後可能會尋問飯店附近景點如何遊玩才能節省必要的移動時間、移動成本等問題，因此本文也提供旅遊行程最適性的規劃。

第三節 研究目的與方法

要解決前述潛在顧客的保留 (Potential Customer Retain) 問題，可以利用資料挖掘技術 (Data Mining) 對過去顧客的交易記錄進行有意義的規則或是模型探詢，資料挖掘是屬於整個知識發現 (Knowledge Discovery in Database, KDD) [11,12]的其中一個步驟，一般而言資料挖掘有下列不同的任務：關聯法則 (Association rules)、分群 (Clustering)、分類 (Classification)、預測 (Forecast) [13] 等，而用來解決資料挖掘問題的方法或工具則有決策樹 (Decision tree)、自動群聚偵測 (Automatic cluster detection)、類神經網路 (Artificial neural networks) [14]，再根據所挖掘的結果進行適當的行銷管理策略[5]。基本上在休旅飯店業者的顧客交易記錄中，業者會根據資料庫 (Data Base) 或是資料倉儲系統 (Data Warehouse System) 的架構設計彙整有關顧客的個人基本屬性資料，例如：姓名、年紀、職業、消費記錄 (包含住宿樣式、消費金額、住房頻率) 等，再利用統計技術[15]，例如：迴歸分析 (Regression Analysis)、變異數分析 (Analysis of Variance) 或是多變量分析 (Multivariate Analysis) 等來進行必要的研究；在這樣的研究議題下，主要的邏輯概念就是透過聚類 (Clustering) 的解析先將顧客區分成不同的屬性群組，接著判斷各屬性群組的貢獻程度 (可依實際的需求考量限制)，再從參考的指標或是限制條件中選定最適合開發的潛在顧客，業者便可以針對這些潛在的顧客實施適切的行銷策略。這樣的思維的考量下，所涉及的技術首要是聚類的問題，近年來類神經網路 (Artificial Neural Networks, ANNs) 在聚類分析的應用領域中展現不錯的成效

[16]，其主要是利用類神經網路中的非監督式學習(unsupervised learning)模式，在眾多的非監督式模式中自適應性共振理論模式(Adaptive Resonance Theory, ART)是一個架構簡易且應用上較有彈性的模式[4]，而且考慮到資料的尺度及型態，故本研究在進行顧客屬性群的聚類解析時將採用類神經網路。此外，經過顧客屬性群的解析後，我們仍無法直接獲得那個屬性群的顧客是最值得進行資源投資與開發，因此本研究亦設計一個簡易的評比程序，透過顧客重住率的指標計算，搭配顧客屬性群的分析，業者將可以很快速地從眾多資料中挖掘出最具投資價值的潛在顧客群。對於新顧客的行銷建議，也是無法直接由顧客屬性群聚解析後獲得，因此本研究也透過屬性比率最高者以及屬性比率的排序思維來提供業者面臨新顧客時採用的行銷建議資訊。而新、舊顧客登入飯店後，服務人員可能會被詢問到顧客想去玩的旅遊景點應如何搭車較方便？怎麼進行是最省錢？這些類型的問題就要考慮到前往各個景點移動時間或者移動費用(成本)的最適性問題，此問題可由基因遺傳演算法來進行最適性行程之搜尋，以滿足顧客所需。

第四節 論文架構

本文探討對新顧客的推薦及舊顧客的保留，以及當顧客住進飯店後可提供飯店附近具最少移動時間及最小移動成本的旅遊景點，論文架構詳見圖 1-1。

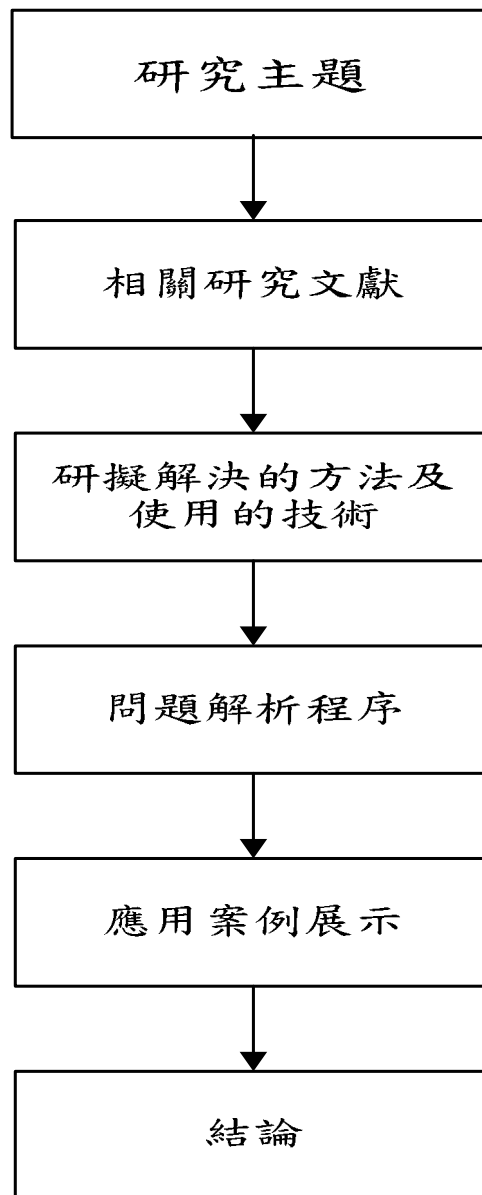


圖 1-1 論文架構圖

第二章 相關研究文獻

第一節 相關研究

目前利用 IT 或資料挖掘技術在旅遊產業上的相關研究中有，以 ECT (Efficient City Tour)、EST (Efficient Spot Tour) 和 ELM (Efficient Leading & Meal) 方法分別找出最短路徑城市、最短路徑景點及最合適的食宿地點，再利用 Apriori 演算法探勘，找出符合最小支持度門檻的項目集合，根據這些項目集合為顧客排定最有效率的旅遊行程[1]。利用關聯法則，來規畫最適性的旅遊行程以及發掘最適性的顧客[2]，以及利用 PAM (Partition Around Medoids) 演算法作分群分析 (Clustering Analysis) 與次序相關分析法 (Sequential Pattern Analysis)，幫顧客尋找最適合推薦的旅遊景點[3]。陳彥匡與屈妃容[8]則應用品質機能展開及資料挖掘技術探討如何在休閒旅館落實顧客關係管理。

第二節 資料挖掘

資料挖掘 (Data Mining) 是一種從整個資料庫裡的資料，利用一種或多種電腦技術來自動分析或擷取知識的過程，資料挖掘的目的是在資料中發現趨勢與樣式。一般而言，資料挖掘有下列不同的功能[11, 12, 13, 14]：

1、描述 (Description):

對資料集 (Data set) 提供一個簡要的屬性描述，以便與其它資料區分。

2、分析 (Analysis):

分析是對一大群的資料加以檢查，以找出其中的某些規則或特徵如出現的頻率、趨勢、變動、相似的順序性及順序的種類等。

3、關聯 (Association) :

對一群資料項進行分析以發現其間的因果關係或關聯性，這些出現頻繁的關係或關聯性多以關聯法則 (Association Rules) 的形式來表示其屬性與數值的情況，如 $X \rightarrow Y$ 意謂凡是滿足條件 X 的資料亦極可能滿足條件 Y 。

4、分類 (Classification) :

對一組訓練資料進行分析以建立一對應之資料分類模型，在此過程中將可產生一決策樹 (decision tree) 或一組分類的法則 (classification rules)，用於對未來或資料庫中的資料進行瞭解及分類各資料的特性。

5、分群 (Clustering) :

一個群體是一個由資料物件 (Object) 相似度高的資料所組成的集合，相似度 (similarity) 可以用距離或其它函數來表示 (群體之間的相似度要低、群體內部的相似度要高)，而分群即是以此對一個資料集加以分析以區分其中各別資料所屬的群體。

6、預測 (Prediction) :

對於某些缺失的資料或一具特定屬性的資料分佈，預測其可能的數值。

資料挖掘過程是涉及多步驟的一個情況，資料挖掘可以依 KDD (Knowledge Discovery in Databases) 的目標不同，採用不同的

類別，由資料集 (Data Sets) 中發現各種關係，再以視覺化的方式 (如 OLAP data mart data cube) 將資料挖掘的結果呈現予 KDD 的使用者。在進行 KDD 的步驟方面 [11, 12, 13]，Fayyad 認為有：

- 1、選擇及取樣欲 KDD 的資料集合 (Data Selection & Sampling)。
- 2、探勘前對目標資料 (Target Data) 的前處理和清洗 (Preprocessing & Cleaning)。
- 3、資料轉換和資料減量 (Transformation & Reduction)。
- 4、進行資料挖掘依欲 KDD 的目標及資料屬性型式選擇適合的資料挖掘演算法以找出資料關係的類型 (patterns)、關聯性 (associations)、改變 (changes)、異常現象 (anomalies)、有意義的結構 (significant structures)。
- 5、挖掘結果的釋意及所得關係類型的評估 (Interpretation & Evaluation) 含知識的呈現 (Knowledge Presentation) 等五階段。

而 KDD 挖掘出的知識，再存於知識庫，作為爾後決策時之參考或對問題及趨勢之預測等用途。而資料挖掘技術的應用 (可包括但不限於下列應用) [14]：

- 1、偵測欺騙行為 (Fraud Detection)
例如：此筆刷卡是否有問題？
- 2、風險管理 (Risk Management)
例如：該給此客戶有多少的貸款？
- 3、價格 (Pricing)
例如：該給此客戶有多少的折扣？

4、混合銷售 (Cross Sell)

例如：有哪些產品可以順便銷售給這位客戶？

5、廣告分析 (Target ads)

例如：該給這個顧客哪一種廣告？

6、追蹤犯罪 (Track Down Criminals)

例如：由犯罪資料分析，可以幫助犯案時找出可疑人口。

7、資訊經濟人 (Information broker)

例如：由大量銷售資料分析顧客的喜好。

8、保留忠誠客戶 (Holding on to good customers)

9、淘汰不好客戶 (Weeding out bad customers)

第三節 旅遊行銷

行銷是企業與消費者之間一種價值 (value) 交換的程序。根據美國行銷學會的定義：「行銷是企劃與執行產品 (product) 訂定價格 (price) 決定通路 (place) 與促銷產品 (promotion)，是服務與表達意見的程序，以交換的方式滿足消費者的需要與欲求，並實現企業目標的過程」。簡單的說，「行銷」是經由交換過程以滿足需求和欲望的社會活動[9]。基本行銷原則如下列所述[10]：

- 1、滿足顧客的需求與慾望：行銷的首要重點，在於滿足顧客的需求與顧客的慾望。
- 2、行銷永續的本質：行銷是一種持續不斷的管理活動，並非一次就做完的決策。
- 3、行銷是連續性步驟：好的行銷是遵循一系列連續性步驟的過程。
- 4、行銷研究的關鍵角色：有效的行銷充分利用行銷研究的結果

來預期與確認顧客的需求與慾望。

- 5、餐飲旅館與旅遊組織間的相互依賴：在此產業中的所有組織，有許多在行銷方面相互合作的機會。
- 6、全體組織內及多部門間的努力：行銷並非只由某個單一部門來全權負責。要想獲得最佳的成果，需要所有的部門或單位共同全力以赴。

旅遊行銷（tourism marketing）的觀念旨在引導一個旅遊產業組織確定目標市場的需要和慾望，並提供業者一種比競爭對手更具效率的行銷管理方式，以及提昇產業效能的經營法則，進而提昇旅遊產品與服務的品質，以取得旅遊市場的潛在商機。旅遊行銷學是為滿足觀光旅遊的消費者的需要，以及達成旅遊組織之經營目標的一種管理哲學，除了引用傳統行銷學的概念與學理基礎之外，並融合觀光產業的獨特性與產品屬性。換言之，旅遊行銷是將行銷的原理應用在旅遊產業之上，而旅遊事業所涵蓋的範圍很廣，除了旅館住宿業、旅行業、餐飲業、運輸業、遊樂服務業、會議規劃等主要產業之外也包括了：金融服務業、藝品店、娛樂事業等周邊的產業，而各事業之間的關係極為密切，也影響彼此的經營[9]，本文導入人工智慧之資訊科技技術，滿足飯店業者及旅客所需，強化服務品質，以利休旅飯店業者永續經營。

第三章 解決的方法及使用的技術

第一節 類神經網路

類神經網路 (Artificial Neural Network, 簡稱為 ANN) 或稱為人工神經網路, 是一種模仿生物神經網路的資訊處理系統, 也是一種計算機系統, 包含了軟體、硬體, 這樣的系統使用了大量邏輯簡單且相連的人工神經元 (Neuron) 來模仿生物神經網路的能力。人工神經元是生物神經元的簡單模擬, 它從外界環境或是其他人工神經元取得資訊, 並經過人工神經元內部簡單的邏輯運算後, 再將其結果輸出到外界環境或是其它人工神經元[4]。近年來, 類神經網路已經被廣泛地應用到許多不同型態的問題領域中, 依其應用領域可分為工業與工程方面、商業與金融方面、科學與資訊方面等等, 相關應用主題有: 辨識 (Recognition)、資料探勘 (Data Mining)、控制 (Control)、最佳化 (Optimization)、預測 (Prediction) 等。根據神經生物學者與醫學學者的研究, 我們腦中有大量的神經元, 而類神經網路發展的主要目的就是為了模擬人腦中神經元的工作方式, 類神經網路發展至今, 已經有數十種知名的網路模式架構, 在眾多的網路架構中, 基本上可分為兩大類型, 一類為實際結果已知的監督式網路 (Supervised Model), 例如: 倒傳遞網路 (Backpropagation neural network, BPNN); 另一類為結果未知的非監督式網路 (Unsupervised Model), 例如: 自組織映設圖網路 (Self-organization Mapping, SOM)、自適應共振理論 (Adaptive Resonance Theory, ART) 等, 其中自適應共振理論因為其架構之簡易, 且採用的機制亦較符合實際的問題需求 (因顧客屬性的不同,

事前無法知道將顧客分成幾類的分類問題), 因此本研究採用非監督式網路中的自適應共振理論來進行顧客屬性分群的解析, 除了可以將既有資料透過屬性來進行聚類解析外, 當有新的資料輸入網路時, 以能迅速地自動對顧客進行近似屬性群的解析。

ART[4]網路是一種非監督式學習網路, 其模式建立的研究在1960年代即已開始, 1976年葛羅斯伯(S.Grossberg)將此模式作了重要整合與擴充提供了正式引介此模式的基礎, 此模型後來又有卡本特(G.A. Carpenter)加以發揚光大。ART一般可分成兩類:ART-1只能用於輸入二元值者, 而較晚發展的ART-2則可用在輸入值為連續值者。近年來又有ART-3的提出。

ART基本原理可追溯自認知學(Cognitive), 人類的記憶系統中儲存著一些已知的事物, 當它記憶新的事物時, 對於以前所記憶的事物又需保留, 有時可能有矛盾, 一個好的記憶系統必需具有二個要求:

- 1、穩定性: 當輸入新的事物, 舊的事物應適當地保留。
- 2、可塑性: 當輸入新的事物, 應快速地學習。

穩定性與可塑性是有些衝突的, ART網路採用「警戒值測試」(vigilance test)來解決此一矛盾, 其基本原理為:

- 1、如果新的事物的特性與「某一個」舊的事物特性夠相似(即通過警戒值測試), 則只修改系統中該舊事物的部份記憶, 使其能同時滿足新舊事物的特性, 使舊的事物可適當地保留, 如此可滿足穩定性的要求。
- 2、如果新的事物的特性與「所有的」舊的事物的特性不夠相似(即不通過警戒值測試), 則系統為此新事物建立全新的記

憶，以快速地學習此新事物，如此可滿足可塑性的要求。

一、網路架構：

輸入層：此層用來輸入訓練範例向量，其處理單元數目依問題而定。使用線性轉換函數，即 $f(x) = x$ ，輸入層中的每一個輸入向量必是 $\{0, 1\}$ 二元值。

輸出層：此層即訓練範例的聚類，其處理單元數目在學習過程之初只有一個，處理單元會隨著學習的進展而逐漸增加，收斂後會穩定在一定的數目，即學習過程結束。

網路連結層：ART 網路中每個輸入層處理單元與每個輸出層處理單元有二條網路連結，說明如下。

- 1、由下而上的加權值網路 w^b ，用來計算範例的輸入向量對一個輸出處理單元的「匹配值」，公式表示如下：

$$\text{net}[j] = \sum_i w^b[i][j] \cdot X[i] \quad (1)$$

i ：代表第 i 個輸入單元

j ：代表第 j 個輸出單元

匹配值：

「匹配值」高的輸出處理單元將會優先被用作「警戒值測試」， w^b 的值為 $[0, 1]$ 連續值。

- 2、由上而下的加權值網路 w^t ，用來計算範例的輸入向量對一個輸出層處理單元的「相似值」之用，公式表示如下：

$$V_j = \frac{\sum_i w^t[i][j] \cdot X[i]}{\sum_i X[i]} \quad (2)$$

i ：代表第 i 個輸入單元

j : 代表第 j 個輸出單元

相似值 :

「相似值」則用來判定一個輸出處理單元是否通過「警戒值測試」。 w^t 的值為 $\{0, 1\}$ 二元值。

當網路達收斂學習完畢後，全部的輸入層處理單元與每一個輸出層處理單元所連結的 w^t 加權值所構成的向量，表示一個訓練範例對應聚類的型態 (Pattern)，而 w^t 與 w^b 的關係式如下：

$$w^b = \frac{w^t_{ij}}{0.5 + \sum w^b_{ij}} \quad (3)$$

i : 代表第 i 個輸入單元

j : 代表第 j 個輸出單元

網路架構如圖 3-1 所示：

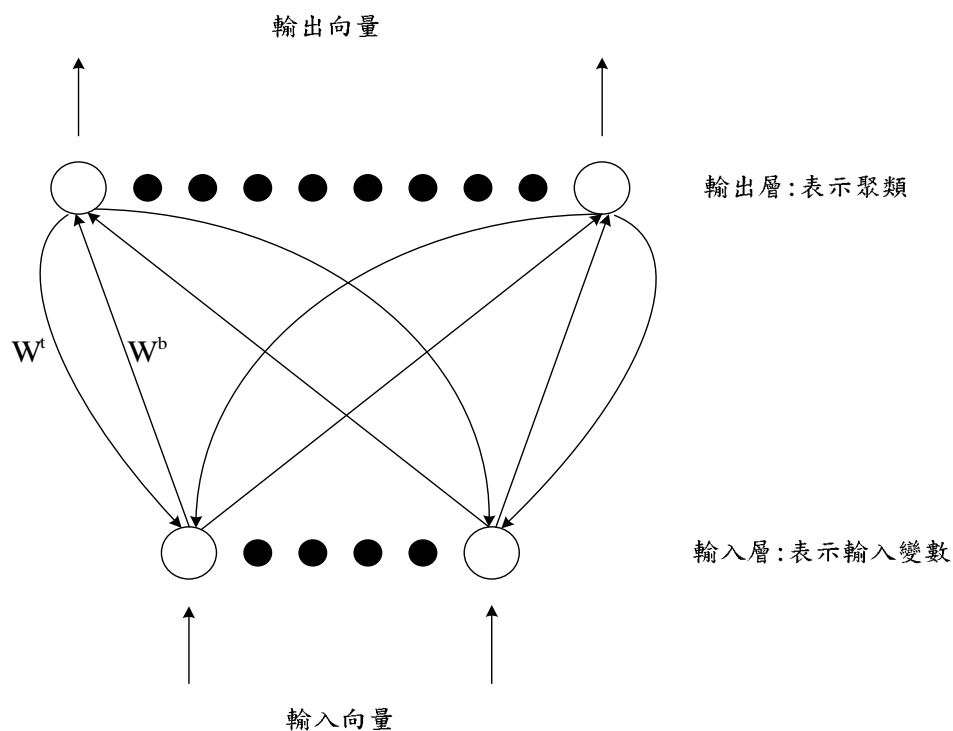


圖 3-1 ART 網路架構

二、學習過程：

- 1、設定輸出層處理單元數為 1
- 2、設定 w^t 、 w^b 加權初始值
- 3、輸入一個訓練範例的輸入向量
- 4、計算匹配值
- 5、找出匹配值最大的輸出層處理單元
- 6、計算匹配值最大的輸出層處理單元之相似值
- 7、測試警戒值

假如 相似值 < 警戒值

則 到步驟 8

否則 到步驟 9

- 8、測試是否還有可用之輸出層處理單元

假如 有 (表示還有輸出層處理單元可供測試相似值)

則 到步驟 5

否則 (表示已無輸出層處理單元可供測試相似值)

(1) 產生新類別

(2) 設定新加權值 w^t 及 w^b

(3) 回到步驟 3

- 9、修正加權值

(1) 修正匹配值最大的輸出層處理單元之加權值 w^t 及 w^b

(2) 如果此時正好完成一學習循環，而且在此循環內未產生

新類別

則 輸出結果而且終止執行

否則 回到步驟 3

顧客群聚產生是利用輸入顧客屬性資料之向量，透過 ART 計算以學習範例的內在聚規則，以觀察其群聚情況（那些值得投資、開發），以及發現其顧客消費型態（可擬定行銷方案、策略）。

第二節 形心

群聚完成後，每一群聚計算其形心，每一群聚之形心代表每一群之中心座標，用來代表每一群，形心用來與新顧客之輸入向量作相似性的計算，以比對其相似性。各群聚中每一個質心座標的每一個向量質心計算如下（4）式：

$$\frac{\sum_{j=1}^{n_k} X_{ijk}}{n_k} \quad (4)$$

n_k ：第 k 群聚中之範例數

X_{ijk} ：第 k 群聚中，第 j 個範例，第 i 個向量值。

$\sum_{j=1}^{n_k} X_{ijk}$ ：第 k 群聚中，所有範例，第 i 個向量值總合。

第三節 歐氏距離 (Euclidean distance)

比對相似性一般可用二個樣本間或各範例間之輸入向量之歐氏距離[4]，新顧客之輸入向量與各群聚之形心座標，做歐氏距離之計算，一個新顧客對每一群聚會產生一個歐氏距離，假如有 5 個群聚，則一個新顧客會對應 5 個歐氏距離（每一群聚一個），在所有的歐氏距離中取最小的歐氏距離，其最小歐氏距離所對應的群聚即新顧客與該群有最高的相似性，該新顧客就歸屬於此群聚中。

新顧客對各群聚的歐氏距離：

$$d_k = \sqrt{\sum (X_i - \frac{\sum_{j=1}^{n_k} X_{ijk}}{n_k})^2} \quad (5)$$

k ：代表第 k 群。

X_i ：代表新顧客屬性值編碼後，第 i 個向量值。

第四節 顧客重住率

對於舊顧客的投資、開發，使用顧客重住率，觀察那些群聚是值得投資、開發，本文是利用顧客屬性中之「住宿頻率」，計算顧客重住率，以保留潛在有用之舊顧客群，顧客重住率如 (6) 所示，如何判斷值得投資、開發的舊顧客群如 (7) 所示。

第 i 個屬性群的重住率 (6)

= (第 i 個屬性群中顧客重住的個數 / 第 i 個屬性群的總顧客數)

整體的重住率

= (整體顧客重住的個數 / 整體總顧客數)

如果 (第 i 個顧客屬性群的重住率 > 整體顧客的平均重住率) (7)

則

第 i 個顧客屬性群將是最適合進行資源投資的屬性群 (其結果也有可能是多個屬性群，原則是重住率越高者越值得開發)

否則

- 1、將各顧客屬性群的顧客重住率由大至小做排序。
- 2、將具有最大的顧客重住率之屬性群視為最適合進行投資的顧客群。

第五節 基因演算法 (Genetic Algorithm, GA)

達爾文 (Darwin Charles) 於 1859 年，在他的著作物種源起 (The

Origin of Species) 中，提出生物進化的理論「物競天擇，適者生存」，說明大自然的生物為了適應環境的變化，透過複製 (Reproduction)、交配 (Crossover) 與突變 (Mutation) 來繁衍適合生存的下一代。基因演算法基本理論是由 John Holland 於 1975 年首先提出[5]，它是基於自然選擇過程的一種最佳化搜尋機制。其基本精神在於仿效生物界中物競天擇、優勝劣敗的自然進化法則，它能夠選擇在母代上具有較好特性的物種，並且隨機性的相互交換彼此的位元 (基因) 資訊，以期望能產生較上一代更優秀的子代，如此重覆下去以產生適應性最強的最佳物種，它是增加個體對環境的適應性來解決問題的演算法。

基因演算法基本的精神[6]為針對所產生的問題，設計適應函數 (Fitness Function)，所謂的適應函數如同一般數理分析中的應變數 (Response Variable)，一開始設定一個集合，此集合叫作母體或叫作族群 (Population)，集合內有 N 個染色體，每個染色體由許多基因 (自變數) 組成， N 個染色體個數可由使用者自定且隨機產生，在每個染色體中，染色體的自變數會先經過編碼，因為每個染色體即是一組編碼，而每一組編碼對應一個應變數，即每一個應變數會對應一個染色體，然後選取適應函數的最佳值即應變數最佳值的染色體放入交配池 (Gene groups) 中；此過程為複製過程，再經由交配 (Crossover) 及突變 (Mutation) 的過程即完成一個世代的基因演算法則，此過程重覆直到滿足終止條件，以產生最適性的子代，即找到滿足應變數的染色體 (編碼)。

一、在基因演算法領域中，有一些專有名詞需先做說明：

1、基因 (Gene)：相當於決策的自變數 (independent variable)，

編碼的最小單位，實際數值因編碼方式不同而異。

- 2、染色體 (Chromosome String): 是由基因所組成的集合，它代表尋找最佳解過程中的一個解。
- 3、母體或族群 (Population): 是由許多染色體所構成的集合，此集合存在尋找最佳解過程中暫時的解，母體或族群的大小可以由使用者自訂。
- 4、世代 (Generation): 從隨機產生母體或族群並計算適合度 (Fitness)，再經由複製、交配、突變過程即完成一世代。
- 5、適應函數 (Fitness Function): 它是一個評估函數，目的是為了賦於每個染色體一個評估值，這個評估值即代表該染色體是否適合繼續生存的量度。
- 6、適應度：即適應函數 (Fitness Function) 所產生的值。

二、而引用基因演算法的參數設定說明如後：

1、編碼型態 (Encoding)

編碼型態隨著問題的不同而有所不同，常見的編碼方式有二進位型式 (Binary) 及非二進位 (Non-Binary) 型式，可依問題的特性、結構選擇。

2、族群大小 (Population Size)

設定的族群太小，較難達到預期成果，收斂較早；族群太大，則會消耗較長的計算時間，通常族群的大小為 30 至 200。

3、交配機率 (Crossover Rate)

控制染色體交配的機率，此參數可由使用者自行調整，其值界於 0, 1 之間，它的功能為控制新物種進入族群的速度，設定太低，收斂太快；設定太高，則會消耗較長的計算時間

才能收斂。

4、突變機率 (Mutation Rate)

突變的頻率視突變機率的大小而定，一般來說，突變機率都定得很低，因為定得太高則與隨機搜尋無異，其功能是為了防止某些新物種沒機會進入族群。

5、適應函數的設計 (Fitness Function)

適應函數設計的原則是根據求解的條件來設計，須能反應出不同物種間適應程度的差異，也須能將次佳物種快速地淘汰，以加速搜尋最佳解的過程。

6、終止條件 (Stopping Conditions)

為了結束基因演算法搜尋最佳解的循環，所設定的停止條件，常見的停止條件有：

- (1) 達到使用者設定的演化世代數。
- (2) 設定固定演化時間。
- (3) 設定無法達到更優解的次數。

三、基因演算法演算程序

複製 (Reproduction) 是依據每一物種的適應程度來決定下一子代是否被複製或被保留的一個運作的過程，如果物種適應程度較高，物種將大量被複製在下一子代中；如果物種適應程度較低，物種將會在下一子代中消失而被淘汰。而適應程度是由適應函數的計算而獲得。複製過程已知有三種型式可供使用：(a) 輪盤法 (Roulette Wheel Selection) (b) 競爭法 (Tournament Selection) 及 (c) 等級輪盤法 (Rank Based Wheel Selection) [6]，經由複製過程將物種的基因放入交配池準備交配。

交配 (Crossover), 此過程是隨機地選取交配池中兩個母代的物種字串 (即染色體), 彼此交換基因的資訊, 進而組合產生另外兩個新物種的染色體, 交配的過程是為了累積前代的優秀的染色體, 期望能產生更優秀的子代染色體。交配過程已知有六種型式可供使用 [6]: (a) 單點交配 (One-Point Crossover) (b) 雙點交配 (Two-Point Crossover) (c) 均勻交配 (Uniform Crossover) (d) 部份對應交配 (Partially Matched Crossover, PMX) (e) 順序交配法 (Order Crossover) 及 (f) 循環交配法 (Cycle Crossover)。

突變 (Mutation), 此過程是隨機的選取一染色體且隨機地選取突變點並改變染色體中的基因, 突變過程已知有五種型式可供使用 [6]: (a) 順序導向突變 (Order Based Mutation) (b) 反轉突變 (Inversion Mutation) (c) 位置導向突變 (Position Based Mutation) (d) 結合突變 (Splice) 及 (e) Last-will-be-First (Lwbf) Mutation, 突變過程為了防止收斂過早, 開發新的搜尋區域, 防止收斂於局部最佳解。

基因演算法應用流程, 詳如圖 3-2 所示。

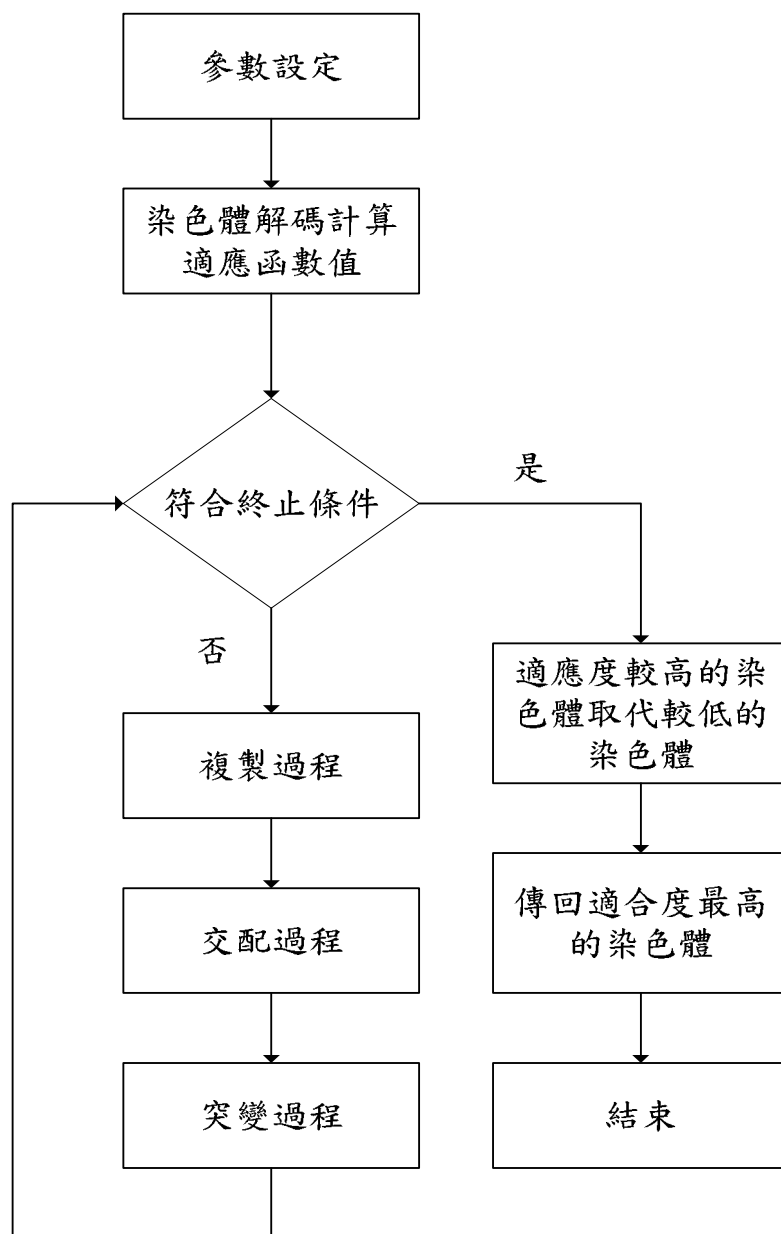


圖 3-2 基因演算法應用流程

第四章 問題解析程序

第一節 新、舊顧客問題解析程序

顧客資料存放於資料庫中，大量的資料必須經過處理才能產生對休旅飯店有利的行銷知識，對於飯店之舊顧客而言，如何投資及開發有利的潛在顧客？以及如何知悉那些顧客有那些消費行為？針對那些潛在的顧客消費習性制定行銷方案、策略，成為休旅飯店提升競爭優勢的必要手段之一，圖 4-1 是針對舊客戶保留的解析程序架構圖；而對休旅飯店業者來說的新顧客，我們會希望能夠很迅速地瞭解其消費屬性，再透過其消費屬性來提供最適切的行銷策略給這些新顧客，圖 4-2 是新顧客推薦架構圖，詳細內容說明如下：

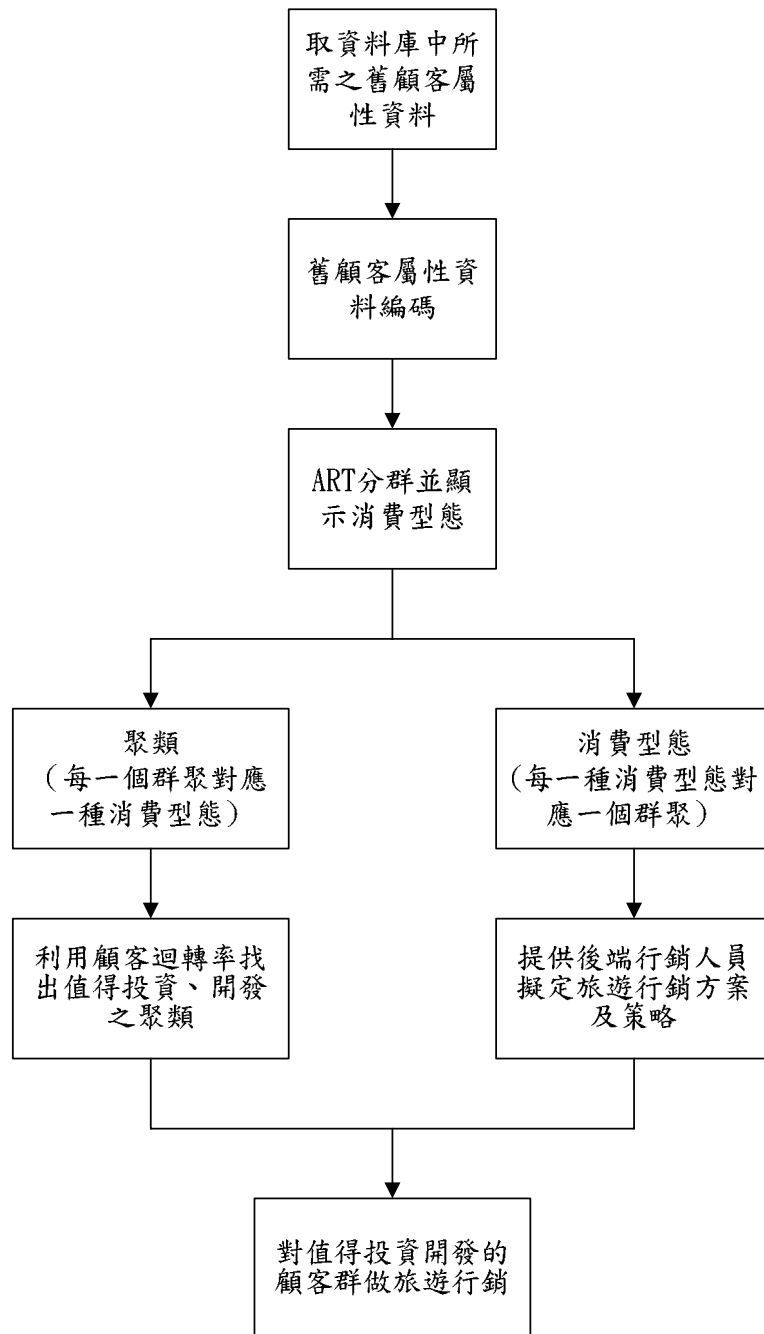


圖 4-1 舊顧客保留解析程序架構圖

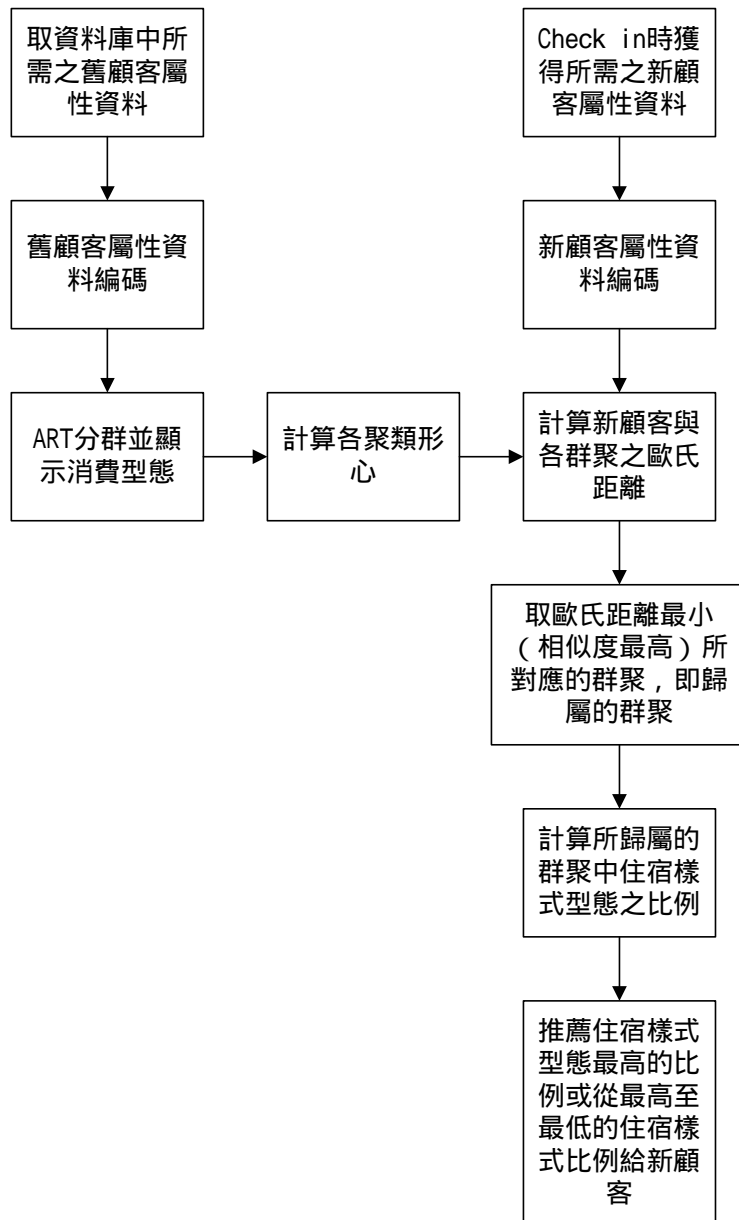


圖 4-2 新顧客推薦架構圖

下列是針對主題一（挖掘具開發投資價值的潛藏舊有的顧客屬性群）及主題二（提供新顧客最適性的行銷資訊）的處理步驟程序。

主題一、挖掘具開發投資價值的潛藏舊有的顧客屬性群：

步驟一：經由前處理過程，從顧客屬性資料庫中將所需的資料欄位及資料取出。

步驟二：將步驟一處理後的顧客屬性欄位資料進行編碼，並輸入至 ART 中，透過 ART 網路的群聚特性產生顧客屬性群聚。

步驟三：計算整體之重住率，以及各顧客群聚之顧客重住率（重住率的公式如公式（8）），接著再判斷那一屬性群或那些屬性群是值得飯店進行投資與開發（相關的判斷規則如公式（9）所示）。

第 i 個屬性群的重住率 （8）

=（第 i 個屬性群中顧客重住的個數/第 i 個屬性群的總顧客數）

整體的重住率

=（整體顧客重住的個數/整體總顧客數）

如果（第 i 個顧客屬性群的重住率 $>$ 整體顧客的平均重住率） （9）

則

第 i 個顧客屬性群將是最適合進行資源投資的屬性群（其結果也有可能是多個屬性群，原則是重住率越高者越值得開發）

否則

- 1、將各顧客屬性群的顧客重住率由大至小做排序。
- 2、將具有最大的顧客重住率之屬性群視為最適合進行投資的顧客群。

步驟四：將飯店值得投資、開發的顧客屬性群取出，並解析每個取出的屬性群對應之消費型態，飯店行銷人員便可依所獲得的資訊制定相關行銷方案與策略。

主題二、提供新顧客最適性的行銷資訊：

步驟一：經由前處理過程，從顧客屬性資料庫中將所需的資料欄位及資料取出。

步驟二：將步驟一處理後的顧客屬性欄位資料進行編碼，並輸入至 ART 中，透過 ART 網路的群聚特性產生顧客屬性群聚。

步驟三：計算各個顧客屬性群的形心結果，再將新顧客屬性資料進行編碼轉換，計算新顧客與各顧客屬性群形心之歐氏距離，找出最小之歐氏距離，新顧客即歸屬於該屬性群。

步驟四：整理分析該屬性群中行銷考量的屬性資料之比率，其具有最高比例者將是推薦給新顧客的最適結果，亦可依據屬性比率的排序訂出行銷推薦的優先順位。

第二節 最適行程問題解析程序

當顧客住進飯店，欲前往飯店附近的旅遊景點遊玩之前，通常會詢問飯店櫃台人員一些有關旅遊行程規劃的問題。例如：如何搭車是較為方便的行程規劃？怎麼的行程規劃才是最省時（針對移動

時間) 最經濟的(針對移動成本)? 雖然這些問題是一般服務人員經常性面臨的問題, 但這些有關旅遊最適性的議題, 對於飯店業服務人員來說, 將會是一個不易提供即時性資訊的狀況。一般來說, 休旅飯店或中心通常只能提供附近旅遊景點的點對點間移動時間預估表及搭車前往旅遊景點的兩旅遊景點車費。但是顧客的需求即有可能是 (a) 怎樣的行程規劃才能從飯店出發用最少的移動時間到達各景點遊玩再回到飯店休息? (b) 怎樣的行程規劃才能從飯店出發用最少的移動費用到達各景點遊玩再回到飯店休息? (c) 怎樣的行程規劃才能從飯店出發用較少移動時間及較少移動費用到達各景點遊玩再回到飯店休息? 如果飯店或是休旅中心能夠提供針對上述顧客要求的最佳旅遊行程路徑, 必能提升飯店對顧客的服務品質。為了解決這種最適性化行程規劃之問題, 我們嘗試引入基因遺傳演算法, 並提供一個整合性的問題解析程序, 說明如下, 圖 4-3 是最適性行程規化架構圖, 圖 4-4 為最適性行程問題解析程序步驟。

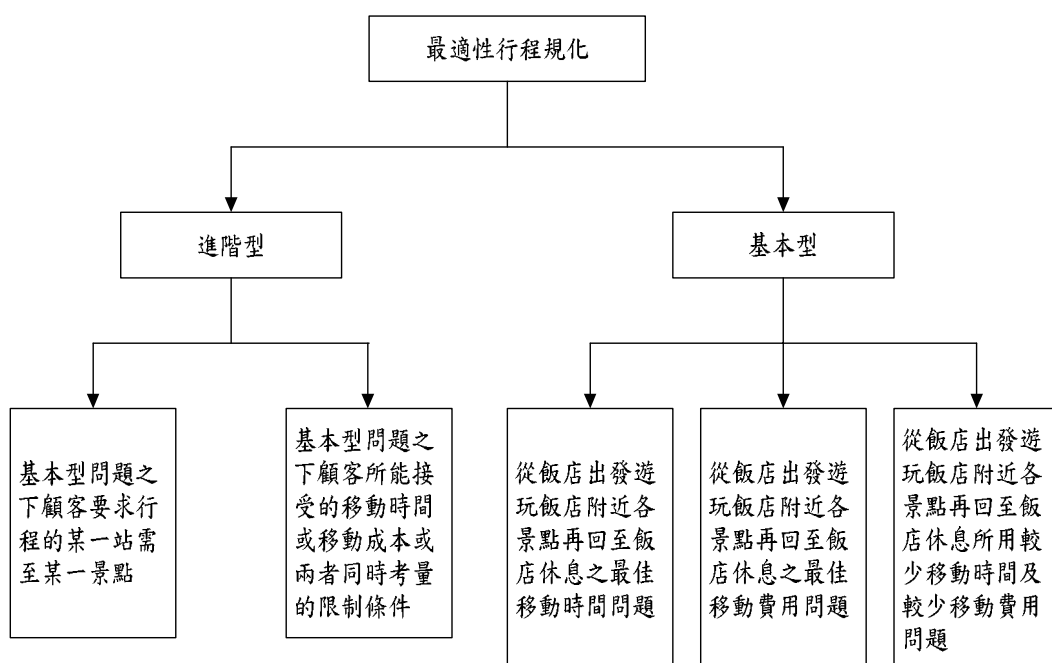


圖 4-3 最適性行程規化架構圖

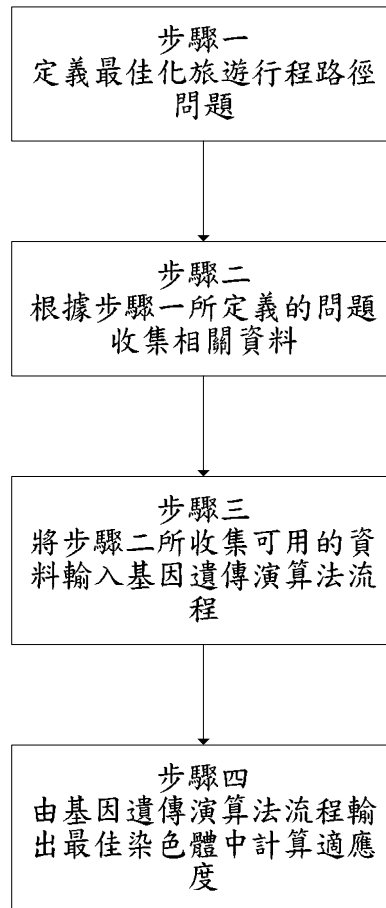


圖 4-4 最適性行程問題解析程序步驟

步驟一：定義最佳化旅遊行程路徑問題，如移動時間或移動成本的問題。

步驟二：根據步驟一所定義的問題收集相關資料，如兩景點間移動時間或成本。

步驟三：將步驟二所收集可用的資料輸入基因遺傳演算法流程。

步驟四：由基因遺傳演算法流程輸出最佳染色體（編碼，即最佳旅遊行程徑路）中計算適應度（如移動總時間、移動總成本），提供顧客參考。

第五章 應用案例展示

第一節 新、舊顧客問題範例展示

本研究配合某飯店業者進行研究程序及方法的可行性分析，從飯店過去歷史性的顧客交易資料中共篩選蒐集約 30000 筆資料（將一些資料記錄不完整者予以剔除），接著將從透過這些資料來解析出：

主題一：

- (a) 潛在的顧客屬性群。
- (b) 透過評估指標來判斷最值得投資與開發的顧客屬性群。

主題二：

- (c) 根據既有的顧客屬性群資料對新顧客提供最適切的行銷建議。

接著將依據本研究的解析程序詳細予以說明。

主題一、挖掘具開發投資價值的潛藏舊有的顧客屬性群：

步驟一：由顧客屬性資料庫，將所需的資料欄位及資料取出處理，表 5-1 是取出處理的顧客資料屬性、對應之屬性值及進行編碼所需之向量數（因取出的資料在前置處理後均為類別型態的資料，故向量數即代表進行編碼所需的位元組），表 5-2 是針對年齡屬性作編碼的說明，其餘各類屬性以此類推。

表 5-1、顧客屬性對應之屬性值及編碼所需向量數

屬性	年齡	性別	職業	住宿樣式	住房頻率	住房月份 (季節)	折扣	住宿樣式 型態
屬性 值	11 至 20 21 至 30 31 至 40 41 至 50 51 至 60 61 至 70 71 以上	男 、 女	士 、 農 、 工 、 商	標準單人 標準雙人 標準家庭 豪華單人 豪華雙人 豪華家庭	1 次 2 至 3 次 4 次以上	1 至 3 (冬) 4 至 6 (春) 7 至 9 (夏) 10 至 12 (秋)	是 、 否	中式 西式 和式 歐式
編碼 數	7	2	4	6	3	4	2	4

表 5-2、年齡屬性編碼設定範例

年齡	ART 網路輸向量編碼	編碼屬性值
11 至 20	向量一	對應每筆記錄成立設為 1 否定 設為 0
21 至 30	向量二	對應每筆記錄成立設為 1 否定 設為 0
31 至 40	向量三	對應每筆記錄成立設為 1 否定 設為 0
41 至 50	向量四	對應每筆記錄成立設為 1 否定 設為 0
51 至 60	向量五	對應每筆記錄成立設為 1 否定 設為 0
61 至 70	向量六	對應每筆記錄成立設為 1 否定 設為 0
71 以上	向量七	對應每筆記錄成立設為 1 否定 設為 0

步驟二：將處理後的顧客屬性欄位資料編碼，帶入 ART 網路中進行群聚分析，在循環次數 2 次後可達收斂，其所使

用的警戒值為 0.75 (進行模式建立時 , 警戒值選取範圍從 0.5 至 0.9 , 增量為 0.5 , 經過多次的試誤後 , 得此適切之警戒值 , 系統執行畫面如圖 5-1 所示) , 可以產生五種顧客屬性群 , 表 5-3 是經過 ART 解析後的屬性群及對應之型態。

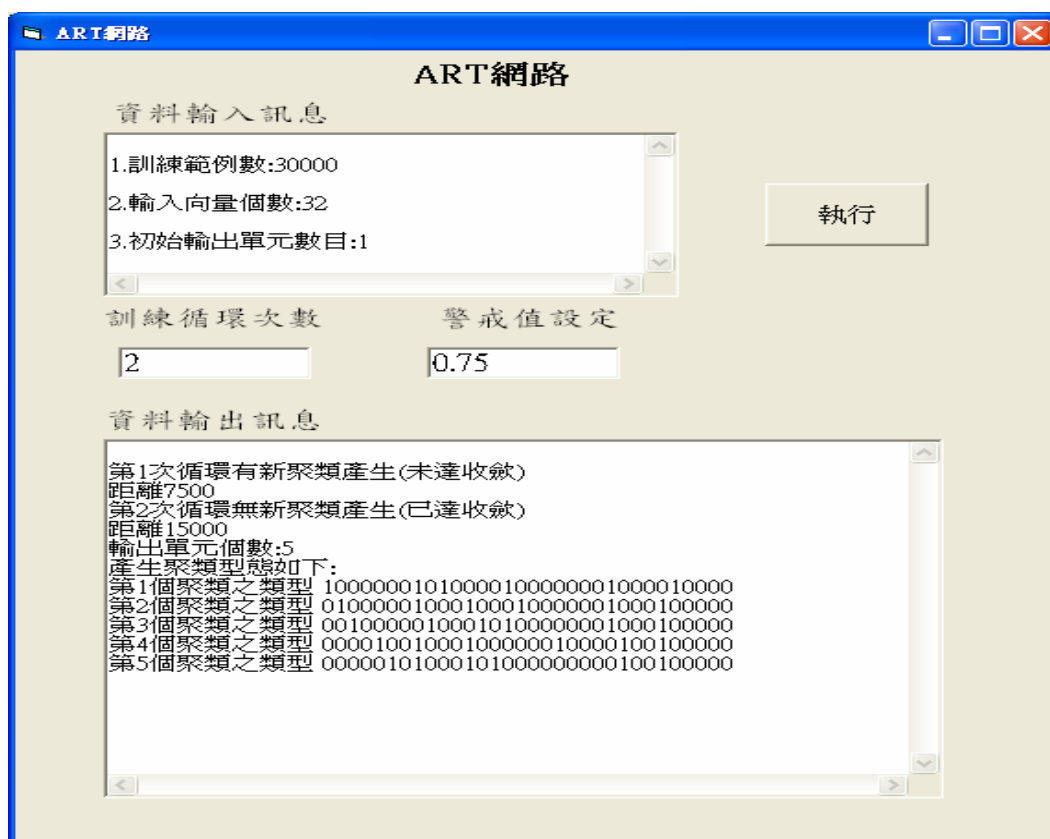


圖 5-1 ART 網路分群結果

表 5-3 顧客屬性群型態解析 (續)

屬性群	編碼型態	對應屬性型態解析	顧客人數
第 1 類型	1 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0	11 歲至 20 歲、男性、 職業為士、住標準雙人 客房、住房月份 1 至 3 月、未用折扣。	3722
第 2 類型	0 1 0 0 0 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0	21 歲至 30 歲、男性、 職業為工、住標準家庭 客房、住房月份 1 至 3	7425

		月、使用折扣。	
第 3 類型	0 0 1 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0	31 歲至 40 歲、女性、 職業為商、住標準雙人 客房、住房月份 1 至 3 月、使用折扣。	3742
第 4 類型	0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0	51 歲至 60 歲、男性、 職業為工、住豪華家庭 客房、住房月份 4 至 6 月、使用折扣。	7477
第 5 類型	0 0 0 0 0 1 0 1 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0	61 歲至 70 歲、男性、 職業為工、住標準單人 客房、住房月份 4 至 6 月、使用折扣。	7634

步驟三：利用住房頻率指標來計算整體之顧客重住率，以及各顧客群聚之顧客重住率，接著判定那一屬性群或那些屬性群是值得飯店進行投資與開發，表 5-4 是整體、各屬性群的重住率，在表 5-4 中第 2、5 屬性群的重住率值已明顯地超過整體的平均重住率值，由第 (9) 式可知落在表 5-4 中第 2、5 群的顧客是最具投資與開發價值者。

表 5-4 整體及各聚類重住率

聚類	重住人數	人數	重住率
整體 (不屬於任何聚類)	15059	30000	0.50196
第 1 類型	1220	3722	0.327781
第 2 類型	6205	7425	0.83569*
第 3 類型	560	3742	0.149653
第 4 類型	783	7477	0.104721
第 5 類型	6291	7634	0.824076*

步驟四：這兩個最值得投資與開發的屬性群，對應表 5-3 中的「對應屬性型態解析」欄位，其結果為「21 歲至 30 歲、男性、職業為工、住標準家庭客房、住房月份 1 至 3 月、使用折扣」(第二屬性群)及「61 歲至 70 歲、男性、職業為工、住標準單人客房、住房月份 4 至 6 月、使用折扣」(第五屬性群)。休旅飯店行銷人員，即可利用所獲得的資訊進行可行的行銷方案、策略。

由上述四個步驟，即可挖掘出問題描述中：(a) 潛在顧客的屬性群聚；(b) 值得投資、開發的潛在顧客，以利飯店的行銷人員對行銷方案、策略的制定，而提高飯店的競爭力。

主題二、提供新顧客最適性的行銷資訊：

步驟一：同主題一問題步驟解析的步驟一。

步驟二：同主題一問題步驟解析的步驟二。

步驟三：計算各顧客屬性群的形心結果，接著依據編碼原則將新顧客資料進行資料編碼(如表 5-5 所示)，再計算新顧客型態資料與各屬性群形心之歐氏距離，找出歐氏距離值最小的屬性群，該新顧客便可歸屬於此屬性群，本例新顧客的屬性資料如表 5-5 所示，與各屬性群形心的歐氏距離結果如圖 5-2 所示，從圖 5-2 可以判斷出新顧客與顧客屬性群二之歐氏距離最小，所以新顧客可歸屬於第二屬性群。

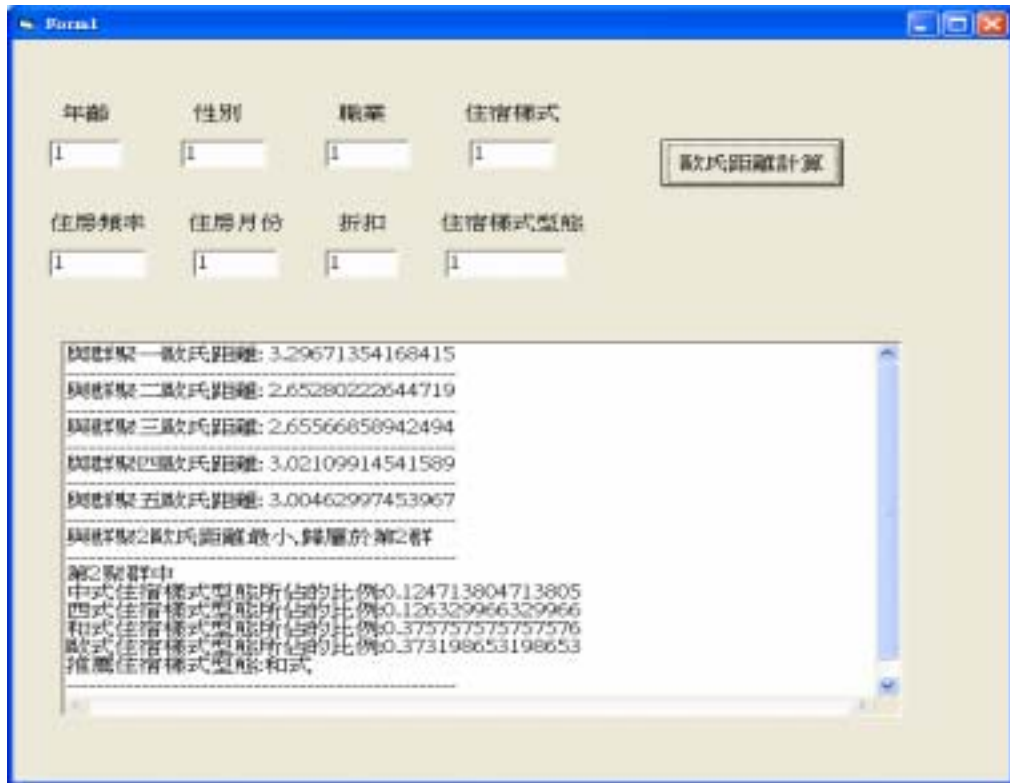


圖 5-2 歐氏距離計算結果與推薦住宿樣式型態

表 5-5 新顧客消費型態及說明

新顧客消費型態	說明
0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 0	年齡為 31 至 40 歲、男性、職業為商、要住標準家庭客房、因為是第一次住房所以住宿頻率為未知、住房月份 7 至 9 月、使用折扣、第一次住房可能住宿樣式型態未定。

步驟四：本案例中行銷人員要進行行銷的考量因素是針對顧客住房型式的最適建議，故在進行最適住房型式建議時，需計算該顧客所屬之顧客屬性群中各住房型式的比率值，選擇比率值最高的住房型式為最適建議。因該顧客是屬於第二屬性群，其住宿型式中具最高比率者為和式住宿（比例如表 5-6 所示），故將其視為該顧

客最適推薦建議；除了可提供最適的建議外，如果新顧客無法認同此最適建議，行銷人員可以透過表 5-6 的比率值高低排訂出住房建議之優先順序（和式 歐式 西式 中式）。

表 5-6 第二群聚中各住宿樣式型態所佔比例

住宿樣式型態	所佔比例（排序順位）
中式	0.1247 (4)
西式	0.1263 (3)
和式	0.3757* (1)
歐式	0.3731 (2)

第二節 最適性行程範例說明

一、問題描述

假設某一家休旅飯店業者周邊有五個知名觀光景點（如表 5-7），各景點間的移動時間（以一般的大眾交通工具，如公車或是計程車，參考數據資料如表 5-8），各景點間搭乘大眾交通工具所需之移動費用（如表 5-9），接下來，我們將展示如何利用本研究所提出的解析程序來針對這兩種類型的行程進行最適性規劃。

表 5-7 飯店附近旅遊景點

景點編號	景點名稱
1	飯店
2	市集
3	海灣
4	觀海點
5	公園
6	文物館

表 5-8 飯店預估兩旅遊景點之間的移動時間（分鐘）

景點編號	1	2	3	4	5	6
1	0	70	65	30	20	75
2	70	0	20	25	40	55
3	65	20	0	20	20	45
4	30	25	20	0	30	40
5	20	40	20	30	0	35
6	75	55	45	40	35	0

表 5-9 兩旅遊景點移動費用（元）

景點編號	1	2	3	4	5	6
1	0	40	40	20	15	45
2	40	0	15	20	30	35
3	40	15	0	17	17	37
4	20	20	17	0	28	33
5	15	30	17	28	0	35
6	45	35	37	33	35	0

二、基本型最適性行程規劃

1、步驟解析

首先以 (a) 怎樣的行程規劃才能從飯店出發用最少的移動時間到達各景點遊玩再回到飯店休息為例來展示最佳路徑規劃之解析。

- (1) 首先設定族群大小、交配機率、突變機率、終止條件、編碼（染色體），而編碼採非二進位（Non-Binary）型式，其形態為：

2	3	4	5	6
---	---	---	---	---

- (2) 適應函數（fitness function）設計為：

染色體最前及最後加入固定值 1，代表從飯店出發最後

回到飯店休息

1	2	3	4	5	6	1
---	---	---	---	---	---	---

其順序性為{1,2}、{2,3}、{3,4}、{4,5}、{5,6}、{6,1}，對應表 5-8，即{1,2}為 70 代表景點編號 1 至景點編號 2 移動時間為 70 分鐘，而{1,2}、{2,3}、{3,4}、{4,5}、{5,6}、{6,1} (70) + (20) + (20) + (30) + (35) + (75) = 250，即總移動時間需 250 分鐘，在此適應函數為從飯店出發到達各景點遊玩再回到飯店休息總移動時間，其最適性是要求總移動時間最小。

- (3) 再透過基因演算法的機制，複製、交配、突變等過程來修改染色體及染色體的順序性而使得總移動時間最小化，最後染色體已達終止條件的收斂結果可能如下：

4	2	3	6	5
---	---	---	---	---

而順序性為

1	4	2	3	6	5	1
---	---	---	---	---	---	---

而{1,4}、{4,2}、{2,3}、{3,6}、{6,5}、{5,1}所計算出的總移動時間為 175 分鐘，也就是說只要走“1423651”此順序，就可以用最短移動時間遊玩各個旅遊景點，可提供顧客作參考。

接著讓我們針對顧客可能的要求：(a) 怎樣的行程規劃才能從飯店出發用最少的移動時間到達各景點遊玩再回到飯店休息？(b) 怎樣的行程規劃才能從飯店出發用最少的移動費用到達各景點遊玩再回到飯店休息？(c) 怎樣的行程規劃才能從飯店出發用較少移動

時間及較少移動費用到達各景點遊玩再回到飯店休息？利用基因演算法的流程及應用程式，找出最佳化旅遊行程路徑的分析。

2、參數設定

- (1) 編碼型態：非二進位 (Binary) 型式，基因數：5 (因為除了飯店外有 5 個旅遊景點要遊玩，如表 5-7)，隨機從 2、3、4、5、6 數字作排序即編碼動作，可能產生 2、3、4、5、6 或 3、2、4、6、5 等染色體 (編碼)。
- (2) 族群大小 (Population Size): 30
- (3) 交配機率 (Crossover Rate): 0.5
- (4) 突變機率 (Mutation Rate): 0.1
- (5) 適應函數的設計 (Fitness Function):
 - (A) 問題設計為總移動時間最小化。
 - (B) 問題設計為總移動費用 (成本) 最小化。
 - (C) 問題設計為總移動時間較少及總移動費用 (成本) 較少之最小化。
- (6) 終止條件 (Stopping Conditions): 執行 100 世代，即停止運作。

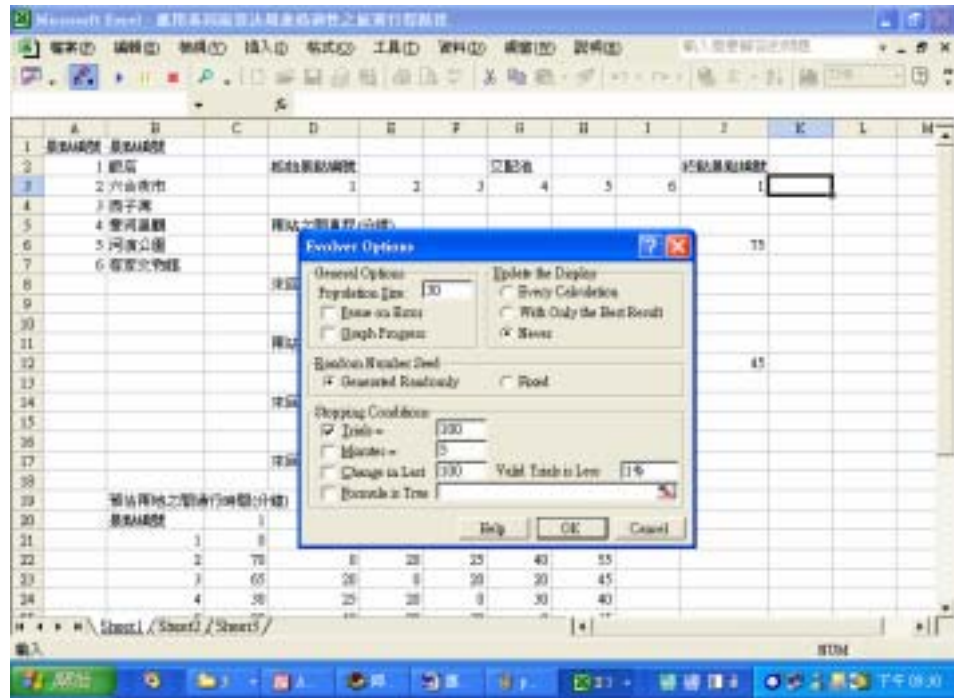


圖 5-3 設定族群大小及終止條件

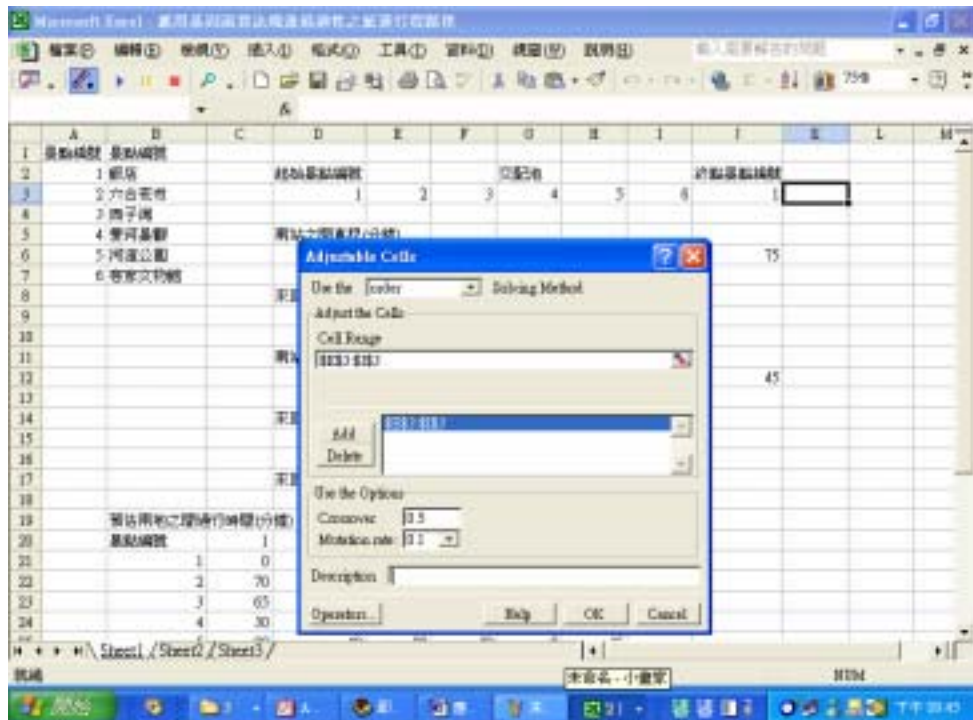


圖 5-4 設定交配機率及突變機率

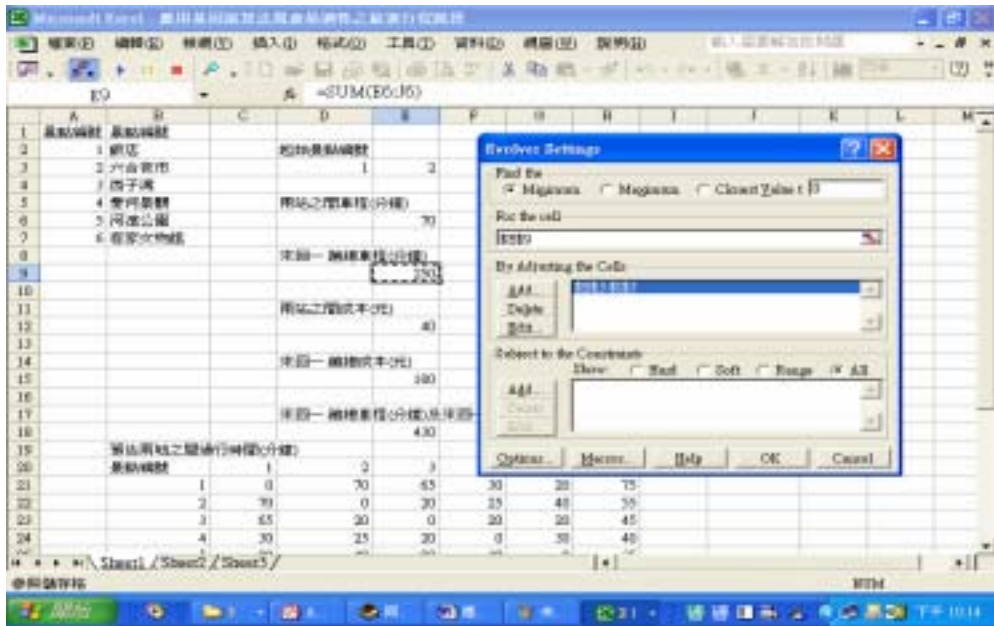


圖 5-5 設定總移動時間最小化

最後經由圖 4-4 的演算流程，計算上述各問題 (a)、(b)、(c) 的最佳解，以提供飯店顧客作旅遊行程路徑規劃參考，以提升飯店的服務品質，分析結果如表 5-10。

表 5-10 基本型最適性行程規劃分析結果

問題	染色體 (編碼)	順序性	適應度	附註
(a)	56324 42365	1563241 1423651	175 分鐘	142 元
(b)	46235 53264	1462351 1532641	135 元	185 分鐘
(c)	43265	1432651	180 分鐘、137 元	

在 (c) 問題中，考慮到飯店服務人員使用上的簡單性，其適應函數將以線性考量設計如下：

$$\text{Fit} = \mathbf{F}(\) + \mathbf{G}(\) \quad (10)$$

Min. Fit

其中，Fit 表示適應函數、F() 表示移動車程時間函數、G() 表示

移動成本函數、 x_{ij} 表示景點 i 和景點 j 之間的移動車程時間、 y_{ij} 表示景點 i 和景點 j 之間的移動成本， $i=1,2,\dots,n$ ， $j=1,2,\dots,n$ ， n 表示景點總數。這種適應函數型式在概念陳述上是可以接受的，但是在適應函數中，總移動時間及總移動成本單位並不相同，因此我們將透過下列方式，將表 5-8、表 5-9 的數據轉換，使適應函數無單位化，避免受到數據尺度大小不一以及因單位不同而產生運算影響。

$$m_{ij} = \frac{x_{ij}}{\text{Max } x_{ij}} , \quad n_{ij} = \frac{y_{ij}}{\text{Max } y_{ij}} \quad (11)$$

故轉換後的適應函數為：

$$\mathbf{Fit}_t = \mathbf{F}_t() + \mathbf{G}_t() \quad (12)$$

Min. \mathbf{Fit}_t

\mathbf{Fit}_t ：表示轉換後的適應函數。

m_{ij} ：表 5-8 轉換後的數值，其值介於 0，1 之間。

x_{ij} ：表 5-8 中所對應的數值， $x_{ij} \geq 0$ 。

$\text{Max } x_{ij}$ ：表 5-8 中最大的數值。

n_{ij} ：表 5-9 轉換後的數值，其值介於 0，1 之間。

y_{ij} ：表 5-9 中所對應的數值， $y_{ij} \geq 0$ 。

$\text{Max } y_{ij}$ ：表 5-9 中最大的數值。

$\mathbf{F}()$ ：表示由起始景點經所有景點再回到起始景點的移動時間函數。

$\mathbf{G}()$ ：表示由起始景點經所有景點再回到起始景點的移動成本函數。

$\mathbf{F}_t()$ ：表示由起始景點經所有景點再回到起始景點轉換後的移動時間函數。

$\mathbf{G}_t()$ ：表示由起始景點經所有景點再回到起始景點轉換後的移動成本函數。

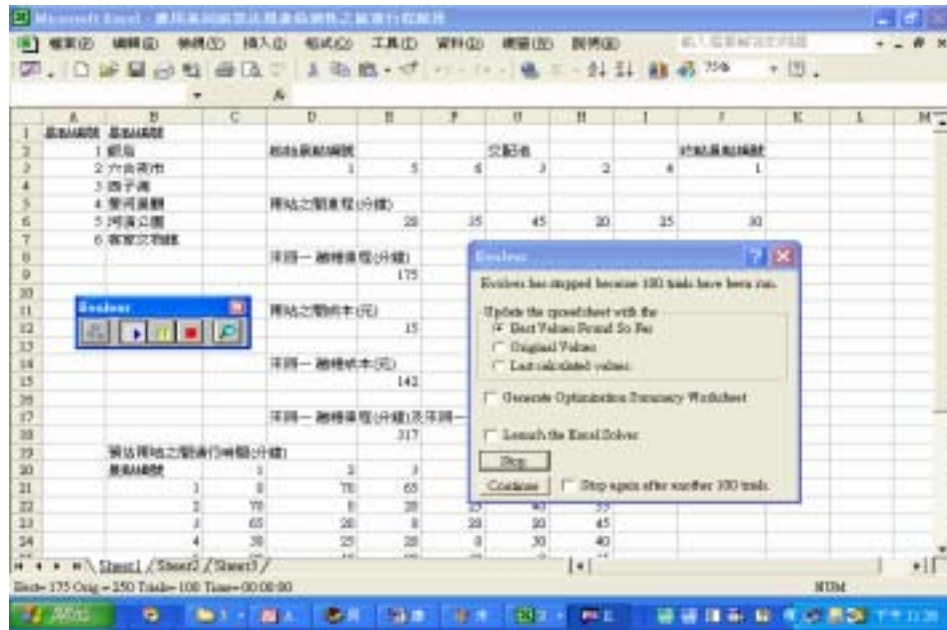


圖 5-6 套裝軟體針對問題 (a) 執行的結果

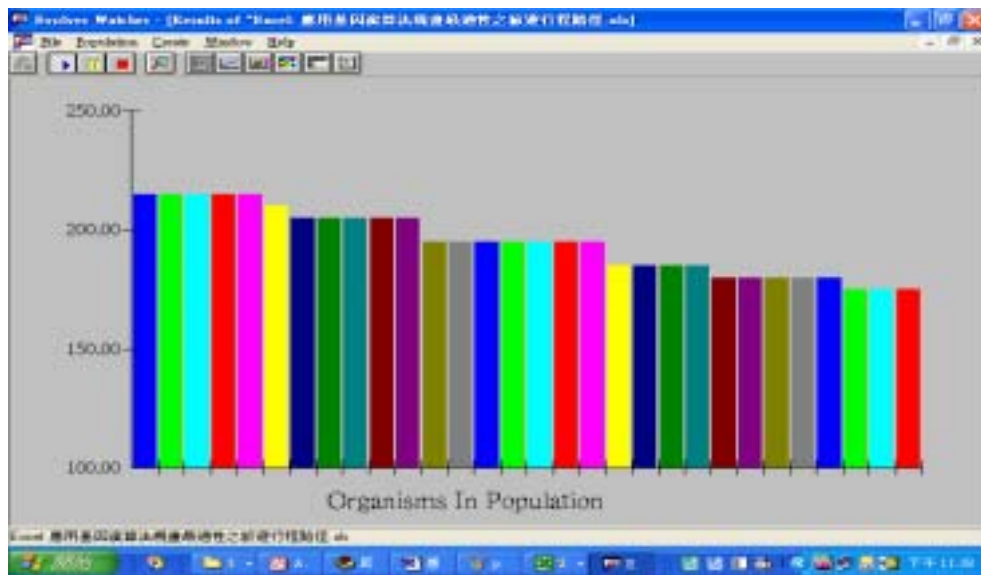


圖 5-7 針對問題 (a) 在最終世代族群中各物種的適應度

由表 5-10 可知針對 (a) 問題而言最後最優的染色體是 56324 或 42365，其對應的順序性即旅遊景點的順序為 1563241 或 1423651，對應表 5-7，即飯店 公園 文物館 海灣 市集 觀海點 飯店或著從飯店 觀海點 市集 海灣 文物館 公園 飯店，此二個路徑是顧客花費最小的移動時間可到達各個景點的路

徑，但搭車走以上二個路徑需花 142 元。對於問題 (b) 顧客可以最經濟的方式 (135 元) 搭車即可玩遍表 5-7 各景點，但須要花 185 分的移動時間，但也產生與問題 (a) 不同的路徑。針對問題 (c) 而言，是對 (a)、(b) 問題取得一個平衡，顧客可花 180 分鐘的移動時間及花 137 元的移動費用玩遍表 5-7 各景點，但路徑與解決 (a)、(b) 問題的路徑不同，以上都可產生最適性的旅遊行程路徑來滿足顧客的要求，並提供顧客作參考，以利顧客對旅程作適當的決策。

三、進階型最適性行程規劃

如果顧客有某些指定條件的需求，例如：(d) 顧客要求行程的某一站需至某景點；以及另一問題 (e) 在顧客能接受的移動時間或移動成本或兩者同時考量的限制條件下，作最適性行程規劃建議，解析步驟、結果如下：

針對 (d) 問題解析之步驟：

步驟一：定義顧客要求行程最適限制性問題 (總移動時間最少；或總移動費用 (成本) 最少；或總移動時間較少及總移動費用 (成本) 較少)。

步驟二：根據步驟一所定義的問題收集相關資料 (表 5-8 及表 5-9)。

步驟三：將步驟二所收集可用的資料輸入基因遺傳演算法流程。

步驟四：由基因遺傳演算法流程輸出最佳染色體中計算適應度 (表 5-11)，提供顧客參考。

現假設顧客要求其行程第一站需先至景點 6 後，再至其他景點遊玩。則利用基因遺傳演算法的程序設定及結果如下：

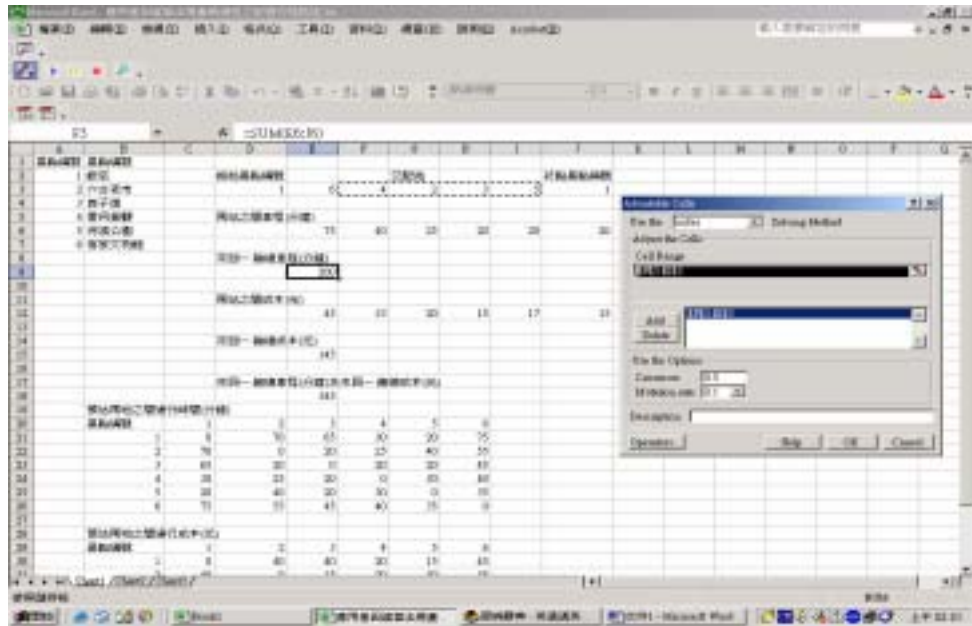


圖 5-8 針對 (d) 問題設定染色體

表 5-11 針對 (d) 問題解析結果

觀點 (第一站至景點 6)	染色體 (編碼)	順序性	適應度	附註
總移動時間最少	64235	1642351	200 分鐘	145 元
總移動費用 (成本) 最少	64235	1642351	145 元	200 分鐘
總移動時間較少及總移動費用 (成本) 較少	64235	1642351	200 分鐘、 145 元	

針對 (e) 問題解析之步驟：

步驟一：定義顧客能接受的移動時間或移動成本或兩者同時考量的限制條件 (假設條件如後：(1) 總移動時間約在 240 分鐘；(2) 總移動費用 (成本) 約為 160 元；(3) 總移動時間約為 180 分鐘及總移動費用 (成本) 約為 140 元)，作最適性行程規化之問題。

步驟二：根據步驟一所定義的問題收集相關資料 (表 5-8 及表 5-9)。

步驟三：將步驟二所收集可用的資料輸入基因遺傳演算法流程。

步驟四：由基因遺傳演算法流程輸出最佳染色體中計算適應度（表 5-12），提供顧客參考。

步驟一之（3）問題考慮到顧客可能偏重總移動時間或偏重總移動費用（成本）或兩者都偏重的關係，故其適應函數修正為：

$$\mathbf{Fit}_m = w_1 * \mathbf{F}_t() + w_2 * \mathbf{G}_t() \quad (13)$$

$$\text{Min. } \mathbf{Fit}_m$$

$$\text{St. } w_1 + w_2 = 1$$

$$w_1 > 0, w_2 > 0$$

其中， w_1 ：表示總移動時間分數之權重、 w_2 ：表示總移動成本分數之權重。此問題可分為三種情況來進行討論：

情況一：兩個選擇具同等重要考量，即 $w_1 = w_2$ ；故取 $w_1 = 0.5$ ； $w_2 = 0.5$

情況二：總移動時間考量較總移動成本為重，即 $w_1 > w_2$ ；故取 $w_1 = 0.8$ ； $w_2 = 0.2$

情況三：總移動成本考量較總移動時間為重，即 $w_1 < w_2$ ；故取 $w_1 = 0.2$ ； $w_2 = 0.8$ 解析結果如表 5-12 所示。

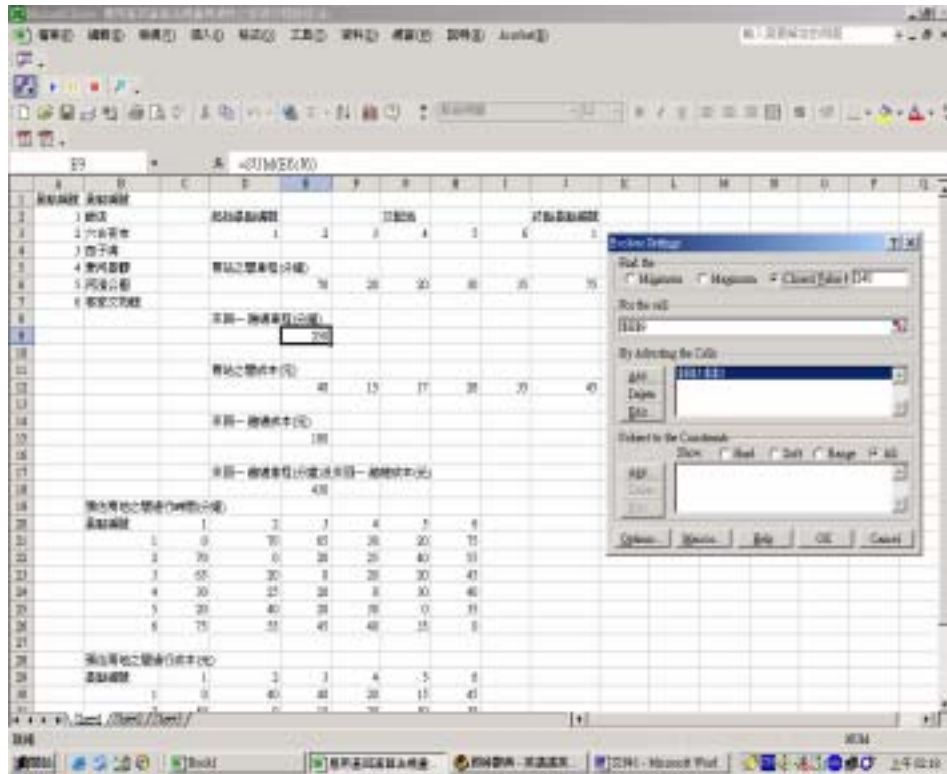


圖 5-9 針對 (e) 問題設定總移動時間之限制條件

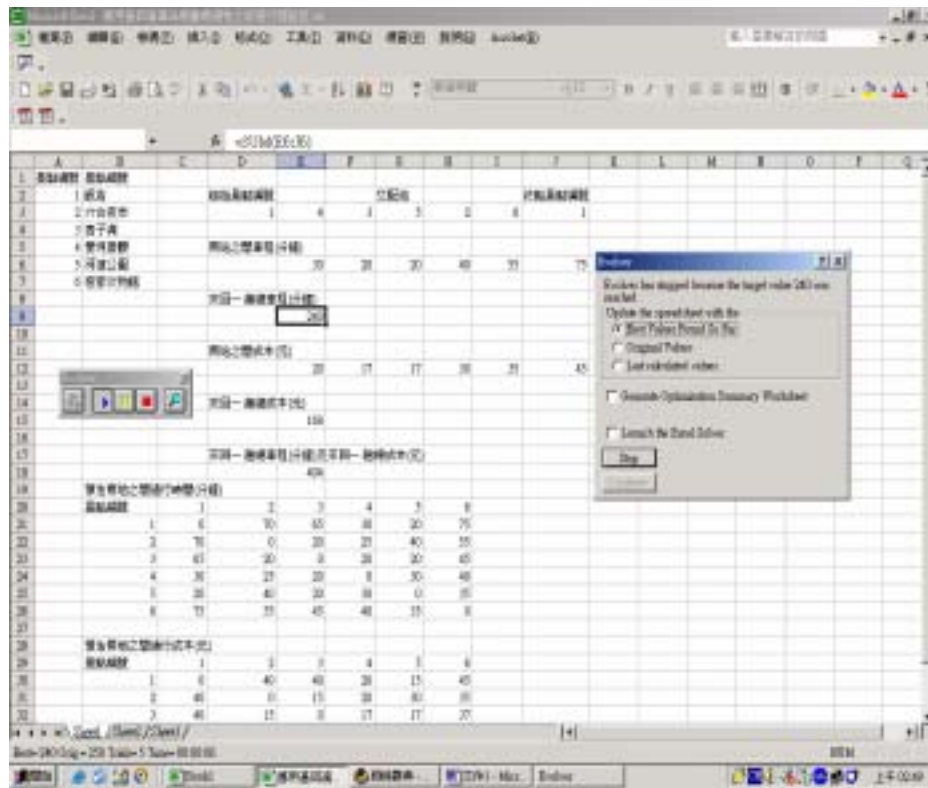


圖 5-10 針對 (e) 問題之總移動時間限制 (240 分鐘) 解析結果其中之一

表 5-12 針對 (e) 問題解析結果

觀點		染色體 (編碼)	順序性	適應度	附註
總移動時間 = 240 分鐘		43526	1435261 *	240 分鐘	164 元
		45236	1452361		175 元
		25634	1256341		179 元
		63254	1632541		175 元
		52643	1526431		170 元
		54362	1543621		172 元
		43652	1436521		179 元
總移動費用(成本)= 160 元		45326	1453261	160 元	230 分鐘
		23564	1235641 *	160 元	215 分鐘
		63245	1632451 *	160 元	215 分鐘
		46532	1465321 *	160 元	215 分鐘
		54236	1542361 *	160 元	215 分鐘
		62354	1623541	160 元	230 分鐘
總移動時間 = 180 分鐘 及總移動費用(成本) = 140 元	情況一	53264	1532641 *	185 分鐘、 135 元	
		42365	1423651 *	175 分鐘、 142 元	
		56324	1563241 *	175 分鐘、 142 元	
	情況二	56234	1562341 *	180 分鐘、 137 元	
		43265	1432651 *	180 分鐘、 137 元	
	情況三	56324	1563241 *	175 分鐘、 142 元	
		42365	1423651 *	175 分鐘、 142 元	

*表示建議之最適性行程規劃

四、最適性行程應用實例範例

休旅飯店業者可應用上述的最適性行程規劃之程序在網路上，如圖 5-11 所示，做為提供旅客在飯店附近景點遊玩最適性行程的推薦以提升飯店之競爭力。



圖 5-11 飯店網站最適性行程規劃應用

點選行程規劃後進入線上最適行程規劃網頁如圖 5-12 所示，可根據使用者考量的因素如移動時間、移動成本、移動時間及移動成本，以建議顧客之最適性行程給使用者做參考。



圖 5-12 線上最適性行程規劃網頁
從「旅遊景點」欄位選擇所需要的旅遊景點至「行程規劃景點」

欄位如圖 5-13 所示,再由使用者考量的因素如移動時間、移動成本、移動時間及移動成本,經由基因遺傳演算法計算出最適行程旅程,以提供使用者作決策參考。



圖 5-13 使用者選擇所需要的旅遊景點

最適性行程規劃中使用者以移動時間考量結果之一如圖 5-14 所示,以移動成本考量結果之一如圖 5-15 所示,考量移動時間及移動成本結果之一如圖 5-16 所示。



圖 5-14 最適性行程規劃中使用者以移動時間考量結果之一



圖 5-15 最適性行程規劃中使用使用者以移動成本考量結果之一



圖 5-16 最適性行程規劃中使用使用者以移動時間及移動成本考量結果之一

在休旅飯店高度競爭的微利時代中，各家飯店業者不斷提升自己本身的服務品質，目前飯店業者在線上最常見只提供線上訂位、訂房等功能，而未推展至可以利用資訊科技的技術，讓旅客選擇旅遊景點並同時考量移動時間、移動成本、移動時間及移動成本之最適性行程，讓飯店業者達成差異化及客製化服務以提升顧客價值，進而提升飯店之競爭力，以提高飯店業者的利潤，促使飯店業者能永續經營。

第六章 結論

在飯店業者高度競爭的微利時代下，保留潛在的顧客及對新顧客的行銷成為各家飯店業者必要的競爭手段之一，本文是利用人工智慧型式的資料挖掘技術，由飯店顧客資料庫中挖掘出顧客消費型態的屬性群，提供後端行銷人員擬定行銷方案、策略，顧客重住率的使用簡單地偵測出可保留的潛在顧客群；對新顧客而言，透過這樣的人工智慧技術可找出它所歸屬的顧客群，並針對所對應的屬性群來推薦適合的消費型態給新顧客，以充分有效地掌握新客源；而當顧客住進飯店後對於週邊景點的旅遊行程規劃與建議，本文也提供另一個智慧型基因遺傳演算法的技術來達成行程最適性規劃，讓顧客可以在基本條件下或是客製化條件下以最少的移動時間及移動成本來滿足其最適性行程規劃，這樣的新技术的導入才能強化飯店業者之競爭力、提高飯店利潤，以利飯店業者永續經營發展。

參考文獻

一、中文部分：

- [1]陳肇男，「旅遊行程安排及探勘分析之實作」，國立雲林科技大學電子工程與資訊工程技術研究所，碩士論文，1999。
- [2]陳垂呈，「應用關聯規則規畫最適性之旅遊行程及行銷策略」，第二屆觀光休閒暨餐旅產業永續經營研討會，國立高雄餐旅學院，2002，241 247 頁。
- [3]曾勇森、陳垂呈，「應用資料探勘技術推薦消費者最適性之旅遊景點」，第二屆觀光休閒暨餐旅產業永續經營研討會，國立高雄餐旅學院，2002，273 282 頁。
- [4]葉怡成，類神經網路模式應用與實作，台北，儒林圖書有限公司，1993 年 8 月。
- [5]Michael J.A.Berry,Gordon Linoff 著，資料探礦-顧客關係管理暨電子行銷之運用，彭文正譯，維科圖書有限公司，台北，2001。
- [6]蘇木春、張孝德，機器學習：類神經網路、模糊系統以及基因演算法則，台北，全華科技圖書股份有限公司，1997 年 12 月。
- [7]陳昭吉，「應用遺傳基因演算法於晶圓製造場之設施佈置問題」，元智大學工業工程與管理研究所，碩士論文，2003。
- [8]陳彥匡、屈妃容，「應用品質機能展開及資料探勘技術於國際觀光旅館之顧客關係管理」，第三屆觀光休閒暨餐旅產業永續經營學術研討會，國立高雄餐旅學院，2003，375 383 頁。
- [9]曹勝雄，觀光行銷學，台北，揚智文化事業股份有限公司，2001 年 1 月。
- [10]Alastair M.Morrison，餐旅服務業與觀光行銷，王昭正譯，弘

智文化事業有限公司，台北，1999年8月。

二、英文部分：

- [11] Usama Fayyad, Gregory, Piatetsky-Shapiro, Padhraic Smyth, “ Knowledge Discovery and Data Mining : Towards a Unifying Framework ”, Proceedings of the Second International Conference on Knowledge Discovery and Data mining, Portland, Oregon, AAAI Press, August 2-4, 1996.
- [12] Usama Fayyad, Paul Stolorz. “Data mining and KDD : Promise and challenges”, Future Generation Computer Systems 13, pp.99-115, 1997.
- [13] Sung Ho Ha, Sang Chan Park. “Application of data mining tools to hotel data mart on the Intranet for database marketing”, Expert Systems With Application 15, pp.1-31, 1998.
- [14] Michael J.A. Berry, Gordon Linoff, Mastering Data Mining-The Art and Science of Customer Relationship Management, Wiley Computer Publishing, New York, 2000.
- [15] Johnson, R. A. and Wichern, D. W., Applied Multivariate Statistical Analysis, Prentice Hall, Inc., 1992
- [16] NeuralWorks Professional II/Plus and NeuralWorks Explorer, Penn Center West: NeuralWare, Inc., 1990