

# 行政院國家科學委員會專題研究計畫 成果報告

## 運用資料探勘求解公車最適靠站型態之研究 研究成果報告(精簡版)

計畫類別：個別型  
計畫編號：NSC 95-2221-E-216-055-  
執行期間：95年08月01日至96年07月31日  
執行單位：中華大學運輸科技與物流管理學系

計畫主持人：林祥生

計畫參與人員：碩士班研究生-兼任助理：謝耀霆、池維竣

處理方式：本計畫可公開查詢

中華民國 96年10月21日

# 行政院國家科學委員會補助專題研究計畫成果報告

## 運用資料探勘求解公車最適靠站型態之研究

計畫類別： 個別型計畫       整合型計畫

計畫編號：NSC95-2221-E-216-055

執行期間：95年08月01日至96年07月31日

計畫主持人：林祥生 博士

共同主持人：

計畫參與人員：謝耀霆、池維竣

成果報告類型(依經費核定清單規定繳交)： 精簡報告

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年  二年後可公開查詢

執行單位：中華大學運輸科技與物流管理學系

中 華 民 國 96 年 08 月 01 日

# 行政院國家科學委員會專題研究計畫成果報告 (精簡版)

## 運用資料探勘求解公車最適靠站型態之研究

### A study of data mining for optimal combination of bus stop operation strategies

計畫編號：NSC 95-2221-E-216-055

執行期限：95年8月1日至96年7月31日

計畫主持人：林祥生 中華大學運輸科技與物流管理學系

計畫參與人員：謝耀霆、池維竣

## 一、中文摘要

大台北地區的公車系統正面臨捷運路網成熟後的功能轉型，業者必須改變過去的經營手法，以科學化的營運技術即時掌握市場情報，並提供更貼近乘客需求的公車服務，以因應日益嚴峻的市場競爭。過去有關公車營運的研究，多偏重於公車排班及人車調度課題，對於因應乘客在不同時空環境的需求分布，是否應提出多元化的靠站方式較少著墨，而受限於乘客實際需求資料取得的困難，前人研究亦多以虛擬的簡例或少數調查資料進行測試。

資料探勘是從大量資料中萃取知識的分析工具，近年來已成為挖掘消費者行為的重要途徑，但是較少見於運輸經營領域。本研究擬採用林祥生等人[2005]的悠遊卡資料處理模式，先將公車乘客的刷卡記錄轉化為OD需求分布矩陣，進而運用資料探勘的群集化技術及關聯法則，構建不同組合的公車靠站方式，以尋求兼顧乘客旅行時間節省及業者營運成本降低的最適服務組合。本研究的成果將可提供業者未來「因地制宜、因時而異」的客製化服務模式，以更貼近公車乘客的品質期待。

**關鍵詞：**資料探勘、悠遊卡、公車靠站型態

### Abstract

Owing to Mass Rapid Transit (MRT) network getting completed, Taipei's bus service system is facing function reforming. Bus operators have to change current service strategies and explore scientific tools to acquire more information of passengers' demand, and then provide tailor-made bus service to keep surviving in competition market. In previous articles, the authors had paid plenty of attentions on bus scheduling and dispatching, rarely research were focused on optimization of bus stop operation strategies.

Data mining is the process of exploring valuable information and useful knowledge from huge amounts of raw transaction data. It is popular for banks and telecom companies to study their consumers' behavior, but little effort is paid in bus transit operation. In this project, at first we will transform the bus transaction records of Taipei smart card to bus OD demand distribution matrix, and then apply clustering and association rule of data mining techniques to construct multiple combinations of different bus stop operation strategies (ex: all-stop, skip-stop, non-stop, etc). With the objectives to minimize the sum of operator's cost and riders' travel time cost, we will try to provide bus operator a more customized service model to facilitate bus

riders' patronage in heterogeneous demand situations.

**Keywords : data mining, Taipei smart card, bus stop operation strategy.**

## 二、前言

近年來隨著大台北地區捷運路網的逐步綿密，傳統公車事業正面臨功能轉型的考驗，業者必須改變過去的經營理念，將自己重新定位(Repositioning)，從原本以服務OD 旅次為主，轉變為捷運系統的接駁互補角色。因此今後公車業者若想永續經營，就必須調整經營策略，以科學化的營運技術即時掌握乘客需求，在兼顧服務品質及成本效益的前提下，提供客製化(Customerization)的公車服務，以更貼近民眾需求的服務型態，扮演捷運系統最稱職的策略夥伴。

過去有關改善公車服務績效的研究，多偏重於公車排班及人車調度的課題，對於因應乘客在不同時間及空間上的需求分布，如何提出各取所需的多元靠站方式較少著墨，而受限於乘客實際需求資料取得的困難，前人研究亦多以虛擬的簡例或少數調查資料進行測試，故其研究成果較難呈現實務上的說服力。

所幸由於大台北地區運輸電子票證的日漸普及，面對人手一張的悠遊卡，裡面記載著乘客詳細真實的搭乘資料，如搭乘日期、刷卡時間、票種及轉乘紀錄等，然而目前公車業者只是以悠遊卡作為票證改革的收費工具，其貢獻僅限於票卡清算、資料統計、營收轉帳等基本功能，對於乘客使用悠遊卡時同步產生的龐大運量記錄卻任其閒置，因此無法發揮將原始資料轉化為管理資訊的功能，因此若能善用悠遊卡裡的交易記錄，運用適當的資料處理技術，將可完整呈現乘客需求在不同時間及空間上的分布，大幅改善以往人工調查OD 旅次代表性不足的缺陷。

在前述問題背景及研究動機之下，本研究擬應用現有的悠遊卡刷卡資料，推導公車乘客的起迄需求分布型態，進而構建不同類型的公車靠站組合，以尋求乘客旅行時間節省及業者營運成本降低的改善效益，故界定本研究之目的如下：

1. 運用悠遊卡刷卡資料及公車衛星定位到站時間，比對乘客刷卡地點，以推估出乘客起迄點需求矩陣 (OD Table)。
2. 運用資料探勘技術之群集分析功能及關聯法則功能，分別求解區間車及直達車營運改善方案。
3. 提出不同需求型態下，可獲致較高績效的營運改善方案。

## 三、文獻回顧

### 3.1 資料探勘的定義

資料探勘(Data Mining)是從龐大的資料量中，透過理解、萃取、整合、查核、除誤、轉化、測試、解釋等階段，逐步挖掘出有價值的隱藏資訊，以作為企業決策的重要參考，因此麻省理工學院的 Technology Review 雜誌將其選為改變未來世界的十大創新科技之一[MIT, 2001]。資料探勘在過去經常被用來分析信用卡客戶或手機用戶的消費行為，以輔助經營者從龐雜的客戶帳單資料中，挖掘出有助於未來精準行銷的決策資訊。

而在資訊科技的推波助瀾下，不僅企業競爭的強度與速度倍數於以往，激增

的市場交易也使得各企業所需儲存與處理的資料量愈來愈龐大。所以，企業如何因應外界的競爭，快速且有效地從資料庫中取得有用資訊，並反映市場或消費者的需求，已成為各企業重視的焦點；而資料探勘技術不但能輔助企業建立自動預測顧客行為的模型，還可與商業資料倉儲(Commercial Data Warehouse)結合，發展出更有價值的商業用途。

關於資料探勘的定義有許多不同版本，其中較具有代表性者包括，Fayyad 等人[1996]指出「資料探勘能將資料簡化成正確的、未曾被知道的、有用的以及最容易了解之規則的一連串重要處理動作」，Berry 等人[1997]認為「資料探勘是針對大量的資料，利用自動化或半自動的方式進行分析，以尋找出有意義的關係或法則」，Han[2000]則定義「資料探勘是從大量資料中萃取出來的知識」，而尹相志[2003]認為「資料探勘是利用統計以及機械學習的演算法，啟發性地從大量資料中找尋隱藏具有商業價值的知識與規則，作為自動化商業策略之應用」。關於資料探勘的定義，其他許多學者也有提出相關的解釋，並整理如表 3.1 所示。

表 3.1 資料探勘的定義

作者	定義
Frawley (1991)	資料探勘是從資料中萃取出隱藏的、先前未知的有用資訊。
Grupe and Owrang (1995)	資料探勘乃是從現存資料中剖析出新事實及發現專家們尚未知曉的新關係。
Fayyad <i>et al.</i> (1996)	資料探勘是知識發現(Knowledge Discovery in Database ; KDD)其中的一個步驟。資料探勘透過演算法，將資料作一分析與應用，以找出其特性(pattern)與模式(model)的過程。
Cabena <i>et al.</i> (1997)	資料探勘是將未知且有效的資訊從大型資料庫抽出的過程，並且將萃取出來的有用資訊提供給主管做決定性的決策。
Kim (2000)	資料探勘是提煉未知且可能有用的知識來幫助決策的一系列流程。
Shaw <i>et al.</i> (2001)	資料探勘是尋找和分析資料的一個過程，主要的目的是找出隱含在裡面有用的資訊。
Roger <i>et al.</i> (2002)	資料探勘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與去擷取知識的過程。

資料來源：本研究整理

### 3.2 資料探勘的主要功能

有關資料探勘的主要功能，Berry 等人[1999]將其區分為六大類，這些功能

的意義及可使用的技巧可概述如下：[陳麗君，民 91]

#### 1. 分類 (Classification)

分類是指檢視新物件的特性，然後將其指定到預先定義好的類別中，即利用一連串的輸入變數來預測「類別變數」。而所要檢視的物件，通常是指資料庫中的紀錄，包含更新資料紀錄及標上類別編號等。例如在信用卡的分析上，發卡銀行會利用顧客的資產狀況及繳費記錄，將信用卡申請人分類為低、中、高風險群，藉此分類方式對不同族群給予不同的額度及服務。常用的方法有決策樹(Decision Tree)或類神經網路(Neural Network)等。

#### 2. 推估 (Estimation)

推估善於處理連續性的數值，憑著一些輸入資料，可用來推估一些未知的連續性變數，例如按照信用申請者之教育程度及行為別，來推估其信用卡消費量。相關的使用技術包括統計方法上之相關分析、迴歸分析及類神經網路方法等。

#### 3. 預測 (Prediction)

預測是推估未來的數值以及趨勢，歷史資料可用來建立模型以檢視近來觀察值的資料，例如由過去行銷活動所產生的反應，來預測未來新活動的回應率，或是由顧客的職業、年齡、收入等人口屬性特質及其消費行為來預測可能的流失率等。使用的相關技術包括迴歸分析、時間序列分析(Time Series Analysis)、類神經網路及案例庫推理(Case-Based Reasoning)等。

#### 4. 關聯法則 (Association rule)

關聯法則主要描述在龐大資料庫中某些資料項目間彼此之關聯性，其形式為  $X \rightarrow Y$ ，其中  $X$  及  $Y$  分別表示資料庫中不同之項目組。關聯法則最早應用於超市購物籃分析(Market Basket Analysis)，藉由顧客的交易記錄，找出商品間彼此的關聯性，以做為超市商品擺設及進貨庫存之參考。

#### 5. 群集化 (Clustering)

群集化就是將一群異質的個體，區隔為同質性較高的群體或是子群。它與分類不同的是，群集化沒有依靠事先明確定義的類別來進行分類，資料是根據自身的相近性群聚在一起。因此，群集化可說是分類的前置作業，它也是進行市場區隔的第一步。

#### 6. 序列型樣 (Sequential Pattern)

序列型樣技術的重點是考慮時間的因素，利用此方法分析不同時間點上各事件的關聯性。序列型樣主要分為順序性型樣與週期性型樣兩種，順序性型樣乃考慮事件發生之時間先後關係，而週期性型樣考慮時間區段的變化，分析時間區段內所發生的事情，是否其他相同時間區段內也會發生。這兩種方法雖不同，但對使用者而言，隨著時間的多樣變化，找出有用的規則已日形重要。

### 3.3 公車靠站型態之相關研究

葉祥海[民 72]針對都市公車，利用電腦模擬方法，分別建立普通車、區間車與直達車三種派車策略之電腦模擬模式，進行各種組合派車策略之模擬實驗。主

要目的是在分析同一公車路線上行駛直達車、區間車與普通車在各種組合比例下之系統績效。作者把實地調查 5 路及 10 路公車之數據資料及路線基本資料，當作模擬程式之起始輸入值，再進行各種組合派車之模擬實驗。研究中對於直達車運作方式之停靠站並未說明其選擇方法，僅以每分鐘上下人數總和大於 1.5 人之站位，作為直達車之停靠站位。而有關於區間車服務之區段選擇方法也未加說明。模擬結果顯示，直達車或區間車愈多時，車輛迴轉率愈高，但將造成乘客等車時間顯著增加及載客數愈不均勻之不良現象。

Santhakumar 和 Hariharan[1991]運用電腦模擬模式，探討在各種運輸系統管理 (TSM) 策略對於公車路線服務績效之影響，包括：改變行車速度、改變加速度或減速度、減少站牌數、改變公車靠站型態 (探討直達車)、公車車型的配置、重新安排公車站牌位置等六種。研究中利用四條公車路線之實際資料，透過電腦模擬結果顯示，當施實直達車之停靠站方式，總行車時間約減少 30%，可有效提升服務績效，不過研究中並未說明。

趙瑞芳[民 85]針對公車運輸系統運作策略之特性、優缺點與適用條件，求解單一公車路線之最佳運作策略組合。該研究建立一公車排班及載客模擬模式，以系統總成本 (包括業者營運成本、乘客車上時間成本與等車時間成本) 最小為目標式，限制條件有車輛容量限制、班車最大班距、原路線乘客總等車時間等，並以道路交通環境、公車路線基本資料、尖離峰乘客需求資料為資料起始輸入值，並用卡方檢定確認模擬產生之觀測值與輸入的資料屬於相同之分配型態，再進行排班及載客模擬分析。策略組合包括每站皆停與直達車、每站皆停與區間車、每站皆停與分區停車 (skip stop) 之運作組合。最後以 36 路公車資料進行實例應用與分析，結果顯示若乘客起迄分佈型態集中於路線部分區段，則實施區間車與每站皆停運做組合策略有較佳的成果。

郭月萍[民 89]依據公車路線乘客起迄資料及公車排班的計算原理，利用啟發式方法建構公車路線最適營運策略組合模式，並以系統總成本最小為總評估指標，尋求營運者成本及旅客旅行時間成本同時下降最多之營運策略組合。研究中假設了六種公車問題，分別為：數量且乘客 OD 量均勻分布、數量小且乘客 OD 量均勻分布、數量且乘客 OD 量集中於部分停靠站、數量且乘客 OD 量集中於後段停靠站、數量小且乘客 OD 量集中於部分停靠站、數量小且乘客 OD 量集中於後段停靠站。估計出各營運策略組合下各類班車之運量，並實際求出各營運策略組合下各類班車之停靠站、班距、總車輛數、業者營運成本及乘客旅行時間成本等資料。

### 3.4 智慧卡產出資料之相關研究

林祥生[民 80]依據台北市聯營公車之營運特色，研擬公車票證電腦化作業之規劃目標、設計要求及系統架構，並依據此一規範篩選出適用於台北市的自動化收費設備，最後提出一套計便利乘客用票又可提升業者管理績效之票證制度及票務作業流程。其研究發現，公車票證電腦化的實施，除了直接增進業者營運管理效益，同時亦提供具記錄具公信力的營收資料，可作為政府今後補貼政策之依據。更重要的是，研究單位可依據票卡紀錄內容，推估乘客起迄運量表，以取代

目前高成本、低效率的運量調查方式。

林祥生等人[民 94]根據公車紅 29 的刷卡資料，自行構建一分析軟體，分別從總體面及個體面挖掘公車乘客的需求特性及搭乘行為。在總體面，本研究分析平日與假日的載客量比較、一日中各時段的尖離峰現象、不同平常日及不同身份別的尖峰時段差異、捷運轉乘行為等，而在個體面更進一步比較全票及學生乘客的搭乘習性，讓經營者能清楚看見個別顧客的消費者行為，一方面可針對目標乘客實施精準行銷策略，另一方面亦能有效安排車輛定檢及人員輪休。凡此過去業者無從得知的需求特性，皆可由本研究所構建的資料探勘模式逐一呈現，使業者今後能以科學數據進行更有效的營運策略。

Bagchi 和 White[2005]發現智慧卡 (smart card) 將成為可用來分析旅行者行為的新運輸資料來源。智慧卡裡所紀錄之資料，可以幫忙克服現存運輸資料來源的不足及讓業者從統計數字中了解乘客的需求。透過智慧卡系統，業者將可取得：1. 大量的個人旅行資料；2. 可以連接到個人卡或是使用者的資料；3. 比現存運輸資料來源可以獲得更長期的連續旅次資料；4. 知道他們的忠誠顧客是誰。可是若要從現存運輸資料，作為分析乘客行為來源的資料是有困難，因為智慧卡所記錄的內容不夠完整，如不知乘客確切起迄點、旅次目的等，而這些缺漏資料可能還需要藉由調查或其他方法來克服彌補。不過業者仍可再在了解資料上的限制後，尋找一種實用的方法來分析和應用。此外，公車業者也可因應現在潮流，根據資料分析結果發展出適合路線特性之行銷策略。

## 四、起訖需求推估邏輯概念

### 4.1 起迄需求推估邏輯概念

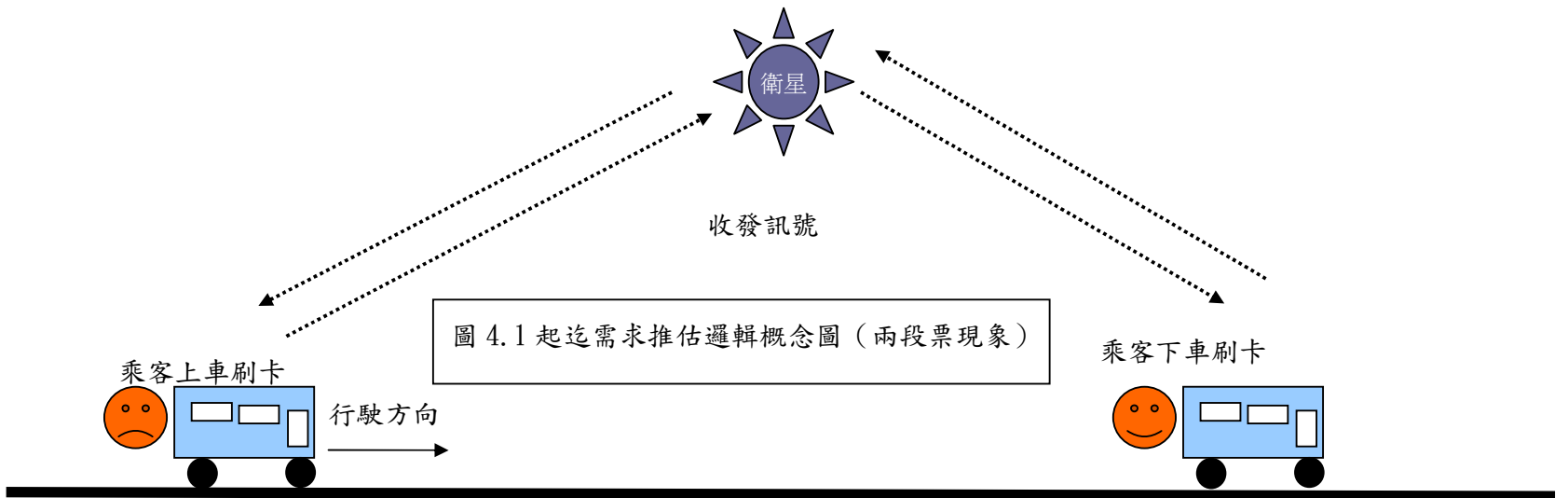
欲建立乘客之起迄運量分布，要先能得知乘客之起迄點位置。但目前衛星定位系統尚未與悠遊卡系統整合，所以我們無法直接從乘客交易資料中得知乘客的刷卡位置，因此將先了解現存資料那些紀錄內容，可將原始資料透過處理及轉換過程，進而推導出乘客刷卡站位，最後再由乘客刷卡站位試圖推估出乘客起迄需求位置，建立一乘客起迄需求運量表。

圖 4.1 所示，此行駛路段範圍為兩段票，乘客刷卡方式為上車刷一段，下車再刷一段。目前台北縣政府 e-bus 之衛星定位系統，以站牌為單位，會紀錄該站位即將要進站、離站之公車車牌號碼，該公車之行駛路線和其進站時間（站牌前 50 公尺）及離站時間（站牌後 80 公尺）。如圖 1 所示，在捷運新埔站該站站牌，它記錄了 238-AC 公車，將於 08:00:00 進站，且於 08:02:30 離站，該輛公車行駛路線為 802。同理，當公車行駛至樹林市衛生所站時，在樹林市衛生所站該站站牌，它記錄了 238-AC 公車，將於 09:00:00 進站，且於 09:01:30 離站，該輛公車行駛路線為 802。而在悠遊卡系統裡，當乘客上車刷卡時，會紀錄該位乘客所使用悠遊卡之卡號 ID 及上車刷卡時間，而在下車刷卡付費時，亦同樣紀錄該位乘客所使用悠遊卡之卡號 ID 及下車刷卡時間。我們可經由比對乘客上車刷卡時間及衛星定位系統裡公車當時位置，對應出乘客上車位置（乘客起點），同理，可經由比對乘客下車刷卡時間及公車當時位置，對應出乘客下車位置（乘客迄點），



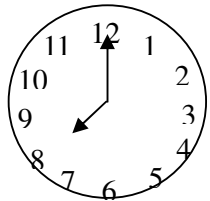
即可推估出乘客於該趟旅次之起迄點位置，並進而建構出乘客起迄需求矩陣。

一段票刷卡行為因單筆交易僅能紀錄某一位置資料（如搭乘起點或迄點），其推估邏輯為經由持續觀察多天乘客的搭乘行為，比對出乘客較固定之刷卡地點，以上車刷卡為例，乘客往程會記錄上車位置（起點），返程也會紀錄上車位置（起點），經由多筆相同記錄，可推估乘客的往程下車位置為返程的上車位置，進而推估乘客該趟旅次之起迄點位置，所推估出之乘客起迄點需求矩陣如表 4.1 所示



衛星定位系統紀錄資料  
 公車路線：802  
 公車車牌：238-AC  
 公車位置：捷運新埔站  
 公車進站時間：08:00:00

悠遊卡系統紀錄資料  
 卡片 ID：123456789  
 刷卡時間：08:01:30



乘客上車刷卡時間 v.s.  
 公車停靠位置

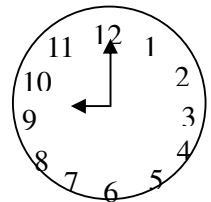


比對出乘客上車位置  
 (起點)



推估出乘客起迄點

悠遊卡系統紀錄資料  
 卡片 ID：123456789  
 刷卡時間：09:00:50



比對出乘客下車位置  
 (迄點)



衛星定位系統紀錄資料  
 公車路線：802  
 公車車牌：238-AC  
 公車位置：樹林市衛生所站  
 公車進站時間：09:00:00

乘客下車刷卡時間 v.s.  
 公車停靠位置









## 4.2 研究路線特性

本研究對象為首都客運所提供之 802 路線（如圖 4.2）。802 為兩段票公車路線，從三峽出發，沿途經過台北大學三峽校區附近、柑園工業園區、樹林火車站、丹鳳、輔仁大學、捷運新埔站，在抵達捷運新埔站於板橋花市原線折返回到三峽，全線共設有 61 個停靠站，往返共有 116 個站牌，路線往返總長度為 48 公里，平均往返行車時間約為 2 小時 40 分鐘。營運時間從上午 5 時 20 分至晚上 9 時，尖峰時段每 10~15 分鐘發一班車，離峰時段每 20~40 分鐘發一班車。資料範圍為 94 年 10 月份乘客搭乘資料。

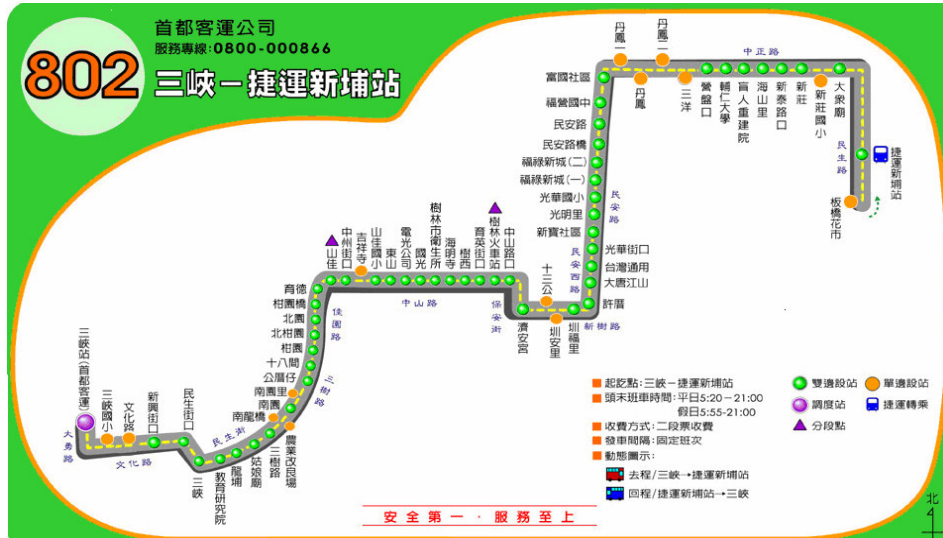


圖 4.2 802 營運路線圖

## 4.3 智慧卡資料特性

智慧卡卡公司所提供 802 路線原始乘客交易資料裡，資料欄位共有 33 項，本研究經過篩選資料內容及評估資料正確性後，發現真正能保留下來且應用的資料只剩 5 欄：卡片交易序號、票卡身份別、交易時間、晶片號碼、個人優惠，剩下 28 個欄位資料多為台北智慧卡公司及公車業者作為清算管理的資訊。所保留的 5 個欄位原因如下：

1. 序號：從交易序號編碼，可了解乘客交易行為及搭乘 802 公車的分布。
2. 票種：辨識乘客身分種類。
3. 交易時間：得知乘客刷卡時間。
4. 晶片號碼：乘客的 ID 記號，可作為計算乘客出現次數得知忠誠度用。
5. 個人優惠：可用來辨別乘客是否有轉乘行為。

此外，由於 802 路線為兩段票，刷卡方式分為：上車刷卡、下車刷卡，兩段票之乘客還會新增有上車刷一次，下車再刷一次等三種方式，因此，每位乘客所紀錄交易資料也會隨其搭乘段數及刷卡方式而不同，其刷卡方式行為的差異也會造成之後在推估乘客起迄需求時的困難度。

關於推估乘客站位之相關資訊，如表 4.2 所示，若乘客搭乘一段票路線，為上車刷卡，我們可得知乘客大約上車時間，不過無法得知乘客上車地點、下車時

間及下車地點。反之為下車刷卡，我們可得知乘客大約下車時間，不過無法得知乘客下車地點、上車時間及上車地點。若是兩段票路線，且為上車刷一次，下車再刷一次者，我們則可得知乘客上、下車時間，不過無法得知乘客上、下車地點。

表 4.2 不同刷卡行為方式所得資訊

	一段票		二段票		
	上車 刷卡	下車 刷卡	上車 刷卡	下車 刷卡	上車刷、 下車再刷
上車 時間	◎		◎		◎
上車 地點					
下車 時間		◎		◎	◎
下車 地點					
代表 涵義	得知乘客大 約上車時間	得知乘客大 約下車時間	得知乘客大 約上車時間	得知乘客大 約下車時間	得知乘客大約 上下車時間

#### 4.4 衛星定位資料特性

台北縣政府交通局之 e-bus 公車動態資訊系統，所記錄之公車到站資料是以站牌為單位，亦即由每各站牌（位置）去記錄每輛公車經過站牌之時間，而並非以公車為單位，記錄公車何時經過何地。

當公車行經站牌時，該站牌會記錄兩筆資料，第一筆是公車距離站牌前 50 公尺的時間（即為進站時間），第二筆是駛過站牌後 80 公尺的時間（即為離站時間）。原始紀錄的內容包括：車號、日期、到站時間及站牌位置，如圖 4.3 所示，第二站之站牌位置，於 2005/10/1 上午 05:56:52 有輛 FP-830 公車即將要進站，而於 2005/10/1 上午 05:57:06 駛離第二站站牌。而 FP-830 公車隸屬 802 路線之配置公車，所以在公車路線欄位顯示為 802 路線之公車。

序號	路線	車號	車號代碼	日期	到站時間	站牌位置
1	802	FP-830	2110003	2005/10/1	上午 05:56:52	2
2	802	FP-830	2110003	2005/10/1	上午 05:57:06	2
3	802	FP-828	2110001	2005/10/1	上午 06:27:01	2
4	802	FP-828	2110001	2005/10/1	上午 06:27:32	2
5	802	FP-829	2110002	2005/10/1	上午 06:41:5	2
6	802	FP-829	2110002	2005/10/1	上午 06:42:1	2
7	802	車號 0005		公車行經時間		公車位置

圖 4.3 衛星定位計錄資料

基於上述原因，若使用公車動態資訊系統所紀錄之原始資料格式，直接與乘客刷卡時間進行比對之資料庫，其撰寫比對邏輯之程式會顯得繁雜，與乘客刷卡

時間進行比對時，也會耗費較大的電腦運轉資源和較多的時間。因此本研究將對記錄資料內容進行加工及轉換，建立一以車號為單位之公車到站時間資料庫，流程及內容說明如下：

#### 1. 抓取各站牌之 802 公車到站時間資料

至台北縣交通局抓取 802 路線於 94 年 10 月份之各站到站時間紀錄檔案，802 路線沿路共設有 116 個站位，因此共有 116 個檔案

#### 2. 依車號別重新匯整檔案

配置行駛 802 路線之公車共有 13 輛，不過僅有 10 輛公車有架設衛星收發器，換言之，僅有 10 輛公車有衛星定位到站時間紀錄檔案。因此資料取得後，每各站牌之到站時間檔案再依照車號排序，接著再把相同車號之各站牌到站時間做合併，即由 116 個各站牌檔案重新匯整成 10 個車號檔案。

#### 3. 排序各公車行經各站時間

匯整完 10 個依車號排序之公車行經各站之到站時間檔案後，每個檔案再依照到站時間做排序，檔案之資料內容即成為以車號為單位，紀錄某車在何時經過哪個站牌。

#### 4. 檢查資料完整性

由於衛星訊號會受到氣候及其他因素干擾，因此接收不到公車到站時間及位置之訊息而造成資料缺漏，因此，此步驟要檢查公車站牌位置是否連續，若不連續或是班次時間內無紀錄到公車到站時間及位置之欄位，則填入 333 代碼表示遺漏值。

### 4.5 資料轉換及前處理過程

由於台北縣政府交通局之公車動態資訊系統，所紀錄之到站時間是以站牌為單位，亦即每支站牌會記錄每輛將要進站及已離站公車之時間及路線，並非以公車為單位。且台北智慧卡公司目前只提供公車業者有關票卡清算、營收轉帳等服務，因此各客運公司並未對資料欄位之內容，做詳細正確的資料紀錄動作，因此在原始刷卡資料中至少存在下列問題：

#### 1. 資料內容不正確：

雖然資料欄位有紀錄乘客刷卡方式（上車刷卡或是下車刷卡），不過由於司機疏忽及乘客行為難以規範，因此，該欄位所紀錄之刷卡方式不正確，這將造成日後推估乘客需求起迄點時，較難用程式判斷。

#### 2. 系統訊號受干擾：

衛星定位系統之收發訊號，會受到天氣及其他因素干擾。

#### 3. 系統未整合連線：

目前衛星定位系統尚未與刷卡系統整合，只紀錄公車位置及時間，和乘客刷卡時間，無法直接得知乘客刷卡地點。

上列因素會影響到資料分析的結果，因此本研究必須對資料內容進行簡化及轉換，流程及內容說明如下：

#### 1. 取得 802 路線乘客交易資料：

首都客運提供 94 年 10 月份整月份原始乘客交易資料。

#### 2. 資料轉換過程：



把連續型刷卡記錄轉換成離散型資料，變成易判讀、統計的資料格式，可以縮小資料檔案的空間及大小。轉換的資料有刷卡時間的離散化及分辨有無轉乘轉次，此兩項的資料經過轉換後，可當作往後推估乘客起迄點的參考資訊。

### 3. 資料前處理過程：

在此步驟可新增欄位，使資料庫內容更趨完整，增加往後在資料探勘時的資訊豐富性，新增的欄位有營運日、星期別及刷車站別。而刷車站別是運用衛星定位系統所紀錄之公車位置及時間，與乘客刷卡時間比對，可得出乘客刷卡位置，其比對概念如圖 4.4。

圖 4.4 係以下車刷卡為例，公車於第一站離站時間為 08:00:10，於第二站離站時間 08:05:20，於第三站離站時間 08:10:30，以此類推。若今天有位乘客其刷卡時間為 08:08:50，其刷卡時間已超過第二站離站時間的 08:05:20，但未超過第三站離站時間 08:10:30，因此推導出該位乘客其刷卡地點應於第三站。

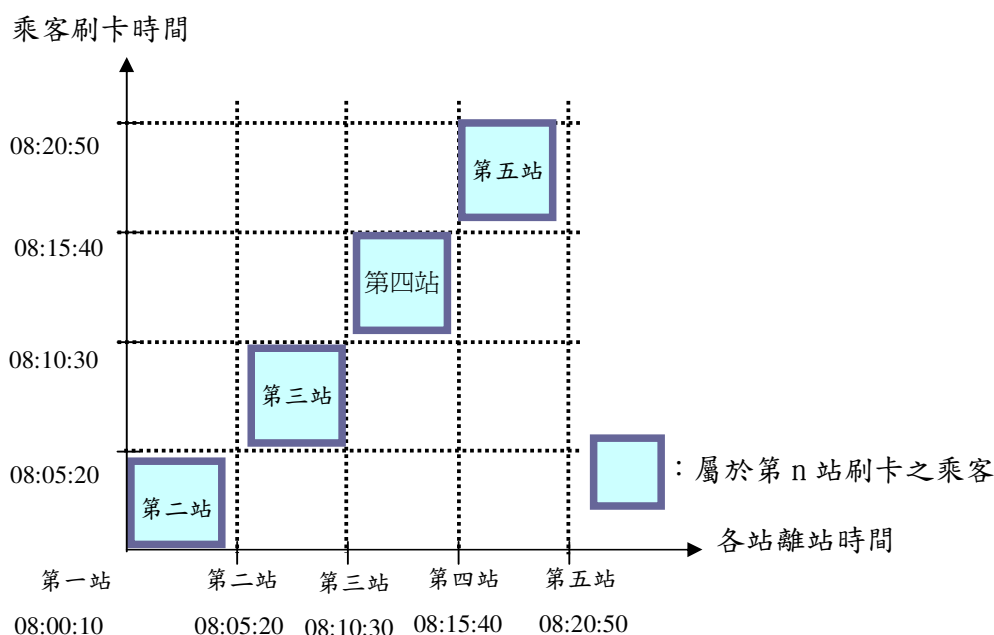


圖 4.4 乘客刷車站位示意圖

## 4.6 起迄需求推估方式

經由乘客刷卡時間、公車到站位置及時間等資料比對得到乘客刷車站位後，接下來，可藉由乘客交易資料判斷乘客的搭乘行為，以推估出乘客之起迄點需求。推估的步驟如下，結果如圖 4.5、圖 4.6 所示：

步驟一：經由晶片號碼可查詢出該位乘客於 10 月份間之交易資料。

步驟二：判讀交易資料之站別位置關係，可推導出該位乘客於該時段可能之起迄點位置。

如表 4.3 及圖 4.5 所示，某位乘客搭乘兩段票路段，其刷卡方式為上車刷一次，下車再刷一次，經由與衛星定位系統時間比對，推導出上車時間之上車位置，下車時間之下車位置，即可推估出該位乘客於該趟旅次之起迄點。若為一段票路

段，其刷卡方式為往程時上車刷卡，返程時亦為上車刷卡（如圖 4.6 所示），經由乘客刷卡時間與衛星定位系統時間比對，推導出其往返程之上車位置，且由交易紀錄發現有規律之搭乘行為，其上下午之需求點與吸引點有方向性，即上午為上班旅次，下午為返家旅次，則可推估該位乘客起點為往程之上車位置，迄點為返程之上車位置。

表 4.3 推估需求起迄點概念

	可取得之交易資料	推導出乘客刷卡站位	可推估之站位關係
一 段 票	往程上車刷一次， 返程上車刷一次	往程刷卡時間，往程上車位置， 返程刷卡時間，返程上車位置	<ul style="list-style-type: none"> <li>➤ 兩點位置</li> <li>➤ 刷卡時間</li> <li>➤ 乘客之起迄點</li> </ul>
	往程下車刷一次， 返程下車刷一次	往程刷卡時間，往程下車位置， 返程刷卡時間，返程下車位置	<ul style="list-style-type: none"> <li>➤ 兩點位置</li> <li>➤ 刷卡時間</li> <li>➤ 乘客之起迄點</li> </ul>
	往程上車刷一次， 返程下車刷一次	往程刷卡時間，往程上車位置， 返程刷卡時間，返程下車位置	<ul style="list-style-type: none"> <li>➤ 刷卡時間</li> <li>➤ 皆為同一點，無法推估</li> </ul>
	往程下車刷一次， 返程上車刷一次	往程刷卡時間，往程下車位置， 返程刷卡時間，返程上車位置	<ul style="list-style-type: none"> <li>➤ 刷卡時間</li> <li>➤ 皆為同一點，無法推估</li> </ul>
二 段 票	往程上車刷兩次， 返程上車刷兩次	往程刷卡時間，往程上車位置， 返程刷卡時間，返程上車位置	<ul style="list-style-type: none"> <li>➤ 兩點位置</li> <li>➤ 刷卡時間</li> <li>➤ 乘客之起迄點</li> </ul>
	往程下車刷兩次， 返程下車刷兩次	往程刷卡時間，往程下車位置， 返程刷卡時間，返程下車位置	<ul style="list-style-type: none"> <li>➤ 兩點位置</li> <li>➤ 刷卡時間</li> <li>➤ 乘客之起迄點</li> </ul>
	往程上車刷兩次， 返程下車刷兩次	往程刷卡時間，往程上車位置， 返程刷卡時間，返程下車位置	<ul style="list-style-type: none"> <li>➤ 刷卡時間</li> <li>➤ 皆為同一點，無法推估</li> </ul>
	往程下車刷兩次， 返程上車刷兩次	往程刷卡時間，往程下車位置， 返程刷卡時間，返程上車位置	<ul style="list-style-type: none"> <li>➤ 刷卡時間</li> <li>➤ 皆為同一點，無法推估</li> </ul>
	上車刷一次， 下車再刷一次	單程之上、下車刷卡時間 單程之上、下車位置	<ul style="list-style-type: none"> <li>➤ 兩點位置</li> <li>➤ 刷卡時間</li> <li>➤ 乘客之起迄點</li> </ul>

營運日	平假日	星期	晶片號碼	票種	刷卡時間	刷卡時段	有無轉乘	公車編號	公車離站時間	刷卡站別
2005/10/1									2005/10/1 上午 08:04:37	3
2005/10/1									2005/10/1 上午 08:55:11	50
2005/10/3									2005/10/3 上午 08:01:05	3
2005/10/3									2005/10/3 上午 08:55:53	50
2005/10/3									2005/10/3 下午 10:27:11	64
2005/10/3									2005/10/3 下午 11:13:26	114
2005/10/4									2005/10/4 上午 07:59:24	3
2005/10/4									2005/10/4 上午 08:50:53	50
2005/10/4									2005/10/4 下午 10:27:09	64
2005/10/4									2005/10/4 下午 11:10:17	114

**刷卡行為：上車刷一次、下車再刷一次**  
**可得：上車時間 → 上車地點**  
**下車時間 → 下車地點**  
**結果：乘客起點在第 3 站**  
**乘客迄點在第 50 站**

圖 4.5 推導二段票乘客起迄點示意圖

營運日	平假日	星期	晶片號碼	票種	刷卡時間	刷卡時段	有無轉乘	公車編號	公車離站時間	刷卡站別
2005/10/3		0	1	725419012	2	2005/10/3 上午 06:17:21	6	0 PF831	2005/10/3 上午 06:18:22	2
2005/10/3		0	1	725419012	2	2005/10/3 下午 04:59:33	6	0 PF831	2005/10/3 下午 04:59:33	86
2005/10/4									2005/10/4 上午 06:17:50	2
2005/10/4									2005/10/4 下午 04:38:26	86
2005/10/5									2005/10/5 上午 06:20:07	2
2005/10/5									2005/10/5 下午 04:34:15	86
2005/10/6									2005/10/6 上午 06:19:06	2
2005/10/6									2005/10/6 下午 04:41:18	86
2005/10/7									2005/10/7 上午 06:20:14	2
2005/10/7									2005/10/7 下午 05:24:07	86
2005/10/8									2005/10/8 上午 11:56:55	2
2005/10/8									2005/10/8 上午 11:56:55	2
2005/10/12									2005/10/12 上午 06:19:47	2
2005/10/12									2005/10/12 下午 04:41:21	86
2005/10/13									2005/10/13 上午 06:19:28	2
2005/10/13									2005/10/13 下午 01:00:04	85

**刷卡行為：上車刷一次**  
**可得：往程上車時間 → 往程上車地點**  
**返程上車時間 → 返程上車地點**  
**乘客往程起點在第 2 站**  
**乘客返程起點在第 86 站**  
**推估：乘客往程起點在第 2 站，往程迄點為返程**  
**起點（第 86 站）之對向站牌，為第 27 站。**  
**結果：乘客起點在第 2 站，迄點在第 27 站。**

圖 4.6 推導一段票乘客起迄點示意圖

## 五、感知反應時間分析

本章將根據 802 路線之乘客交易資料，在經過資料前處理轉換等過程，建構出可供資料探勘用之資料庫後，使用資料探勘技術，產出 802 公車路線於區間車及直達車之較佳公車靠站路線之方案。在探勘時段方面，分別從資料庫中選取較具代表性之平日上午尖峰時段、平日下午尖峰時段、平日離峰時段及假日時段等四個時段，連續三班車之乘客需求資料作為研究範圍。

從資料探勘之產出方案及各方案之評比計算式運算結果，可判斷 802 公車路線是否適合行駛區間車或是直達車，在何種時段較適合行駛何種靠站型態，和其較佳行駛範圍。

### 5.1 區間車

區間車為應用群集化分析之 K-means 演算法概念，嘗試將乘客依照其搭乘特性做一分群，經由整理產出方案並計算其評比計算式後，可判斷 802 路線是否適合行駛區間車，及區間車的行駛範圍。探勘區間車過程，共有 5 個產出方案，方案的設定標準依照 OD 矩陣每 n 站合併為一單位 (n=2、4、6、8、10)，而每個方案又會再分成二群、三群、四群各進行群集分析，以下為各時段之探勘結果。

#### 1. 平日上午尖峰時段

圖 5.1 為各分群的方案最佳解之評比計算式績效比較，總體看來平日上午尖峰實施區間車皆能節省成本，二分群 (m=2) 中之方案四-2 (n=8) 為相對較佳解，三分群 (m=3) 中之方案五-3 (n=10) 為相對較佳解，四分群 (m=4) 中之方案一-4 (n=2) 為相對較佳解。總體而言，仍以四分群結果最佳，其最佳營運路線為往程：福路新城 (二) 到捷運新埔站，返程：捷運新埔站到丹鳳 (二)，

如表 5.1。

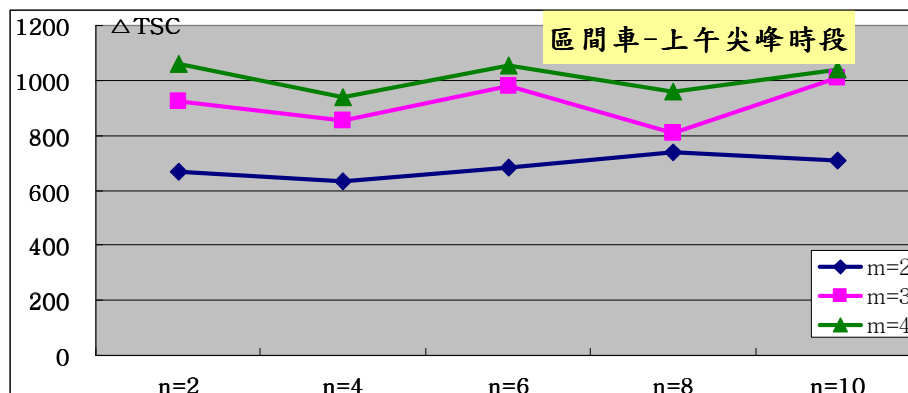


圖 5.1 上午尖峰時段各方案成本

表 5.1 上午尖峰時段最佳行駛路線

區間車													
往程			返程			往程			返程				
編號	站牌	路徑	編號	站牌	路徑	編號	站牌	現況	最佳解	編號	站牌	現況	最佳解
1	三峽站		116	三峽站		28	樹林火車站			85	樹林火車站		
			115	三峽國小		29	中山路口			84	中山路口		
			114	文化路		30	濟安宮			83	濟安宮		
2	新興街口		113	新興街口		31	圳安里			82	十三宮		
3	民生街口		112	民生街口		32	圳福里			81	圳福里		
4	三峽		111	三峽		33	許厝			80	許厝		
5	教育研究院		110	教育研究院		34	大唐江山	▽		79	大唐江山	△	
6	龍埔		109	龍埔		35	台灣通用	▽		78	台灣通用	△	
7	姑娘廟		108	姑娘廟		36	光華街口	▽		77	光華街口	△	
8	三樹路		107	三樹路		37	新寶社區	▽		76	新寶社區	△	
9	農業改良所		106	南龍橋		38	光明里	▽		75	光明里	△	
10	南園		105	南園		39	光華國小	▽		74	光華國小	△	
			104	南園里		40	福祿新城(1)	▽		73	福祿新城(1)	△	
11	公厝仔		103	公厝仔		41	福祿新城(2)	▽	▼	72	福祿新城(2)	△	
12	十八間		102	十八間		42	民安路橋	▽	▼	71	民安路橋	△	
13	柑園		101	柑園		43	民安路	▽	▼	70	民安路	△	
14	北柑園		100	北柑園		44	福營國中	▽	▼	69	福營國中	△	
15	北園		99	北園		45	富國社區	▽	▼	68	富國社區	△	
16	柑園橋		98	柑園橋		46	丹鳳	▽	▼	67	丹鳳一	△	

17	育德		97	育德		47	三洋	▽	▼	66	丹鳳二	△	▲
18	山佳		96	山佳		48	營盤口	▽	▼	65	營盤口	△	▲
19	中州街口		95	中州街口		49	輔仁大學	▽	▼	64	輔仁大學	△	▲
			94	吉祥寺		50	盲人重建院	▽	▼	63	盲人重建院	△	▲
20	山佳國小		93	山佳國小		51	海山里	▽	▼	62	海山里	△	▲
21	東山		92	東山		52	新泰路口	▽	▼	61	新泰路口	△	▲
22	電光公司		91	電光公司		53	新莊	▽	▼	60	新莊	△	▲
23	國光		90	國光		54	新莊國小	▽	▼				
24	樹林衛生所		89	樹林衛生所		55	大眾廟	▽	▼	59	大眾廟	△	▲
25	海明寺		88	海明寺		56	捷運新埔站	▽	▼	58	捷運新埔站	△	▲
26	樹西		87	樹西		57	板橋花市	▽	▼				
27	育英街口		86	育英街口									

## 2. 平日下午尖峰時段

最佳營運路線為捷運新埔站到丹鳳（二）。

## 3. 平日離峰時段

最佳營運路線為往程：輔仁大學到捷運新埔站，返程：捷運新埔站到輔仁大學。

## 4. 假日時段

最佳營運路線為往程：輔仁大學到捷運新埔站，返程：捷運新埔站到輔仁大學。

## 5. 現況分析

下列各表為各時段最佳解方案與現況 802 區間車績效之比較。總體而言，無論任何時段探勘結果之績效，皆比業者目前行駛路線能節省更多成本。

表 5.2 平日上午尖峰時段之現況與探勘結果績效比較

上午尖峰	△TBLC	△TPWAC	△TPBSC	△TSC
	營運里程	等車時間	車上時間	總節省成本
<b>方案一-4</b>	<b>1696</b>	<b>-645</b>	<b>9.28</b>	<b>1060.3</b>
現況分析	1351.5	-575	8.54	785.0
績效差異	344.5	-70	0.74	275.24

表 5.3 平日下午尖峰時段之現況與探勘結果績效比較

下午尖峰	△TBLC	△TPWAC	△TPBSC	△TSC
	營運里程	等車時間	車上時間	總節省成本
<b>方案四-4</b>	<b>1908</b>	<b>-145</b>	<b>25.026</b>	<b>1788.0</b>
現況分析	1351.5	-25	31.84	1358.3
績效差異	556.5	-120	-6.814	429.686

表 5.4 平日離峰時段之現況與探勘結果績效比較

平日離峰	△TBLC	△TPWAC	△TPBSC	△TSC
	營運里程	等車時間	車上時間	總節省成本
<b>方案四-4</b>	<b>1908</b>	<b>-300</b>	<b>0.37</b>	<b>1608.4</b>
現況分析	1351.5	-255	0.55	1097.1
績效差異	556.5	-45	-0.18	511.32

表 5.5 假日時段之現況與探勘結果績效比較

假日	△TBLC	△TPWAC	△TPBSC	△TSC
	營運里程	等車時間	車上時間	總節省成本
<b>方案四-4</b>	<b>1908</b>	<b>-400</b>	<b>0.94</b>	<b>1508.9</b>
現況分析	1351.5	-270	3.09	1084.6
績效差異	556.5	-130	-2.15	424.35

## 5.2 直達車

直達車應用關聯法則之 Apriori 演算法概念，嘗試依照各站牌間之需求強度特性，經由整理產出方案並計算其評比計算式後，可判斷 802 路線是否適合行駛直達車，及直達車的行駛範圍。以下為各時段之探勘結果。

### 1. 平日上午尖峰時段

經由探勘過程中設定不同支持度可產出不同方案，總體看來平日上午尖峰實施直達車皆能節省成本，以方案一之成果最佳，如圖 5.2 所示。沿路僅需停靠 29 個站位，如表 5.6 所示。

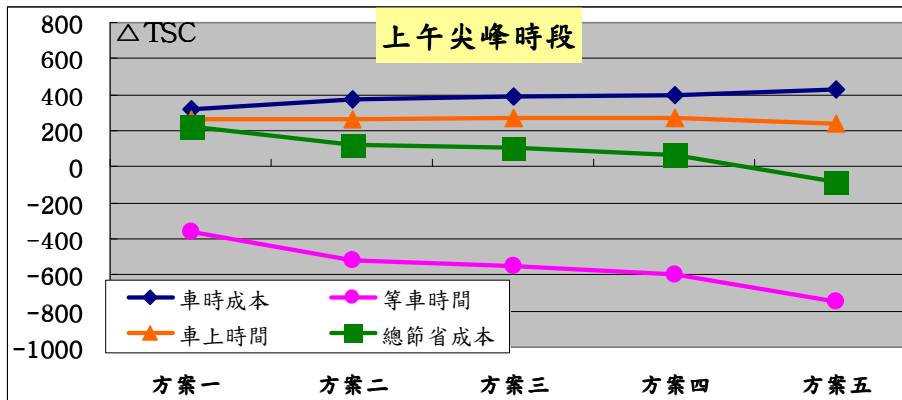


圖 5.2 平日上午尖峰直達車各方案成本

表 5.6 平日上午尖峰直達車最佳營運行駛路線

直達車										
往程			返程			往程			返程	
編號	站牌	路徑	編號	站牌	路徑	編號	站牌	路徑	編號	站牌
1	三峽站	▼	116	三峽站	▲	29	中山路口		84	中山路口

			115	三峽國小		30	濟安宮	▼	83	濟安宮	
			114	文化路		31	圳安里		82	十三宮	
2	新興街口	▼	113	新興街口		32	圳福里		81	圳福里	
3	民生街口	▼	112	民生街口		33	許厝		80	許厝	
4	三峽	▼	111	三峽	▲	34	大唐江山		79	大唐江山	
5	教育研究院	▼	110	教育研究院		35	台灣通用		78	台灣通用	
6	龍埔		109	龍埔		36	光華街口		77	光華街口	
7	姑娘廟		108	姑娘廟		37	新寶社區		76	新寶社區	
8	三樹路		107	三樹路		38	光明里		75	光明里	
9	農業改良所		106	南龍橋		39	光華國小	▼	74	光華國小	
10	南園		105	南園		40	福祿新城(1)	▼	73	福祿新城(1)	
			104	南園里		41	福祿新城(2)	▼	72	福祿新城(2)	
11	公厝仔		103	公厝仔		42	民安路橋	▼	71	民安路橋	
12	十八間		102	十八間		43	民安路		70	民安路	
13	柑園	▼	101	柑園		44	福營國中	▼	69	福營國中	
14	北柑園		100	北柑園		45	富國社區		68	富國社區	
15	北園		99	北園		46	丹鳳	▼	67	丹鳳一	
16	柑園橋	▼	98	柑園橋		47	三洋		66	丹鳳二	
17	育德	▼	97	育德		48	營盤口		65	營盤口	
18	山佳	▼	96	山佳		49	輔仁大學		64	輔仁大學	▲
19	中州街口		95	中州街口	▲	50	盲人重建院		63	盲人重建院	▲
			94	吉祥寺		51	海山里		62	海山里	
20	山佳國小		93	山佳國小		52	新泰路口	▼	61	新泰路口	
21	東山		92	東山		53	新莊	▼	60	新莊	▲
22	電光公司		91	電光公司		54	新莊國小				
23	國光	▼	90	國光		55	大眾廟		59	大眾廟	▲
24	樹林衛生所		89	樹林衛生所		56	捷運新埔站	▼	58	捷運新埔站	▲
25	海明寺		88	海明寺					57	板橋花市	
26	樹西		87	樹西							
27	育英街口	▼	86	育英街口	▲						
28	樹林火車站	▼	85	樹林火車站	▲						

## 2. 平日下午尖峰時段

平日下午尖峰時段除方案一可實施直達車能節省成本外，其他方案皆無法節省成本，沿路僅需停靠 53 個站位。

## 3. 平日離峰時段

平日離峰時段實施直達車皆能節省成本，不過結果以方案一之成果最佳，沿路僅需停靠 17 個站位。

#### 4.假日時段

平日下午尖峰時段除方案一可實施直達車能節省成本外，其他方案皆無法節省成本，沿路僅需停靠 25 個站位。

## 六、結論與建議

本研究利用資料探勘技術建構公車路線較佳靠站組合，並以系統總節省成本最大為總評估指標，尋找公車營運成本及乘客旅行時間成本節省最多之靠站方式組合，綜合整個研究過程，可歸納成以下之結論與建議。

### 6.1 結論

1. 悠遊卡刷卡資料若運用適當的資料處理技術，除可完整呈現乘客需求在不同時間及空間上的分布，了解乘客總體需求特性外，本研究首創應用於推估公車乘客起迄需求矩陣，進而發現其資料所潛藏的市場情報及旅運資訊，尚有可觀的開發價值。此舉不但克服以往需耗費龐大的人力、時間及金錢才能建構的 OD 矩陣，且可發揮目前悠遊卡交易紀錄之資訊價值。
2. 過去運用資料探勘技術於大眾運輸之研究甚為少見，本研究嘗試以群集分析及關聯法則，應用於求解區間車及直達車的公車營運方案。本研究應用群集化分析之 K-means 演算法概念，嘗試在上午尖時段、下午尖峰時段、平日離峰時段及假日時段，把代表各時段乘客需求之公車站牌做一分群，進而產出不同的區間車方案，再經由評比計算式評選後，提出各時段之最佳方案解。本研究之區間車最佳行駛路線解皆優於業者現況行駛路線，此舉可改善目前公車業者主觀決定區間車方案的缺失。
3. 本研究應用關聯法則之 Apriori 演算法概念，嘗試在上午尖峰時段、下午尖峰時段、平日離峰時段及假日時段，把依照乘客對於各站牌間的需求強度特性與站位及站位間的關聯性，設定不同支持度及信賴度進行探勘，進而產出不同的直達車方案，再經由評比計算式評選後，提出各時段之最佳方案解，此舉可提供公車業者設計直達車方案的規劃架構。
4. 經由四個時段的方案比較後，發現區間車改善效果以下午尖峰時段最佳，直達車改善效果則以上午尖峰時段最佳，這也反映業者可依照路線乘客的需求特性，可因時制宜的調整其經營方式，提供客製化的服務。不過以本研究所探討的公車 802 而言，無論在哪一時段，行駛區間車之總體績效均優於直達車。

### 6.2 建議

1. 本研究為國內首次使用資料探勘技術求解公車靠站組合之研究。由於大眾運輸本身為資料密集之產業，而資料探勘技術本身也善於從大量資料庫探勘出有價值及未知之資訊，不過過去公車業者未善用資料，所以尚未從中找到寶貴之商業經營智慧。然而，資料探勘技術有多種分析工具及應用形式，礙於研究時間無法將所有功能皆應用於該研究課題，後續研究者可再深入了解其他方法論，並找出更有效率及客觀的求解方法。
2. 拜科技進步所賜，隨著悠遊卡系統及公車動態資訊系統的陸續問世，悠遊卡



- 可成為用來分析乘客行為的新運輸資料來源，公車動態資訊系統可彌補悠遊卡系統未記錄地點的缺憾。業者應要善用系統紀錄的資料，使其可轉換成有用、寶貴的資訊及商業智慧。
3. 過去業者在經營公車路線時，過於依賴每站皆停的靠站型態，不過在大台北地區捷運路網的建構完成後，勢必會衝擊公車業者，因此業者應想辦法讓路線的經營方式更有彈性，如提供更客製化的靠站型態，根據不同時段的乘客需求提供因時因地制宜的靠站型態。國內有關公車營運的研究已不在少數，但多研究公車排班及人車調度領域，鮮少有關公車靠站型態研究，未來可多加研究靠站型態建構之相關課題。
  4. 本研究係以單一路線為研究課題，未來後續研究可善用悠遊卡資料，建構出大眾運輸乘客的總體需求特性及其相關課題改善方案。
  5. 802 公車路線乘客使用悠遊卡付費比率約為 60%，再加上並非每輛公車皆裝有衛星定位系統及乘客刷卡行為差異，使得乘客起迄需求推估率約 70%，因此尚無法代表整條路線乘客之需求，未來若因業者詳加規定刷卡方式和每輛公車皆能裝有衛星定位系統，而政府也能加以推廣行銷使用悠遊卡，提高悠遊卡使用比率，則可建構出完整正確之乘客需求起迄表，供業者及政府進行相關研究使用。
  6. 本研究所構建之各單一時段的最佳路線解，為考量業者實務容易操作，乘客容易辨別，後續研究者可將本模式擴充為考量多時段方案評比模式，以更符合實務需求。
  7. 本研究對於區間車或直達車之方案產出，未考慮剩餘全程班次之政策班距及容量問題，後續可將其列為限制條件，可使本模式更趨嚴謹和合理。

## 參考文獻

- 尹相志(民 92)，「SQL 2000 Analysis Service 資料採礦服務」，維科圖書。
- 林祥生、邱詩淳、劉益豪(民 94)，「應用悠遊卡資料挖掘公車乘客之需求特性」，中華民國運輸學會第 20 屆論文研討會。
- 林祥生(民 80)，「票證電腦化實施後台北聯營公車的票務系統規劃」，運輸學刊，第 12 期，頁 15~26。
- 陳麗君(民 92)，「應用資料探勘技術於信用卡黃金級客戶之顧客關係管理」，元智大學工業工程與管理研究所碩士論文。
- 葉祥海(民 72)，「普通車直達車與區間車組合派車之模擬研究」，國立交通大學交通運輸研究所碩士論文。
- 趙瑞芳(民 85)，「單一公車路線選擇最佳運作策略之研究」，國立台灣大學土木工程學研究所碩士論文。
- Berry, M.J.A. and Linoff, G. (1997), "Data Mining Technique for Marketing, Sale, and

- Customer Support, "Wiley Computer.
- Berry, M.J.A. and Linoff, G. (1999), "Mastering Data Mining: The Art and Science of Customer Relationship Management, Sale, and Customer Support," Wiley Computer.
- Cabena, P., Hadjinaian, P. O., Stadler, D. R. J., Verhees, J. and Zanasi, A. (1997), "Discovering Data Mining from Concept to Implementation," Prentice Hall.
- Fayyad, U., Piatetsky, G. and Smith, P. (1996), "From data mining to knowledge discovery in databases," AI Magazine, pp. 37-54.
- Feeldersa, A., Danielsa, H. and Holsheimer, M. (2000), "Briefings methodological and practical aspects of data mining," Information & Management 37, pp. 271-281.
- Grupe, F. H., Owrang, M. M. (1995), "Database Mining Discovering New Knowledge and Cooperative Advantage," Information Systems Management, Vol. 12, No. 4, pp. 26-31.
- Han, J., Kamber, M. (2000), "Data Mining : Concepts and Techniques," Morgan Kaufmann Publisher, Inc.
- M. Bagchi, P.R. White (2005), "The potential of public transport smart card data," Transport Policy Volume:12, pp.464-474.
- MIT (2001), "The Technology Review Ten," MIT Technology Review.
- Piatetsky, G., Fayyad, U., and Smith, P. (1996), "From data mining to knowledge discovery: An overview. In Advances in Knowledge Discovery and Data Mining," AAAI Press, pp. 1-34.
- Roger, R.J., Geatz, M.W. (2002), "Data Mining : A Tutorial-based Primer."
- Shaw, M. J., Subramaniam, C. and Tan, G. W. (2001), "Knowledge management and data mining for marketing," Decision Support Systems, 31, pp. 127-137.
- S.Moses Santhkumar and P.Hariharan (1991), "Transportation systems management options to improve urban bus performance using computer simulation," TRR1338, pp.22-27.

<http://www.cs.sfu.ca>

## 成果自評

1. 研究內容與原計畫相符程度：符合計畫
2. 達成預期目標情況：已達預期目標
3. 研究成果之學術或應用價值：客運路線調整規劃參考、客運業行銷策略擬定、路線車輛班次規劃、降低營運成本、提高營運效率。
4. 是否適合在學術期刊發表或申請專利：適合期刊發表
5. 主要發現或其他有關價值等：由電子票卡運用資料探勘，可分析級推估旅客旅行型態及客源所在。

# 可供推廣之研發成果資料表

可申請專利

可技術移轉

日期：96年08月01日

<p><b>國科會補助計畫</b></p>	<p>計畫名稱：運用資料探勘求解公車最適靠站型態之研究                  計畫主持人：林祥生                  計畫編號：NSC 95-2221-E-216-055                  學門領域：土木工程（交通、測量、建築）</p>
<p><b>技術/創作名稱</b></p>	<p>運用資料探勘求解公車最適靠站型態</p>
<p><b>發明人/創作人</b></p>	<p>林祥生</p>
<p><b>技術說明</b></p>	<p>中文：                  過去國內有關公車營運的研究已不在少數，可是最大的研究障礙往往來自欠缺完整且具代表性的載客紀錄。本研究以悠遊卡系統內之乘客實際交易資料和公車衛星定位系統內之公車到站時間資料，比對出乘客刷卡位置後，經由乘客搭乘紀錄的推估可建立出真實乘客 OD 矩陣。接著運用真實之乘客 OD 矩陣進行公車營運改善求解。本研究採用直達車、區間車等營運路線組合作為研究範疇。並運用資料探勘技術之群集化分析及關聯法則之功能，分別求解區間車及直達車之最佳營運路線。最後建立產出方案之評比計算式，包含業者的車小時成本及車公里成本、乘客的車上時間成本及等車時間成本，模式目標為系統總節省成本最大。</p> <p>英文：                  In the past, the researches about bus operation have had the same shortage, which was always without complete and reliable OD Table. This research constructs real passenger OD Table by Taipei Smartcard System's data and Taipei County e-bus System's data. As well as explore real OD Table to generate bus operation strategies. This research discusses the short-turn service route and express service route. It applied Data Mining technology about clustering and association rules to figure out optimal short-turn service route and optimal express service route, with the objective to save maximum the sum of operator's cost (including traveling time cost and distance cost) and passengers' travel time cost (including in-vehicle time cost and waiting time cost).</p>
<p><b>可利用之產業 及 可開發之產品</b></p>	<p>公路汽車客運業、大眾運輸業</p>

<p><b>技術特點</b></p>	<ol style="list-style-type: none"> <li>1. 從悠遊卡資料整理出公車乘客需求特性</li> <li>2. 推導公車路線的乘客OD 分布矩陣</li> <li>3. 運用資料探勘技術建構公車靠站型態</li> <li>4. 提出不同需求型態下，可獲致較高績效的營運策略</li> </ol>
<p><b>推廣及運用的價值</b></p>	<ol style="list-style-type: none"> <li>1. 客運路線調整規劃參考。</li> <li>2. 客運業行銷策略擬定。</li> <li>3. 路線車輛班次規劃。</li> <li>4. 降低營運成本。</li> <li>5. 提高營運效率。</li> </ol>