

國立交通大學

運輸科技與管理學系

碩士論文

應用 k -NN 模式於市區公車到站時間預測之研究

The Study of Urban Bus Arrival Time Prediction
Using the k -NN Model



研究生：何佳儒

指導教授：王晉元 教授

中華民國九十九年六月

應用 k -NN模式於市區公車到站時間預測之研究
The Study of Urban Bus Arrival Time Prediction
Using the k -NN Model

研究生：何佳儒

Student : Chia-Ju Ho

指導教授：王晉元

Advisor : Jin-Yuan Wang

國立交通大學
運輸科技與管理學系
碩士論文



Submitted to Department of Transportation Technology and Management

College of Management

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Transportation Technology and Management

June 2010

Hsinchu, Taiwan, Republic of China

中華民國九十九年六月

應用 k -NN模式於市區公車到站時間預測之研究

學生：何佳儒

指導教授：王晉元 博士

國立交通大學運輸科技與管理學系碩士班

摘 要

先進大眾運輸系統 (Advanced Public Transportation System, APTS) 的公車到站時間預測能提供大眾運輸系統使用者有用的乘車資訊，降低民眾搭公車的等待時間，藉其提昇民眾搭乘公車意願，減少使用私人運具。在車隊管理方面，亦可提供業者公車到站時間，協助業者有效管理車隊。

市區公車的到站時間預測，需考量路口號誌及站牌上下客的停等時間，因此本研究以公車裝配之全球定位系統(Global Positioning System, GPS)回傳的即時資料為主要資料來源。

本研究先利用 GPS 資料建立歷史資料庫，在資料比對上採用 k -NN 法預測公車到站時間。其中，因為實際的路段資料可能會有遺漏的情形，將歐基里德距離除上可取得的路段資訊數量作調整，以避免因較多的路段資訊造成較大的誤差。在計算路段平均速度時，採用延遲加總的方法，以反映車輛在行進中所遭遇的交通狀況。最後以台北市區公車為實測對象，以隨車調查的實際到站時間作驗證，探討在不同情況本研究方法的表現績效。

就整體結果而言，本研究與顯著優於現有系統之預測表現，表示本研究之預測模式可用，且在少站位數及下午尖峰情境下，本演算法有比較好的表現。

關鍵詞： k -NN 法，市區公車，到站時間預測

The Study of Urban Bus Arrival Time Prediction Using the k-NN Model

student : Chia-Ju Ho

Advisors : Jin-Yuan Wang

Department of Transportation Technology Management
National Chiao Tung University

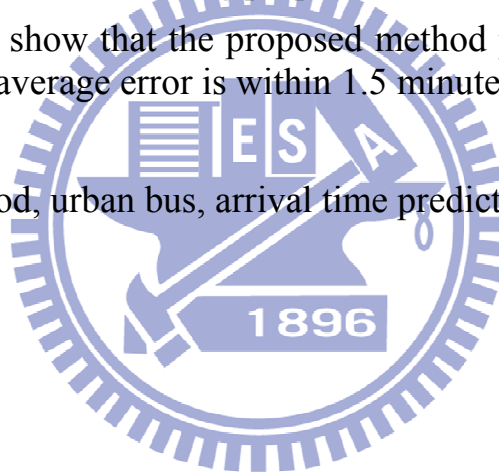
ABSTRACT

The estimation of urban bus arrival time using the GPS data is a fundamental function of Advanced Public Transit Systems (APTS). This is a difficult task due to the complications of traffic, signals, and bus stops.

We first propose a database schema to store the collected historical data. Then, a k-NN based method is developed to derive the real time estimation of bus arrival time. The real world data collected from various routes in Taipei City is used for testing purpose.

The testing results show that the proposed method performs better than the existing methods. The average error is within 1.5 minutes.

Keywords: *k*-NN method, urban bus, arrival time prediction.



誌 謝

首先感謝這兩年給予我指導的王晉元老師，在完成論文的過程中，不厭其煩地幫我釐清問題點，協助我找出解決方法；並給我許多參與計畫的實作經驗，讓我可以從不同的角度觀察問題，亦可學習表達的能力；另外，在待人處世上也給與了我相當大的影響及鼓勵，期望在未來的日子裡依然能夠與老師維持良好的互動。接下來要感謝中華大學的蘇昭銘老師及系上的黃家耀老師，幫我找出了論文上的瑕疵以及一些說明不清之處，讓我有能力將這篇論文順利完成。也要謝謝系上的老師在這兩年給我的指導與體諒，讓我增進不少見聞。

在研究的過程中，感謝學長姐彥佑、思文、小松、葉珮、健綸、黃恆昱光、老總等，在參與計畫的合作、研究或程式等各方面的幫助；謝謝一起努力的夥伴 Daisy、小惠、ANIGI、許大爺、思琪、志霖、怡婷，在為了論文而不眠的夜晚，帶給我許多的支持與幹勁。

另外在 11 樓的生活中，雖然有做研究的苦悶，但因為許多人的陪伴與邀請，讓我能夠參與許多的娛樂活動，將不好的情緒不定期的一掃而盡，豐富我原本可能超乾渴的研究生生活，謝謝 hoho、貽婷學姊、GoGoBoys、小朱、依潔、阿聰師、阿芬、大哥、劉婉、家誼、小薇和其他族繁不及備載的學長姐、同學、學弟妹們，儘管是小小的聊天或興趣上的交流，都是我這兩年不可或缺的調劑品。

最後，也是最重要的，要感謝我的家人，願意在我的求學生涯中，給我無邊無際的支援與支持，讓我能夠勇敢面對所有的問題，所有的難關，所有的所有，讓我知道我有後盾可以讓我依靠，謝謝你們，在此僅以此篇論文獻給在我身邊的你們。

何佳儒 謹 誌于
國立交通大學運管所
中華民國九十九年六月

目 錄

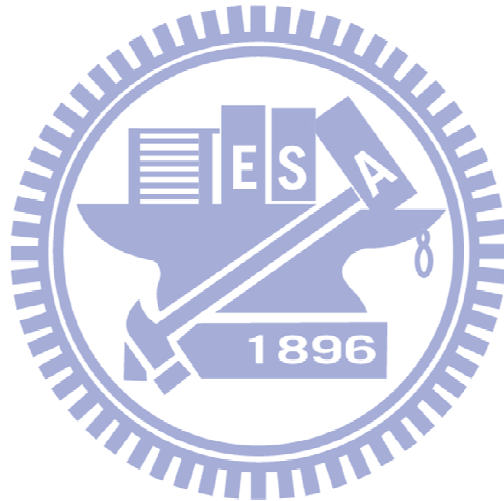
中文摘要.....	i
英文摘要.....	ii
誌 謝	iii
目 錄	iv
圖目錄	vi
表目錄	vii
第一章 緒論	1
1.1 研究動機.....	1
1.2 研究目的.....	1
1.3 研究範圍.....	1
1.4 研究流程.....	2
第二章 文獻回顧	4
2.1 無號誌化路口道路的旅行時間預測研究.....	4
2.2 市區旅行時間預測研究.....	6
2.3 小結.....	8
第三章 公車到站時間預測模式.....	9
3.1 k -NN 模式介紹.....	9
3.2 k -NN 之公車到站時間預測模式構建.....	11
3.3 範例說明.....	15
3.4 研究課題.....	22
第四章 資料處理	24
4.1 資料過濾.....	24
4.2 建立資料庫.....	24
4.2.1 路線分段.....	25
4.2.2 計算路段旅行時間.....	25
4.2.3 劃分時段.....	25
4.3 設定參數.....	25

4.3.1	距離量度	25
4.3.2	不完全路段的旅行時間計算	26
4.3.3	k 值設定	27
4.3.4	時間窗	27
第五章 實證分析		28
5.1	現況說明	28
5.2	資料蒐集	30
5.2.1	交通資料	30
5.2.2	實際到站資料	31
5.2.3	現有系統預測資料	31
5.3	參數設定	32
5.3.1	k 值設定	32
5.3.2	時間窗設定	32
5.4	結果分析	32
5.5	與現有預測系統比較分析	38
5.6	小結	41
第六章 結論與建議		42
6.1	結論	42
6.2	建議	42
參考文獻		44
簡 歷		46



圖目錄

圖 1 研究流程圖	3
圖 2 第一階建立資料庫流程圖	11
圖 3 第二階段 k -NN 預測流程圖	13
圖 4 範例現況說明示意圖	15
圖 5 路線分段示意圖	25
圖 6 計算路段旅行時間示意圖	25
圖 7 範例不完全路段示意圖	27
圖 8 現有系統一、四期廠商之到站時間演算流程	29
圖 9 現有系統二、三期廠商之到站時間演算流程	30
圖 10 k 值與絕對誤差趨勢圖	32
圖 11 預測誤差分配圖	33



表目錄

表 1 範例路段長度	15
表 2 範例歷史資料-路段旅行時間	15
表 3 範例歷史資料-路段平均速度	16
表 4 範例現狀旅行時間	18
表 5 範例距離計算	18
表 6 範例五筆歷史路段旅行時間計算(1)	19
表 7 範例五筆歷史路段旅行時間計算(2)	20
表 8 範例五筆歷史路段旅行時間計算(3)	20
表 9 範例五筆歷史路段旅行時間計算(4)	20
表 10 範例五筆歷史路段旅行時間計算(5)	21
表 11 範例五筆歷史路段旅行時間計算(6)	21
表 12 範例五筆歷史路段旅行時間計算(7)	21
表 13 路段資料欄位說明	24
表 14 路段旅行時間資料	26
表 15 歷史與現狀之路段旅行時間比較	26
表 16 定時回報資料	30
表 17 定點回報資料	31
表 18 測試路線	31
表 19 整體誤差結果	33
表 20 早到/晚到誤差結果	33
表 21 站位數資料分析	34
表 22 距離站位數 ANOVA 分析	34
表 23 站位數屬性差異分析	34
表 24 尖離峰資料分析	35
表 25 尖離峰情境 ANOVA 分析	35
表 26 尖離峰屬性差異分析	35
表 27 平假日資料分析	36
表 28 平假日情境 ANOVA 分析	36
表 29 公車專用道資料分析	36
表 30 有無公車專用道情境 ANOVA 分析	36
表 31 列隊資料分析	37
表 32 有無列隊情境 ANOVA 分析	37
表 33 演算法運算時間	37
表 34 各路線運算時間	38
表 35 整體誤差與現有系統比較結果	38
表 36 整體資料與現有系統之差異分析	38

表 37	早到與晚到資料與現有系統比較結果	38
表 38	早到與晚到資料與現有系統之差異分析	39
表 39	站位數資料分析與現有系統比較結果	39
表 40	站位數資料各與現有系統之差異分析	39
表 41	尖離峰資料與現有系統比較結果	39
表 42	尖離峰資料各與現有系統之差異分析	40
表 43	平假日資料與現有系統比較結果	40
表 44	平假日資料與現有系統之差異分析	40
表 45	有無公車專用道資料與現有系統比較結果	40
表 46	公車專用道資料與現有系統之差異分析	41
表 47	公車列隊資料與現有系統比較結果	41
表 48	公車列隊資料與現有系統之差異分析	41



第一章 緒論

1.1 研究動機

先進大眾運輸系統（Advanced Public Transportation System, APTS）是以先進電子、通訊、電腦等相關科技協助大眾運輸提昇效率與服務水準，使民眾能夠多利用大眾運輸工具，減少使用私人運具，進而降低空氣污染及噪音等對環境所造成的負面衝擊，節省能源消耗。

公車到站時間預測能提供大眾運輸系統使用者有用的資訊，降低民眾搭公車的等待時間與等待期間的不確定感，藉其提昇民眾搭乘公車意願與方便性。除此之外，在車隊管理方面，可提供業者公車到站時間，協助業者有效管理車隊。公車的到站時間預測屬於旅行時間預估預測的一種，旅行時間的預估預測根據道路屬性大致上能夠分為無號誌化路口及有號誌路口；根據預測的空間距離則可分為長距離及短距離。而公車的到站時間預測屬於有號誌化路口的旅行時間預測預測，其中若是班次較為密集的路線，預測的下一班公車為三、五站的短距離預測；若為班次較為稀疏的路線，則目標會延伸為整條路線的長距離預測。

無號誌化路口的旅行時間預測，如近年來蓬勃發展的高速公路旅行時間預測演算法，至今，已有許多學者針對各種預估預測方法作研究探討；有別於高速公路車流同質性高、有明確起迄點的特性，市區的旅行時間預估預測需要加入更多的考量因素，如路口的停等，而預估預測公車的到站時間尚要考慮站牌的停等時間及是否設有公車專用道等差異，因此，市區公車的到站時間預估預測比無號誌預測更困難。

有鑑於過去旅行時間預測文獻大部分以高速公路為研究對象，較少針對市區公車的到站時間預測加以探討，因此本研究以市區公車作為研究對象，並利用公車定位訊息當作資料來源，發展一套市區公車的到站時間預測演算法。

1.2 研究目的

本論文之研究目的為利用 GPS 資料為主要資料來源，發展一套市區公車的到站時間預測演算法，並且能夠廣泛應用到各種型態的公車路線。期望提供精確之公車到站時間，提供民眾及公車業者參考之用。

1.3 研究範圍

研究以市區公車作為研究對象，藉由實際取得的公車到站資訊，當作實際公車到站時間資料進行模式驗證。並進一步分析各種型態的績效表現，探討演算法在不同情境下有無明顯差異。

本研究對象包括有無行駛公車專用道的路線、離峰與尖峰時段、平假日及站位數等屬性。

1.4 研究流程

本研究流程如圖 1 所示，其詳細步驟說明如下：

(1). 確定研究問題與範圍

決定實際案例的研究對象以及預測範圍起迄點，並根據研究背景與目的將問題進行清楚地描述與界定。

(2). 文獻回顧

蒐集國內外研究旅行時間推估的相關文獻，分析市區公車與高速公路旅行時間預測的特性，針對市區公車歸納出各種演算法的優缺點，進而整理出本研究的研究方向。

(3). 建立預測模式

利用公車 GPS 資料建立可行之公車到站時間預測模式。

(4). 資料處理

所得之資料必須先進行過濾，排除非交通狀況所造成的異常資料點，以提高推估模式的精確度。

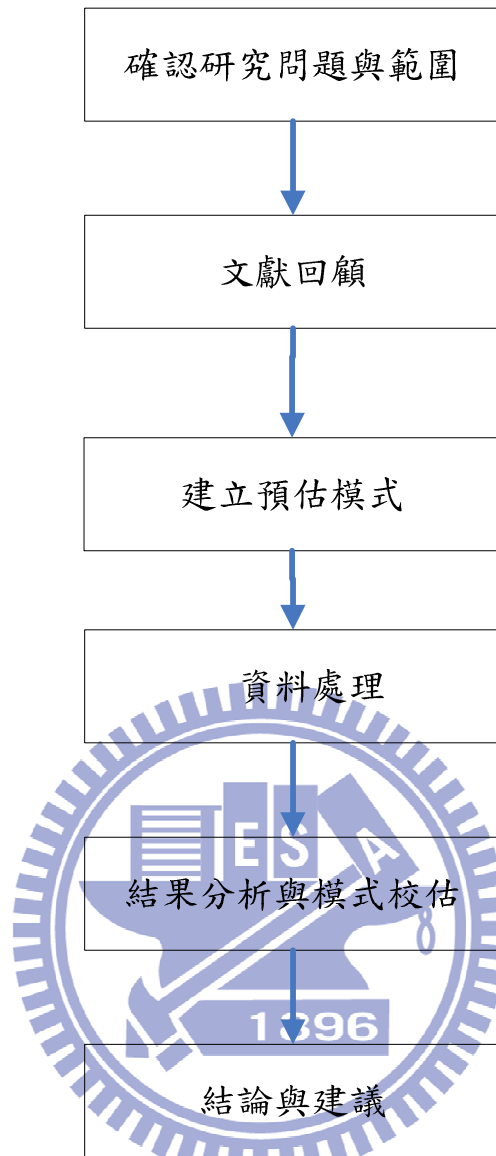
(5). 結果分析與模式校估

將發展之公車到站時間預測模式應用在實際公車行駛路段，以驗證此模式是否適用於現實狀況，並由運作結果對模式作調整與修正。

(6). 結論與建議

對本研究過程與結果提出結論與建議，並提出後續可以進行研究的方向。





第二章 文獻回顧

市區公車的預測到站時間屬於旅行預測時間的一環，就文獻而言，旅行時間預測可分為有號誌化路口及無號誌化部分，而公車到站時間的預測屬於有號誌化路口的預測，以下針對這兩大部分進行文獻回顧：

2.1 無號誌化路口道路的旅行時間預測研究

無號誌化路口部分在高速公路的領域中已有許多學者投入研究，多數之研究以道路偵測器作為主要的資料來源。

Huang[1]、Kwon[2]及 Zhang[3]等人利用上、下游偵測器的點速度與路段長路求得路段中的平均旅行時間，進而依即時偵測到的交通資訊預測目標路徑的未來旅行時間。Huang[1]假設現況的旅行時間及歷史的旅行時間與未來的旅行時間存在自然對數的線性關係，研究以 PeMS 資料庫與實際探針車 GPS 資料作模式驗證，利用均方根誤差(Root Mean Square Error, RMSE)來衡量此模式的預測結果。Kwon[2]假設整段路徑的旅行時間與當下每個路段旅行時間總和存在線性關係，研究以美國 I880 高速公路為對象預測未來 15 分鐘的旅行時間，實驗以平均絕對誤差(Mean Absolute Error, MAE)當作評估指標。Zhang[3]假設目前的旅行時間與未來的旅行時間存在線性關係，利用時變係數模式(time-varying coefficient)來預測未來的旅行時間，研究以美國 I880 高速公路為對象，利用平均絕對誤差(Mean Absolute Percentage Prediction Error, MAPPE)當作評估指標，結果顯示 TVC 的表現比用歷史平均資料預測的表現來得好。

Sun[4]、Li[5]、Kwon[6]等人利用現況與歷史資料先預測出未來時間的交通資料後，再以其資料推估出路段或路徑的旅行時間。Sun[4]利用局部線性迴歸(local linear regression)預測出未來時間點的速度資料，由路段長與速度預測值推估出路段的旅行時間，研究以 US-290NW 高速公路為對象，並以自動車牌辨識(Automatic Vehicle Identification, AVI)監控系統資料庫作為驗證資料。Li[5]假設旅行時間與車流量有關，利用關係式(1)推得旅行時間：

$$T_t = A + B_t(F_t)^n, \quad B_t > 0, n \geq 1 \quad (1)$$

其中 T_t 為 t 時間的旅行時間， A 為自由車流下旅行時間， B_t 為 t 時間的參數值， F_t 為 t 時間的車流量， n 為指數；研究以卡門濾波器(Kalman Filter)預測未來時間的車流量 F_{t+1} 與係數 B_{t+1} 。Kwon[6]假設相同星期別、相同時間點具有相同的流量、佔有率，試著找出佔有率、流量與旅行時間之間的關係，對探針車得到的實際旅行時間做線性迴歸。利用美國 I880 資料庫來建立旅行時間預測模式，實驗以均方根預測誤差(Root Mean Squared Prediction Error, RMSPE)作為評估指標。

Sun 等人[7]利用相鄰三組偵測器的速度推估出合適的二次方速度曲線，再使用積分找出路段的旅行時間，並加入路段速度最大值與最小值的範圍限制。研究對象為美國 I66 高速公路，測試路段長度為 4.82 英哩，實驗結果的相對誤差約為 5%。

Chang 等人[8]將演算法分為旅行時間推估與預測模組兩部分。在路段旅行時

間推估模組中將偵測器所得的流量、佔有率代入群集線性迴歸式路段推估旅行時間，依照上下游針測器的交通擁擠狀況將路段的交通分為四類，再根據不同交通情況建立其線性迴歸模式；而預測模組應用 k -NN(k -Nearest Neighbors Model)與增強時變係數 (Enhanced Time-Varying Coefficient Model)作預測。測試路段為美國高速公路 MD70 至 I695 長 25 英哩，研究結果的預測準確度平均約為 95.8%。

Yeon 等人[9]將路段分為擁塞、非擁塞兩種情況，非擁塞情況下路段旅行時間為自由車流速度下的旅行時間，在擁塞情況下則以 $T(f)=a \times e^{bf}$ 關係式推估路段旅行時間，其中 T 為旅行時間， f 為車流量， a 、 b 為校估參數，收集資料的時間間隔為 1 分鐘。實驗對象為 US202 高速公路，測試路段長為 8 英哩，預測結果以統計檢定檢驗與實際旅行時間是否有差別，而在 99%信賴水準下的檢定結果為推估值與實際值無差別。

Cortes 等人[10]利用反覆推估的方式求得路段的旅行時間，假設路段速度為上下游偵測器速度的線性組合：

$$\bar{S}_{12}(t_1) = \alpha \bar{u}_1(t_1) + (1 - \alpha) \bar{u}_2(t_1 + tt_{12}) \quad (2)$$

$\bar{S}_{12}(t_1)$ 為 t_1 時間的路段速度， $\bar{u}_1(t_1)$ 為在 t_1 時間的上游偵測器速度， tt_{12} 為行駛目標路段的旅行時間， $\bar{u}_2(t_1 + tt_{12})$ 為在 $(t_1 + tt_{12})$ 時間的下游偵測器速度， α 為權重參數。如果路段速度 \bar{S}_{12} 已知，則路段旅行時間為：

$$tt_{12} = X_{12} / \bar{S}_{12}(t_1) \quad (3)$$

其中 X_{12} 為路段長度，由式(2)與式(3)進行反覆地計算，直到所得的 tt_{12} 與上一回合計算所得的 tt_{12} 的差在可接受的範圍內，則輸出為此路段的推估旅行時間。測試路段為美國 I405 高速公路。

Paterson 與 Rose[11]利用排隊理論來描述在路段中的車輛通過情形，將路段的旅行時間分為自由車流下的旅行時間與延滯時間的總和，假設車流中的車輛皆符合先進先出的行為方式，利用上下游偵測器即可計算路段中的車輛數，如果車輛數小於道路容量，則路段旅行時間為自由車流速度下的旅行時間；反之，車輛數大於道路容量，代表此路段發生擁塞情形，則需加入延滯時間。

在 GPS 系統開始發展後，亦有學者開始使用 GPS 當作資料來源預測高速公路的旅行時間。Hellenga[12]使用探測車推估路段的旅行時間，測試抽樣偏差對調查估計的影響。該研究理論基礎為等候理論，可以得到在平均延滯的抽樣估計中，其到達時間的分配，偏差及調查比例這兩項因素將導致計算結果的誤差。此研究的推導與模擬分析顯示，若以調查為基礎的平均旅行時間為基準之模式下，其旅行時間會因為抽樣誤差而導致預測結果偏高。

使用單一交通資料來預測旅行時間會因為資料的屬性及缺漏有所偏誤，因此在多種交通資訊可取得的情況下，發展出利用資料融合的技術來預測旅行時間。Hellenga 及 Gudapati[13]提出多種交通資料來源融合為單一路段旅行時間的推估模式。研究使用 VD、GPS 及駕駛路況回報等資料庫來發展資料融合的模式，如果有多種資料來源則先各自推估旅行時間後，再進行資料融合步驟得出最後所要的路段旅行時間推估值。

Choi 及 Chung[14]使用 VD 與探針車 GPS 資料來推估旅行時間。利用上游偵

測器的速度與路段距離求出一分鐘的路段旅行時間，再用 voting 資料融合方式將一分鐘的資料合併為五分鐘的資料；由 GPS 所蒐集的資料利用地理資訊系統技術找出路段的旅行時間。由兩種交通資料所得的旅行時間運用 Bayesian pooling 方式融合，得出最後所要的路段旅行時間。

Soriguera[15]、Faouzi[16]等人使用 VD 與 ETC 資料來推估旅行時間。Soriguera[15]將 VD 資料分別利用 spot speed 和 cumulative flow balance 方法推估出路段旅行時間後，運用 fuzzy logic 融合兩個推估值為 *ITT*；將 ETC 所得的路段旅行時間當作 *RTT*。利用貝氏定理融合 *ITT* 與 *RTT* 得出最後所要的路段旅行時間。Faouzi[16]將路段上的速度以距離最近的 VD 速度為準，再以路段長除以速度求出路段旅行時間；推估出的路段旅行時間與 ETC 所得的路段旅行時間以 Dempster-Shafer 演算法融合，得出所要的路段旅行時間。

Nanthawichit 等人[17]利用路段長與路段速度求得旅行時間。路段區域的分配原則是將 VD 的位置盡量分配到區域中間，以其速度當作路段的平均速度；若區域中有回傳 GPS 資料，則依照資料的可靠度加權平均。所得的旅行時間推估值運用卡門濾波器(Kalman Filter)做短期間的測模式。

Smith 及 Demetsky[18]對 *k*-NN 法進行績效評估，分析比較以下四種交通流量的預測方法：歷史平均法，時間序列法、類神經網路法與 *k*-NN 法。評估的方法為利用歷史的流量資料預測未來的流量再和實際值做比對，判斷那一個方法的績效較差，結果發現當歷史資料大時，*k*-NN 法所預測的誤差結果比其他三種方法所得到的結果誤差來得小，有較佳的預測能力。Smith、Williams 以及 Oswald[19]將權重(Weight)的想法加入 *k*-NN 的模型，考慮目前的流量可能會與先前的交通流量有關係，但是隨著時間距離的增加，其影響的程度愈小，所以利用權重的方法進行修正，將不同時間所造成的影響納入考慮。

Clark[20]除了流量外，試著利用 *k*-NN 法試著針對其他可蒐集到的交通資訊進行預測，如速度和佔有率。接著進行一一分析和交叉分析，結果發現當同時利用流量、佔有率和速率這三項變數進行分析時，所得到的預測值相對於個別比對而言，誤差有下降的現象，所以愈多變數進行討論可以得到更精確的結論。

2.2 市區旅行時間預測研究

市區旅行時間預測屬於有號誌化路口的部份，市區的平面道路屬於開放性的道路，行駛在路上的車種多樣，所包括行進速度範圍廣泛，且於非限制路段能夠自由駛離車道或在路口轉向，在路上亦會遭遇到號誌停等的狀況，其交通狀況的變化較大。Choi 等人[21]利用 GPS 以及電子地圖來計算路段之動態旅行時間，並針對測量路段旅行時間的技術作比較；另外在都市幹道的網路上，路段旅行時間是最易受到號誌系統與環境交通因素的影響。因此，為得到路段的旅行時間，最常用的方法是浮動車輛法(floating car method)，推估點至點間的旅行時間。實驗利用所蒐集到的實際路網資料與此研究所計算出的旅行時間值進行績效評估，可以驗證結果準確。Robinason 和 Polak[22]利用 VD 資料來預測市區的路段旅行時間，以 *k* 最近鄰點法建立預測模式，研究與回歸方法做比較，在各種情境下皆顯示有較好的預測能力。

在市區旅行時間預測的領域中又以公車的到站時間預測為大宗，公車行走的路線、班次固定，大部分的公車業者皆會在營運中的公車車輛上裝配 GPS 相關系統，記錄公車實際的行車、停等狀況，因此大部分的公車到站時間預測研究以公車的 GPS 訊息當作預測的資料來源。

Shalaby 及 Farhan[23]利用 GPS 及自動乘客計數器 (Automatic Passenger Counting, APC)資料預測市區公車的到站時間。研究模式將到站時間分為路段間的旅行時間與停靠站的乘客上下車時間，分別利用卡曼濾波器來預測停靠站間的車輛行駛時間與旅客到達率；測試資料為美國多倫多地區的 5 號公車路線，實驗結果的 MRE 約為 6%、平方根相對誤差(Root Squared Relative Error, RSRE)約為 7.3%。

Patnaik 等人[24]利用 APC 資料預測市區公車的到站時間。研究路段區域以高需求量的公車停靠站作分別，利用線性迴歸模式預測任兩站間的公車旅行時間，其變數包括兩站之間的距離、兩站之間的累積停等時間、兩站之間的累積停等站數及區別時段的伯努力變數；實驗結果的 RMSE 約為 3%， R^2 介於 0.96 至 0.99 之間，顯示研究的迴歸模式具有良好的預測能力。

Chung 及 Shalaby[25]利用 GPS 資料預測校車的到站時間，利用前五天的歷史資料推估歷史一般狀況，再用現況的每站延誤時間與當天的天氣狀況調整校車的到站時間；研究並考慮到實際應用時，發佈的到站預測時間比實際到達時間晚的話，會造成學生無法搭上校車的情況，因此加入調整機制將發佈的到站預測時間輕微提早。研究方法與歷史移動平均法、迴歸法比較，在各種情境下皆顯示有比較低的預測誤差。

吳佳峰[26]利用車輛歷史資料預測公車的到站時間，將旅行時間分為車輛運行時間與車輛停等時間兩部分，依 GPS 所回傳的實際車輛行駛速度及前車經過路段的旅行時間調整預測車輛運行時間，並隨車輛實際停等狀況調整未經過路段的預測車輛停等時間；最後將所有車輛未經過路段之預測車輛運行時間及預測車輛停等時間加總，即可得到預測車輛旅行時間。研究以國內客運業者資料做測試，結果發現模式在一般時段有不錯的預測能力，而當遭遇非重現性擁塞時，亦能夠過調整機制將預測的旅行時間控制在可接受之誤差範圍內。

陳建名[27]利用 GPS 資料預測市區公車的到站時間，將公車旅行時間切割為車輛運行時間和車輛停等時間，其停等時間包括於交叉路口停等號誌之停等時間，及公車在停靠站牌載客上下車所發生之停等時間。預測車輛運行時間在公車正常情況下行駛時，使用歷史平均速度作預測，當交通狀況出現異常狀況時，則使用前車資訊作預測。在停等時間之預測，以改變點分析將歷史資料庫依車速高低之不同型態切割成數個不同時段，而在不同車速之時段，此研究將以不同預測模式來預測公車於交叉路口停等號誌之停等時間。

2.3 小結

在本章之文獻回顧可發現，國外實作之旅行時間預測系統主要多針對城際間之高速公路，而較少針對市區路網。在高速公路的環境下，旅行時間之變動因素主要來自車流速度(此速度與車流量以及道路容量有關)，以及突發之事故狀況，為較單純之系統環境，大部分使用 VD 資料建立預測模式。反之，預測市區內旅次之旅行時間，則必須多加考慮路口號誌造成的停等情況，若是預測對象為市區公車的到站時間，則需考量到站牌上下客的停等時間，大部分研究以 GPS 訊息當作資料來源，以取得車輛實際的行走、停等的時間及位置。

在有關旅行時間預測方法的部分，迴歸法是常用的預測方法，可運用在預測旅行時間週期性的變化上，但因為迴歸法缺少彈性變化，無法即時反應旅行時間的變化情形。最近興起的 k -NN 方法就其精神亦是文獻常用的相似歷史資料比對的方法，在非號誌化路口的路網上利用 k -NN 法預測旅行時間的文獻亦日益增多，其研究結果也證實預測方法有不錯效果。對於國內實際的市區公車的預測到站時間由於大部分公車班距比較短，預測的空間距離比較短，短時間內交通狀況變化不會太大，因此大部分並沒有採用預測的方法，而是以現狀的推估為主；但將其運用在公車班距長、預測的空間距離長的路線上，則會因為交通狀況改變而容易產生比較大的預測誤差。因此本研究嘗試引進在無號誌化路口之旅行時間預測上有良好表現的 k -NN 法，用於市區的公車到站時間預測上，檢視是否能夠把預測的準確度更加以提升。



第三章 公車到站時間預測模式

本研究的目的是在於如何利用即時的交通資訊進行旅行時間之預測，由於旅行時間會隨著路上交通情況的不同而有不同的現象，若車輛遭遇相似的交通狀況，則車輛行駛的時間亦相似；因此本研究將利用 k -NN 法找出相似交通資訊的歷史資料預測旅行時間。以下將針對 k -NN 模式進行介紹，並且說明模式所需要的參數。

3.1 k -NN 模式介紹

k 最近鄰法(k -Nearest Neighbor Method, k -NN Method)是一種利用比較歷史資料與即時資料，找出擁有相同特性的資料的方法。最初是由 Benedetti[28]、Stone[29] 及 Tukey[30]這些學者提出了最近鄰法(nearest neighbor method)的概念發展而來。這些學者利用一元位置估計(Univariate Location Estimators)，建立了無母數迴歸式的模型，進而延伸出最近鄰法的概念，利用輸入的資訊找尋出與資料庫中最相似的資料。

Altman[31]將以上的研究進行整理，將一元位置估計延伸至多元位置估計(Multivariate Location Estimators)並且提出了 k -NN 法，該方法將歷史資料區分群，再將輸入資料與分群後的歷史資料作比對，比對出該資料與那 k 組歷史資料較為類似，即可利用 k 組歷史資料預測未來資料。

Robinson 及 Polak[22]提出了建立 k -NN 模型時應該先設定的條件。

(1) 決定模式中的特徵向量。

k -NN 法中最重要的第一步即是要決定要比對的內容為何，如果選取的特徵向量太少，則無法有足夠的特徵去分辨歷史資料，可能會有不同的輸出值卻有相似的特徵向量；反之，若是選擇太多的特徵向量，則可能產生維度的問題，讓不相關的特徵向量影響距離量度，而選到不好的歷史資料。

(2) 設定距離量度(distance metric)。

距離量度為兩組資料相似程度之指標。常見得距離量度主要有下列三種：

◇ 歐幾里得距離(Euclidean distance)

假設兩個事件 x_i 、 x_j ，每個事件有 m 個屬性， x_{ip} 為 x_i 事件的第 p 個屬性， x_{jp} 為 x_j 事件的第 p 個屬性，則 x_i 事件與 x_j 事件間的歐幾里得距離 d_{ij} 如(4)式所示：

$$d_{ij} = \left[\sum_{p=1}^m (x_{ip} - x_{jp})^2 \right]^{1/2} \quad (4)$$

◇ 馬式距離(Mahalanois distance)

馬式距離 D^2 可用矩陣如(5)式所示：

$$D_{ij}^2 = (X_i - X_j)^T C_w^{-1} (X_i - X_j) \quad (5)$$

其中 X_i 為 x_i 事件的屬性矩陣， C_w 為聯合組內共變異數矩陣 (pooled within-group covariance matrix)。馬式距離為歐幾里得距離平方的一種延伸：

$$\begin{aligned} D_{ij}^2 &= (X_i - X_j)^T C_w^{-1} (X_i - X_j) \\ &= [x_{i1} - x_{j1} \quad x_{i2} - x_{j2} \quad x_{i3} - x_{j3} \quad \dots \quad x_{im} - x_{jm}] \begin{bmatrix} x_{i1} - x_{j1} \\ x_{i2} - x_{j2} \\ \vdots \\ x_{im} - x_{jm} \end{bmatrix} \\ &= \sum_{p=1}^m (x_{ip} - x_{jp})^2 \end{aligned} \quad (6)$$

意即在求 d_{ij}^2 時， $C_w^{-1} = C_w = I$ (單位矩陣)，換言之，當各軸的單位長度相等，且各軸互為垂直時，歐幾里得距離平方與馬式距離相同。

◇ 街道區距離(city block distance)

假設每個事件有 m 個屬性，則第 i 個事件與第 j 個事件間的街道區距離如(7)式所示：

$$d_{ij} = \sum_{p=1}^m |x_{ip} - x_{jp}| \quad (7)$$

然而距離量度的判斷目前並沒有一個很好的準則來對各種資料進行距離量度的判斷，在面對不一樣的環境條件，分析人員必須自行建立一套準則來做處理，因此隨著不同的特徵分群，將會採用不同的距離量度以判斷即時資料與那一個歷史資料類似。

(3) 決定判斷類似樣本的數量。

決定 k 值是決定預測誤差大小的關鍵步驟，若選取太大或太小的 k 值，容易產生較大的預測誤差。 k 值的決定通常為測試選取不同 k 值所得出的誤差結果，由此選出使誤差最小的 k 值，以此設定 k 值，以求出預測誤差可盡量減少。

由以上 k -NN 法的介紹可知， k -NN 法是利用輸入資料比對歷史資料，找出擁有相同特性的歷史資料，預測未來值；且若要使用 k -NN 演算法必須蒐集交通資訊，建立比對的歷史資料庫，設定相關參數包括比對特徵向量、距離量度計算。以下將先針對 k -NN 的演算機制建構適合本研究的公車到站預測演算法，接著根據上述文獻及模式建立過程，提出本研究所需要探討的研究課題，並提出研究設計的解決方法。

3.2 k -NN 之公車到站時間預測模式構建

本研究利用 k -NN 方法建立公車到站時間預測模式，希望可以利用該方法尋找與即時樣本相類似的歷史交通狀況。假設在相同的交通趨勢變化下會有相同的到站時間，利用即時的交通資訊比對歷史交通資訊的方法，能夠辨識出哪些歷史時間下擁有類似的交通資訊，並由該歷史時間所對應的到站時間預測目前的到站時間。

進行 k -NN 預測前需先建立歷史資料庫，以提供演算法的資料比對過程，圖 2 為第一階段建立資料庫的流程圖：

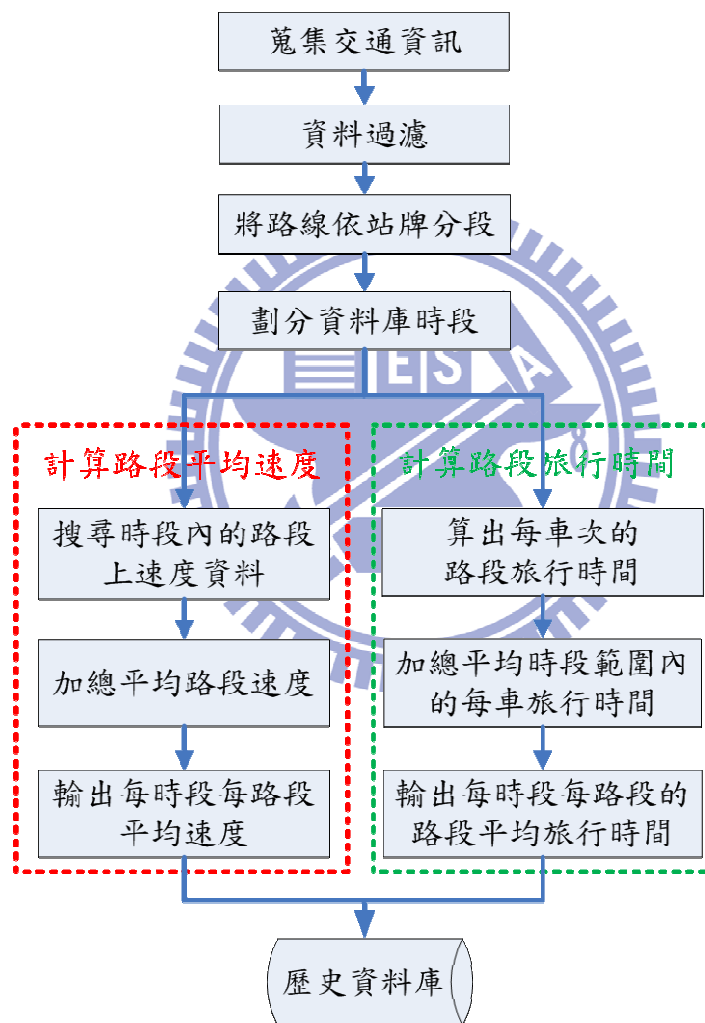


圖 2 第一階段建立資料庫流程圖

詳細步驟說明如下：

◇ 步驟 1：蒐集交通資料

取得市區公車的即時資料。

◇ 步驟 2：資料過濾

過濾不合理的資料，以避免異常的歷史資料造成預測的誤差。

◇ 步驟 3：將路線依站牌分段：

以站牌為切割點，將路線分為數個路段。

◇ 步驟 4：劃分資料庫時段

時段內的資料要能代表當時的交通狀況，以交通狀況改變幅度不大的時段為資料畫分的範圍。

◇ 步驟 5：計算路段平均速度

◎ 步驟 5.1：搜尋時段內的速度資料

依劃分的時段，依序搜尋非進離站範圍的路段速度資料。

◎ 步驟 5.2：加總平均速度資料

將搜尋出來的速度資料加總平均，得到路段平均速度。

◎ 步驟 5.3：輸出路段平均速度資料

輸出每時段每路段的路段平均速度資料。

◇ 步驟 6：計算路段平均速度

◎ 步驟 6.1：計算每車次的路段旅行時間

將每車次依進離站資訊計算路段旅行時間，亦即相鄰兩站的進站時間差為路段的旅行時間。

◎ 步驟 6.2：加總平均時段內的每車旅行時間

依照劃分時段，將時段內的每車旅行時間加總平均，得到每時段的路段旅行時間。

◎ 步驟 6.3：輸出路段旅行時間資料

輸出每時段每路段的路段旅行時間。

◇ 步驟 7：儲存至歷史資料庫

計算出來的路段平均速度及路段旅行時間，依資料格式儲存至資料庫。

第二階段為依照上述 k -NN 的演算法構建市區公車的到站時間預測模式。當有一筆即時資料輸入時，從歷史資料庫中先搜尋出符合時間窗限制的歷史資料，再比對即時與歷史資料，得出各筆歷史資料與即時資料的距離量度，從各距離量度中尋找與即時資料最鄰近的 k 筆歷史資料，並將此 k 筆資料的路段旅行時間作加總平均，再依延遲加總的機制算出目標的總旅行時間，即為所得之預測結果。圖 3 為本研究構建的 k -NN 預測流程圖：

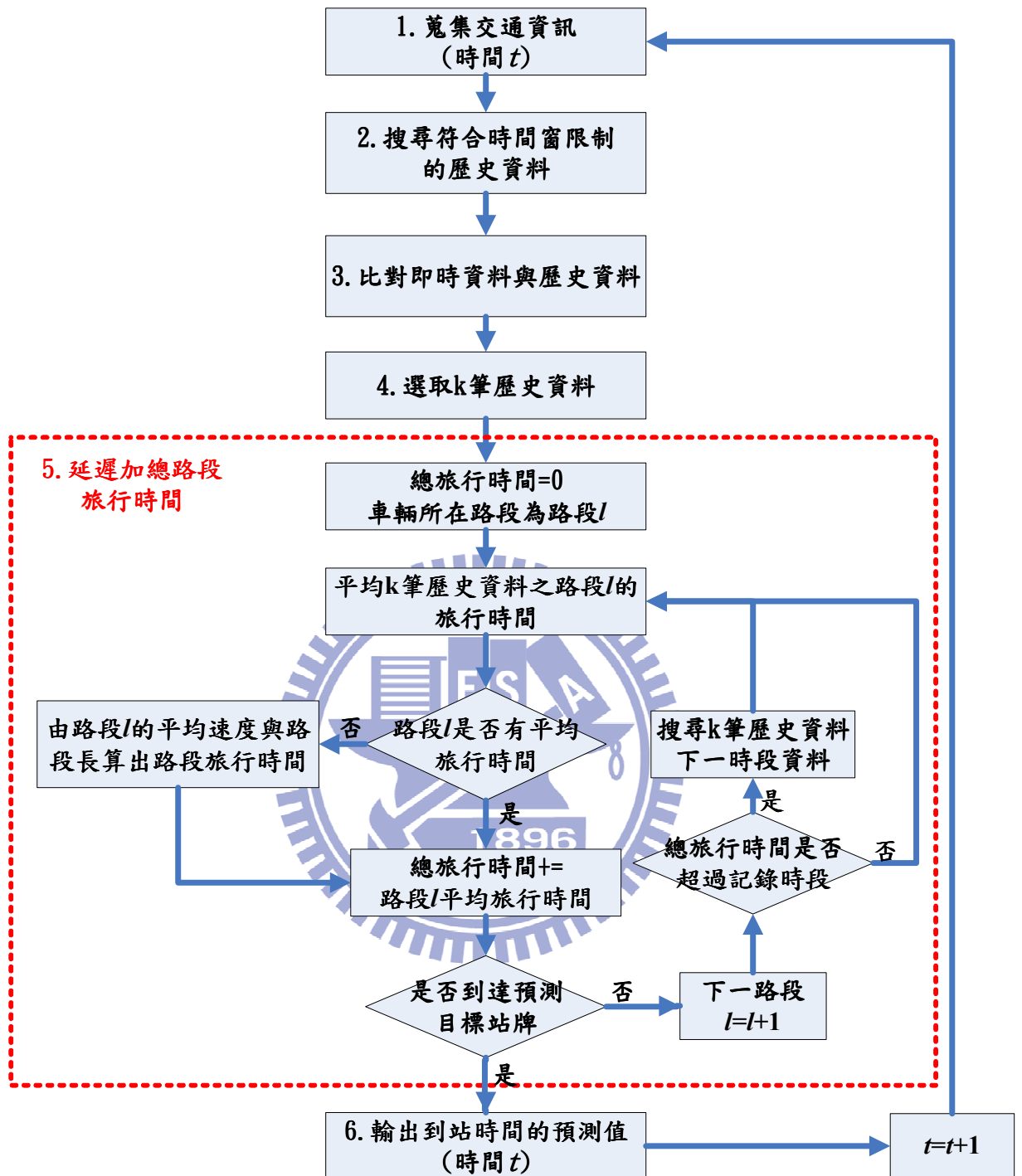


圖 3 第二階段 k-NN 預測流程圖

詳細步驟說明如下：

◇ 步驟 1：蒐集交通資料

取得市區公車的即時資料。假設某路線現況的路段旅行時間為 tt_t^l ， t 為預測時間， l 為路段。

◇ 步驟 2：搜尋符合時間窗限制的歷史資料

訂定時間窗範圍，搜尋預測時間前後時間段內的資料。利用 k-NN 法的歷史資料庫通常都很龐大，若是每次計算過程都要比對全部的歷史資料，則需要花費許多的運算時間；因為每日同時段的交通狀況相似，故本研究訂定搜尋時間窗範圍，只搜尋預測時間前後時間段內的資料。

假設歷史旅行時間為 $tt_{d_m t_n}^l$ ， d_m 為日期， t_n 為時間， l 為路段。搜尋 $|t-t_n| < tw$ 的歷史資料， tw 為時間窗範圍。

◇ 步驟 3：比對即時資料與歷史資料

利用距離量度比對即時與歷史資料，以得到各筆歷史資料與即時資料的差異程度。其距離公式為

$$dist_{d_m t_n} = \sqrt{\sum_{i=1}^l (tt_t^i - tt_{d_m t_n}^i)^2} \quad (8)$$

◇ 步驟 4：選取 k 筆歷史資料

依照距離量度的大小排序，選取 k 筆最小距離的歷史資料。假設 $M = \{h_1, h_2, h_3, \dots, h_k\}$ 為 k 筆距離最小的歷史資料集合， h_1 為距離最小的歷史資料。

◇ 步驟 5：延遲加總路段旅行時間

因為到站時間預測為預測未來的到站時間，車輛在行進的時候時間隨之改變，所儲存的路段旅行時間亦因為時段不同，若是從頭到尾只用同一時段的資料，則無法準確反映車輛所遭遇的交通狀況。因此本研究採用延遲加總方法，從車輛所在位置，依序加總路段旅行時間，檢察是否有超過資料庫的記錄時段，調整所參考的歷史路段旅行時間。

- ◎ 步驟 5.1：在 t 時間預測車輛至目標站牌的總旅行時間， T_l 為 0，路段 l 為車輛目前的所在路段。
- ◎ 步驟 5.2：將 k 筆歷史資料中的路段 l 的旅行時間加總平均，得路段 l 之歷史旅行時間預測值， tt_l 。

$$tt_l = \frac{1}{k} \sum_{i \in M} tt_{h_i}^l \quad (9)$$

- ◎ 步驟 5.3：若由於 k 筆歷史資料之路段 l 的旅行時間皆為空值，而無法得到路段 l 的平旅行時間，則利用歷史資料的路段平均速度與路段長度算出路段 l 的歷史旅行時間。
- ◎ 步驟 5.4：累加總旅行時間，則 $T_l += tt_l$ 。
- ◎ 步驟 5.5：若預測路段已到達目標站牌，則停止延遲加總計算；反之，則繼續下一路段的計算。
- ◎ 步驟 5.6：檢查總旅行時間是否超過歷史資料庫的記錄時段，若是超過時

段長度，則搜尋 k 筆歷史資料所對應的下一個時段資料作為下個路段的旅行時間。

◇ 步驟 6：輸出到站時間預測值

到站時間的預測值預測時間點(t)加上總旅行時間，將預測的旅行時間輸出，並將此旅行時間預測值存入歷史資料庫中，提供給下一個時段的旅行時間預測模式作使用。

3.3 範例說明

以下舉例說明模式演算流程：

假設下圖 4 為上午 08:30 某路線上之公車行進狀況示意圖，假設路線上共有 14 個站牌分為 13 個路段，各路段長度如表 1 所示，即時資料回傳了路線上共有兩輛公車，黑色虛線為各輛公車在 5 分鐘內行駛的軌跡，分別為在路段 2 上的 1 車(過去 5 分鐘完整經過路段 1，在路段 2 的最新位置上，與上游站距離 150 公尺，與下游站距離 210 公尺)、在路段 13 上的 2 車(過去 5 分鐘完整經過路段 8 至路段 12，最新位置在路段 13 上)。現在要輸出 1 車到預測目標站牌的到站時間。

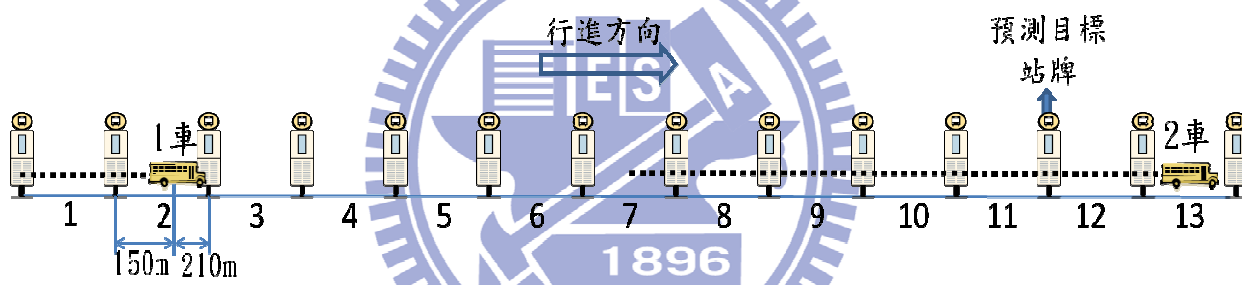


表 1 範例路段長度

路段距離長度(公尺)												
1	2	3	4	5	6	7	8	9	10	11	12	13
345	360	410	350	280	410	340	390	320	540	300	295	490

◇ 範例假設：

(1) 歷史資料格式如下表 2 及表 3 所示，歷史資料庫中有 3 天各路段的平均旅行時間及平均速度，5 分鐘記錄一筆，以下為 8:00 到 9:00 的歷史資料。

表 2 範例歷史資料-路段旅行時間

日期	時間	路段旅行時間(秒)												
		1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:00	56	44	23	29	23	58	59	56	41	51	55	34	38
Day1	08:05	48	40	53	53	28	20	51	59	46	53	36	25	41
Day1	08:10	60	33	44	60	38	49	18	48	54	20	32	17	35
Day1	08:15	30	26	24		36	25	17	31	41	19	52	22	21

Day1	08:20	46	48	15	16	56	49	48	30	26	40	24	58	40
Day1	08:25	53	39	21	35	35	51	42	37	58	25	55	48	25
Day1	08:30	32	22	54	29	28	44	52	26	22	57	18	17	48
Day1	08:35	44	25	21		54	49	35	46	33	37	27	35	35
Day1	08:40	23	37	52	31	45	22	21	39	39	35	36	21	28
Day1	08:45	27	43	46	22	57	30	42	25	51	38	33	43	52
Day1	08:50	32	57	35	32	60	25	24	18	24	28	18	25	15
Day1	08:55	26	23	28	36	48	59	44	22	48	20	26	42	57
Day2	08:00	50	44	37	54	30	15	59	37	58	17	16	58	37
Day2	08:05	50	26	54	52	53	34	33	48	58	26	25	33	32
Day2	08:10	23	58	26	37	56	43	59	46	18	29	20	54	41
Day2	08:15	51	57	40		45	56	34	26	25	49	38	49	26
Day2	08:20	27	39	45	56	26	53	36	46	15	33	18	44	47
Day2	08:25	19	21	54	53	55	35	22	41	21	43	23	47	44
Day2	08:30	20	47	56	33	42	57	15	44	23	17	44	55	42
Day2	08:35	16	27	40	37	25	17	57	51	38	48	41	16	18
Day2	08:40	33	43	21	25	54	42	51	49	57	39	17	26	35
Day2	08:45	36	53	29		20	39	52	43	38	41	47	44	38
Day2	08:50	48	24	19	60	17	55	36	45	30	22	51	56	27
Day2	08:55	33	38	17	58	38	26	40	53	21	34	42	19	53
Day3	08:00	41	50	51	57	21	47	36	35	48	59	30	24	43
Day3	08:05	31	60	44	18	50	20	35	53	53	35	30	28	15
Day3	08:10	16	32	26	33	53	44	40	42	40	47	45	23	50
Day3	08:15	38	37	27	26	24	23	31	15	60	22	50	21	17
Day3	08:20	23	38	24	34	40	16	46	26	28	29	28	34	36
Day3	08:25	17	53	49	46	37	20	43	47	17	19	38	54	24
Day3	08:30	23	60	20	38	26	45	37	31	40	59	35	45	53
Day3	08:35	26	40	37	26	22	25	20	57	44	16	39	36	37
Day3	08:40	22	40	55	31	27	46	55	32	55	21	50	55	21
Day3	08:45	22	17	26		49	57	31	49	16	22	60	24	27
Day3	08:50	41	54	37	56	46	29	28	26	37	42	54	39	22
Day3	08:55	43	44	36	59	24	47	32	28	18	38	23	33	45

表 3 範例歷史資料-路段平均速度

日期	時間	路段平均速度(公里/小時)												
		1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:00	28	22	29	6	18	17	8	47	6	18	20	16	49
Day1	08:05	20	21	28	24	33	47	25	46	39	15	41	43	11
Day1	08:10	13	16	19	47	27	28	50	38	27	7	20	47	6
Day1	08:15	49	10	9	27	6	31	11	37	38	42	34	23	27

Day1	08:20	38	34	8	24	43	35	38	20	30	14	11	19	34
Day1	08:25	28	11	45	36	9	44	16	28	30	20	40	19	13
Day1	08:30	46	20	5	44	26	27	38	43	22	34	49	35	26
Day1	08:35	47	49	22	50	49	19	15	14	8	18	45	9	35
Day1	08:40	45	19	19	29	18	43	35	10	14	9	14	13	32
Day1	08:45	45	26	14	15	42	14	22	43	5	37	41	40	20
Day1	08:50	50	38	26	34	23	50	49	27	23	6	49	18	21
Day1	08:55	48	43	35	10	37	21	10	22	7	29	36	45	17
Day2	08:00	22	14	35	42	43	21	33	23	8	10	48	42	12
Day2	08:05	8	40	27	36	34	8	7	36	26	33	21	47	23
Day2	08:10	14	17	49	5	12	5	39	41	14	32	31	38	13
Day2	08:15	48	22	14	50	18	29	46	24	15	31	33	6	41
Day2	08:20	8	35	40	21	30	36	50	34	29	41	9	45	13
Day2	08:25	23	47	50	36	38	9	35	38	11	35	22	13	47
Day2	08:30	50	23	34	31	35	25	47	14	49	48	32	8	40
Day2	08:35	8	14	20	44	25	15	30	9	42	25	32	14	41
Day2	08:40	11	19	16	34	13	24	30	27	25	26	17	11	22
Day2	08:45	33	21	46	6	8	50	13	44	22	50	19	16	27
Day2	08:50	44	46	14	46	40	10	14	18	38	49	45	7	24
Day2	08:55	5	45	24	28	37	21	40	36	17	16	12	16	33
Day3	08:00	50	21	19	11	30	7	8	5	30	15	42	8	18
Day3	08:05	30	43	27	19	48	20	8	47	48	18	18	43	12
Day3	08:10	49	16	44	38	12	34	42	16	31	18	5	12	12
Day3	08:15	19	5	31	19	49	9	31	9	15	25	23	32	33
Day3	08:20	28	41	28	41	48	5	29	12	44	46	42	29	11
Day3	08:25	38	38	11	28	28	32	24	11	49	48	39	38	23
Day3	08:30	18	33	15	50	8	30	40	6	43	43	37	17	49
Day3	08:35	45	28	41	34	8	14	7	24	23	43	50	9	10
Day3	08:40	45	42	31	50	33	20	13	26	11	38	36	50	36
Day3	08:45	35	18	11	46	35	35	36	15	25	5	46	39	11
Day3	08:50	22	15	21	14	11	40	15	27	40	23	8	19	44
Day3	08:55	35	7	8	49	19	49	9	40	13	19	25	44	28

- (2) 搜尋資料庫的時間窗範圍設為前後 15 分鐘。
- (3) 比對特徵向量為路段的旅行時間。
- (4) 距離量度使用歐基里德距離。
- (5) k 值設為 5。
- (6) 不完整路段由車輛與上下游站牌的距離比內插出至下游站牌的旅行時間，

即(加權平均後的路段旅行時間)×(車輛與下游站的距離佔整個路段長的比
例)。

◇ 演算流程：

- 步驟 1：將蒐集到的即時資料依照進離站時間，計算出 3 輛車在過去 5 分鐘內所經過路段的路段旅行時間，結果如下表 4 所示：

表 4 範例現狀旅行時間

		路段旅行時間(秒)												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
現況	08:30	54							33	20	35	50	25	

- 步驟 2：搜尋歷史資料庫中 08:30 前後 15 分鐘資料，共 21 筆資料，如表 2 的灰底資料，找出路段 1、路段 8 至路段 12 的路段旅行時間。
- 步驟 3：與現況的路段旅行時間資料作比對，計算其距離量度，如下表 5 所示：

表 5 範例距離計算

日期	時間	路段旅行時間(秒)													距離
		1	2	3	4	5	6	7	8	9	10	11	12	13	
Day1	08:15	30	26	24		36	25	17	31	41	19	52	22	21	35.9166
Day1	08:20	46	48	15	16	56	49	48	30	26	40	24	58	40	43.5775
Day1	08:25	53	39	21	35	35	51	42	37	58	25	55	48	25	45.9891
Day1	08:30	32	22	54	29	28	44	52	26	22	57	18	17	48	45.9239
Day1	08:35	44	25	21		54	49	35	46	33	37	27	35	35	32.7261
Day1	08:40	23	37	52	31	45	22	21	39	39	35	36	21	28	39.6232
Day1	08:45	27	43	46	22	57	30	42	25	51	38	33	43	52	48.7442
Day2	08:15	51	57	40		45	56	34	26	25	49	38	49	26	31.6070
Day2	08:20	27	39	45	56	26	53	36	46	15	33	18	44	47	48.0833
Day2	08:25	19	21	54	53	55	35	22	41	21	43	23	47	44	50.6656
Day2	08:30	20	47	56	33	42	57	15	44	23	17	44	55	42	50.4579
Day2	08:35	16	27	40	37	25	17	57	51	38	48	41	16	18	49.2240
Day2	08:40	33	43	21	25	54	42	51	49	57	39	17	26	35	56.3205
Day2	08:45	36	53	29		20	39	52	43	38	41	47	44	38	33.9706
Day3	08:15	38	37	27	26	24	23	31	15	60	22	50	21	17	48.6313
Day3	08:20	23	38	24	34	40	16	46	26	28	29	28	34	36	40.9268
Day3	08:25	17	53	49	46	37	20	43	47	17	19	38	54	24	53.0566
Day3	08:30	23	60	20	38	26	45	37	31	40	59	35	45	53	50.6557
Day3	08:35	26	40	37	26	22	25	20	57	44	16	39	36	37	50.3885
Day3	08:40	22	40	55	31	27	46	55	32	55	21	50	55	21	57.8446
Day3	08:45	22	17	26		49	57	31	49	16	22	60	24	27	39.5727

例：即時資料與 Day1 08:15 的距離量度=

$$\sqrt{(30-54)^2 + (31-33)^2 + (41-20)^2 + (19-35)^2 + (52-50)^2 + (22-25)^2} = 35.9166$$

- 步驟 4：選出距離量度中最小的前五筆資料的歷史資料，分別為 Day1 08:15、Day1 08:35、Day2 08:15、Day2 08:45、Day3 08:45，如表 5 灰底資料。
- 步驟 5：延遲加總路段旅行時間

以 1 車到達目標站牌的到站時間為對象進行預測：

- ◎ 步驟 5.1：設 1 車至目標站牌的總旅行時間為 0，路段 2 為車輛目前的所在路段。
- ◎ 步驟 5.2：將五筆歷史資料中的路段 2 的旅行時間加總平均，得路段 2 之歷史旅行時間為 35.6 秒，如表 6 所示。對 1 車而言，路段 2 為不完全路段，因此利用 1 車與上下游站牌距離比內插出剩餘距離的旅行時間 = $35.6 \times [210 / (150 + 210)] = 20.8$ 秒。

表 6 範例五筆歷史路段旅行時間計算(1)

		路段旅行時間												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:15	30	26	24		36	25	17	31	41	19	52	22	21
Day1	08:35	44	25	21		54	49	35	46	33	37	27	35	35
Day2	08:15	51	57	40		45	56	34	26	25	49	38	49	26
Day2	08:45	36	53	29		20	39	52	43	38	41	47	44	38
Day3	08:45	22	17	26		49	57	31	49	16	22	60	24	27
平均旅行時間			35.6											
總旅行時間			20.8											

- ◎ 步驟 5.3：路段 2 可以五筆歷史資料求出路段旅行時間，不進行此步驟。
- ◎ 步驟 5.4：累加總旅行時間，則總旅行時間 = $0 + 20.8 = 20.8$ 秒。
- ◎ 步驟 5.5：尚未到達目標站牌，繼續下一路段的計算。
- ◎ 步驟 5.6：20.8 秒小於資料庫的記錄時段 5 分鐘，因此使用相同的五資料計算下一路段的旅行時間。

路段 3 重覆步驟 5.2 至 5.6 的計算結果如下表 7 所示：

表 7 範例五筆歷史路段旅行時間計算(2)

		路段旅行時間(秒)												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:15	30	26	24		36	25	17	31	41	19	52	22	21
Day1	08:35	44	25	21		54	49	35	46	33	37	27	35	35
Day2	08:15	51	57	40		45	56	34	26	25	49	38	49	26
Day2	08:45	36	53	29		20	39	52	43	38	41	47	44	38
Day3	08:45	22	17	26		49	57	31	49	16	22	60	24	27
平均旅行時間			35.6	28										
總旅行時間			20.8	48.8										

對路段 4 進行計算：

- ◎ 步驟 5.2：路段 4 的歷史資料皆為空值，無法求出路段平均旅行時間。
- ◎ 步驟 5.3：路段 4 無法由五筆歷史資料求出路段旅行時間，因此搜尋對應的路段速度資料，並計算路段 4 之平均速度為 35.8 公里/小時，如表 8 所示。

表 8 範例五筆歷史路段旅行時間計算(3)

		路段平均速度(公里/小時)												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:15	49	10	9	27	6	31	11	37	38	42	34	23	27
Day1	08:35	47	49	22	50	49	19	15	14	8	18	45	9	35
Day2	08:15	48	22	14	50	18	29	46	24	15	31	33	6	41
Day2	08:45	33	21	46	6	8	50	13	44	22	50	19	16	27
Day3	08:45	35	18	11	46	35	35	36	15	25	5	46	39	11
路段平均速度					35.8									

◆ 路段 4 的路段旅行時間=350(公尺)/35.8(公里/小時)=35.2 秒

表 9 範例五筆歷史路段旅行時間計算(4)

		路段旅行時間(秒)												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:15	30	26	24		36	25	17	31	41	19	52	22	21
Day1	08:35	44	25	21		54	49	35	46	33	37	27	35	35
Day2	08:15	51	57	40		45	56	34	26	25	49	38	49	26
Day2	08:45	36	53	29		20	39	52	43	38	41	47	44	38
Day3	08:45	22	17	26		49	57	31	49	16	22	60	24	27
平均旅行時間			35.6	28	35.2									
總旅行時間			20.8	48.8	84									

◎ 步驟 5.4：累加總旅行時間，則總旅行時間=48.8+35.2=84 秒。

◎ 步驟 5.5：尚未到達目標站牌，繼續下一路段的計算。

路段 5 到路段 10 重覆步驟 5.2 至 5.6 的計算結果如下表 10 所示：

表 10 範例五筆歷史路段旅行時間計算(5)

		路段旅行時間(秒)												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:15	30	26	24		36	25	17	31	41	19	52	22	21
Day1	08:35	44	25	21		54	49	35	46	33	37	27	35	35
Day2	08:15	51	57	40		45	56	34	26	25	49	38	49	26
Day2	08:45	36	53	29		20	39	52	43	38	41	47	44	38
Day3	08:45	22	17	26		49	57	31	49	16	22	60	24	27
平均旅行時間			35.6	28	35.2	40.8	45.2	33.8	39	30.6				
總旅行時間			20.8	48.8	84	124.8	170	203.8	242.8	273.4				

對路段 10 進行計算：

- ◎ 步驟 5.2：將五筆歷史資料中的路段 10 的旅行時間加總平均，得路段 10 之歷史旅行時間為 33.6 秒，如表 11 所示

表 11 範例五筆歷史路段旅行時間計算(6)

		路段旅行時間(秒)												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:15	30	26	24		36	25	17	31	41	19	52	22	21
Day1	08:35	44	25	21		54	49	35	46	33	37	27	35	35
Day2	08:15	51	57	40		45	56	34	26	25	49	38	49	26
Day2	08:45	36	53	29		20	39	52	43	38	41	47	44	38
Day3	08:45	22	17	26		49	57	31	49	16	22	60	24	27
平均旅行時間			35.6	28	35.2	40.8	45.2	33.8	39	30.6	33.6			
總旅行時間			20.8	48.8	84	124.8	170	203.8	242.8	273.4	307			

- ◎ 步驟 5.3：路段 10 可以五筆歷史資料求出路段旅行時間，不進行此步驟。
- ◎ 步驟 5.4：累加總旅行時間，則總旅行時間=273.4+33.6=307 秒。
- ◎ 步驟 5.5：尚未到達目標站牌，繼續下一路段的計算。
- ◎ 步驟 5.6：307 秒超過資料庫的記錄時段 5 分鐘，因此搜尋五筆歷史資料的下一時段資料進行下一個路段的平均旅行時間計算，下表 12 為下一時段的歷史資料。

表 12 範例五筆歷史路段旅行時間計算(7)

		路段旅行時間(秒)												
日期	時間	1	2	3	4	5	6	7	8	9	10	11	12	13
Day1	08:20	46	48	15	16	56	49	48	30	26	40	24	58	40
Day1	08:40	23	37	52	31	45	22	21	39	39	35	36	21	28
Day2	08:20	27	39	45	56	26	53	36	46	15	33	18	44	47
Day2	08:50	48	24	19	60	17	55	36	45	30	22	51	56	27
Day3	08:50	41	54	37	56	46	29	28	26	37	42	54	39	22
平均旅行時間			35.6	28	35.2	40.8	45.2	33.8	39	30.6	33.6	36.6		
總旅行時間			20.8	48.8	84	124.8	170	203.8	242.8	273.4	307	343.6		

對路段 11 進行計算：

- ◎ 步驟 5.2：將五筆歷史資料中的路段 11 的旅行時間加總平均，得路段 11 之歷史旅行時間為 36.6 秒
- ◎ 步驟 5.3：路段 11 可以五筆歷史資料求出路段旅行時間，不進行此步驟。
- ◎ 步驟 5.4：累加總旅行時間，則總旅行時間=307+36.6=343.6 秒。
- ◎ 步驟 5.5：已預測到目標站牌，停止延遲加總計算。

步驟 6：輸出預測的到站時間

1 車預測的到站時間=08:30:00+343.6 秒=08:35:43

3.4 研究課題

由前幾節的 k -NN 流程介紹能夠整理出以下幾項研究課題，將在之後的章節逐一針對各項課題提出本研究設計的處理方法：

(1) 資料過濾

收集回來的公車 GPS 資料不完全都是合理的資料點，例如路段旅行時間過長、進站停等的速度點資料、號誌停等的速度點資料等，若是採用所有的資料點，則會影響計算的平均速度及平均旅行時間，故需要將原始資料作適當的過濾，其後建立之資料庫及計算所得的數值才能夠具有有效性。

(2) 建立歷史資料庫

k -NN 法是利用比對歷史資料找出相似資料特性的方法，故建立合適的資料是很重要的前置作業。本研究的預測對象為公車的到站時間，因為公車的到站時間與道路的交通狀況與站牌停等情況有很大的關係，公車所遭遇的交通狀況會因為路段的不同而有所不同，亦會因為尖離峰的時間因素有不同的交通狀況，因此建立預測公車到站時間的歷史資料庫，需要探討路段與時段的分法，以適當表現當地當時的交通狀況。

● 分路段

本研究將利用市區公車能夠得到的資訊當作 k -NN 比對的特徵向量，分的好壞會影響到比對的品質。若是切割的路段數少，則無法精確描述整條路線的交通狀況；反之，若是將路線切割成太小的單位，則可能造成許多路段資料發生空值的情況，因此，需要針對取得交通資料的特性決定路段的分法。

● 分時段

若是粗略切割成很大的時段，則無法保證此時段內的交通狀況不變；若是分小時段，則比對的歷史資料量會很大。因此時段的分法亦是需要考慮的設計要件。

(3) 參數校估

- 距離量度

k -NN 的目標是在找尋最接近即時資訊的歷史資訊，需要依照比對資料的特性選取合適的距離量度，若是距離量度無法表現資料的相似程度，則可能會因為取得差異較大的歷史資料作預測，而造成較大的偏誤。

- 時間窗

進行資料比對的歷史資料筆數將影響演算法的運算時間。若是搜尋整個歷史資料庫會使得演算法在比對的過程中花費較多時間，因此必須定出搜尋的時間窗範圍，以減少搜尋時間。

- k 值

k 值的設定將決定輸出的預測值，若是選取較大的 k 值，則會過度平滑歷史資料；若是選取較小的 k 值，選取的值中有比較大的偏誤，則會造成預測值有較大的誤差。

(4) 不完全路段的旅行時間計算

在市區公車的行走時間包括停等時間，若是較長的路段，則估算路段剩餘距離的行駛時間的誤差可能會超出可忍受的誤差範圍。因此，需要依照資料特性設計適當的剩餘旅行時間估算方法。

以上為本研究建立 k -NN 公車到站時間預測模式時需要解決的課題，資料過濾與資料庫建立的前置作業處理，及關於影響模式準確度的參數估計、不完全路段的旅行時間計算及加權平均方法將於第四章資料處理做詳細介紹；最後將應用實際資料作測試，並針對結果作改善與調整。

第四章 資料處理

本研究以 k -NN 法作為資料分析的演算法，由上一章節對 k -NN 演算流程的介紹能夠得知資料過濾、建立資料庫及設定參數皆為演算法中重要的設計要件，以下將針對各項設計要件加以說明。

4.1 資料過濾

在建立歷史資料庫或引用即時資料的時候，若是直接將資料依照設計的流程建立歷史資料庫或當作演算法的輸入資料，則可能會因為不適當的資料點造成演算法的誤差，此時必須先進行資料過濾的步驟，才能進行資料庫建立及演算法的運算流程，以下將列出需要過濾掉的資料情形：

- (1) 受站牌上下客影響：當車輛經過站牌時，會因為上下車的乘客而使速度為零，若是因為上一輛公車的停等行為讓車輛速度明顯降低，故在進離站範圍內的速度點資料將予剔除。
- (2) 受紅燈停等影響：當車輛接近路口時，若是遇到紅燈使得速度為零，則會影響路段平均速度，因此在進離站範圍外且接近路口區域，連續數個速度點為零的資料，在計算路段平均速度將予剔除。
- (3) 受急踩煞車影響：當車輛位置在進離站範圍外，且非接近路口區域時，若有急踩煞車的情形產生，則會使得車速突然降為零，故將在進離站範圍外且非路口附近，突然為零的速度點資料剔除。

4.2 建立資料庫

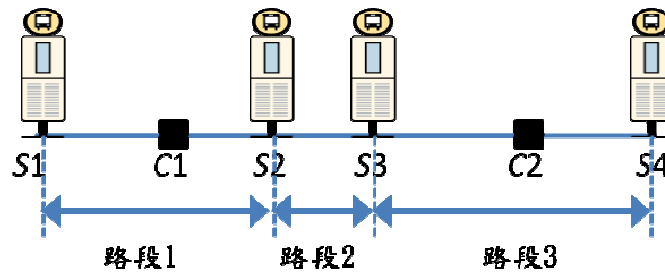
本研究以 k -NN 模式進行公車到站時間預測，以路線上的各路段旅行時間代表路線的交通狀態，將現況的路段旅行時間與歷史的路段旅行時間進行比對，找出與現狀路段旅行時間分配最相似的歷史資料，亦即與現狀有相似的交通狀況。因此建立適當的資料庫為演算法重要的前置作業，研究設計的路段資料庫欄位內容如表 13 所示，主要儲存處理過的路段旅行時間、速度：

表 13 路段資料欄位說明

欄位內容	欄位說明
路線	路段所屬的公車路線編號
去回程	為去程或回程的資料(0為去程，1為回程)
上游站	路段的上游站牌編號
下游站	路段的下游站牌編號
路段旅行時間	路段的平均旅行時間(秒)
路段速度	路段的平均速度(公里/小時)
時段	資料所屬的時段(包括日期及時間)

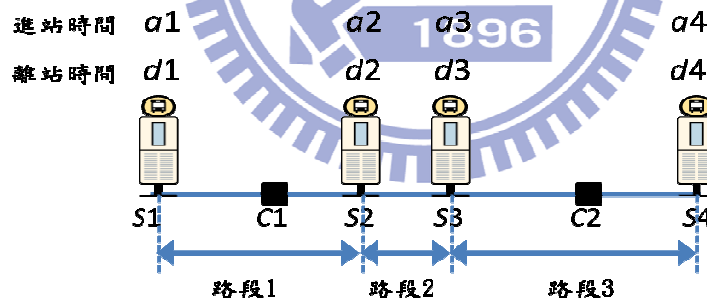
4.2.1 路線分段

本研究將各個公車路線以站牌為切點分為數個路段，下圖為路線分段的示意圖，假設路線上共有 4 個公車站牌分別為 $S1$ 、 $S2$ 、 $S3$ 及 $S4$ ，與 2 個路口分別為 $C1$ 及 $C2$ ，以站牌為切點則可分出路段 1、路段 2 及路段 3：



4.2.2 計算路段旅行時間

確定每個路段的起迄點後，接著計算每個路段的旅行時間。以每車的進離站資訊計算路段旅行時間，將站內的等待乘客上下車之停等時間併入下一路段的路段旅行時間，意謂路段兩端站牌的進站時間差即為路段旅行時間。如圖 6 所示，公車依行進方向前進，於站牌 $S1$ 的進站時間為 $a1$ 、離站時間為 $d1$ ，站牌 $S2$ 的進站時間為 $a2$ 、離站時間為 $d2$ ，則路段 1 的路段旅行時間為 $a2-a1$ 。



4.2.3 劃分時段

在劃分歷史資料時段時，因各個路段之車流尖峰與離峰時段不盡相同，因為市區交通狀況在 5 分鐘內擁有比較相似的情況，故本研究將劃分時段設為 5 分鐘，加總平均 5 分鐘內的每車次路段旅行時間。

4.3 設定參數

4.3.1 距離量度

由 3.1 節 k -NN 模式介紹中，可知常用的距離量度是將現況資料與歷史資料

中的每個路段旅行時間作比較，算出每路段的距離後，再把每一段的距離依定義的公式加總計算出每一筆歷史資料與現況資料的距離量度；其中隱含每個比對的路段都必須要有值，否則會因為資料缺漏而沒有加總到總距離中，使得總距離量度值變小，在演算的過程被視為與現狀有比較相似的交通狀況。

由於使用公車的路段旅行時間作比對，而每個時段中並非每個路段都會有公車經過，因此所擷取出的即時與歷史資訊產生缺漏的比例高，以表 14 為例，歷史資料 1 中有值的為路段 1、路段 2 及路段 3，而歷史資料 2 中有值的為路段 1、路段 3、路段 4、路段 5 及路段 6。

表 14 路段旅行時間資料

		路段 1	路段 2	路段 3	路段 4	路段 5	路段 6
旅行 時間	現況	150	171	170	22	72	29
	歷史資料 1	101	115	117			
	歷史資料 2	119		118	69	100	71

距離量度以歐幾里得距離作計算，可得出歷史資料 1 與狀況的距離為 91.356，歷史資料 2 與現狀的距離為 91.77。由結果可知歷史資料 1 之距離小於歷史資料 2 之距離，代表歷史資料 1 的交通狀況與現狀比較相似。

但觀察每路段的旅行時間差值如表 15 所示，歷史資料 2 的每路段差值都小於歷史資料 1 中路段 2 的差值 72，表示歷史資料 2 筆歷史資料 1 更接近現況；卻因為能夠計算的資訊較多而使得總距離變大，此與篩選歷史資料的機制相異。

表 15 歷史與現狀之路段旅行時間比較

		路段 1	路段 2	路段 3	路段 4	路段 5	路段 6
與現況之絕對差	歷史資料 1	49	56	53			
	歷史資料 2	31		52	47	28	42

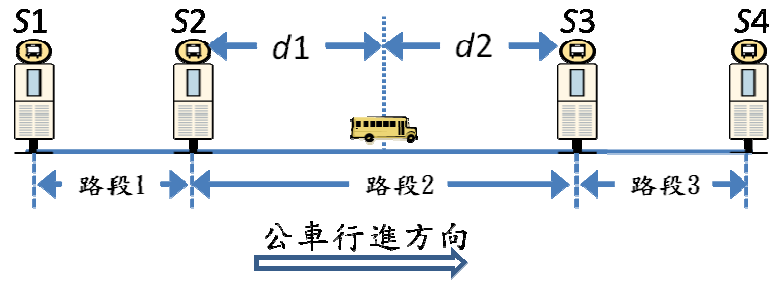
因此本研究為了因應市區公車之路段資料缺漏的特性，將距離量度的計算公式調整為「每路段的均方根距離」，如(10)式所示， m 為可比較的路段個數：

$$d_{ij} = \left[\sum_{p=1}^m (x_{ip} - x_{jp})^2 / m \right]^{1/2} \quad (10)$$

將上例依設計之距離公式重新計算，可得出歷史資料 1 與狀況的距離為 52.744，歷史資料 2 與現狀的距離為 41.041。由結果可知歷史資料 2 之距離小於歷史資料 1 之距離，表示歷史資料 2 比歷史資料 1 更為接近現況，此與 k-NN 的篩選機制相符。

4.3.2 不完全路段的旅行時間計算

因為市區公車的現況位置並不一定位於路段的端點上，可能會產生不完全路段，如圖 7 所示，如果預預測公車到站牌 S4 的旅行時間，則對公車而言，路段 3 為完整路段，可由歷史資料直接獲得路段 3 的旅行時間；相對而言，路段 2 為不完整路段，需估計公車現狀位置至下游站牌 S3 的旅行時間。



本研究利用公車的現況位置與前後站牌的相對位置估計剩餘路段的旅行時間。如所示，假設公車的現況位置與上下游站牌距離 $d1$ 與 $d2$ ，由歷史資料得出路段 2 的旅行時間為 t ，則公車至下游站 S2 的旅行時間為 $t \times d2 / (d1 + d2)$ 。

4.3.3 k 值設定

本研究將透過實驗資料比較在不同的 k 值下，所得的到站時間的平均絕對誤差，選取適當的 k 值當作之後預測演算法的固定參數值。

4.3.4 時間窗

本研究將搜尋歷史資料的時間窗範圍設為公車最大的班距。因為公車依照時間表行駛時，會保持在前後兩班公車的行駛範圍內，因此時間窗選擇最大的班距，即可包括公車行駛速度快慢所可能出現的時段。



第五章 實證分析

根據前面第四章所建構 k -NN 市區公車到站時間預測模式，本章節將以台北市區公車作為測試對象，與現狀預測系統的預測績效進行評估。

5.1 現況說明

台北之市公車動態資訊系統共分為四期，分別由不同廠商建置，採用不同的公車到站時間預測方法，以下將詳細敘述各期廠商的公車到站時間預測演算法。

方法一：一、四期建置廠商

當通訊伺服器接收到車機定位訊號後，則計算車輛於該營運路線之到離站資訊，再傳送至演算中心進行預估到站時間計算。經由營運路線狀況分析可知，台北市公車發車間距短則 5 分鐘內，長則 1 小時，亦有固定班次之路線，故研擬兩種預估到站時間演算邏輯。

演算法以路線特性劃分，針對台北市班次密集及較長班距之路線分別設計演算邏輯，例如班次在 20 分鐘以內之路線採前 3 班之平均旅行時間做為預估到站時間之參考；20 分鐘以上及固定班次之路線採歷史資料（以路線、站牌、星期、時、行駛方向劃分）推估方式。預估到站時間演算邏輯分述如下：

(1) 前車資料推估演算法

當接收到即時定位資訊時（包含定時或是發送點所傳回的資訊），會主動通知預估到站模組，此時預估到站模組將即時計算每一傳回位置點所屬的路段，以提供預估到站用。前車旅行時間資料以所屬路段之到達及離開時間計算每個站牌間之旅行時間；因此，每一個路段都會有 0 至多筆的時間資訊。

每當有模組需要該路段的前車資訊時，則模式會取預測時間前 3 筆資訊加總平均，得到該路段的前車旅行時間並回傳。

(2) 歷史資料推估演算法

接收到的每一筆即時定位資訊，都會被儲存成歷史資料，以供計算歷史平均旅行時間用。當資料被存成歷史資料後，系統會根據每天累積的資訊，計算每個星期別中每個時段（以 1 小時為間距）的歷史平均速度表。並利用當天的資料來調整之前所計算出的歷史平均旅行時間表；調整的方式為將新的資料與已算出的平均旅行時間表做平均。

其預估到站演算法流程如下圖：

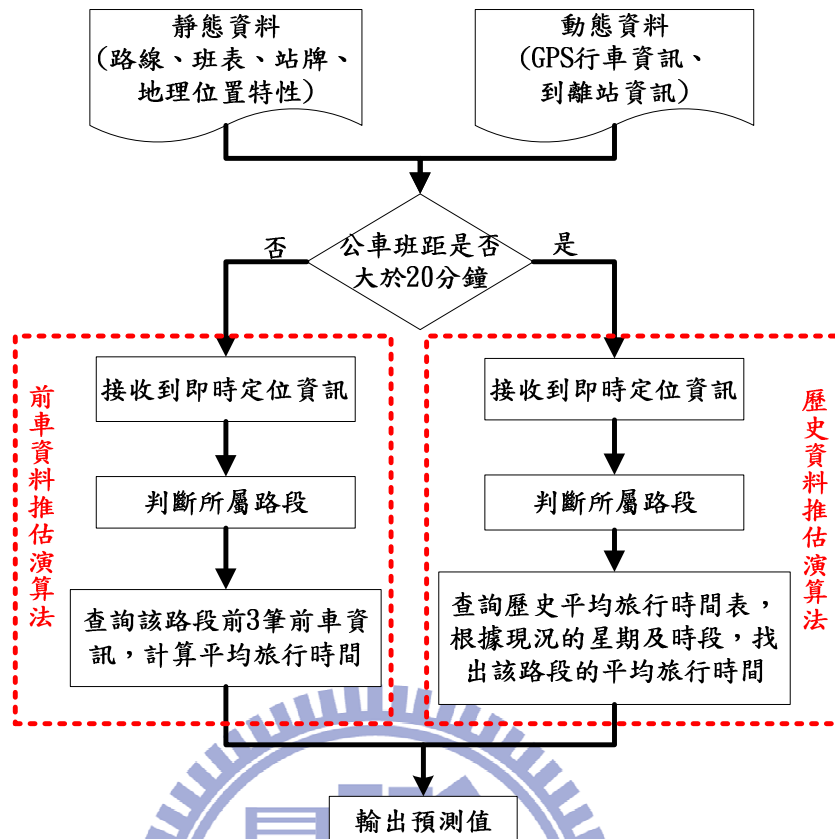


圖 8 現有系統一、四期廠商之到站時間演算流程

方法二：二、三期建置廠商

當系統接收到每筆車機定位資訊時(包含定時或是發送點所傳回的資訊)，會先自動先判斷屬於該路線的那一個路段，以提供預估到站用。每當有模組需要該路段的歷史旅行時間資訊時，則模組會取目前日期時間之所屬星期及時段之平均旅行時間，得到該路段的歷史旅行時間並回傳。

車機設備預先儲存站為經緯度位置及行駛路線，公車客運行駛時，固定週期回報定位資訊，當公車客運進站前，即回報到站資訊，系統根據這些資訊進行公車到站時間預估，在預估演算方面分為歷史行車時間演算法與即時路況兩部分。

在歷史行車時間演算法的方式，將每日行車資訊依站位、平日/假日、星期、時段及車輛平均行駛速度分類，建立歷史行車統計資料。當公車到達某一站牌時，系統依據與前一路段間的平均行駛速度、平日/假日、星期及時段查詢歷史行車統計資料。此外在站間距離較長的路線，將在適當距離加入虛擬站點增加演算的次數。

其預估到站演算法流程如下圖：

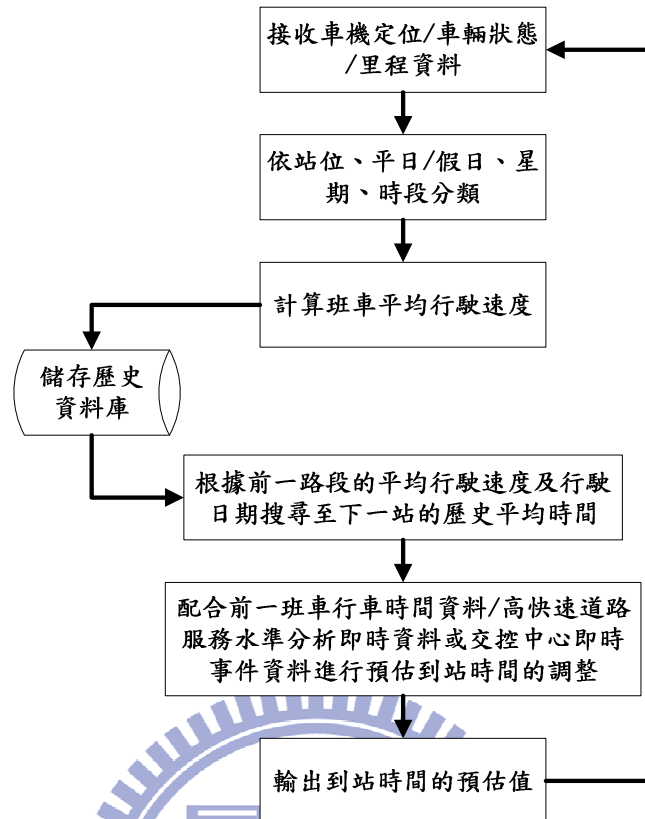


圖 9 現有系統二、三期廠商之到站時間演算流程

5.2 資料蒐集

5.2.1 交通資料

以台北市公車 GPS 資料當作演算法的輸入資料，由台北市公車動態資訊系統之第一、四期廠商取得相關資料，包括定時回報資料及定點回報資料，資料格式內容如表 16 及表 17 所述，定時回報資料之用途為固定時間回傳各車輛之行車狀況，定點回報為車輛進離站時回報各車輛之行車狀況；用 2009 年 12 月至 2010 年 1 月共兩個月的資料建立歷史資料庫，2010 年 2 月 6 日到 2010 年 2 月 12 日的資料當作績效評估之用。

表 16 定時回報資料

主要欄位內容	欄位說明
車輛代碼	公車 GPS 代碼
勤務狀態	公車執行勤務的狀況(包括正常、開始、結束)
行車狀況	公車的行車狀況(包括正常、車禍、故障、塞車等)
路線代碼	公車執行的路線代碼
去返程	為路線的去程或回程
車輛經度座標	車輛位置的經度座標

車輛緯度座標	車輛位置的緯度座標
車輛速度	車輛行駛之瞬間速度
車輛方位角	車輛的行駛方向
GPS 時間	資料點回傳的 GPS 時間

表 17 定點回報資料

主要欄位內容	欄位說明
車輛代碼	公車 GPS 代碼
勤務狀態	公車執行勤務的狀況(包括正常、開始、結束)
行車狀況	公車的行車狀況(包括正常、車禍、故障、塞車等)
路線代碼	公車執行的路線代碼
去返程	為路線的去程或回程
站牌代碼	公車進離站的站牌代碼
進站離站	為進站或離站資料
GPS 時間	資料點回傳的GPS時間

5.2.2 實際到站資料

本研究由台北市公共運輸處取得路線隨車調查資料，資料以隨車調查方式收集公車實際到站時間，並記錄平假日、尖離峰、是否經過公車專用道、進站時是否有列隊情形等相關屬性。因為隨車調查資料只記錄公車車牌號碼，而定點回報資料記錄公車 GPS 代碼；因此將實際到站的時間比對取得的公車定時回報資料，找出每筆調查車輛的公車 GPS 代碼，經過篩選可比較分析的路線作為測試對象，結果如下表 18 所示：

表 18 測試路線

名稱	每日 發車數	發車間隔(分鐘)		總往返 里程(km)	起站	迄站	建置 廠商
		尖峰	離峰				
棕 9	80	04~06	05~10	27.6	東湖	圓環	1
518	99	04~06	05~10	25.6	麥帥新城	圓環	4
612	58	12~15	20~30	48	松德站	大同之家	4
299	55	12~15	15~20	43	輔大	永春高中	4
513	72	12~15	15~20	22.4	輔仁大學	捷運台大醫院站	4
72	55	12~15	15~20	27.7	大直	捷運麟光站	4
279	28	20~30	40~60	41.6	天母	南港國宅	4

5.2.3 現有系統預測資料

透過台北市公共運輸處取得現有系統績效分析資料，資料以網路查詢的方式

進行，記錄每次的查詢時間及公車到所有下游站分鐘數的預測值，透過查詢時間加上預測分鐘數可得公車到每一下游站的到站預測時間。接著將到站預測時間與實際隨車調查時間做比對，即得到現有系統的預測績效資料。

5.3 參數設定

5.3.1 k 值設定

由 4.3.3 小節可知， k 值的設定需透過預測誤差的觀察分析，因此本研究利用 2010 年 2 月 6 日的資料比較在不同的 k 值下，所得之到站時間的平均絕對誤差，其各路線在不同 k 值下所得出的平均絕對誤差如圖 10 所示。由圖中可觀察出當 $k=40$ 時，各路線的誤差由較大的起伏趨勢逐漸降為平緩的狀態，表示當 $k>40$ 時有比較穩定的預測誤差，因此本研究將 k 值設為 40。

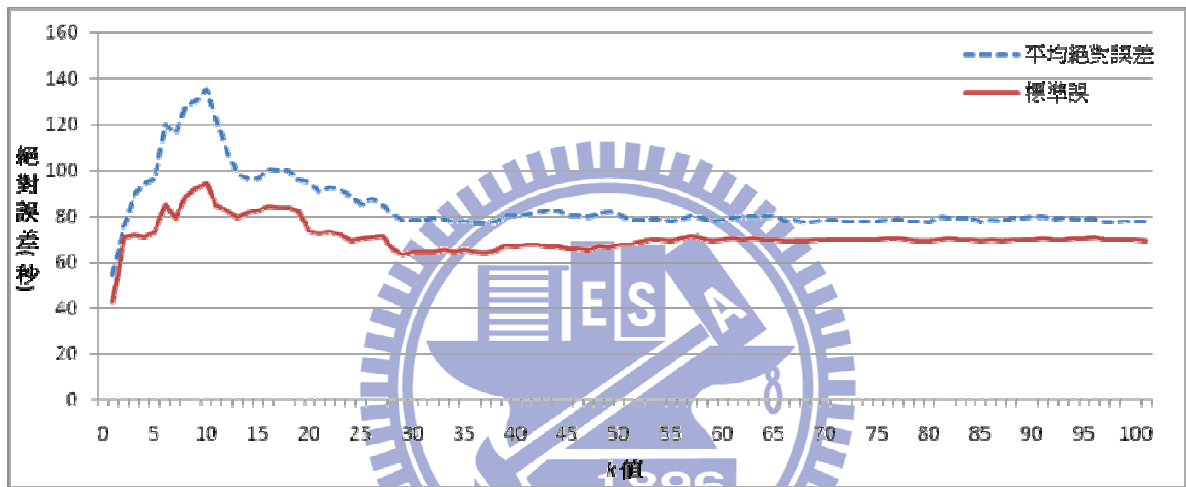


圖 10 k 值與絕對誤差趨勢圖

5.3.2 時間窗設定

依據本研究對象台北市公車的公車班距為 5~40 分鐘，因此將時間窗設為最大的班距 40 分鐘。

5.4 結果分析

本研究以 5 分鐘進行一次預測演算法，針對每一站牌輸出即將到站公車之到達時間預測值。以下是本研究演算法的預估結果分析，分別比較本研究在各種情境下的績效表現、本研究與現有系統的績效表現及演算時間的績效。

預測誤差以兩個績效指標做分析，包括平均絕對誤差(Mean absolute deviation, MAD)及平均絕對百分比誤差(Mean Absolute Percentage Error, MAPE)：

$$MAD = \frac{1}{n} \sum_{i=1}^n |e_i - o_i| \quad (11)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|e_i - o_i|}{o_i} \times 100\% \quad (12)$$

其中 e_i 為預測值， o_i 為實際值， n 為資料筆數。

各情境的績效表現分析：

(1) 整體分析

本研究的預估結果與實際停靠站時間進行比對，共可以產生 2567 筆預測資料，其結果分析如表 19 所示。就整體而言，本研究公車預估到站時間預測的平均絕對誤差約為 1.5 分鐘，平均絕對百分比誤差約為 25%。

表 19 整體誤差結果

指標	個數	最小值	最大值	平均數	標準差
絕對誤差	2567	0	630	81.62	87.64
絕對百分比誤差	2567	0	13.2	0.25	0.76

(2) 早到與晚到分別分析

如將資料分為早到與晚到（正負值）分別加以探討，則可個別分析這兩類資料之平均數與資料散佈情況，其結果如表 20 所示。

這兩者資料個數約為總資料數之 46%及 54%。早到與晚到的平均誤差分別為 73.01 秒與 88.87 秒，顯示早到的誤差小於晚到的誤差。

表 20 早到/晚到誤差結果

指標	情境	個數	最小值	最大值	平均數	標準差
絕對誤差	正值/早到	1172	0	630	73.01	87.16
	負值/晚到	1395	1	428	88.87	87.41
絕對百分比誤差	正值/早到	1172	0	13.2	0.37	1.1
	負值/晚到	1395	0	1	0.15	0.15

下圖 11 為誤差次數分配情形，由圖中可看出正負誤差的分配對稱，顯示預測的結果沒有太大的偏誤，且小於正負 3 分鐘的誤差皆落在兩倍標準差的範圍內，表示研究設計之預測演算模式可用。

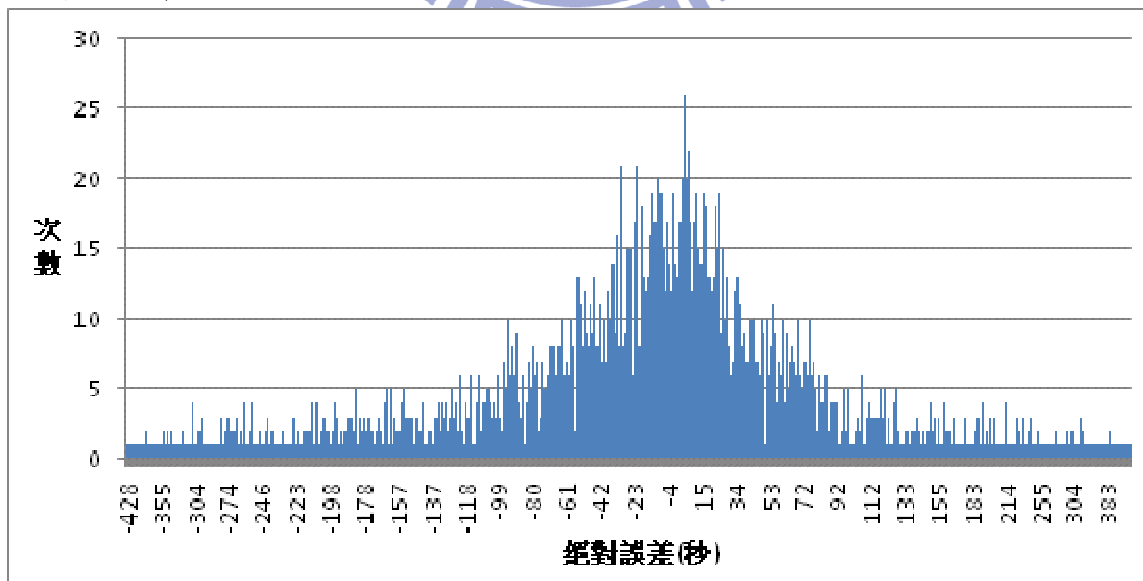


圖 11 預測誤差分配圖

(3) 距離站位數分析

本部分主要探討車輛與目標站牌的距離對誤差所可能造成的影響。為求精簡起見，本項分析以多、中及少站位數分類，多站位數為公車位置距離目標站牌 10 站以上，中站位數為距離 5~10 站，少站位數為距離 5 站以內。相關的分析結果如表 21 所示。

表 21 站位數資料分析

指標	情境	個數	最小值	最大值	平均數	標準差
絕對誤差	多站位數	641	0	554	144.41	108.82
	中站位數	1001	0	630	72.07	73.07
	少站位數	925	0	587	48.46	58.43
絕對百分比 誤差	多站位數	641	0	1.1	0.1	0.09
	中站位數	1001	0	11.52	0.15	0.43
	少站位數	925	0	13.2	0.48	1.14

● 距離站位數情況差異分析

針對三種情境作 ANOVA 分析，檢查不同情境下是否存在差異，如表 22 所示，結果為 $F(2, 22952) = 289.96$ ， $p < 0.05$ ，顯示不同情境下存在差異。接著利用 Tukey 法分析各情境的差異比較，可得表 23。由結果可知三種情境皆有顯著的差異，表示演算法在少站位數的情況下有最好的預測表現，隨著預測站位數增加，所累加的誤差亦隨之增加。

表 22 距離站位數 ANOVA 分析

	平方和	自由度	平均平方和	F 檢定	顯著性
組間	3635430.041	2	1817715.021	289.96	0.000
組內	16073349.69	2564	6268.857		
總和	19708779.73	2566			

表 23 站位數屬性差異分析

情境	情境	平均數差異	標準誤	顯著性	95% 信賴區間	
					下限	上限
少站位數	中站位數	-23.61671(*)	3.61106	0.000	-32.0849	-15.1485
	多站位數	-95.94940(*)	4.06902	0.000	-105.4915	-86.4073
中站位數	少站位數	23.61671(*)	3.61106	0.000	15.1485	32.0849
	多站位數	-72.33269(*)	4.0053	0.000	-81.7254	-62.94
多站位數	少站位數	95.94940(*)	4.06902	0.000	86.4073	105.4915
	中站位數	72.33269(*)	4.0053	0.000	62.94	81.7254

附註：(*) 在 0.05 水準下有顯著的差異

(4) 尖離峰分析

本部分主要探討尖離峰對誤差所可能造成的影響，因此資料分析以尖離峰加以分類。尖峰分為上午尖峰（晨峰）7~10點、下午尖峰（昏峰）16~20點，離峰為10~16點。相關分析結果如表 24 所示。

表 24 尖離峰資料分析

指標	情境	個數	最小值	最大值	平均數	標準差
絕對誤差	上午尖峰	1176	0	513	79.32	88.86
	下午尖峰	101	0	479	55.49	71.96
	離峰	1290	0	630	85.77	87.27
絕對百分比 誤差	上午尖峰	1174	0	10.45	0.21	0.5
	下午尖峰	100	0	10.5	0.37	1.09
	離峰	1286	0	13.2	0.28	0.9

● 尖離峰情況差異分析

針對三種情境作 ANOVA 分析，檢查不同情境下是否存在差異，如表 25 所示，結果為 $F(2, 2564) = 6.371$ ， $p < 0.05$ ，顯示不同情境下存在差異。接著利用 Tukey 法分析各情境的差異比較，可得表 26，由結果可知三種情境皆有顯著差異，表示演算法在下午尖峰的時段最好的預測表現。

表 25 尖離峰情境 ANOVA 分析

	平方和	自由度	平均平方和	F 檢定	顯著性
組間	97456.274	2	48728.137	6.371	0.002
組內	19611323.46	2564	7648.722		
總和	19708779.73	2566			

表 26 尖離峰屬性差異分析

情境	情境	平均數差異	標準誤	顯著性	95% 信賴區間	
					下限	上限
上午尖峰	下午尖峰	23.83458(*)	9.06829	0.023	2.5688	45.1003
	離峰	-6.45314	3.52608	0.16	-14.722	1.8158
下午尖峰	上午尖峰	-23.83458(*)	9.06829	0.023	-45.1003	-2.5688
	離峰	-30.28772(*)	9.03655	0.002	-51.479	-9.0964
離峰	上午尖峰	6.45314	3.52608	0.16	-1.8158	14.722
	下午尖峰	30.28772(*)	9.03655	0.002	9.0964	51.479

附註：(*)在 0.05 水準下有顯著的差異

(5) 平假日分析

本部分主要探討平假日對誤差所可能造成的影響，因此資料分析以平假日分類，平日為星期一至星期五，假日為星期六與星期日。相關分析結果如表 27 所示。

表 27 平假日資料分析

指標	情境	個數	最小值	最大值	平均數	標準差
絕對誤差	平日	1799	0	630	83.69	91.8
	假日	768	0	513	76.79	76.87
絕對百分比 誤差	平日	1799	0	13.2	0.24	0.73
	假日	768	0	11.6	0.28	0.81

● 平假日情況差異分析

針對兩種情境作 ANOVA 分析，檢查不同情境下是否存在差異，如表 28 所示，結果為 $F(1, 2565) = 3.338$ ， $p > 0.05$ ，顯示兩種情境之預測誤差沒有顯著的差異。

表 28 平假日情境 ANOVA 分析

	平方和	自由度	平均平方和	F 檢定	顯著性
組間	25612.805	1	25612.805	3.338	0.068
組內	19683166.93	2565	7673.749		
總和	19708779.73	2566			

(6) 公車專用道資料

本部分主要探討公車專用道對誤差所可能造成的影響，因此以是否有公車專用道進行資料分析，可得到有無公車專用道對於預估時間準確度之影響。相關分析結果如表 29 所示。

表 29 公車專用道資料分析

指標	情境	個數	最小值	最大值	平均數	標準差
絕對誤差	有公車專用道	212	0	435	87.72	87.1
	無公車專用道	2355	0	630	81.08	87.69
絕對百分比 誤差	有公車專用道	212	0	11.6	0.28	0.92
	無公車專用道	2355	0	13.2	0.25	0.74

● 公車專用道情況差異分析

針對兩種情境作 ANOVA 分析，檢查不同情境下是否存在差異，如表 30 所示，結果為 $F(1, 2565) = 1.117$ ， $p > 0.05$ ，顯示兩種情境間無顯著差異，表示就有無公車專用道屬性資料而言，演算法並無不會因為公車有無行駛公車專用道而造成誤差。

表 30 有無公車專用道情境 ANOVA 分析

	平方和	自由度	平均平方和	F 檢定	顯著性
組間	8576.473	1	8576.473	1.117	0.291
組內	19700203.26	2565	7680.391		
總和	19708779.73	2566			

(7) 列隊資料分析

本部分主要探討有無列隊（前方是否有其他公車列隊）對誤差所可能造成的影響，因此以是否有公車列隊之情形進行資料分析，可得公車列隊對於預估時間準確度之影響。相關分析結果如表 31 所示。

表 31 列隊資料分析

指標	情境	個數	最小值	最大值	平均數	標準差
絕對誤差	有列隊	231	1	479	72.43	74.6
	無列隊	2336	0	630	82.53	88.79
絕對百分比 誤差	有列隊	231	0	4.48	0.21	0.41
	無列隊	2336	0	13.2	0.26	0.78

● 列隊情況差異分析

針對兩種情境作 ANOVA 分析，檢查不同情境下是否存在差異，如表 32 所示，結果為 $F(1, 2565) = 2.794$ ， $p > 0.05$ ，顯示兩種情境間無顯著差異。

表 32 有無列隊情境 ANOVA 分析

	平方和	自由度	平均平方和	F 檢定	顯著性
組間	21447.696	1	21447.696	2.794	0.095
組內	19687332.04	2565	7675.373		
總和	19708779.73	2566			

(8) 演算法運算效率

研究以 VB.NET 2005 進行演算法的撰寫，資料庫使用 MS SQL 2005，運作環境為 AMD Athlon(tm) 64 X2 DualCore Processor 5200+，2.61GHz，2.00GB RAM。其每條路線進行一次到站時間預測的運算時間如下表 33 所示，平均運算時間為 7.92 秒，每次預測平均輸出 20 個預測值。

表 33 演算法運算時間

預測次數	最小值	最大值	平均值
604	1	23	7.92

各路線的演算效率如表 34 所示，由平均運算時間對應路線的單趟站數，可發現站數多，所需的演算時間越久，因為在延遲加總的過程中，是從起始路段開始依序計算進入下一路段的時間，再判斷是否要用下一時段的資料，因此，如果站數越多，所需計算的路段旅行時間越多。由表 34 中可知運算時間最少的為 513 路線，平均 3.54 秒；最多的為 299 路線，平均 10.83 秒。

表 34 各路線運算時間

路線	預測次數	最小值	最大值	平均值	路線單趟站數
棕 9	86	1	15	7.8	32
518	81	3	14	8.69	35
621	82	1	20	10.79	51
299	58	3	23	10.83	49
513	148	1	10	3.54	18
72	58	2	12	6.91	33
279	91	3	20	10.67	53

5.5 與現有預測系統比較分析

在本小節中，將分別針對各個調查屬性進行分析，以絕對誤差比較本研究演算法與現有預測系統的預測績效。

(1) 整體分析

就整體資料而言，如表 35，本研究之平均絕對誤差比現有系統的平均絕對誤差少了 36.47 秒鐘，以 ANOVA 針對兩個系統作分析，如表 36 所示，結果為 $F(1, 21175) = 254.984$ ， $p < 0.05$ ，顯示本研究顯著優於現有系統的整體預測誤差。接著將針對各種情境進行比較，分析本研究與現有系統在各種情境下表現之優劣。

表 35 整體誤差與現有系統比較結果

	個數	最小值	最大值	平均數	標準差
現有系統	18610	0	1099	118.09	111.03
本研究	2567	0	630	81.62	87.64

表 36 整體資料與現有系統之差異分析

	平方和	自由度	平均平方和	F 檢定	顯著性
組間	2999647.681	1	2999647.681	254.984	0.000
組內	249104392.1	21175	11764.08		
總和	252104039.8	21176			

(2) 早到與晚到分別分析

如將資料分為早到與晚到（正負值）分別加以探討，其結果如表 37 所示，並以 ANOVA 針對兩個系統作分析，如表 38 所示。由表可知本研究在早到與晚到資料中皆有比較好的表現。

表 37 早到與晚到資料與現有系統比較結果

		個數	最小值	最大值	平均數	標準差
正值/早到	現有系統	11273	0	919	111.99	96.41
	本研究	1172	0	630	73.01	87.16
負值/晚到	現有系統	7337	1	1099	127.46	129.79
	本研究	1172	1	428	88.87	87.41

表 38 早到與晚到資料與現有系統之差異分析

情境	自由度		F 檢定	顯著性
	組間	組內		
正值/早到	1	12443	176.652	0.000
負值/晚到	1	8730	113.556	0.000

(3) 距離站位數分析

本部分主要探討車輛與目標站牌的距離對誤差所可能造成的影響。為求精簡起見，本項分析以多、中及少站位數分類，多站位數為公車位置距離目標站牌 10 站以上，中站位數為距離 5~10 站，少站位數為距離 5 站以內。相關的分析結果如表 39 所示，並以 ANOVA 針對兩個系統作分析，如表 40 所示。由表可知本研究在三種情境中皆有比較好的表現。

表 39 站位數資料分析與現有系統比較結果

		個數	最小值	最大值	平均數	標準差
少站位數	現有系統	6244	0	1026	83.47	81.81
	本研究	925	0	587	48.46	58.43
中站位數	現有系統	8161	0	999	120.54	105.66
	本研究	1001	0	630	72.07	73.07
多站位數	現有系統	4205	0	1099	164.74	137.8
	本研究	641	0	554	144.41	108.82

表 40 站位數資料各與現有系統之差異分析

情境	自由度		F 檢定	顯著性
	組間	組內		
少站位數	1	7167	157.555	0.000
中站位數	1	9160	198.912	0.000
多站位數	1	4844	12.751	0.000

(5) 尖離峰分析

本部分主要探討尖離峰對誤差所可能造成的影響，因此資料分析以尖離峰加以分類。尖峰分為上午尖峰（晨峰）7~10 點、下午尖峰（昏峰）16~20 點，離峰為 10~16 點。相關的分析結果如表 41 所示，並以 ANOVA 針對兩個系統作分析，如表 42 所示。由表可知本研究在三種情境中皆有比較好的表現。

表 41 尖離峰資料與現有系統比較結果

		個數	最小值	最大值	平均數	標準差
上午尖峰	現有系統	5429	0	1026	106.32	96.8
	本研究	1176	0	513	79.32	88.86
下午尖峰	現有系統	6911	0	898	131.34	114.17
	本研究	101	0	479	55.49	71.96

離峰	現有系統	6270	0	1099	113.67	117.37
	本研究	1290	0	630	85.77	87.27

表 42 尖離峰資料各與現有系統之差異分析

情境	自由度		F 檢定	顯著性
	組間	組內		
上午尖峰	1	6603	77.355	0.000
下午尖峰	1	7010	44.325	0.000
離峰	1	7558	65.449	0.000

(6) 平假日分析

本部分主要探討平假日對誤差所可能造成的影響，因此資料分析以平假日分類，平日為星期一至星期五，假日為星期六與星期日。相關的分析結果如表 43 所示，並以 ANOVA 針對兩個系統作分析，如表 44 所示。由表可知本研究在兩種情境中皆有比較好的表現。

表 43 平假日資料與現有系統比較結果

		個數	最小值	最大值	平均數	標準差
平日	現有系統	11847	0	1099	122.6	107.88
	本研究	1799	0	630	83.69	91.8
假日	現有系統	6763	0	1079	110.19	115.92
	本研究	768	0	513	76.79	76.87

表 44 平假日資料與現有系統之差異分析

情境	自由度		F 檢定	顯著性
	組間	組內		
平日	1	13644	210.857	0.000
假日	1	7529	60.719	0.000

(7) 公車專用道資料

本部分主要探討公車專用道對誤差所可能造成的影響，因此以是否有公車專用道進行資料分析，可得到有無公車專用道對於預估時間準確度之影響。相關的分析結果如表 45 所示，並以 ANOVA 針對兩個系統作分析，如

表 46 所示。由表可知本研究在兩種情境中皆有比較好的表現。

表 45 有無公車專用道資料與現有系統比較結果

		個數	最小值	最大值	平均數	標準差
有公車專用道	現有系統	11634	0	1099	126.21	119.55
	本研究	212	0	435	87.72	87.1
無公車專用道	現有系統	6976	0	932	104.55	93.58
	本研究	2355	0	630	81.08	87.69

表 46 公車專用道資料與現有系統之差異分析

情境	自由度		F 檢定	顯著性
	組間	組內		
有公車專用道	1	11844	21.765	0.000
無公車專用道	1	9329	114.309	0.000

(8) 列隊資料分析

本部分主要探討有無列隊（前方是否有其他公車列隊）對誤差所可能造成的影響，因此以是否有公車列隊之情形進行資料分析，可得公車列隊對於預估時間準確度之影響。相關的分析結果如表 47 所示，並以 ANOVA 針對兩個系統作分析，如表 48 所示。由表可知本研究在兩種情境中皆有比較好的表現。

表 47 公車列隊資料與現有系統比較結果

		個數	最小值	最大值	平均數	標準差
有列隊	現有系統	1438	0	786	134.52	126.72
	本研究	231	1	479	72.43	74.6
無列隊	現有系統	17172	0	1099	116.71	109.5
	本研究	2336	0	630	82.53	88.79

表 48 公車列隊資料與現有系統之差異分析

情境	自由度		F 檢定	顯著性
	組間	組內		
有列隊	1	1667	52.512	0.000
無列隊	1	19506	208.913	0.000

5.6 小結

由前面 5.4、5.5 節可知，本研究所設計的公車到站 k -NN 預測演算法與現有系統比較有比較好的預測表現。就整體而言，本研究之平均絕對誤差為 81.62 秒，相對於現有系統之平均絕對誤差 118.09 秒改善了 30.8% 的誤差；在其他屬性分析中，本研究的表現亦明顯優於現有預測系統，其中，以下午尖峰的誤差改善幅度最大，與現有系統相比改善了 57.7%，其次為少站位數，與現有系統相比改善了 41.9% 表示 k -NN 演算法用於下午尖峰、短距的到站時間預測有不錯的效果。

第六章 結論與建議

本研究應用 k -NN 演算法預測公車到站時間，利用 7 條台北市區公車路線作實例測試，並與現有預測系統作比較，由分析結果可知本研究所設計的演算法具有良好表現績效，能夠提供使用者更精確的預測結果。

6.1 結論

- (1) 本研究利用市區公車 GPS 資料，以 k -NN 方法預測公車到站時間，整體資料之平均絕對誤差為 81.62 秒，相對於現有預測系統往下修正 30.8% 的平均誤差；且在各種情境下與現有系統比較皆有比較小的誤差，表示本研究所建構的預測模式能夠適用於各種情境。
- (2) 由各個屬性的資料分析，可知在站位數為少站位數、時段為下午尖峰的情境下，本演算法有比較好的表現。
- (3) 由運算時間結果可發現，當路線上的公車站數越多，其所需的運算時間越長，此與演算法之延遲加總機制有關，但其平均運算時間皆在可接受範圍內。
- (4) 在設定 k 值的實驗中，可發現 $k > 40$ 的誤差趨於平穩狀態，顯示 k -NN 依靠大量歷史資料庫有穩定預測表現。

6.2 建議

- (1) 研究在資料過濾的處理上，以簡單的判斷方法剔除不合理的資料點，未來研究可針對搜集到交通資料特性，從現有的資料過濾演算法中找出有可有效過濾不合理資料點的方法，以提高歷史資料的可靠度。
- (2) 在建立路段歷史資料庫的時，因為以路線上五分鐘的車輛進離站資料得出路段旅行時間，在班距較稀疏的路線，可能會因為只有一輛車的資料，而造成許多路段出現空值的情形。未來若能藉由其他輔助資訊，得到路段的旅行時間，減少空值的產生，則可獲得較多的比對資訊，增加預測的準確性。
- (3) 在運算時間的結果分析中可發現，當站數越多所需的運算時間越長，若將來欲應用此方法到站數較多的路線上，則可改善延遲加總的運算流程，以提高運效率。
- (4) 本研究以同一條路線的路段旅行時間當作比對向量，若公車路線有行駛相同路段的部分，則可加入重疊路段的其他路線資料，減少路段資料缺漏的情形，獲得較多的比對資訊，同時增加預測之準確性。
- (5) 目前僅建立兩個月的歷史資料庫，並用一個星期的實際調查資料作實證。未來研究能夠取得更多的歷史資料，如持續搜集一年以上的歷史資料，則能夠反映更多樣的交通狀況，如季節變化、連續假期等；另外一方面若能蒐集較長時間的實際調查資料，則能夠 k -NN 模式校估得更為準確，得出更好的預測值。若建立更龐大的資料庫則需考慮到歷史資料搜尋的時間，可觀察資料

特性將其分類，例如星期別、特別假日等，只搜尋相似的歷史資料以減少運算時間。

- (6) 本研究未針對事件或事故發時進行預測分析，但此時為等車民眾最需要知道到站時間的時機；未來若能取得市區交通事故資料，則可針對事故發生時，公車到站時間的變化為何來調整預測演算法。
- (7) 本研究在進行路段的延遲加總時，將全部的k筆旅行時間作加總平均，未來可考慮路段的旅行時間的變異程度，剔除變異較大的資料，或是給予權重作調整，以提高預測的精準性。



參考文獻

- [1] L. Huang, and M. Barth, "A Novel Loglinear Model For Freeway Travel Time Prediction", the 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China, October 12-15, 2008.
- [2] J. Kwon and K. Petty, "A Travel Time Prediction Algorithm Scalable to Freeway Networks with Many Nodes with Arbitrary Travel Routes", Presented at 84th TRB Annual Meeting at Washington D.C., Forthcoming in Transportation Research Record, 2005.
- [3] X. Zhang and J. Rice, "Short-Term Travel Time Prediction", Transportation Research Part C, Vol. 11, pp. 187-210, 2003.
- [4] H. Sun, H. Liu, H. Xiao, and B. Ran, "Short Term Traffic Forecasting Using the Local Linear Regression Model", the 82nd Annual Meeting of the Transportation Research Board, Washington, DC, Jan. 2003.
- [5] Y. Li, "Short Term Prediction of Motorway Travel Time Using ANPR and Loop Data", Journal of Forecasting, Vol. 27, pp. 507-517, 2008.
- [6] J. Kwon, B. Coifman, and P. Bickel, "Day-to-Day Travel-Time Trends and Travel-Time Prediction from Loop-Detector Data", Journal of the Transportation Research Board, Vol. 1717, pp. 120-129, 2000.
- [7] L. Sun, J. Yang, H. Mahmassani, "Travel Time Estimation Based on Piecewise Truncated Quadratic Speed Trajectory", Transportation Research Part A, Vol. 42, pp. 173-186, 2008.
- [8] G. L. Chang, N. Zou, and J. Wang, "Development and Field Evaluation of a Real-Time Travel Time Prediction System", Maryland Department of Transportation State Highway Administration, Final Report, 2006.
- [9] J. Yeon, L. Elefteriadou, and S. Lawphongpanich, "Travel Time Estimation on a Freeway Using Discrete Time Markov Chains", Transportation Research Part B, Vol. 42, pp. 325-338, 2008.
- [10] C. E. Cortés, R. Lavanya, J. S. Oh, and R. Jayakrishnan, "General Purpose Methodology for Estimating Link Travel Time with Multiple Point Detection of Traffic", Transportation Research Record: Journal of the Transportation Research Board, Vol. 1802, pp. 181-189, 2002.
- [11] D. Paterson and G. Rose, "A Recurse, Cell Processing Model for Prediction Freeway Travel Times", Transportation Research Part C, Vol. 16, pp. 432-453, 2008.
- [12] B. Hellenga, G. Knapp, "Automatic Freeway Incident Detection using Travel Time Data from AVI Equipped Vehicles", 6th World Congress on Intelligent Transport Systems, Toronto, Canada, 1999.
- [13] B. Hellenga, and R. Gudapati, "Estimating Link Travel Times For Advanced Traveler Information Systems", the Canadian Society of Civil Engineers, 3rd Transportation Speciality Conference. London, Ontario, 2000.
- [14] K. Choi and Y. Chung, "A Data Fusion Algorithm for Estimating Link Travel Time", Intelligent Transportation Systems, Vol. 7, pp. 235-260, 2002.
- [15] F. Soriguera, D. Abejón, L. Thorson, and F. Robuste, "Highway Travel Time Data Fusion", the TRB 87th Annual Meeting. Washington, D.C., Transportation Research Board, 2008.
- [16] N. Faouzi, L. Klein, O. Mouzon, "Improving Travel Time Estimates from Loop and

- Toll Collection Data with Dempster-Shafer Data Fusion”, TRB Annual Meeting CD-ROM, 2009.
- [17] C.Nanthawichit, T.Nakatsuji, H.Suzuki, “Application of Probe Vehicle Data for Real-Time Traffic State Estimation and Short-Term Travel Time Prediction on a Freeway”, RB Annual Meeting CD-ROM, 2003.
- [18] B.L Smith, and M.J. Demetsky, “Traffic Flow Forecasting: Comparison of Modeling Approaches”, Journal of Transportation Engineering, Vol.123, pp.261-266, 1997.
- [19] B.L Smith, B.M. Williams, and R.K Oswald, “Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting”, Transportation Research Part C, Vol.10, No.4, pp.303-321, 2002.
- [20] S. Clark, “Traffic Prediction Using Multivariate Nonparametric Regression”, Journal of Transportation Engineering, Vol.129, No.2, pp.161-168, 2003.
- [21] K.Choi, C.H.Shin, and I.Park, “An Algorithm for Calculating Dynamic Link Travel Times Using GPS and A Digital Road Map” 5th ITS World Congress, 1998.
- [22] S. Robinson, and J. Polak, “Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the K-NN Method”, Journal of the Transportation Research Board, Vol.1935, pp.47-56, 2005.
- [23] A.Shalaby and A.Farhan, “Prediction Model of Bus Arrival and Departure Times Using AVL and APC Data”, Journal of Public Transportation, Vol.7, No.1, pp.41-61, 2004.
- [24] J.Patnaik, S.Chien, and A.Bladikas, “Estimation of Bus Arrival Times Using APC Data”, Journal of Public Transportation, Vol.7, No.1, pp.1-20, 2004.
- [25] E.H. Chung, and A. Shalaby, “Expected Time of Arrival Model for School Bus Transit Using Real-Time Global Positioning System-Based Automatic Vehicle Location Data”, Journal of Intelligent Transportation Systems, Vol.11, No.4, pp.157-167, 2007.
- [26] 吳佳峰, 「有 GPS 資訊提供下之車輛旅行時間預估模式之研究」, 國立交通大學運輸工程與管理系碩士論文, 民國九十年。
- [27] 陳建名, 「有 GPS 資訊提供下之公車旅行時間之研究」, 國立交通大學運輸科技與管理系碩士論文, 民國九十二年。
- [28] J. Benedetti, “On the Nonparametric Estimation of Regression Functions”, Journal of the Royal Statistical Society, Series B, Vol. 39, pp. 248-253, 1977.
- [29] C. Stone, “Consistent Nonparametric Regression”, The Annals of Statistics, Vol. 5, pp. 595-645, 1977.
- [30] J.Tukey, “Exploratory Data Analysis”, Reading, MA: Addison Wesley, 1977.
- [31] N.S. Altman, “An Introduction to Kernel and Nearest Neighbor Nonparametric Regression”, The American Statistician, Vol. 46, NO. 3, pp. 175-185, 1992.

簡 歷



姓名：何佳儒

籍貫：臺灣省南投縣

生日：民國 74 年 12 月 25 日

學歷：民國九十九年七月國立交通大學運輸科技與管理學系碩士班畢業

民國九十七年七月國立交通大學運輸科技與管理學系學士班畢業

電郵信箱：flora96321@gmail.com

