

建構公車預估到站時間模式之研究

朱松偉¹ 歐俊輝²

摘要

本研究利用現行裝有 GPS 以及 GPRS 之客運車輛做為探針車輛，透過其所傳回大量真實 GPS 時空資料，發展一套資料的處理流程，先劃分停等區域及行駛區域後推估路段速率，以做為路段速率資訊提供之用。研究時間兩個月的資料，共計二十六天之資料，超過兩百萬筆為研究母體，取桃園縣 1 路公車中之一路段作為樣本，藉由發展出的資料處理模式，將 GPS 所回傳之原始資料加以整合分析，並且嘗試運用新的資訊科技來探討市區公車的旅行時間。本研究利用了 SQL SERVER 2005 所內建之資料探勘功能，企圖使用大量時空資料並且結合決策樹來建構推估模式，其中模式一共分割出了十二個迴歸式；並且由資料採礦散佈圖得知，在不同的兩點平均速率上代用不同的迴歸式其準確度在兩站旅行時間為 95 秒以下時預測值與實際值相同，並且在九十五秒以上的旅行時間之預測準確度也尚可接受。

關鍵詞：全球衛星定位系統、旅行時間推估、資料採礦、決策樹。

The Study to Build Bus Arriving Model

Song-Wei Chu¹ Chun-Hui Ou²

Abstract

This study uses the buses that attached GPS and GPRS as probe vehicle, and to develop a procedure of data processing by way of these a mount of data. First, we divides the waiting area and the travel area and then estimates the driving speed. The population is two million datum, that take from 1 line in Taoyuan county. Moreover, we also use the decision tree that developed from SQL Server 2005 to develop a model, and base on this model, we verify the dependability.

Keywords : GPS, Travel time estimation, Data Mining, Decision Tree.

一、緒論

1.1 研究背景與動機

收集即時旅行資訊有兩項基本的方法。一種是每隔一段距離，在固定點設置車輛偵測器(Vehicle Detector, VD)，以固定的頻率來偵測車輛，此方法可收集到車流量、時間平均速度(time mean speeds)、車間距離(headway)、車輛種類、車道容量(lane

¹清雲科技大學行銷與流通管理系助理教授（聯絡地址:320 桃園縣中壢市健行路 229 號;電話:03-4581196 轉 7501;E-mail: swchu@cyu.edu.tw）

²清雲科技大學經營管理碩士（聯絡地址:320 桃園縣中壢市健行路 229 號;電話:03-4581196 轉 7501;E-mail: M9434012@cyu.edu.tw）

occupancy)等資料。另一種方法是，利用在交通流(traffic flow)中行駛之裝有全球衛星定位系統(Global Positioning System, GPS)配備之車輛來收集資料，此種裝有GPS配備之車輛也就是探針車輛(probe vehicle)，探針車輛與控制中心之間保持固定頻率之通訊，藉以追蹤探針車輛在路線上的所在位置與瞬時速度。

公車是大眾交通運輸載具之一，其到站的時間經常是搭乘者無法掌握的，長久以來傳統的市區公車站牌上所能提供的班車相關資訊並不能夠滿足搭乘者的需要，搭乘者如要搭乘服務班次較少或車班相距較長的客運路線時，必須花費更多不明確的時間等候，因而耗費許多寶貴時間。所以如果能夠提供搭乘者正確的乘車資訊，例如預估到站時間以及預估旅行時間，這將能使搭乘者節省較多的等候時間，也會增加民眾搭乘大眾運輸載具的意願。

1.2 研究目的

本研究主要是運用現行裝有全球衛星定位系統以及整合封包無線電服務(General Packet Radio Service, GPRS)之車輛做為探針車輛，透過其所傳回之資料，發展一套資料的處理流程。另外，藉由大量之歷史時空資料的解析可以真實呈現道路狀況而不需藉由撰寫模擬程式的方式來進行資料收集，再者，透過新的IT技術來從歷史之旅行時間作車輛旅行時間之推估。同時針對車輛實際運行狀況對推估之車輛旅行時間修正，以期能夠提供使用者準確且符合即時現況之車輛旅行時間推估。故本研究希望藉由大量的真實歷史資料建構出一推估到站時間之模式，期望利用即時更新的班車動態資訊以及預測鄰近班車到站時間服務搭乘者。

二、文獻探討

綜觀過去旅行時間預測的研究可以歸納出大致皆有兩個步驟，首先為預測各段區間的旅行時間，另一步驟則連接各段區間時間以獲得距離較長之旅行距離時間預測值。在大多數之文獻中皆將各區間之旅行時間視為互相獨立，因此所得之旅行時間預測值乃為個別旅行時間相加總而求得。對於預測旅行時間之概念較為常見的是利用即時或事後所偵測之交通參數資料來進一步分析流量與旅行時間之預測值。

2.1 國外相關旅行時間預測

國外實作之旅行時間預估系統主要多針對城際間之高速公路，而較少針對市區路網。Jasperse(1999)的研究中認為對於搭乘者而言，為了提供有用且準確的交通資訊，必須發展出一套用以計算旅行時間與等候長度的演算法。其演算法大致的步驟為：資料的前處理、長期間流入與流出之校正、車輛估計數、旅行時間估計。

Ding et al.(1999)的研究裡將公車動態到達旅行時間的預測方法分成兩種，一為以路段為準(Link-based)的模式，計算出兩個公車停靠點間所有路段之旅行時間，並將之儲存加總而求得，另一則為停靠點為準(Stop-based)的預測模式，採用兩點間之交通參數資料(如交通流量、速度與延滯之平均數與變異數)。

Choi(1998)運用了GPS與數位化過後的道路地圖來計算路段之動態旅行時間。此研究中除了利用GPS與地理資訊系統(Geographic Information System, GIS)來獲得都會

區街道之動態路段旅行時間之外，並在研究中比較了不同的測量路段旅行時間的技術，研究中認為最常用的技術是流動車輛法(floating car method)，其用以得到點至點的旅行時間，不過此方法需要具備較多的交通資料，例如每個車道、方向以及時段的交通參數資料。

Frederick W. Cathey et al.(2001)在其研究中使用大眾運輸載具作為探針車輛，展示了一個新的演算法，用以判斷經過一條公路或是動態走廊所需的旅行時間與速率。此外，利用自動車輛定位系統在大眾運輸載具經過的每個路段建立「速率偵測器」。將所經路段定位設立標籤，再利用GIS來組織車輛現在狀態的估計值。最後使用這些所得的資料庫來推估旅行時間。Frederick W. Cathey et al.(2004)延續其在2001年所作的研究成果，此研究中是將自動車輛定位系統所收集到的資料，利用卡門濾波器(Kalman Filter)先行去掉極端值後，再將這些資料對應GIS上的各路段，最後加總各路段的「距離／時間」，即可推估出旅行時間。

2.2 國內相關旅行時間預測

在國內的旅行時間相關研究方面，目前大多都以國道高速公路為研究對象，主要是利用不同模式或演算法來構建高速公路旅行時間預測模式，並嘗試結合固定式偵測器與移動式偵測器等不同偵測單元，以及時間、事件發生、幾何設計等不同行車環境。

在旅行時間的推估上，其相關研究如：吳佳峰(2001)由於目前國內車輛偵測器數量並非足夠，故其研究中利用GPS之定位資訊來推估車輛之旅行時間。以實際國內客運業者車輛旅行資料對推估模式作實例測試，從測試結果發現模式在未遭遇非重現性之擁塞時，推估旅行時間有著不錯之準度，而當遭遇擁塞時，模式之推估旅行時間誤差亦能透過模式推估旅行時間之調整機制在可接受之誤差範圍內。

其後續研究，如：張惠汶(2001)仍使用相同方法，利用公車在行駛中傳回之GPS定位資料，發展一套資料處理方式，其分析的結果，在資料過濾模式與資料切割模式方面均能達到所期望之功能。以公車做為探針車輛，在速率推估方面，有低估或高估的現象，較不穩定，而由於實例調查所記錄的資訊並不完整，無從推測其造成原因。此研究所發展之模式，有助於提升現裝設有GPS設備支車隊以及將來加裝GPS之車隊其GPS的附加價值，進而取代偵測器的設置並增加路況收集的涵蓋率。

李穎(2002)運用國道客運班車GPS資料、車輛偵測器資料以及事件資料等真實資料，透過類神經網路的手法尋找各項資料來源其參數與旅行時間之關係，構建出國道1號西螺至永康交流道路段之旅行時間預測模式。在最後以真實車流資料對模式輸出結果進行充分驗證，組合各路段預測模式的方式，企圖滿足了實際生活中各個區間使用者需求、匝道進出車流影響與不同路段長度與特性的考驗。

綜觀以上國內外研究中可發現，由於搭乘運輸載具旅行習慣以及道路屬性不同，造成使用者的需求也隨之不同，國外研究所建構出之旅行時間預估系統其大多都是針對城際間之高速公路，而較少針對市區路網建構旅行時間預估模式。

再者國內研究所使用的研究資料大多都是使用現有或者自行撰寫模擬程式來產生研究所需的交通資料，後續驗證的部分則是使用少許的真實資料來驗證其所建構出來的模式。綜觀國內雖有著許多動態資訊系統，但至今仍能未見有人利用

各資料庫中所長期蒐集而來的大量時空資料，故本研究本著使用大量真實時空資料來加以分析、處理，期望能夠建構出更符合真實路況之模式，以造福廣大的大眾交通運輸載具之搭乘者。

三、研究方法

本研究主要是運用探針車輛實際地在研究路段上運行，並且利用 GPS 及 GPRS 等技術將其行駛的時空資料回傳至監控中心，後續則透過許多不同的工具及處理手法來進行資料處理以及分析。在本研究中，完全採取由探針車所回傳的真實資料，藉由大量的歷史資料其所能夠解讀之資訊，相信並不是少許資料或以人工記錄方式所獲得之資料可以比擬。以下則詳敘本研究中所使用之研究方法及資料處理之流程。

3.1 研究對象與範圍

本研究以推估旅行時間為主，將已裝設有全球衛星定位系統及配備GPRS之公車做為探針車輛，並分析其固定每十五秒所回傳之GPS定位資料為主。所使用之探針車資料係採用清雲科技大學行車運籌管理研究中心之商車營運系統資料庫，進行空間資料探勘作業，該資料庫主要配合客運行駛班次以及路線作為基本資料，在公車動態系統的部份目前共有八十輛探針車在不同之路段行駛，然而本研究根據業者所提供之實際班表資料並對照資料庫中之歷史行車資料，將有行駛經研究範圍之探針車挑出，共計有二十三輛探針車輛。

本研究的資料時間範圍為九十八年四月到九十八年五月共兩個月的資料，其全部資料共計三百多萬筆，在考慮公車營運特性及一般大眾搭乘之習慣，取其每週二、三、四共計二十六天之資料，兩百餘萬筆為研究母體，進而切割路段選取研究範圍作為樣本。故本研究排除每週一、五、六、日有可能影響一般大眾搭乘特性之資料。

3.2 資料處理流程

本研究特別研究出一套適用的資料處理流程包含了如圖1的資料庫匯出後的原始資料前處理：將檔案格式轉換成Arc View可接收之檔案格式、整併各車機及日期資料、剔除異常值及極端值、切割研究路段、定義停等範圍、觀察經緯度及方向角變化區分南北向等等。

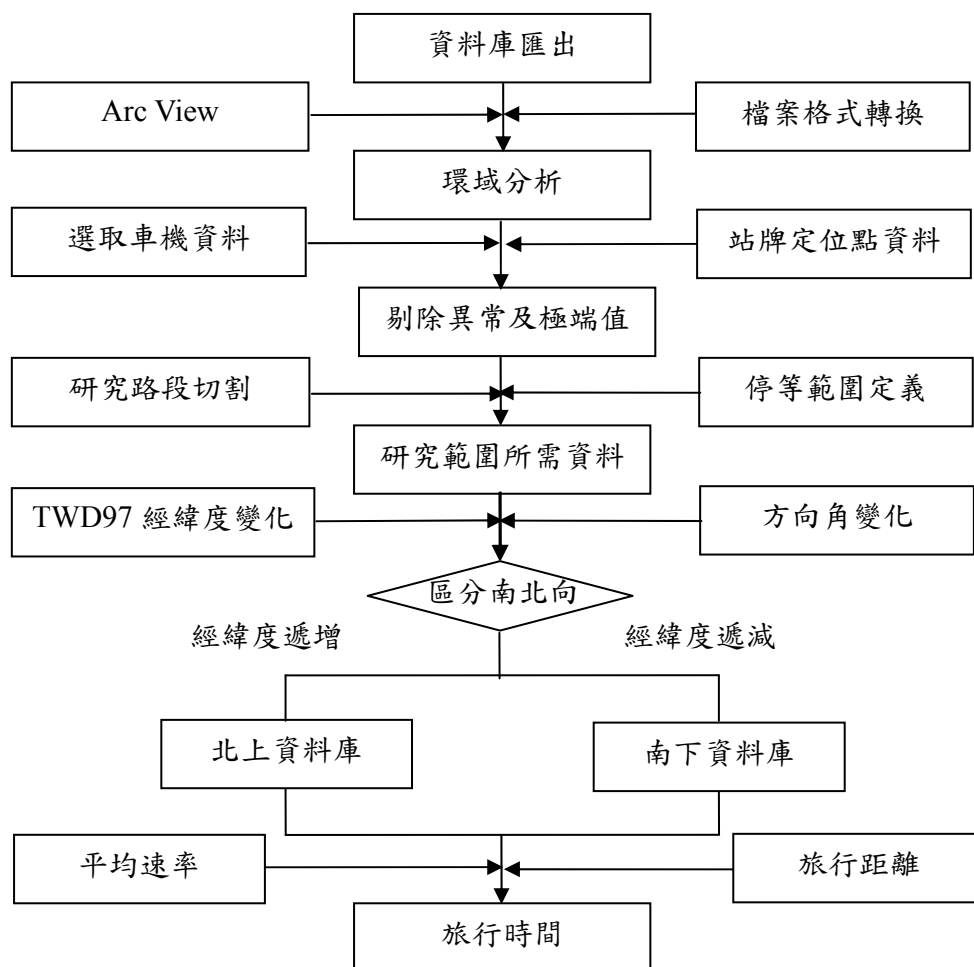


圖 1 資料前處理流程圖

在得出研究原始資料庫及研究範圍資料庫後，明確定義路段定義及劃分停等區域並產生路段資料庫以及停等區資料庫，再者藉由座標系統的轉換(WGS84轉換成TWD97)即可計算出兩點資料的距離進而觀察經緯度及方向角變化來判別南北向以及判斷探針車經過停等車時是否發生停等。

3.3 資料探勘

資料探勘自發展以來，一直是個發展快速的領域，各種相關的研究報告、系統應用或是雛型的發展歷史來看，資料倉儲系統與資料探勘工具的整合在系統中也開始出現，逐漸成為一種趨勢，這樣的結合可以方便於不同類型資料庫進行探勘，如：主動式資料庫、物件導向式資料庫、時間和空間資料等。

資料探勘(Data Mining)為目前學術界與實務業界廣泛運用來處理並分析大量資料的新技術。資料探勘能在大量無明顯特性的資料中，運用各種有系統的方法將隱藏於這些資料中可用的資訊及知識尋找出來，進而可以利用產生的結果建立相關的分類或預測模式，提供使用者作為各類決策的參考依據。隨著資料探勘技術相關研究的出現，不同學者對於資料探勘的定義也提出各自不同的解釋。一般而言，Data Mining 技術可達到下列五項功能：分類(classification)，推估(estimation)，預測(prediction)，關聯分組(affinity grouping)以及叢集或稱同質分組(clustering)。

3.4 資料分析方法

資料採礦是商業智慧相當重要的關鍵技術，透過演算法與產業知識的合作無間，可從龐大的歷史資料中挖掘出未來的脈絡，進而做為決策之參考。SQL Server 2005所擁有完整且強大的演算法，包括：決策與迴歸樹、時間序列、群集與時序群集、關聯規則、貝氏機率分類、類神經網路、文字採礦、線性迴歸，以及BASEL信用評分卡標準演算法羅吉斯迴歸等。

本研究之資料分析方法，除了研擬出一套資料處理流程之外並希望能夠運用新的資訊科技來解決問題，故本研究在資料採礦所使用的軟體技術為2005年底由微軟所推出的SQL SERVER 2005內所內建之資料探勘功能，並在其所提供之九個資料探勘手法中選定以決策數為探勘方法。

四、實證分析

資料前處理的重要性在資料採礦中佔了非常重要的地位，資料的品質與真實性將會直接的影響到後續的分析結果，將研究所需資料處理完畢後匯入資料庫中即可運用 SQL SERVER 2005 內的資料探勘功能進行決策樹分析。

4.1 建構決策樹

本段說明在SQL中決策樹的分割方法，決策樹是通過遞迴分割（recursive partitioning）建立而成，遞迴分割是一種把資料分割成不同小的部分的疊代過程。如果有以下情況發生，決策樹將停止分割：1.該群資料的每一筆資料都已經歸類到同一類別。2.該群資料已經沒有辦法再找到新的屬性來進行節點分割。3.該群資料已經沒有任何尚未處理的資料。

4.2 決策樹分析

本研究的目的是推估旅行時間，故設定旅行時間為預測之變數，輸入變數則選擇日期、兩點平均速率、時段以及日期，決策樹建構完成後，其結果共分割成三層(如圖2)，末端有六個分支，各節點各字經由運算產生不同的迴歸式，總計共得出十五個迴歸式。以下就各分支分別作說明，其運算時間單位為秒，速率單位為公里/小時，：

- (一) 在全部資料的情況下一共輸入 1638 筆資料所得出的旅行時間迴歸式為 $31.091-1.88*(\text{兩點平均速率}-2.094)$ 。
- (二) 當兩點平均速率 ≥ 24.058 及 < 29.795 時，所得出的旅行時間迴歸式為 $23.924-0.894*(\text{兩點平均速率}-27.437)+0.002*(\text{時段}-9.316)$ 。
- (三) 當兩點平均速率 ≥ 47.006 時，所得出的旅行時間迴歸式為 $13.082+0.001*(\text{時段}-7.442)-0.245*(\text{兩點平均速率}-50.162)$ 。
- (四) 當兩點平均速率 ≥ 35.532 及 < 47.006 時，所得出的旅行時間迴歸式為 $16.572-0.401*(\text{兩點平均速率}-39.705)-0.0009*(\text{時段}-8.967)$ 。

- (五) 當兩點平均速率 ≥ 6.847 及 < 12.584 時，所得出的旅行時間迴歸式為
 $66.56 + 0.62 * (\text{時段} - 9.291) - 6.975 * (\text{兩點平均速率} - 10.080)$ 。
- (六) 當兩點平均速率 ≥ 9.142 及 < 9.716 時，所得出的旅行時間迴歸式為
 $69.202 + 0.001 * (\text{時段} - 8.773) - 7.34 * (\text{兩點平均速率} - 9.454)$ 。
- (七) 當兩點平均速率 ≥ 9.716 及 < 12.584 時，所得出的旅行時間迴歸式為
 $58.653 - 0.003 * (\text{時段} - 9.282) - 5.3 * (\text{兩點平均速率} - 11.214)$ 。
- (八) 當兩點平均速率 ≥ 6.847 及 < 9.142 時，所得出的旅行時間迴歸式為
 $80.082 - 10.137 * (\text{兩點平均速率} - 8.223)$ 。
- (九) 當兩點平均速率 ≥ 12.584 及 < 18.321 時，所得出的旅行時間迴歸式為
 $44.802 * (\text{兩點平均速率} - 14.78) - 0.001 * (\text{時段} - 9.236)$ 。
- (十) 當兩點平均速率 ≥ 13.158 及 < 13.731 時，所得出的旅行時間迴歸式為
 $48.863 + 0.0005 * (\text{時段} - 9.45) - 3.637 * (\text{兩點平均速率} - 13.387)$ 。
- (十一) 當兩點平均速率 ≥ 13.731 及 < 18.321 時，所得出的旅行時間迴歸式為
 $41.648 - 2.613 * (\text{兩點平均速率} - 15.812) + 0.003 * (\text{時段} - 8.923)$ 。
- (十二) 當兩點平均速率 ≥ 12.584 及 < 18.321 時，所得出的旅行時間迴歸式為
 $50.66 - 3.948 * (\text{兩點平均速率} - 12.914) + 0.0002 * (\text{時段} - 10)$ 。
- (十三) 當兩點平均速率 < 6.847 時，所得出的旅行時間迴歸式為 $176.795 - 52.168 * (\text{兩點平均速率} - 4.54)$ 。
- (十四) 當兩點平均速率 ≥ 29.795 及 < 35.532 時，所得出的旅行時間迴歸式為
 $20.036 - 0.614 * (\text{兩點平均速率} - 32.742) + 0.0002 * (\text{時段} - 9.141)$ 。
- (十五) 當兩點平均速率 ≥ 18.321 及 < 24.058 時，所得出的旅行時間迴歸式為
 $30.713 - 0.006 * (\text{時段} - 9.153) - 1.448 * (\text{兩點平均速率} - 21.420)$ 。

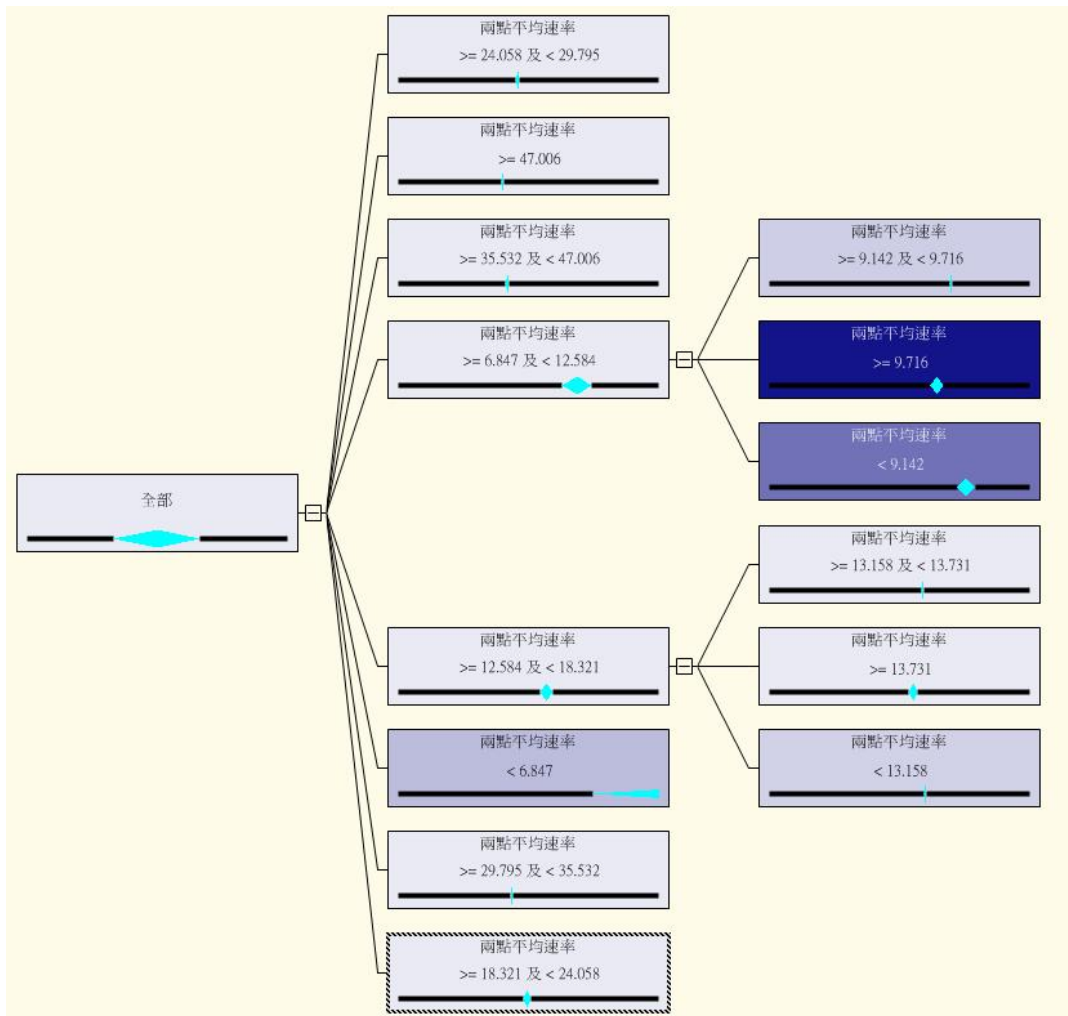


圖 2 決策樹分割

由於決策樹分割成三層，並由決策樹中得知兩點平均速率共被分成十二種狀況，故在迴歸式的使用上可依兩點平均速率的不同而運用不同之迴歸式推算兩站旅行時間。

4.3 相依性分析

藉由相依性網路可以得知自變數和應變數之關係，當關聯性越強則其相依性越強，由圖3可以得知當將關聯性遞增時自變數-時段對於兩站旅行時間的限消失，也就是兩點平均速率對於兩站旅行時間的關連性高於時段對於兩站旅行時間的關連性。故迴歸式中兩點平均速率的係數較時段之係數為大。

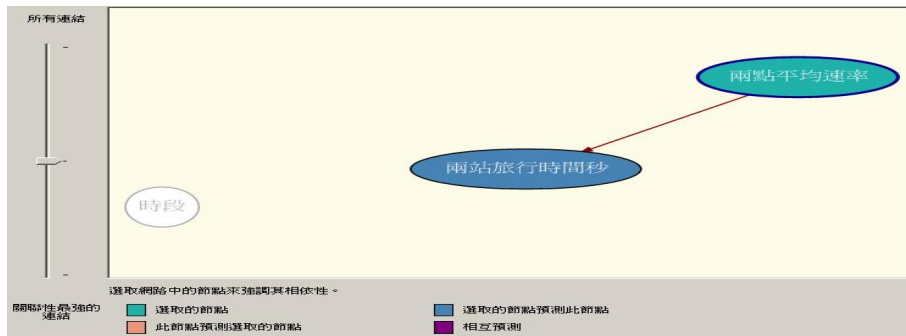


圖 3 相依性網路

4.4 散佈圖分析

如圖4，藉由資料採礦的散佈圖可以看出決策樹所建構出的模式是否準確，由圖中可知所建構出的決策樹在實際值旅行時間為95秒以下時其模式所預測之預測值與實際值相同，故本研究之準確度尚可接受。

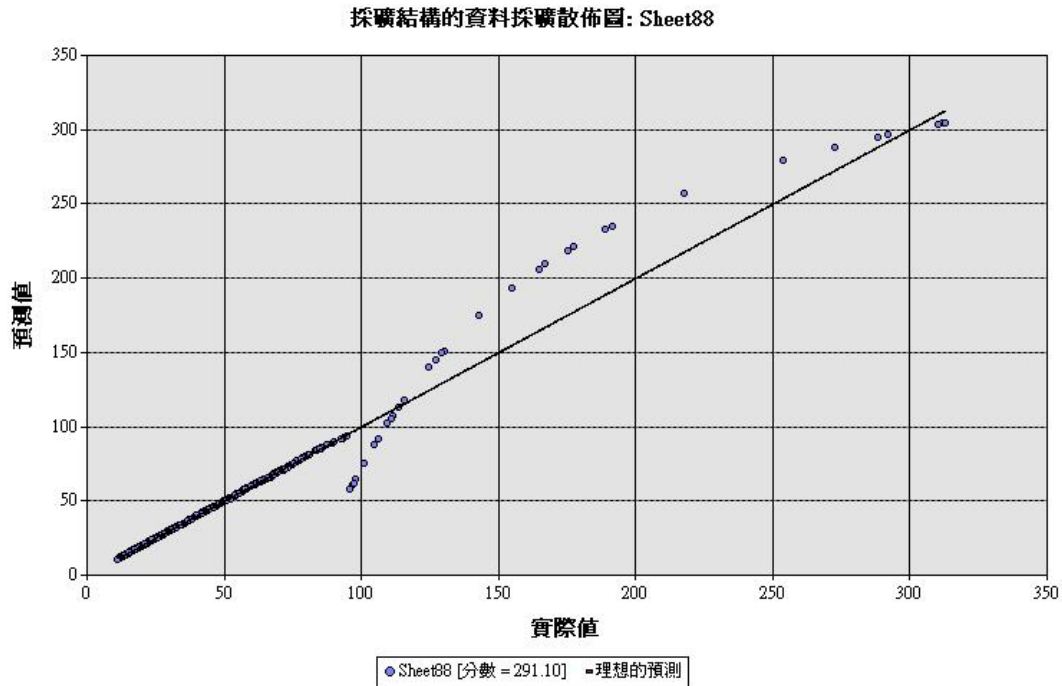


圖 4 資料採礦散佈圖

五、結論與建議

綜合前述實證分析之結果，總結本研究之結論，最後則提出未來研究建議，以供後續研究者參考。

5.1 結論

本研究以桃園客運之1路公車為研究對象，藉由發展出的資料處理模式，將GPS所回傳之原始資料加以整合分析，並且嘗試運用新的資訊科技來探討市區公車的旅行時間。本研究利用了SQL SERVER 2005所內建之資料探勘功能，企圖使用大量時空資料並且結合決策樹來建構推估模式，其中模式一共分割出了十二個迴歸式；並且由資料採礦散佈圖得知，在不同的兩點平均速率上代用不同的迴歸式其準確度在兩站旅行時間為95秒以下時預測值與實際值相同，並且在九十五秒以上的旅行時間之預測準確度也尚可接受。

5.2 研究限制與後續研究建議

本研究對象為桃園客運1路公車，行經範圍道路特性為基本路段，由於站牌點與交通號誌點的距離相距甚近，故本研究不另將交通號誌納入停等點之列。由於研究的過程中，受到時間的限制，雖然有將路段加以切割，但僅以公車營運路段之兩站牌間作為研究對象，代表性可能有所不足，因此建議後續研究可以擴大研究範圍，將營運路段依照站牌間即停等區分割成數段，並產生不同之模型，以使研究結果更具說服力；此外本研究僅使用單一模型作為探討，建議後續研究者可利用其餘不同之資料採礦模型加以分析比較。

參考文獻

1. Choi, D. B., Ko, H. S., and Ahn, B. H., 1998, On Multisensor Data Fusion using Attribute Association for Intelligent Traffic Congestion Information Inference, *5th ITS World Congress*.
2. Ding, Y., Chien, S. I., and Wei, C. H., 2000, Dynamic Transit Arrival Time Prediction Using Link-Based and Stop-Based Artificial Neural Networks (Revised Version), *Submit to Transportation Research, October*.
3. Jasperse, D. and Toorenburg, J. V., 1999, Real-Time Estimation of Travel-Times and Queue-Lengths A Practical Study, *6th ITS World Congress*.
4. Frederick W. Cathey and Daniel J. Dailey, 2004, Estimating Corridor Travel Time by Using Transit Vehicles as Probes, *Transportation Research Record 1855*, pp. 60-65.
5. Frederick W. Cathey and Daniel J. Dailey, 2001, Transit Vehicles as Traffic Probe Sensors, *Transportation Research Record 1804*, pp. 23-30.
6. 李穎，2002，類神經網路應用於國道客運班車旅行時間預測模式之研究，成功大學，交通管理科學研究所碩士論文。
7. 吳佳峰，2001，有GPS資訊提供下之車輛旅行時間預估模式之研究，交通大學，交通運輸工程研究所碩士論文。
8. 張惠汶，2002，利用公車GPS定位資料推估路段行車速率之研究，交通大學，運輸科技與管理學系碩士論文。