

國立陽明交通大學
運輸與物流管理學系
碩士論文

Department of Transportation and Logistics Management
National Yang Ming Chiao Tung University
Master Thesis

以改良式k-NN法預測高速公路旅行時間之研究
The Study of Predicting Freeway Travel Time By Using
Improved k-NN Method

研究生：丁楚文 (Ting, Chu-Wen)
指導教授：王晉元 (Wang, Jin-Yuan)

中華民國一一〇年六月
June 2021

以改良式k-NN法預測高速公路旅行時間之研究
The Study of Predicting Freeway Travel Time By Using
Improved k-NN Method

研究生：丁楚文 Student：Chu-Wen Ting

指導教授：王晉元 Advisor：Jin-Yuan Wang

國立陽明交通大學

運輸與物流管理學系

陽明交大
碩士論文

A Thesis
Submitted to Department of Transportation and Logistics Management
College of Management
National Yang Ming Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master
in
Traffic and Transportation

June 2021
Hsinchu, Taiwan, Republic of China

中華民國一一〇年六月

誌謝

時光飛逝，兩年的生活始於騎著機車，被新竹的風吹得左搖右晃，終於平穩穿梭在快慢車道間，在疫情肆虐下繼續前行。這些日子，最要感謝的是我的指導教授 王晉元老師，於啟垣廳相遇的十分鐘、開學前螢幕中的 OCW、第一次發抖進入的 8 樓辦公室、往返新竹宜蘭的車上、到遠距的視訊口試，研究所求學的日子，老師總是無時無刻地出現在我的生活，在學業上給予我幫助、於計劃案的耐心指導、更使學生的碩士論文順利完成，跟老師的討論，總能反思並發現自己不足的地方，從中學習如何有邏輯的去思考、清楚的表達想法、且能夠有找出問題並解決的能力，遇到問題時，老師都會給方向去思考，但其實知道我的問題在哪。我想，也即將離開有老師叮嚀的日子了，我是個時常讓您操心，添麻煩的事一點也沒少的學生，但還是想跟您說，能成為您的學生，是件幸福的事。

在論文口試期間，感謝 盧宗成老師和 鍾易詩老師撥冗閱讀我的論文與擔任口試委員，並於口試時提供學生寶貴的建議與指教，使本篇論文更臻完善，感謝學期間教導我過的所有老師，讓我收穫更多知識的走出校園。

感謝王 LAB 吵吵鬧鬧的夥伴，無時無刻說著很吵卻習慣的幹話、在乾杯拍下勁爆合照的士軒，一起熬夜拉緊報、在 LAB 開發遊戲的 77，能夠訴說心事、一起討論股票的 carol，一起下課去酒吧、唱歌到天亮的阿敏，手作天才拔辣，口才很好的明展，還有碩一時照顧我們的學長姐允哥、鏡芳、子勤、佑芳，碩二時目送我們離開的學弟妹子賢、日龍、語芳，以及能打聽到不少故事的品蓁學姊，有你們的 LAB 生活，我會非常懷念的。

感謝這些日子一同前行的朋友，像姊姊般溫暖而笑聲爽朗的菜包，讓我串門子的佳臨，一起回家卻又一同趕不上車的美食專家暨時間管理大師 Toro，統一獅冠軍就要付錢的祐德，按摩超舒服的宣理，見證喝酒錄下黑暗史在一起腸胃炎的宗翰，爬山唱歌自帶嗨嗨效果采馨，同為草莓多多愛好者的二代團員洋洋，三不五時就會來點尖叫、容易一起嚇到的 hyu，看劇常看到哭、每晚離開前都看得到的晟華，喜歡益全卻害羞不敢說出來的蝦堡 Twei，外表冷酷但真的很酷的雅方，熟到被人以為很不熟、睡覺睡到東倒西歪的老朋友庭嘉，隨著咪聽浮上浮下的深海魚先維，帶了好多包辣條回來的安然，住在可愛動物區的貞瑩，很會吃也是袋鼠餅乾來源的家瑄，位置常被我佔領、希望以後能常見到的鄰居家瑜，聖誕禮物獲得馴鹿的 Anita，研所的日子有你們陪伴，一起在統艙玩桌遊、打麻將、喝酒、夜唱、新館舊館來回打羽球、出去吃飯、夜唱，有你們的日子耍廢或認真總是不無聊，多虧你們，這兩年才會如此充實與快樂。

感謝在我身邊的妳，貼心叮嚀我日常小事，也暖暖照顧像小孩的我，總是跟妳在耍廢卻很開心，未來的日子也許忙碌，但有妳在身邊足以。最後，感謝最愛的爸媽，你們是我內心最安穩的力量，謝謝你們讓我衣食無慮，我愛你們，只希望你們身體健康、平平安安，謝謝我的妹妹，胖妞，謝謝妳從小到大的陪伴，雖然，最後還是沒能參加我的畢業典，但一定要過得快樂，也要常來夢裡找我玩嘍。

丁楚文
2021.07 於新竹

以改良式 k-NN 法預測高速公路旅行時間之研究

學生：丁楚文

指導教授：王晉元

國立陽明交通大學運輸與物流管理學系碩士班

摘要

旅行時間預測為我國推動智慧型運輸系統主要發展項目之一，用以解決交通壅塞對環境造成的空氣污染與行程延誤所致之經濟損失，近年研究多蒐集一定數量之數據，並以不同機器學習算法進行預測。然而，影響旅行時間因素甚多，過往研究僅以單一或少數特徵進行預測。本研究期望發展一套即時預測方法，考量多個與旅行時間有關特徵去預測現況時間下之交流道間旅行時間，應用於我國高速公路此類無號誌化路段上，資料蒐集上使用了 VD 與 ETC 兩種資料，並以更為準確之方式推估旅行時間資訊，方法選擇上則考量預測效果與資料適用性，以 k-NN 法進行預測，並針對距離計算方式、估計預測值方法、k 值變動方式進行改良，預測結果可作為民眾出行依據與道路主管機關管理參考。

本研究以高公局資料進行實測，找出最佳特徵向量組合，結果顯示在此組合下與單一特徵向量相比能有效提升預測準確度，改良 k-NN 的部分則顯示本研究所提之距離計算方式同樣能提升預測效果，估計預測值方法上則以平均法進行預測之效果為佳，而 k 值變動與否則在準確度提升方面成效有限。在本研究最佳設定之下平均之 MAPE 為 2.57%、MAE 為 63.4、RMSE 為 118.9，具備準確預測高速公路旅行時間之能力。

關鍵詞：旅行時間預測、k-NN、監督式學習、車輛偵測器資料、電子收費

The Study of Predicting Freeway Travel Time

By Using Improved k-NN Method

Student: Chu-Wen Ting

Advisor: Dr. Jin-Yuan Wang

Department of Transportation and Logistics Management

National Yang Ming Chiao Tung University

Abstract

Travel time prediction is one of the main development projects to promote intelligent transportation systems in Taiwan, to solve air pollution and economic losses caused by traffic congestion. Researches in recent years often use a certain amount of data and different machine learning algorithms to make predictions. There are many factors that affect travel time, but previous studies only used single or a few features to predict. This research hopes to develop a real-time prediction methods, considering multiple features related to travel time in order to predict travel time between interchanges under current conditions, apply it on the segment of freeway which without signalize intersections. Two types of data, VD and ETC, are used in data collection, and estimated travel time information in a more accurate way. It takes into account that choosing k-NN method is about the prediction accuracy and the applicability of the data, and this research improve the method of distance calculation, estimating the predicted value and dynamic k. The results can be used as a reference for road authorities and users.

This research is tested with the data of Freeway bureau, MOTC, the results show that the best feature vector combination we find can effectively improve the prediction accuracy compared with single feature vector, and it shows that the distance calculation method proposed in this research can also improve the prediction accuracy, in terms of the method of estimating the predicted value, the average method is better, but the dynamic k has limited effect on accuracy improvement. Under the best setting of this research, the average MAPE is 2.57%, MAE is 63.4, RMSE is 118.9, shows that it has the ability to accurately predict freeway travel time.

Keywords: Travel Time Prediction, k-NN, Supervised Learning, VD Data, ETC

目錄

摘要.....	i
Abstract.....	ii
目錄.....	iii
圖目錄.....	v
表目錄.....	vi
第一章 緒論.....	1
1.1 研究背景與動機.....	1
1.2 研究目的.....	2
1.3 研究範圍.....	2
1.4 研究流程.....	2
第二章 文獻回顧.....	4
2.1 參數法預測模式.....	5
2.2 非參數法預測模式.....	5
2.3 k-NN 模式.....	8
2.4 CNN 模式.....	11
2.5 小結.....	13
第三章 研究方法.....	15
3.1 研究架構.....	15
3.2 資料蒐集.....	17
3.3 資料處理.....	18
3.3.1 車輛偵測器資料.....	18
3.3.2 ETC 資料.....	20
3.3.3 VD 速率資料差補.....	21
3.3.4 延遲加總.....	22
3.4 歷史資料庫建置.....	23
3.5 模式預測.....	24
3.5.1 標準化.....	24
3.5.2 獲取現況資料.....	24
3.5.3 模式參數設定.....	24
3.5.4 評估成效.....	28
3.5.5 演算範例.....	30
第四章 實測分析.....	33
4.1 實測時空間範圍.....	33
4.2 資料處理.....	33
4.2.1 遺失值比例.....	33
4.2.2 異常值比例.....	34

4.2.3	VD 速率資料差補	35
4.3	實測結果	37
4.4	統計分析	46
第五章	結論與建議	49
5.1	結論	49
5.2	建議	50
參考文獻	51

陽明交大
NYCU

圖目錄

圖 1 研究流程圖	3
圖 2 常見旅行時間預測模式之分類.....	4
圖 3 K-NN 法用於類別型問題	8
圖 4 K-NN 法用於連續型問題	8
圖 5 K-NN 法應用於交通預測問題.....	9
圖 6 CNN 模式架構	11
圖 7 卷積運算 (一)	12
圖 8 卷積運算 (二)	12
圖 9 池化運算	12
圖 10 研究架構圖	16
圖 11 車輛偵測器資料處理流程.....	18
圖 12 ETC 資料處理流程.....	20
圖 13 以 VD 速率資料進行差補情境	21
圖 14 以 VD 速率資料進行差補	22
圖 15 延遲加總概念.....	22
圖 16 變動 K-NN 法架構.....	27
圖 17 預測流程	30
圖 18 預測之空間範圍.....	33
圖 19 各 VD 遺失率	34
圖 20 ETC 資料獲取範圍.....	35
圖 21 前端旅行時間差補.....	36
圖 22 後端旅行時間差補.....	36

表目錄

表 1 模式分類各自優缺點	4
表 2 常見預測方法優缺點比較.....	7
表 3 常見樣本權重計算方式.....	10
表 4 文獻回顧重點整理	13
表 5 特徵向量選取說明	14
表 6 單一 VD 每五分鐘一筆原始資料.....	17
表 7 ETC 單一區間每五分鐘一筆原始資料.....	17
表 8 合併車道後單一車輛偵測器資料.....	20
表 9 各路段旅行時間資訊	23
表 10 歷史資料庫	23
表 11 現況資料	24
表 12 特徵向量與旅行時間之相關係數.....	26
表 13 範例標準化後其值.....	26
表 14 RANK-BASED WEIGHT 估計方式.....	26
表 15 本研究與過往 K-NN 文獻之差異	28
表 16 平均絕對百分比誤差計算.....	29
表 17 演算範例之現況資料.....	30
表 18 演算範例之歷史資料.....	30
表 19 演算範例之特徵向量權重.....	31
表 20 歷史資料與現況資料之距離.....	31
表 21 歷史資料與 X 之距離.....	31
表 22 不同 K 值下之預測.....	32
表 23 不同 K 值下之預測誤差.....	32
表 24 K=3 預測現況資料旅行時間.....	32
表 25 嚴重遺失的車輛偵測器.....	34
表 26 異常值佔比高的車輛偵測器.....	34
表 27 實測選定時段.....	37
表 28 不同 K 值下之誤差指標.....	37
表 29 特徵向量搭配.....	38
表 30 整體預測情形.....	38
表 31 特徵向量重組.....	38
表 32 重組後整體預測情形.....	39
表 33 固定/變動 K 整體預測情形.....	39
表 34 不同估計預測值方法比較.....	40
表 35 本研究所提之權重計算方式效果評估.....	40
表 36 本研究所得之最佳結果預測情形.....	41
表 37 統計分析所比對之樣本 (一)	46

表 38 統計分析所比對之樣本（二）	47
表 39 統計分析所比對之樣本（三）	47
表 40 統計分析所比對之樣本（四）	48

陽明交大
NYCU

第一章 緒論

1.1 研究背景與動機

旅行時間預測是一件很重要的事，為我國近年推動智慧型運輸系統（Intelligent Transportation System, ITS）發展項目之一。根據公路總局公佈的機動車輛登記數統計顯示，我國私有運具持有數仍持續上升，在道路環境供給不變之下，運輸需求的增加使得交通壅塞問題日益嚴重，不僅導致空氣污染，用路人也會因為行程的延誤造成經濟損失。要是能提前預測不同時間、空間之下的旅行時間，將可提供道路主管機關管理之參考，提前進行匝道儀控、路肩開放、號誌控制等措施，也可作為民眾出行之依據，調整出行時間、選擇替代道路甚至是更改其運具選擇，進而舒緩塞車問題。

旅行時間預測十分困難，道路本身狀況總是瞬息萬變，不同時、空間之下的變化是非常大的。我們往往使用電子收費系統（Electronic Toll Collection, ETC）、車輛偵測器（Vehicle Detector, VD）、探針車（Probe Vehicle）、自動車輛辨識（Automatic Vehicle Identification, AVI）資料來進行預測。從這些資料中去篩選出與旅行時間有關的速度、流量等資訊（劉軒寧, 2016）。

過去已有相當多關於旅行時間預測的研究，然而都有其侷限。國內外皆有許多學者利用各種演算法如支持向量機（Long et al., 2019）、線性迴歸模式（Kwon et al., 2000）、時間序列模式（Billings et al., 2006; Alghamdi et al., 2019）、神經網路模式（Ma et al., 2017）、K 最近鄰法（Habtemichael et al., 2016）進行預測。然而仍有其不足之處，許多研究係利用車機、AVI、VD 資料（蔡繼光, 2009; 朱志杰, 2013; 劉軒寧, 2016），並以推估之方式來計算旅行時間，而 Oh 等人（2002）研究中發現，以平均點速度推估出旅行時間之方法，在尖峰時段與實地蒐集旅行時間數據誤差達 20%，容易造成低估情況，而過往研究也較少針對所有與旅行時間有關因素（如流量、速率、佔有率及大型車佔比）進行完善考量，多以單一或少數因素進行預測（劉軒寧, 2016; 陳志華, 2017; Alghamdi, 2019）。

綜合上述觀點，本研究期望發展一套即時預測方法，預測現況時間下之交流道間旅行時間，應用於如我國高速公路此類無號誌化道路上。在資料蒐集方面同時使用 VD 資料與 ETC 資料，採取不同於過往旅行時間獲得之方式，透過 ETC 門架獲取各區段不同時間點所有車輛平均通過時間，於部分 ETC 無法獲得資料之路段則以 VD 資料進行差補。此外，本研究欲考量多個與旅行時間有關之因素，試圖以不同組合進行預測，觀察使用多元特徵向量是否能較單一特徵向量提升其預測準確率。方法選擇上則於後續章節進行探討，期盼預測結果能作為民眾出行

之依據與道路主管機關管理之用，進而舒緩交通壅塞所導致空氣污染、行程延誤等問題。

1.2 研究目的

為了舒緩 1.1 節所提到之交通壅塞問題，本研究目的為建構一無號誌化道路之旅行時間即時預測模式。預測現況時間下交流道間之旅行時間，以混合資料來源（ETC、VD）獲得旅行時間資訊，並同時考量多個與旅行時間有關之因素進行預測，探討不同因素組合下預測結果。盼能有效提升模式準確度，作為民眾出行之依據與道路主管機關管理之參考。

1.3 研究範圍

本研究提出之方法適用於我國無號誌化道路，如國道一號、國道五號。無號誌化道路不具備混合車流之特性，禁止機車行駛。且道路主線上不具備號誌，僅在出入口匝道處設有管制號誌用來控管進入主線之車輛數。

本研究係利用無號誌化道路上所佈設之車輛偵測器與電子收費門架資料進行分析。車輛偵測器型式多為埋設於路面下方之環路線圈偵測器，電子收費門架則架設於路面上方。兩者皆由交通部高速公路局所提供，資料間隔為每五分鐘一筆。

1.4 研究流程

本研究之研究流程如圖 1 所示，各步驟之說明如下：

1. 界定問題

了解研究背景與所現有之問題，並確認研究目的與範圍。

2. 文獻回顧

蒐集有關旅行時間預測之文獻，針對所使用之資料、輸入與輸出之變數、研究場域、預測結果等項目進行探討，找出適合應用在本研究之預測方法。

3. 研究方法建立

設計一預測模式，並說明模式如何進行旅行時間預測。

4. 數據處理

將蒐集之原始資料進行分析與處理，並轉換為旅行時間預測模式所需之格式。

5. 實證研究

以我國高速公路為實測場域，利用本研究所設計之模式進行旅行時間預測。

6. 準確度評估

使用常見誤差指標對實證研究之結果進行準確度評估與統計檢定，分析本研究所設計之模式預測可行性。

7. 結論與建議

將實證結果進行整理，提出結論與後續研究建議。

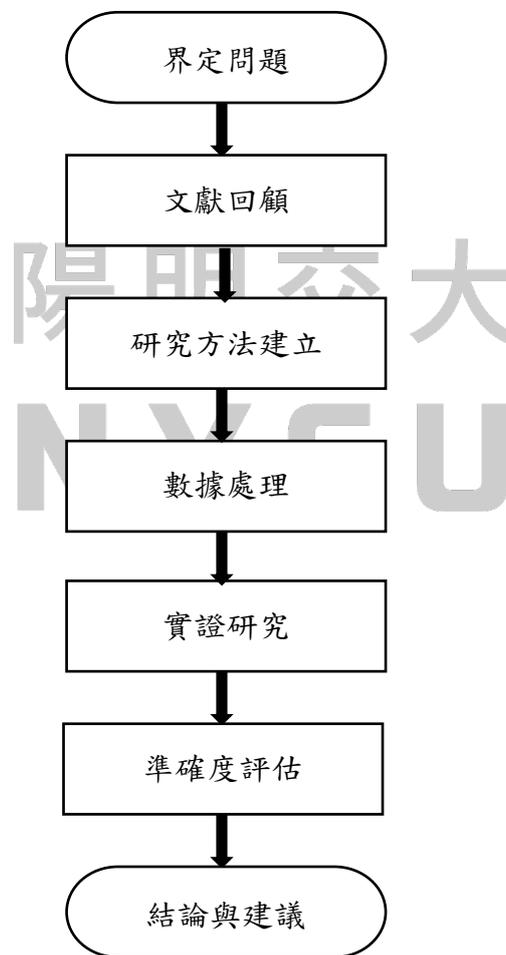


圖 1 研究流程圖

第二章 文獻回顧

旅行時間預測屬於交通狀況預測的一部分，而在過去數十年間，此類預測之研究已被廣泛的討論 (Vlahogianni et al., 2014)。Hinsbergen (2007) 將預測所使用的模式大致分為單純預測法 (Naïve Methods)、參數法 (Parametric Models) 與非參數法 (Non-Parametric Models)。單純預測法像是歷史平均法，以過去相同條件下之資料取平均進行預測，參數法包含時間序列、線性迴歸兩大模式，時間序列係依據過去觀測值與誤差項進行模式建構，線性迴歸則探討自變數與應變數間之線性關係，非參數法則包含 k 最近鄰法、神經網路、支持向量機、決策樹等模式，k 最近鄰法以最相似的 k 個樣本進行預測，神經網路則模仿生物神經網路結構與功能去建構預測模型，支持向量機則以分割線去預測，決策樹則透過樹狀結構產生不同路徑的預測規則。預測模式之分類如圖 2 所示。

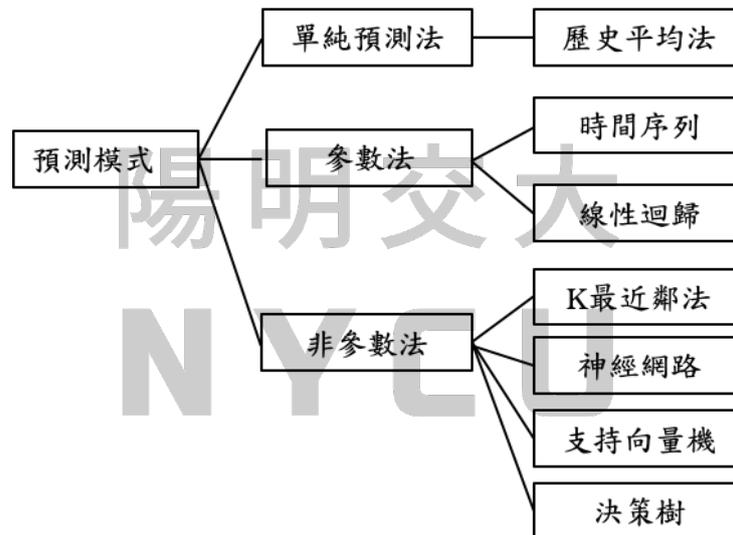


圖 2 常見旅行時間預測模式之分類

過往文獻也針對三類預測方法進行優缺點分析 (如表 1 所示)。單純預測法沒有任何模式假設，準確率較低，儘管易於計算，現今研究大多以參數法與非參數法為主 (Hinsbergen, 2007)，參數法優點為簡單快速，可以很快地從數據中學習，不需要大量數據，並容易解釋模式結果，但模式本身受到限制，僅能考慮特定地點情況，適合應用在較為簡單的問題上 (Habtemichael et al., 2016)，非參數法不對目標函數作出強烈的假設，擁有較大的使用彈性，應用範圍較廣，並善於處理大量數據，但訓練過程較為緩慢 (Jang, 2019)。

表 1 模式分類各自優缺點

分類	優點	缺點
單純預測法	<ul style="list-style-type: none"> 易於計算 	<ul style="list-style-type: none"> 準確率較低
參數法	<ul style="list-style-type: none"> 預測速度快 不需大量數據 	<ul style="list-style-type: none"> 僅能考慮特定地點情況，應用範圍不廣

	<ul style="list-style-type: none"> ▪ 容易解釋結果 	
非參數法	<ul style="list-style-type: none"> ▪ 有較大使用彈性 ▪ 善於處理大量數據 	<ul style="list-style-type: none"> ▪ 訓練過程緩慢

2.1 參數法預測模式

參數法主要包含線性迴歸、時間序列等模式，此類預測模型的結構是固定的，且只需要使用數據即可找出模型的參數 (Hinsbergen, 2007)。Hinsbergen (2007) 指出線性迴歸之假設為預測函數為其共變量 (Covariates) 的線性組合，參數則表示各共變量對預測函數之貢獻，雖然該模型過於簡單，但預測速度非常快，在某些情況下仍有一定效果。Qiao (2013) 指出常見的時間序列預測有自回歸模式 (AR)、移動平均模式 (MA)、結合前兩者的自回歸整合移動平均模式 (ARMA)、平穩化處理的自回歸整合移動平均模式 (ARIMA) 及考量週期性的季節性整合自回歸移動平均模式 (SARIMA)，而又以 ARIMA 模式為時間序列模式中最廣泛被使用的，也常被用於進行交通狀況預測 (黃冠維, 2018)。

Kwon 等人 (2000) 使用 I-880 數據庫 (包含奧克蘭以南 10 公里高速公路線圈資料與探測車往返該路段資料) 當中的車流量、佔有率、速率來預測旅行時間，提出線性迴歸模式 $\tau_i = g(\chi_i) + \varepsilon_i$, $i = 1 \text{ to } n$, χ_i 為跟旅行時間 τ_i 有關之參數，包含 O_i (佔有率)、 F_i (流量)、 S_i (出發時間) 及 W_i (星期幾)，發現此種較為簡單的線性迴歸模式對於短期旅行時間預測有效，也僅適用於較短的預測範圍 (距離)，更長的距離上，需更複雜的模式來預測。

Billings 等人 (2006) 以時間序列模式研究幹道短時間的旅行時間預測，將明尼蘇達 194 號州際公路中最壅擠之路段設定為研究場域，利用 GPS 探針車收集旅行時間數據，由於旅行時間數據的非平穩性，使用 ARIMA 模式進行預測，並利用殘差分析與卡方檢定 (portmanteau test) 進行模式驗證，結果表明其預測具有合理準確性。

Alghamdi 等人 (2019) 蒐集佈設在加利福尼亞州主要都會區高速公路傳感器 2017 年 7 月至 10 月的車流數據，並對數據做一些預處理，用 Kolmogorov-Smirnov、Anderson-Darling 等方法檢驗資料是否為常態分佈，若否，則用 KPSS 測試將數據進行轉換，並將 70% 作為訓練資料，30% 作為測試資料，利用 ARIMA 模式進行預測，調整不同參數來獲得最佳準確度，得出平均絕對標度誤差 (MASE) 為 0.1263 的結果。

2.2 非參數法預測模式

非參數法包含 K 最近鄰法 (K-Nearest Neighbor, k-NN)、類神經網路法、支持向量機 (Support Vector Machine, SVM)、決策樹等模式。非參數法不代表模式之中沒有參數，而是參數的設定上是彈性的，與模型結構都必須透過數據來決定，較參數法而言更需要大量數據 (Hinsbergen, 2007)。K 最近鄰法需建立歷史

數據庫，並從中找出接近當前交通狀況的 K 個事件進行預測，類神經網路法則可描述非線性和動態過程，並衍生許多不同的改良模型 (Hinsbergen, 2007)，支持向量機則將數據映射至更高維度的空間，構建一個具有最大餘量 (Margin) 的超平面 (Separating Hyperplane)，Wu (2004) 將此概念應用於旅行時間預測問題上，決策樹則為一訓練快速卻不穩定之算法，而近年來學者發展集成學習演算法，以多個決策樹來提高預測能力 (林崇翔, 2018)。

蔡繼光 (2009) 以 k -NN 法預測高速公路旅行時間，蒐集國道三號樹林至後龍收費站之 ETC 與 VD 資料，以星期別進行分群，流量、速率作為特徵向量，由於速率與流量每單位對旅行時間的影響不同，因而設立權重，研究希望於尖峰時段仍具有準確性，故參考 Clark (2003) 用尖峰時的平均速率與平均流量進行標準化後再進行比對，此外採用多時間點 k -NN 法，分析在一段時間的推移下旅行時間的改變 (預測上午九時的旅行時間，則以今日上午八時三十分至九時的即時資訊，與前後兩小時之歷史資料進行比對)，研究隨機從民國 97 年 3 月抽取六天的資料進行實測分析，得出誤差為 7.16% 之結果。

Habtemichael 等人 (2016) 提出 k -NN 法預測未來 15 至 90 分鐘的高速公路流量，以日期別進行分群，透過找出 k 個與歷史資料最相似的特性進行預測。研究使用 Guo (2014) 的車流量數據 (包含英國高速公路 12 組 VD 資料集與美國 24 組 VD 資料集)，由於數據於不同時間點高低起伏很大，先利用局部加權迴歸散點平滑法 (Locally Estimated Scatterplot Smoothing) 找出特性將原始資料降噪，針對極端值的處理也不予以刪除，使用 Winsorization 方法將其保留並降低影響， k -NN 法中也需決定一些變數，讓誤差最小化，對於相似性部分使用加權歐式距離 (Weighted Euclidean Distances) 計算，也採用 Rank-based Weight 的方法，使得加權歐式距離越遠，權重越低，並以過去一小時的資料作為滯後時間 (Lag duration)， K 值決定為 10，參考 Guo (2014) 以服務水準和一天中的時間進行預測準確性的評估，與 Guo (2014) 提出數個時間序列模式進行比較，在 MAPE、MAE、RMSE 皆有較為出色之表現。

Ma 等人 (2017) 利用卷積神經網路 (CNN) 進行大範圍運輸路網速率之預測，研究利用裝設有 GPS 定位車輛，來獲取特定時間特定位置的速率，並產製時間、空間二維矩陣，再進一步轉換成時空圖像，丟入 CNN 模式中訓練。研究以北京市二環線與城中心東北方之路網作為實測場域，蒐集配備 GPS 出租車每分鐘所在之位置、行進方向、行使數據等資訊，而該研究認為每分鐘數據波動過大，因此匯整為兩分鐘的時間間隔，並設定四個情境 (不同的預測時階長度、不同的歷史資料時階長度) 以 CNN 模式進行預測，再透過均方誤差 (Mean Squared Error, MSE) 來驗證模式的有效性，最後與其他方法 (OLS、KNN、RF、ANN、SAE、RNN、LSTM、NN) 進行比較，在 MSE、準確度方面於兩路網、四個不同情境之下皆有最佳之結果，而模式的訓練時間也在可接受合理範圍內。

陳志華等人(2017)運用卷積神經網路(CNN)進行旅行時間預測,以國道五號與其替代道路(台2線、台2丙線、台9線)作為實驗場域,於瑞濱、福隆、暖暖、平溪等地架設eTag偵測器,蒐集車輛的旅行時間,共蒐集2016年10至12月資料,前兩個月作為訓練資料,後一個月作為測試資料,再以平均絕對百分比差(Mean Absolute Percentage Error, MAPE)計算誤差值,得出MAPE值為19%之結果,並優於線性迴歸、神經網路、整體神經網路模式。

黃冠維(2018)利用卷積神經網路(CNN)預測號誌化道路未來5至30分鐘之速率,蒐集車輛偵測器(VD)資料,並選取VD編號、日期與時間、對應之平均速率欄位,轉換成輸入CNN模式之時空矩陣圖,並利用佈設於桃園市號誌化道路之VD進行實測研究,計算2016年各月份遺失值比例,取最低的三個月作為實測資料,70%作為訓練資料,30%作為測試資料。研究分析不同的實測情境,考量不同的預測範圍(VD數目差異)、不同空間範圍(VD距離遠近差異)、不同的歷史資料時階長度對預測之影響、不同的預測時階長度(單時階預測、多時階預測),發現CNN模式穩定且預測能力優良。

林崇翔(2018)利用梯度上昇回歸樹演算法(Gradient Boosting Regression Trees, GBRT)預測高速公路旅行時間與壅塞排除時間,結合boosting集成學習框架與決策樹模型的優點來提高預測能力,研究蒐集高公局公開的資料庫,包含ETC門架資料與車輛偵測器(VD)資料,透過調整GBRT參數提高準確率,再將預測結果與類神經網路(NN)進行比較,發現該研究訓練出來的GBRT模式於旅行時間預測與壅塞解除時間預測上皆有較低誤差。

Long等人(2019)蒐集中國湖南省G5513高速公路收費站2018年2月的收費數據、氣象監測站天氣數據、高速公路壅塞報告及交通管制措施報告,使用支持向量機(SVM)提出基於以上數據庫的旅行時間預測模式,為了提高SVM的準確度,應用人工魚群演算法優化模式內之參數,最後將優化後的SVM模式(OSVM)與一般SVM模式、後向傳播神經網路(BPNN)模式進行比較,結果表明OSVM準確度分別較其他兩種模式提升了16.44%與17.27%。

Jang(2019)回顧數個常見旅行時間預測方法(卡爾曼濾波、人工神經網路、時間序列、K最近鄰法)並進行優缺點分析(如表2),發現k-NN較其他方法更具普遍性(Universality),可應用於較多情境,並可收集新的資料來訓練,不斷的更新,因此全世界越來越多人應用k-NN進行旅行時間預測。

表 2 常見預測方法優缺點比較

方法	優點	缺點
卡爾曼濾波	不需要大量的歷史數據	無法預測多時階旅行時間
人工神經網路/時間序列	具有理論基礎/準確	不考慮旅行時間模式的變化
K最近鄰法	可應用於較多情境	需要較長的計算時間

研究也提到多數 k-NN 文獻皆使用歐幾里德距離 (Euclidean distance) 作為距離計算公式，但歐幾里德距離只考慮當前值與歷史值接近程度，無法考慮到趨勢走向，因此使用相關係數 (Correlation coefficient) 來計算距離，並以迴歸方程 (Regression equation) 來調整當前值與歷史值之差異。最後再以韓國平澤市號誌化道路作為實測場域，收集探針車數據驗證其距離計算方式使誤差下降 3.7 個百分點。

與參數法模式相比，非參數法模式的表現往往較好，更適合用於大量與複雜的數據，能於不同情境下進行預測 (Habtemichael et al., 2016)，而近年隨著機器學習的發展與計算能力的提升，相關預測已將重心由參數法轉移至非參數法 (黃冠維, 2018)。本研究將於下一小節分別對非參數法中最常被使用的 k-NN、CNN 進行回顧。

2.3 k-NN 模式

k 最近鄰法核心思想為如果一個樣本 X_1 在特徵空間中最接近的 K 個樣本多數屬於某一類別，則判斷該樣本 X_1 也屬於同一類別，以圖 3 為例，若要對藍色星號進行分類，在 $K=1$ 情況下最近的黃色正方形即為分類結果，而 $K=3$ 情況下，最近的三個樣本分別為兩個紅色三角形與一個黃色正方形，取多數決結果則將藍色星號判斷為紅色三角形。除預測類別型資料問題外，同樣可用於預測連續型數值，藉由找出某一樣本 X_1 最接近 K 個樣本，將這些樣本屬性的平均值給予樣本 X_1 ，即可得到預測值，以圖 4 為例，若要預測藍色星號數值， $K=1$ 情況下則為最近樣本數值 6， $K=3$ 情況下則將最近三個樣本數值取平均，即 $(6+5+4)/3=5$ 。



圖 3 k-NN 法用於類別型問題

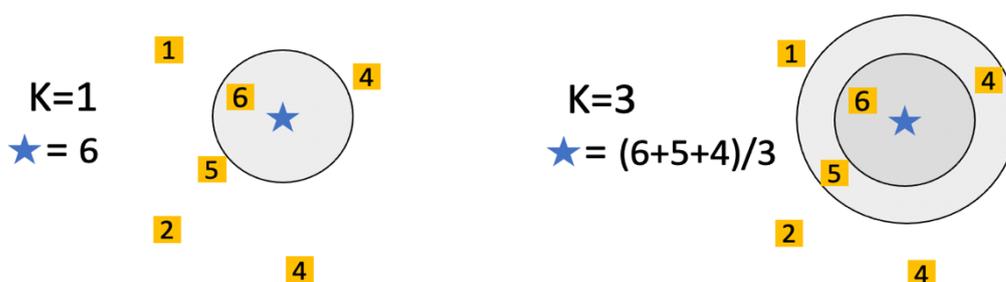


圖 4 k-NN 法用於連續型問題

於交通預測問題上則假設在相似的歷史交通狀況下，會有相似的情況於未來的時間點發生，流程如圖 5 所示，需搜集相關資訊，包含所比對的特徵向量與要預測的旅行時間，建立歷史資料庫，當我們要預測新的一筆資料時，即將該筆資料與歷史資料庫進行比對，找出歷史資料庫中最相近的 k 筆資料，並用這 k 筆資料進行預測。

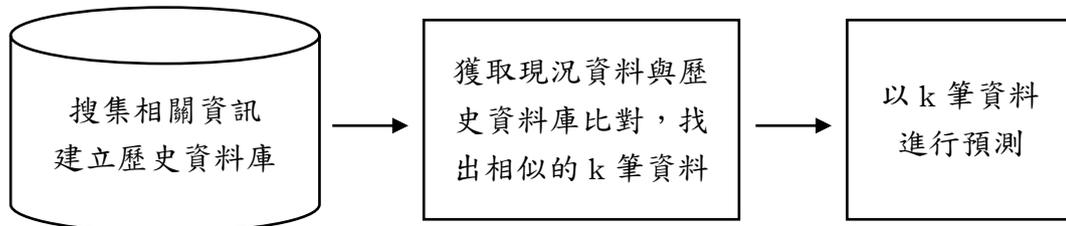


圖 5 k-NN 法應用於交通預測問題

而建構一個 k-NN 模式需要進行參數設定，Robinson 與 Polak (2005) 提出建立一 k-NN 模式需要考慮以下四個重要參數：

1. 模式中所使用的特徵向量

特徵向量的選取對於準確率有著相當程度之影響，如果使用的特徵向量太少，模式可能無法準確區分現況與歷史資料相似情況，使用的特徵向量如果太多，可能包含許多無關特徵向量，同樣會降低預測準確度。

2. 距離量測方式

k-NN 被稱作 k 最近鄰法，即為所預測的現況資料與歷史資料庫相「近」程度，而我們必須定義何謂「近」，會利用各種距離測量方式來衡量相似性，找出預測目標與歷史資料之距離，距離越大代表越不相似，越小則代表相似性越高，因而找出距離最小的 k 筆資料。常見的距離量測方式包含歐式距離 (Euclidean Distance)、曼哈頓距離 (Manhattan Distance)、切比雪夫距離 (Chebyshev Distance)。

3. 如何估計預測值

找出 k 個最相似樣本與其對應之數值後，即可對此作出預測。在預測連續型數值問題上，平均法常被拿來使用，即將 k 樣本中數值取平均作為預測結果，也有考量權重的加權平均法，讓距離越近的樣本則給予較高權重。

4. k 值

k 值的大小也同樣對準確率有著相當程度影響，k 值太小會導致模式過於敏感，容易受極端值影響，而 k 值太大容易使模式過於簡單，無預測意義。

然而，隨著 k-NN 發展日益成熟，但其模式仍有其侷限性，Wu 等人 (2010) 指出 k-NN 有三個主要缺陷：

1. 模式準確率與 k 值高度相關與敏感

Wu 等人 (2010) 指出模式準確度與 k 值之間關係十分敏感，有一種做法即為嘗試不同 k 值，並選擇一最佳 k 值。劉軒寧 (2016) 提到目前研究亦無針對 k 值應如何設定進行探討，因此提出改良式 k-NN，即 k 值不再維持單一固定數值，並以變動 k 值之 k-NN 法預測高速公路旅行時間，先以一般 k-NN 法得到初始 k 值，依據車流狀態調整其值，在 k 值設定上更符合當下情況，並於高速公路尖峰時段進行實測，與固定 k 值之 k-NN 法進行比較，發現其提出的方法具有較高的準確度。

2. 標準歐式距離 (Standardized Euclidean Distance) 有時不為最佳量測方法

Wu 等人 (2010) 指出 k-NN 準確性可能會受到不相關屬性的影響，但許多研究所使用的標準歐式距離是參考所有屬性進行計算的，而當無關屬性增加時，則會使預測不準確。Clark (2003) 以流量、速率、佔有率來預測即時交通資訊，觀察數據集的分布狀況，並假設十分鐘平均流量為 1000 輛、速率為 100 公里/小時、佔有率為 15%，依此給予不同特徵向量權重，即佔有率差 1% 之距離相當於速率差 6.67 公里/小時之距離及流量差 66.67 輛之距離。

3. 需建立更準確地預測方式取代簡單投票法

Wu 等人 (2010) 指出以簡單投票法進行預測會使結果不準確，並使用 Inverse distance weights 法，算出 k 個樣本各自與預測目標的距離，給予距離較近的樣本較高的權重，詳細權重計算方式如表 3 所示。Habtemichael 等人 (2016) 採用 Rank-based weights 法，針對 k 個樣本各自對於預測目標的影響進行改善，依據其相似性進行排名，排名越前面 (數字越小) 即代表越相似，給予較高的權重 (對預測值影響較大)，詳細權重計算方式如表 3 所示。

表 3 常見樣本權重計算方式

作者/年份	Habtemichael (2016)	Wu (2010)
方法	Rank-based weights	Inverse distance weights
計算方式	$W_i = \frac{(K - r_i + 1)^2}{\sum_{j=1}^K (K - r_j + 1)^2}$	$W_r = \frac{1}{d(X_r, X_q)^2 + 0.001}$
備註	W_i 為權重， r_i 為排名	W_i 為權重， $d(X_r, X_q)^2$ 為與預測目標之距離

2.4 CNN 模式

CNN 全名為 (Convolutional Neural Network)，被稱為卷積神經網路，起源最早可追溯至 1962 年，神經生理學家 Hubel 和 Wiesel 對於貓咪大腦中視覺系統之研究，提出了兩種型態的細胞 (simple cells 和 complex cells)，而受到上述之啟發，1980 年日本電腦科學家 Kunihiko Fukushima 提出 Neocognitron (神經感知機)，為現今 CNN 架構之源頭，而 1989 年被稱作卷積網路之父的法裔電腦科學家 Yann Lecun 提出了 CNN 的首個框架 LeNet-5，驗證了 CNN 可行性，而往後數十年 Yann Lecun 與眾多科學家們持續致力於 CNN 之研究，如今 CNN 已廣泛應用於影像辨識、自動駕駛、資訊安全等研究。

CNN 主要架構分為卷積層 (Convolution layer)、池化層 (Pooling layer) 及全連接層 (Fully Connected layer)，首先針對上述結構 (如圖 6) 分別進行說明。

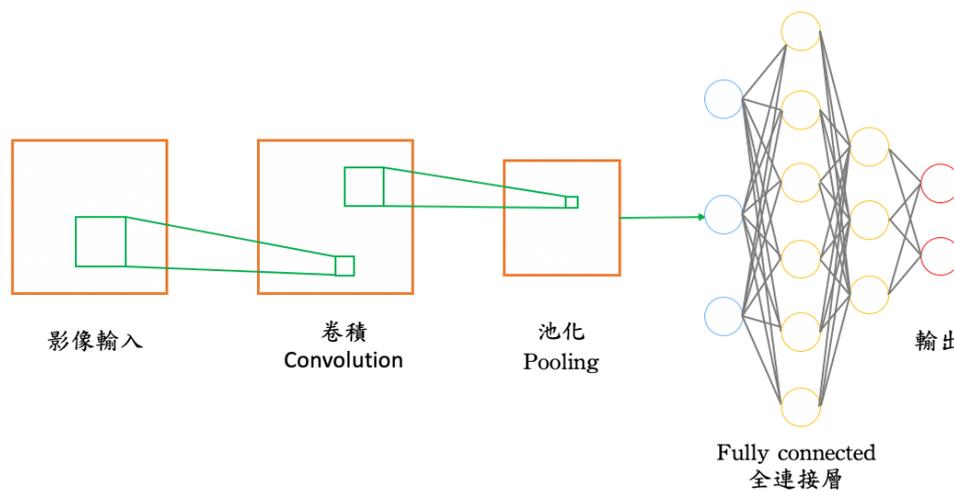


圖 6 CNN 模式架構

2.4.1 卷積層 (Convolution Layer)

卷積之目的為對圖片做出擷取特徵的動作，藉由兩步驟滑動加上內積組成之運算，利用 kernel 卷積核 (或稱之 filter 過濾器) 在輸入的圖片上滑動並持續進行矩陣內積，得到的圖片被稱之為 feature map，而卷積過後圖片大小會根據設定的 kernel 大小 (K)、移動步長 (S) 與是否有填充 (Padding) 有關，輸出大小如下式 1 所示。

$$\frac{(W-K+2P)}{S}+1 \quad \text{式 1}$$

圖 7 與圖 8 為卷積運算之過程，假設有一 5x5 之圖像輸入，並有一 3x3 之卷積核 (紅底處)，移動步長為 1，無任何填充 (Padding) 之情況，會依序將卷積核進行移動並得出 feature map。

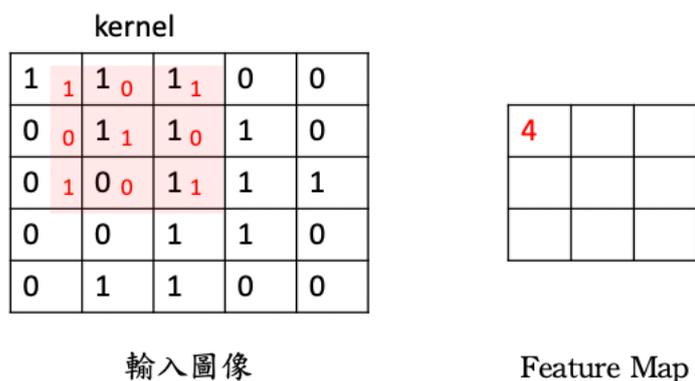


圖 7 卷積運算 (一)

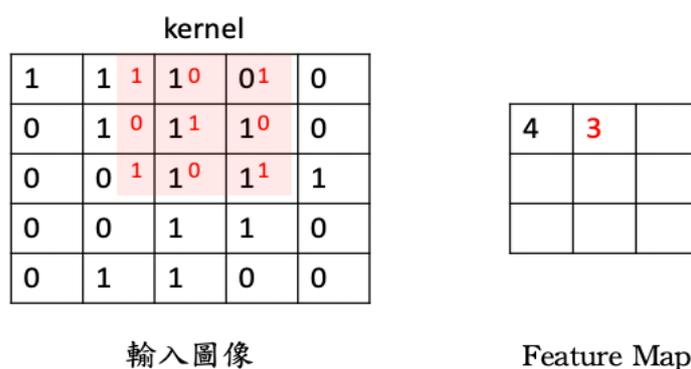


圖 8 卷積運算 (二)

2.4.2 池化層 (Pooling Layer)

池化層則對卷積產生之 Feature Map 做池化之動作，得到降維特徵圖，常見的方法有 Max pooling 及 Average pooling，池化層之目的為減少圖片尺寸，參數減少，防止 Overfitting，讓計算成本隨之下降，同時具有特徵不變的特性，研究假設有一個 2x2 的 kernel，並且移動步長為 2，使用 Max pooling 進行池化運算 (如圖 9)。

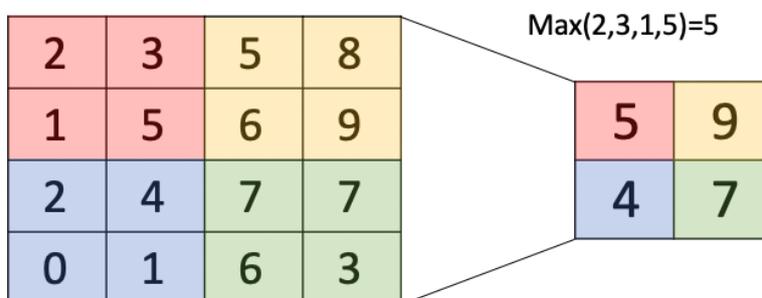


圖 9 池化運算

2.4.3 全連接層 (Fully-connected Layer)

經由池化運算後，會進行攤平 (Flatten)，形成一維向量後，即可輸入全連接層，全連接層即是一般的神經網路，由輸入層 (Input Layer)、隱藏層 (Hidden Layer)、輸出層 (Output Layer) 組成，輸入層表示輸入之變數，隱藏層負責變數之間的交互影響，輸出層則輸出變數。

2.5 小結

本研究依不同的預測方法進行回顧，並針對預測所使用的資料來源、考慮的特徵向量及欲預測之目標進一步探討，並統整如表 4。由表 4 可得知過往研究較常使用 VD、探針車資料，來預測速率、流量、佔有率、旅行時間之資訊，且超過一半比例之文獻只使用單一特徵向量進行預測 (如 Billings, 2006; Habtemichael, 2016; 劉軒寧, 2016; 陳志華, 2017; Ma, 2017; 黃冠維, 2018; Jang, 2019; Long, 2019; Alghamdi, 2019)，而其餘文獻僅考慮部分特徵向量，如蔡繼光 (2009) 以速率與流量進行預測，Kwon (2003)、Clark (2003)、林崇翔 (2018) 以速率、流量、佔有率來預測。

表 4 文獻回顧重點整理

作者/年份	方法	資料來源	考慮特徵向量				預測目標
			速率	流量	佔有率	旅行時間	
Kwon (2000)	線性迴歸	VD 探針車	○	○	○		旅行時間
Clark (2003)	k-NN	數據集	○	○	○		速度 流量 佔有率
Billings (2006)	ARIMA	探針車				○	旅行時間
蔡繼光 (2009)	k-NN	ETC VD	○	○			旅行時間
Habtemichael (2016)	k-NN	VD		○			流量
劉軒寧 (2016)	k-NN	VD	○				旅行時間
陳志華 (2017)	CNN	ETC VD				○	旅行時間
Ma (2017)	CNN	探針車	○				速率
黃冠維 (2018)	CNN	VD	○				速率
林崇翔 (2018)	GBRT	ETC /VD	○	○	○		旅行時間
Jang (2019)	k-NN	探針車				○	旅行時間

Long (2019)	SVM	ETC				O	旅行時間
Alghamdi (2019)	ARIMA	VD		O			流量

Clark (2003) 也發現使用多元特徵向量有助於提高預測準確性，因此本研究期盼蒐集比過往研究更加多元資訊(如表 5)。其中包含文獻回顧中常見的速率、流量、佔有率資訊，此外，本研究也做了以下兩點推論：

1. 旅行時間可能與重車比有關

若駕駛人行駛於重車比高的環境中，可能會因鄰近重車產生的壓迫感，使其更為謹慎地駕駛，或因重車之高度影響視距，減少變換車道、超車等行為，進而影響整體旅行時間，因此將重車比納入，作為模式所使用的特徵向量之一。

2. 旅行時間可能與趨勢有關

假設兩樣本於某時間點速率皆為 60 公里/小時，但其中一樣本於該時間點之五分鐘前速率為 80 公里/小時，另一樣本為 20 公里/小時，兩樣本 60 公里/小時所代表之意義將有所不同，前者在未來時間點速率可能會下降，逐漸造成壅塞，後者在未來時間點速率可能會上升，車流逐漸順暢。因此本研究認為趨勢性可能為影響旅行時間重要因素，將 5 分鐘、10 分鐘、15 分鐘、20 分鐘速率、流量變化納入，同樣作為模式所使用的特徵向量。

表 5 特徵向量選取說明

模式所使用之特徵向量	納入原因
速率	過往文獻使用
流量	
佔有率	
重車比	本研究所提出之推論
5 分鐘速率變化	
5 分鐘流量變化	
10 分鐘速率變化	
10 分鐘流量變化	
15 分鐘速率變化	
15 分鐘流量變化	
20 分鐘速率變化	
20 分鐘流量變化	

第三章 研究方法

本研究於上一章節中回顧旅行時間預測常用模式。過往研究已指出對於大量數據而言，非參數法較參數法有著更為良好的預測準確度（Habtemichael et al., 2016），且近年來趨勢大多以非參數法為主（黃冠維, 2018），非參數法中又以 k-NN、CNN 兩種方法最被廣泛使用，兩種方法間，k-NN 相較於 CNN 在資料預處理上更具優勢，面對資料缺失、異常等情況不做過多的假設與處理，能保留原始資料最真實之狀態（陳建旻, 2008），而本研究所使用的 VD 資料常因車輛偵測器本身通訊問題、天候狀況造成資料遺失或是數值異常（邱逸彥, 2014），恰好為缺失、異常率比例高的一種資料，其適用性較 CNN 為佳。朱志杰（2013）已證明 k-NN 在足夠歷史資料與使用不同特徵向量之下進行分析能提高預測準確性，此外 k-NN 還有應用廣泛（Jang, 2019）、已被證明能夠用於旅行時間預測等特點（劉軒寧, 2016）。因此本研究將以 k-NN 作為基礎進行預測，並加以改良。

本章節主要可分為六個小節，分別說明如下，第一小節說明本研究之架構，第二至第五小節則分別針對研究架構中各步驟進行說明，第六小節則設計一簡化範例進行模式預測說明。

3.1 研究架構

本研究之架構如圖 10 所示，其說明如下：

1. 步驟一：資料蒐集

資料蒐集來源有二種，一為 VD 資料，用以獲得流量、佔有率及速率資訊，二為 ETC 資料，用以獲得各門架間旅行時間資訊。

2. 步驟二：資料處理

接著針對步驟一所蒐集到之資料分別進行處理。

(1) VD 資料

VD 資料首先須對缺漏、異常值進行處理，由於原始 VD 資料將資訊以車道進行細分，需進行車道欄位合併，合併速率、流量、佔有率資訊，並計算出大型車佔比、5、10、15、20 分鐘速率、流量變化趨勢之資訊，作為本研究所使用之 12 種特徵向量。

(2) ETC 資料

ETC 資料同樣須先進行缺漏、異常值處理，而原始 ETC 資料將資訊以車種進行細分，需進行車種欄位合併，且門架多位於交流道至交流道中間，但本研究所預測之目標為交流道間旅行時間，若只採用 ETC

資料來計算交流道間旅行時間會造成誤差，因此本研究加入 VD 速率資料進行填補，再以延遲加總方式得出更為準確之交流道間旅行時間，作為預測目標。

3. 步驟三：歷史資料庫建置

完成上述兩步驟，即可求得欲使用之 12 種特徵向量與預測目標（交流道間旅行時間），可建立模式所需之歷史資料庫，本研究以平日、假日進行分群，將建置二個不同的歷史資料庫，依據現況資料情境分別以不同歷史資料庫進行預測。

4. 步驟四：模式預測

建置完歷史資料庫後，須將歷史資料庫內資料進行標準化，以解決單位不同所造成之偏差，接著進行 k-NN 模式的參數設定，參數設置完成後即可將同樣標準化後之現況資料帶入歷史資料庫進行比對，並輸出預測值。最後則將輸出之預測值與真實值進行誤差比對，利用常見之誤差指標進行成效分析，判斷模式之可行性。

依照上述之步驟，繪製研究架構圖。

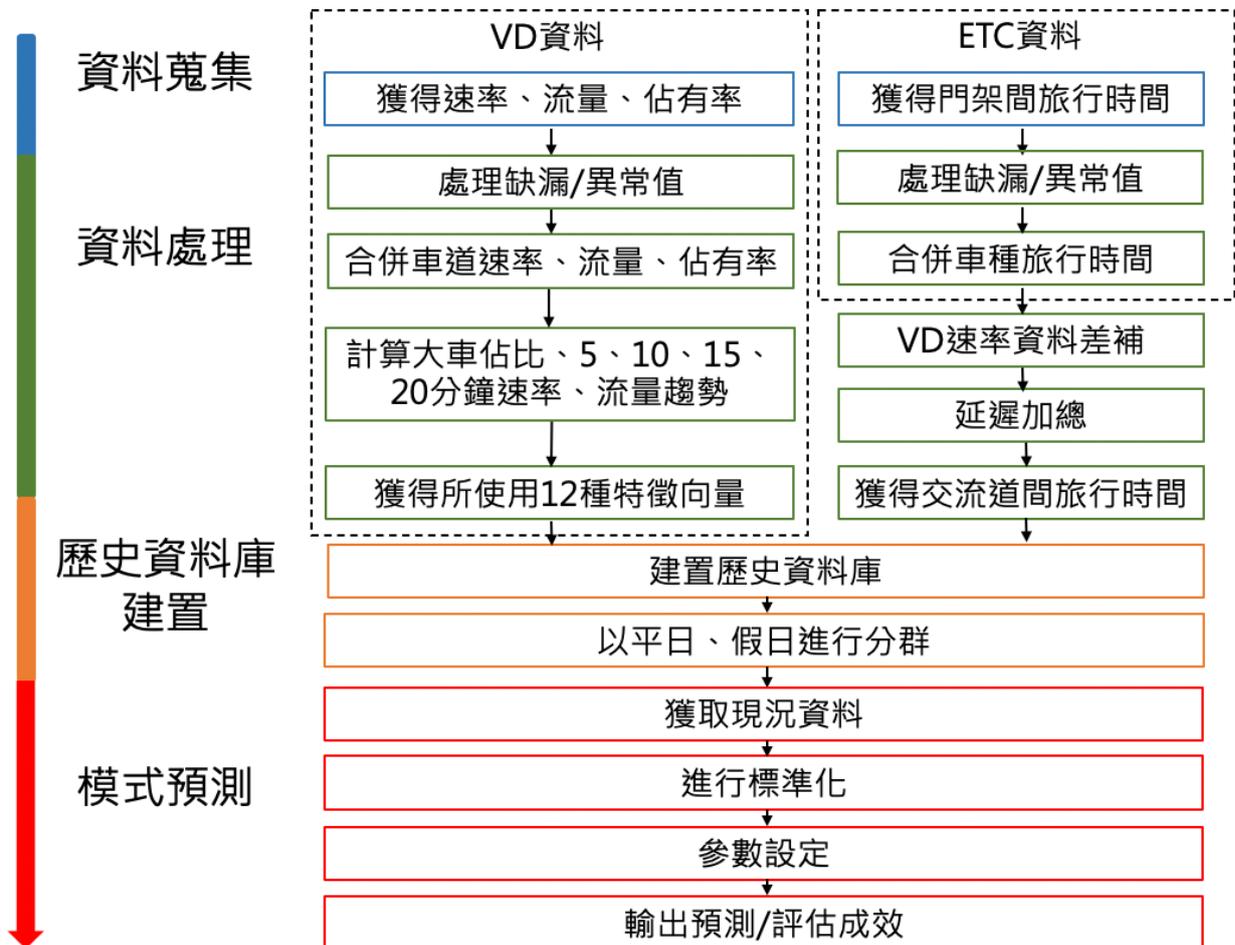


圖 10 研究架構圖

3.2 資料蒐集

本研究所使用的速率、流量、佔有率資訊之來源為車輛偵測器 (VD) 資料，而旅行時間資訊來源為電子收費門架資料 (ETC) 資料，皆由交通部高速公路局所提供，並於交通部數據匯流平台所申請，每五分鐘會產生一筆統計資料。

車輛偵測器資料所蒐集之資訊會依照車道進行細分，因此某偵測器所在位置若有三個車道，每五分鐘便會統計該車輛偵測器三車道共三筆資料，本研究蒐集包含各車輛偵測器代碼、車道代碼、平均速率、統計修正後小車流量、統計修正後連結車流量、統計修正後大車流量、統計修正後總流量、佔有率及資料蒐集時間欄位 (如表 6 所示)。

表 6 單一 VD 每五分鐘一筆原始資料

車輛偵測器設備代碼	車道代碼	平均速率偵測值 (kph)	統計修正後小車流量	統計修正後連結車流量	統計修正後大車流量	統計修正後總流量	佔有率 (單位%)	資料蒐集時間
VD1	1	111.96	24	0	3	27	1.6	2020/8/2 00:05
VD1	2	96.88	7	8	3	18	2.5	2020/8/2 00:05
VD1	0	121.38	5	0	11	16	1.2	2020/8/2 00:05
VD1	0	122.76	4	0	13	17	1.2	2020/8/2 00:10
VD1	2	96.36	7	3	2	13	1.75	2020/8/2 00:10
VD1	1	110.86	27	0	1	28	1.6	2020/8/2 00:10

ETC 資料所蒐集之資訊會依照不同車種進行細分，所記錄的車種包含 (31 小客車、32 小貨車、41 大客車、42 大貨車、5 連結車)，因此兩門架之間每五分鐘會紀錄上述五車種所經過的平均旅行時間共五筆資料，本研究同樣蒐集包含上游偵測站編號、下游偵測站編號、車種、交通量、平均旅行時間 (秒) 及資料代表時間欄位 (如表 7 所示)。

表 7 ETC 單一區間每五分鐘一筆原始資料

上游偵測器編號	下游偵測器編號	車種	交通量	平均旅行時間 (秒)	資料代表時間
ETC01	ETC02	31	62	143	2020-08-02 00:00
ETC01	ETC02	32	11	148	2020-08-02 00:00
ETC01	ETC02	41	4	155	2020-08-02 00:00
ETC01	ETC02	42	2	161	2020-08-02 00:00
ETC01	ETC02	5	4	176	2020-08-02 00:00

ETC01	ETC02	31	51	142	2020-08-02 00:05
ETC01	ETC02	32	14	136	2020-08-02 00:05
ETC01	ETC02	41	2	154	2020-08-02 00:05
ETC01	ETC02	42	0	0	2020-08-02 00:05
ETC01	ETC02	5	0	0	2020-08-02 00:05

3.3 資料處理

本小節將針對 3.2 小節所蒐集之原始資料後，其處理過程進行說明。

3.3.1 車輛偵測器資料

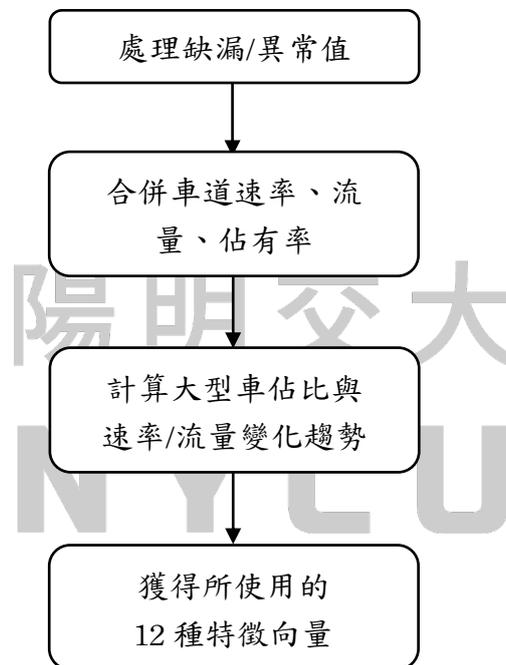


圖 11 車輛偵測器資料處理流程

車輛偵測器資料處理流程如圖 11 所示，首先針對各 VD 缺漏與異常情況進行整理，本研究對於 VD 資料缺漏之定義為於資料蒐集之步驟時，即無獲得任何資料（即本身為空值），而 VD 資料異常之定義為有資料，但為異常之情況，本研究對於 VD 異常情況之判斷標準為速率、流量、佔有率為負值或連續兩筆五分鐘資料為零（參考交通部運研所對 VD 動態資訊連續相同資料不合理判斷條件之極限為 6 分鐘，換算成五分鐘一筆資料形式可為一筆或兩筆，本研究採取較為寬鬆但仍有依據之標準，故以兩筆為限）。此外，王晉元（2013）提到資料遺漏率在 20% 時，經過濾差補後精準度可達 90%，遺漏比例在 20% 以上時，不適合預測旅行時間，因此本研究在整理缺漏與異常值階段設立一門檻值 20%，若一 VD 缺漏或異常情況達 20%，即將該 VD 刪除。此外 k-NN 一大優勢為在資料缺漏或異常情況下，不必做進一步處理（如以歷史平均法填補缺漏值）即可進行預測，因此本研究對於異常值處理方式為將該欄位原先異常之數值以空值取代。

接著需將車道欄位進行合併，以求得該車輛偵測器於該時間點之資訊。需要合併之欄位為速率、流量及佔有率。速率為計算所有車輛通過之總體平均速率(如式 2 所示)，流量則由三個車道總流量相加(如式 3 所示)，佔有率則為三個車道取平均(如式 4 所示)。本研究以表 5 進行說明，車輛偵測器 VD1 在 2020/8/2 00:05 為例，合併後速率 s_m 為第一個車道之速率 111.96kph 乘以第一個車道總流量 27，加上第二個車道之速率 96.88kph 乘以第二個車道總流量 18，加上第三個車道之速率 121.38kph 乘以第三個車道總流量 16，加總後再除以三車道流量加總 61，得出 s_m 為 109.98kph，合併後流量 f_m 為第一個車道之流量 27，加上第二個車道之流量 18，再加上第三個車道之流量 16，得出 f_m 為 61，合併後佔有率 o_m 為第一個車道佔有率 1.6，加上第二個車道佔有率 2.5，加上第三個車道佔有率 1.2，加總後除以總車道數 3，得出 o_m 為 1.767。

$$s_m = \sum_{l=1}^R (s_L \cdot f_L) / f_m \quad \text{式 2}$$

$$f_m = \sum_{l=1}^R f_L \quad \text{式 3}$$

$$o_m = \frac{\sum_{l=1}^R o_L}{R} \quad \text{式 4}$$

s_m ：合併後速率

f_m ：合併後流量

o_m ：合併後佔有率

R：該 VD 之車道數

s_l ：第 L 個車道之速率

f_l ：第 L 個車道之流量

o_l ：第 L 個車道之佔有率

陽明交大
NYCU

完成車道欄位合併後，即可依據合併後流量、速率計算大型車佔比與五至二十分鐘速率、流量變化趨勢。本研究定義的大型車為聯結車與大車兩車種，大型車佔比即為兩車種(聯結車、大車)流量總和佔總流量之比例，五至二十分鐘速率與流量變化計算皆須與前後幾個時間點所蒐集之資料進行比較，本研究以表 8 進行說明，車輛偵測器 VD1 在 2020/8/2 00:20 的大型車佔比為聯結車流量 4 與大車流量 10 之總和佔總流量 57 之比例，即為 24.56%，五分鐘速率變化為 00:20 之速率 118.9 減去 00:15 之速率 110.4，為 8.5 公里/小時(代表過去這五分鐘行駛速率提高了 8.5 公里/小時)，五分鐘流量變化為 00:20 之流量 57 減去 00:15 之流量 64，為負 7(代表過去這五分鐘流量減少了 7)，十分鐘速率變化則為 00:20 之速率 118.9 減去 00:10 之速率 108.1，為 10.8 公里/小時，十分鐘流量變化為 00:20 之流量 57 減去 00:10 之流量 76，為負 19，十五分鐘速率、流量變化，二十分鐘速率、流量變化同樣依此概念計算。

表 8 合併車道後單一車輛偵測器資料

車輛偵測器設備代碼	車道代碼	平均速率偵測值 (kph)	統計修正後小車流量	統計修正後連結車流量	統計修正後大車流量	統計修正後總流量	佔有率 (單位%)	資料蒐集時間
VD1	合併	109.98	36	8	17	61	1.767	2020/8/2 00:05
VD1	合併	108.1	57	5	14	76	2	2020/8/2 00:10
VD1	合併	110.4	52	3	9	64	1.5	2020/8/2 00:15
VD1	合併	118.9	43	4	10	57	1.25	2020/8/2 00:20

由於 Clark (2003) 指出使用多元特徵向量有助於提高預測準確度，加上文獻回顧過程中發現大多研究僅使用單一或少數特徵向量進行預測，因此本研究以過往研究常使用的速率、流量特徵向量為基礎，再多考慮了佔有率、大型車佔比、以及變化趨勢作為特徵向量，相較過往研究蒐集之資訊，本研究在此方面著墨較多，蒐集速率、流量、佔有率資料，並藉由資料處理步驟，得到每個 VD 的 12 種資訊（速率、流量、佔有率、大型車佔比、五分鐘速率變化、五分鐘流量變化、十分鐘速率變化、十分鐘流量變化、十五分鐘速率變化、十五分鐘流量變化、二十分鐘速率變化、二十分鐘流量變化），作為本研究所使用的特徵向量。

3.3.2 ETC 資料

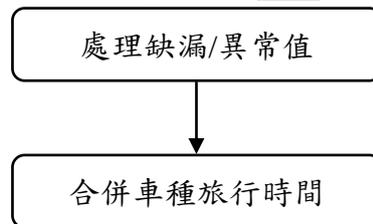


圖 12 ETC 資料處理流程

ETC 資料處理流程如圖 12 所示，首先對各門架之間旅行時間資訊（ETC 資料）進行缺漏值與異常值處理，本研究對於 ETC 資料缺漏之定義為於資料蒐集之步驟時，即無獲得任何資料（即本身為空值），而對於 ETC 資料異常值定義為交通量、平均旅行時間為負值或零，為負值情況本身極不合理，為零之情況則無法於後續計算出平均旅行時間，本研究對此之處理為將異常值設為空值。

而由於 ETC 所蒐集之資訊是依照不同車種進行細分的，因此需將 ETC 資料進行車種欄位合併，本研究目的為預測旅行時間作為一般民眾出行之依據，考量民眾大多以駕駛小客車、小貨車為主，連結車、大貨車、大客車這三種車種多為商業用途，較少為一般民眾所使用，故考慮小型車輛較符合目的，因此需將表 6 「車種」欄位中取出 31、32 資料進行合併，方式如式 5 所示，以表 6 為例，門

架編號 ETC01 行駛至門架編號 ETC02 旅行時間為小客車交通量 62 乘以小客車平均旅行時間 143 加上小貨車交通量 11 乘以小貨車平均旅行時間 148 秒，加總後再除以兩車種交通量 73，以此方式可求得 ETC01 行駛至 ETC02 旅行時間 143.75 秒。

$$tt_{12} = \frac{fc_{12} \cdot ttc_{12} + ft_{12} \cdot tt_{t12}}{fc_{12} + ft_{12}} \quad \text{式 5}$$

tt_{12} : 門架 1 到 ETC 門架 2 旅行時間

fc_{12} : 門架 1 到門架 2 小客交通量

ttc_{12} : 門架 1 到門架 2 小客車平均旅行時間

ft_{12} : 門架 1 到門架 2 小貨車交通量

tt_{t12} : 門架 1 到門架 2 小貨車平均旅行時間

3.3.3 VD 速率資料差補

實務上民眾需透過交流道進出高速公路，以此為考量，預測交流道間旅行時間將更加有用，而 ETC 門架往往與交流道有段距離，不像 VD 能佈設於交流道上，因此需要將這段距離以 VD 速率資料來進行差補，不使用 ETC 資料差補之原因為 ETC 門架間距離較 VD 長，於前後端差補時容易考量到過遠的門架。本研究以圖 13 為例，若我們要預測 T 交流道至 K 交流道旅行時間，需先行找出距離欲預測的兩個交流道最近之 ETC 門架（即 ETC 門架 A 與 ETC 門架 F），蒐集 ETC 門架 A 與 F 所涵蓋範圍之 ETC 資料（即 ETC 門架 A-F 兩兩相鄰門架間旅行時間），而 T 交流道至 ETC 門架 A 與 K 交流道至 ETC 門架 F 這兩段即為需差補之路段。

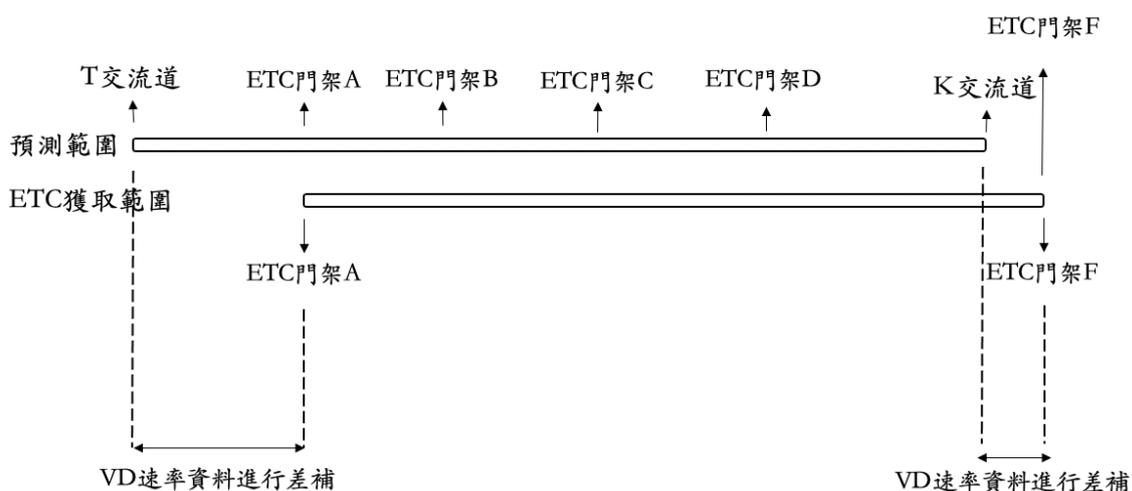


圖 13 以 VD 速率資料進行差補情境

由圖 13 可得知有兩個路段需差補，本研究以其中一路段（T 交流道至 ETC 門架 A）進行差補說明（如圖 14），先將此路段以 VD 分佈進行切割，可分為差補一、差補二、差補三，以兩兩相鄰 VD 之平均速率除以該路段長度來推估其旅行時間，以圖 14 為例，差補一旅行時間為 VD1 平均速率 80km/hr 與 VD2 平均速率 90km/hr 取平均（85km/hr），除以 T 交流道與 VD2 之距離 300 公尺，得出差補一旅行時間約為 12.7 秒，差補二、差補三旅行時間計算方式與差補一相同。

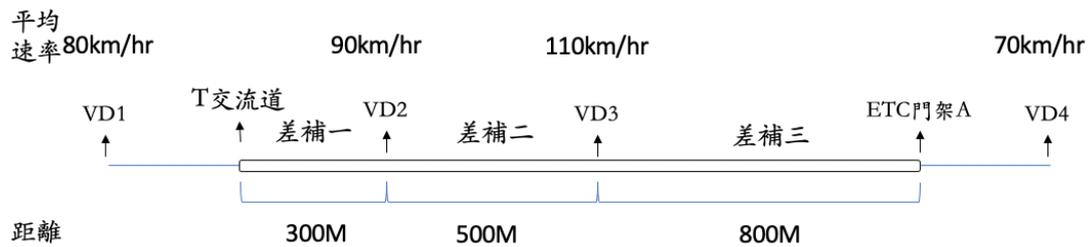


圖 14 以 VD 速率資料進行差補

3.3.4 延遲加總

而各路段旅行時間在時間的推移之下，皆會有所改變，但用路人不太可能於五分鐘內行駛過所有路段，因此我們需要進行延遲加總（如圖 15 所示）。

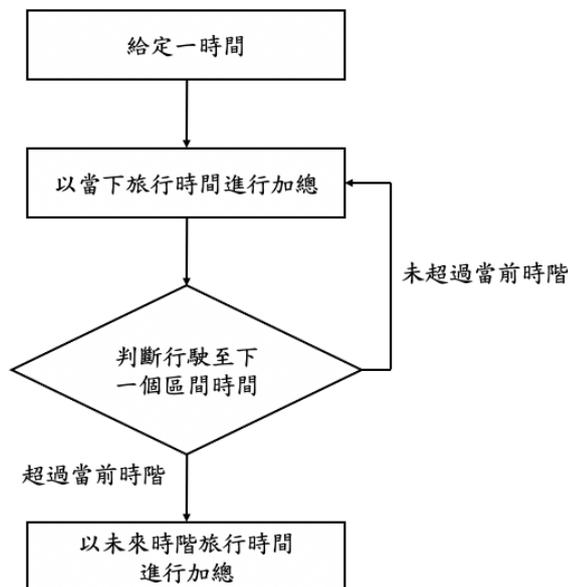


圖 15 延遲加總概念

以圖 13、圖 14 情境進行說明，T 交流道至 K 交流道旅行時間由差補一路段、差補二路段、差補三路段、門架 A 至門架 B、門架 B 至門架 C、門架 C 至門架 D、門架 D 至 K 交流道數個路段所組成，各自行駛時間如表 9 所示。本研究以 00:00 由 T 交流道欲出發至 K 交流道為例，車輛從交流道 T 出發時時間為 00:00，因此差補一路段須參考 00:00 分之資訊，而通過 VD2、VD3、ETC 門架 A 時時間

皆位於 00:00 至 00:05 區間，因此差補二、差補三、門架 A 至門架 B 這三個路段同樣參考 00:00 資訊，而當通過門架 B 時，時間已超過 00:05 但未超過 00:10，因此須門架 B 至門架 C 的旅行時間需參考 00:05 資訊，以此類推計算出總旅行時間（為表 9 灰底處加總），進行完延遲加總後，可獲得兩交流道間旅行時間。

表 9 各路段旅行時間資訊

路段 時段	差補一 旅行時間	差補二 旅行時間	差補三 旅行時間	A to B 旅行時間	B to C 旅行時間	C to D 旅行時間	D to K 旅行時間
00:00	10秒	18秒	35秒	480秒	180秒	720秒	480秒
00:05	18秒	25秒	40秒	600秒	120秒	840秒	420秒
00:10	15秒	21秒	36秒	540秒	60秒	960秒	360秒
00:15	13秒	20秒	32秒	600秒	300秒	780秒	300秒
00:20	12秒	19秒	33秒	360秒	240秒	600秒	480秒
00:25	15秒	19秒	35秒	480秒	180秒	480秒	360秒

3.4 歷史資料庫建置

完成 3.3 小節步驟後，即可建立預測所需要的歷史資料庫，可產出一由時空間資訊所組成的二維矩陣。縱軸為五分鐘回傳的一筆資料，橫軸則為該筆資料所有需要比對的特徵向量（空間範圍內每個 VD 的速率、流量、佔有率、大型車佔比、五分鐘速率變化、五分鐘流量變化、十分鐘速率變化、十分鐘流量變化、十五分鐘速率變化、十五分鐘流量變化、二十分鐘速率變化、二十分鐘流量變化等資訊），以及預測目標（該時間點出發所需之旅行時間）。

本研究將歷史資料庫進行分群，以平、假日做為區隔，建立兩個歷史資料庫，視預測情況不同分別與不同歷史資料庫進行比對分析，如預測目標為平日，則放入平日歷史資料庫。本研究設計一情境進行說明，假設本研究欲預測 6/1 00:30 由 A 至 B 出發的旅行時間，並蒐集過去半小時資料作為歷史資料庫，以速率、流量作為需比對的特徵向量，而 A 至 B 的空間範圍內有三個 VD，建置情況如表 10 所示。

表 10 歷史資料庫

時間	VD1 速率	VD1 流量	VD1 佔有率	VD2 速率	VD2 流量	VD2 佔有率	VD3 速率	VD3 流量	VD3 佔有率	旅行 時間
6/1 00:00	60	85	0.4	70	102	0.4	60	85	0.45	120
6/1 00:05	70	110	0.6	40	100	0.5	70	105	0.55	140
6/1 00:10	45	60	0.35	85	65	0.4	85	35	0.15	95
6/1 00:15	95	50	0.25	95	45	0.2	105	20	0.05	80
6/1 00:20	90	35	0.15	90	30	0.1	100	45	0.2	85
6/1 00:25	100	30	0.1	100	65	0.15	90	30	0.1	70

3.5 模式預測

本小節將說明進入模式前準備工作及模式如何進行預測，及輸出結果後如何評估成效。

3.5.1 標準化

本研究考慮了不同特徵向量，可能會產生單位不同所造成之偏差，為了解決此問題，需進行標準化，轉成純量後再進行預測。本研究在建立完歷史資料庫後，會先以最小值最大值正規化（Min-Max Normalization）方法，將歷史資料庫資料等比例縮放到 0-1 區間，其公式如式 6 所示， X_n 為標準化後新值， X_o 為原始數值， X_{max} 與 X_{min} 分別為該特徵向量最大值與最小值，而後續獲取現況資料後，同樣也需要將現況資料進行標準化。

$$X_n = \frac{X_o - X_{min}}{X_{max} - X_{min}} \quad \text{式 6}$$

3.5.2 獲取現況資料

將歷史資料庫標準化後，即可獲取現況資料，現況資料所需欄位與歷史資料庫大致相同，唯一差別在於預測目標（旅行時間）未知。接續上一小節情境為例，目標為預測 6/1 00:30 由 A 至 B 出發的旅行時間，因此獲取 6/1 00:30 的現況資料（如表 11 所示），再依式 6 將其標準化。

表 11 未標準化前之現況資料

時間	VD1 速率	VD1 流量	VD1 佔有率	VD2 速率	VD2 流量	VD2 佔有率	VD3 速率	VD3 流量	VD3 佔有率
6/1 00:30	80	90	0.2	75	90	0.4	80	80	0.4

3.5.3 模式參數設定

本研究以 Robinson 與 Polak（2005）提出 k-NN 模式需要設定之四個重要參數分別進行說明：

1. 決定模式中所使用的特徵向量

在特徵向量的選擇上，有別於過往文獻至多使用 3 個特徵向量，本研究預計使用 12 個特徵向量。包含速率、流量、佔有率、大型車佔比、五分鐘內速率變化、五分鐘內流量變化、十分鐘內速率變化、十分鐘內流量變化、十五分鐘內速率變化、十五分鐘內流量變化、二十分鐘內速

率變化、二十分鐘內流量變化。

2. 決定距離量測方式

本研究使用 12 個特徵向量，有別於過往使用單一特徵向量之 k-NN 文獻，需衡量不同特徵向量對於旅行時間之影響，因此需修改距離量測之方式，考慮權重設定問題。本研究自行提出一結合加權曼哈頓距離 (Weighted Manhattan distance)，與皮爾森相關係數 (Pearson Correlation Coefficient) 概念之距離計算方式。其算法如式 7 所示，相較一般曼哈頓距離考慮了權重概念，而權重數值 (R_{Fy}) 則由皮爾森相關係數決定 (如式 8 所示)。

皮爾森相關係數主要用於衡量兩個變數之間相關程度，其值 (R_{Fy}) 介於 -1 與 1 之間，數字越大則代表兩者間趨近於正相關，數字越小則意味兩者間關係趨近於負相關，數字接近 0 則代表兩者間無關連性。由於本研究使用相當多特徵向量，但欲預測之旅行時間不一定與每個特徵向量有關，因此本研究先分析各特徵向量 (F) 與旅行時間 (y) 之相關係數，並將計算出的值 (R_{Fy}) 取絕對值作為加權曼哈頓距離之權重，即 R_{Fy} 若接近 1 或 -1 代表與旅行時間較有關聯，因此給予該特徵向量較高權重。

$$Dis = W_F \sum_{F=1}^i |N_F - H_F| / \sum_{F=1}^i W_F \quad \text{式 7}$$

Dis : 距離

i : 需比對的特徵向量數

F : 比對的特徵向量

W_F : 該特徵向量權重

N_F : 該特徵向量的現況值

H_F : 該特徵向量的歷史值

$$W_F = |R_{Fy}| = \frac{N \sum Fy - (\sum F)(\sum y)}{\sqrt{[N \sum F^2 - (\sum F)^2][N \sum y^2 - (\sum y)^2]}} \quad \text{式 8}$$

F : 各特徵向量

Y : 旅行時間

R_{Fy} : 特徵向量與旅行時間之相關係數

本研究設計一範例進行距離量測方式說明，若我們考慮三種特徵向量，分別為速率、流量及佔有率，其與旅行時間之皮爾森相關係數如表 12 所示，而現況資料 N 與歷史資料 A、B 標準化後之數值如表 13 所示，欲比對現況資料 N 與歷史資料 A 之距離，其計算方式為，兩者的速率差值 $|0.8-0.9|$ 乘以相關係數 0.8，加上流量差值 $|0.4-0.2|$ 乘以相關係數 0.6，加上佔有率差值 $|0.3-0.2|$ 乘以相關係數 0.2，加總後再除以總比對相關係數之和 (0.8+0.6+0.2)，即為 0.1375，而現況資料 N 與歷史資料 B

之距離為速率差值 | 0.8-0.2 | 乘以相關係數 0.8，加上佔有率差值 | 0.3-0.7|乘以相關係數 0.2，加總後除以總比對相關係數之和 (0.8+0.2)，即為 0.56，兩者對照之下，由於現況資料與歷史資料 A 距離較短，因而判斷兩者間較為相似。

表 12 特徵向量與旅行時間之相關係數

特徵向量	速率	流量	佔有率
相關係數	0.8	0.6	0.2

表 13 範例標準化後其值

時間	標準化後速率	標準化後流量	標準化後佔有率
現況資料N	0.8	0.4	0.3
歷史資料A	0.9	0.2	0.2
歷史資料B	0.2	空值	0.7

3. 決定如何估計預測值

Wu 等人 (2010) 指出以簡單投票法進行預測會使結果不準確，本研究預計加入 Rank-based weights 方法 (式 9)，找出 K 個樣本後，依距離遠近給予排序，距離最近之樣本排序為 1，距離最遠之樣本排序為 K，再依式 9 計算出其權重。舉一範例如表 14 所示，在 K 值為 3 情況下若有三個最相近樣本 A、B、C，距離分別為 5、8、10，各自旅行時間分別為 100、80、60 秒，估計預測值即為樣本 A 權重 9/14 乘以樣本 A 旅行時間 100 秒，加上樣本 B 權重 4/14 乘以樣本 B 旅行時間 80 秒，加上樣本 C 權重 1/14 乘以樣本 C 旅行時間 60 秒，為 91.4 秒，本研究並會在後續探討 Rank-based weight 是否能有效解決簡單投票 (平均) 法預測不準確之問題。

$$W_i = \frac{(K-r_i+1)^2}{\sum_{j=1}^K (K-r_j+1)^2} \quad \text{式 9}$$

r_i ：該樣本之排序

K：欲預測之樣本數 (K 值)

表 14 Rank-based weight 估計方式

樣本	距離	排序	權重	旅行時間 (秒)
A	5	1	$3^2 / 3^2+2^2+1^2 = 9/14$	100
B	8	2	$2^2 / 3^2+2^2+1^2 = 4/14$	80
C	10	3	$1^2 / 3^2+2^2+1^2 = 1/14$	60

4. 決定 k 值

k 值的選擇上，為了解決模式準確率與 k 值高度相關與敏感之缺陷，

劉軒寧（2016）提出一方法，建立兩層歷史旅行時間資料庫，需要兩倍的實測資料量來達成 k 值變動目的， k 值變動能降低若 k 值選擇不佳，模式準確度所受到之影響。本研究將劉軒寧（2016）所提出來的變動 k 進行改良，主要差異為調整歷史資料庫數量，由原先兩個減少至一個，降低實測資料所需要的數量，並加入本研究提出的距離計算方式與 Rank-based weight 的概念，且新增一 k 值區間，使 k 值於變動過程中不至於使用過多樣本或只以單一樣本進行預測。而本研究以固定 k 值進行測試，找出誤差最小 k 值作為變動 k 值的初始設定 k 。預測過程如圖 16 所示，詳細說明如下：

1. 將現況資料 Y 與歷史資料庫進行比對，並將歷史資料庫設定為兩小時範圍(若要預測晚上 7 點旅行時間，則比對晚上 5 點至 9 點資料)。
2. 從歷史資料庫中找出距離量度最小之資料 X 。
3. 將 X 與歷史資料庫其餘資料進行比對，找出與 X 最相近的 K 、 $K+1$ 、 $K-1$ 筆資料。
4. 以 K 、 $K+1$ 、 $K-1$ 筆資料預測歷史資料 X 的旅行時間。
5. 將 X 實際旅行時間與 K 、 $K+1$ 、 $K-1$ 筆資料所預測 X 之旅行時間進行比對，找出三種不同 K 值中誤差最小之 K 。
6. 判斷 K 是否位於設定區間內（是，則接續過程 8）
7. 不合理則使用上次預測之 K 筆資料預測現況資料 Y 旅行時間。
8. 從歷史資料庫取出更新後 K 筆資料預測現況資料 Y 旅行時間。
9. 更新 k 值

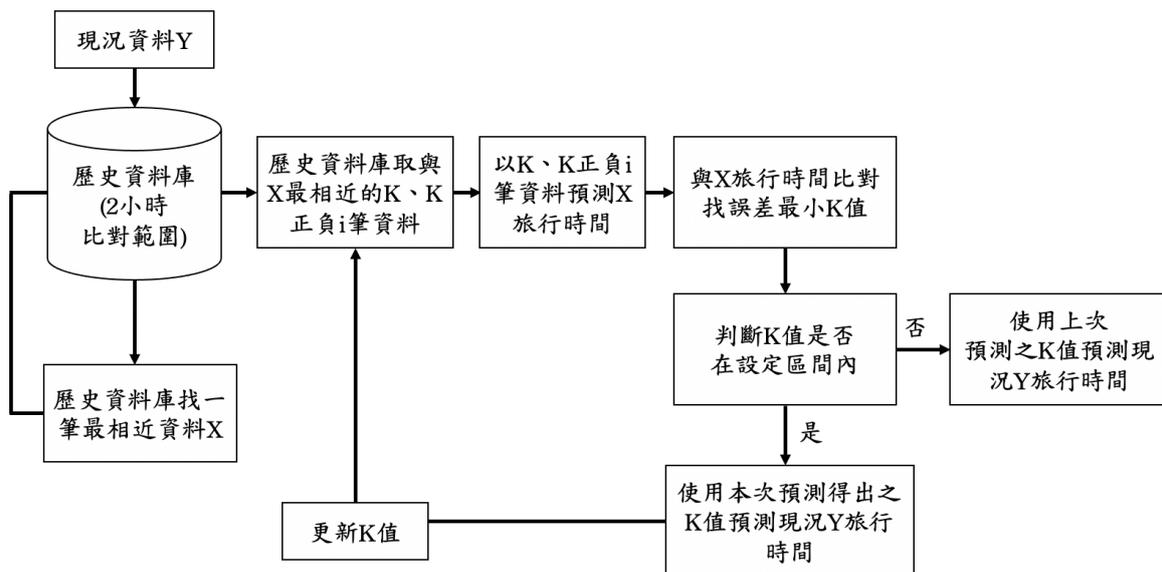


圖 16 變動 k -NN 法架構

本研究與過往 k-NN 文獻之差異如表 15 所示。有別於多數文獻 k 值皆為特定值，本研究與劉軒寧（2016）同樣採取變動 k 值，並針對歷史資料庫數量進行調整，而相較於過往多數文獻（劉軒寧,2016; Ma,2017; 黃冠維, 2018）只考慮速率單一特徵向量，本研究考慮速率、流量、佔有率等 12 種特徵向量，並以各特徵向量與旅行時間之相關性給予權重，提出新的距離量測方式，而在估計預測值則採用 Rank-based weight 法。

表 15 本研究與過往 k-NN 文獻之差異

k-NN文獻	K值	特徵向量權重計算	估計預測值
Clark (2003)	固定	觀察資料分布 進行標準化	無
蔡繼光 (2009)	固定	以尖鋒平均速率及流量 進行標準化	無
Habtemichael (2016)	固定	無	Rank-based weights
劉軒寧 (2016)	變動	無	簡單平均法
Jang (2019)	固定	無	以相關係數決定
本研究 (2021)	變動	以皮爾森相關係數決定	Rank-based weights/ 簡單平均法

3.5.4 評估成效

為了衡量本研究所設計模式之準確度，將比較模式預測旅行時間與實際旅行時間之誤差，以常見誤差量測方法來評估模式成效，包含平均絕對百分比誤差（Mean Absolute Percentage Error, MAPE）、均方根誤差（Root Mean Squared Error, RMSE）及平均絕對誤差（Mean Absolute Error, MAE）。

1. 平均絕對百分比誤差（Mean Absolute Percentage Error, MAPE）

平均絕對百分比誤差公式如式 10 所示，MAPE 以百分比來衡量預測值與實際值之誤差，不受預測值與實際值單位及大小影響，更可以客觀獲得實際差異程度，如表 16 第二與第三列儘管誤差都為 1，但 MAPE 分別為 1%及 20%，MAPE 越低代表模式越準確。

$$MAPE = \frac{1}{S} \sum_{i=1}^S \left| \frac{x(i)-y(i)}{x(i)} \right| \times 100\% \quad \text{式 10}$$

S : 預測數

$X(i)$: 第 i 次預測之真實值

$y(i)$: 第 i 次預測之預測值

表 16 平均絕對百分比誤差計算

實際旅行時間	預測旅行時間	誤差之絕對值	MAPE
100 秒	99 秒	1	1%
5 秒	4 秒	1	20%

2. 均方根誤差 (Root Mean Squared Error, RMSE)

均方根誤差公式如式 11 所示，為將其誤差取平方加總後取平均再開根號，其值與誤差之平方有關，模式一但出現預測誤差過大的情況，RMSE 會迅速提升，主要用來判斷預測穩定度，RMSE 越低即代表模式越可靠，穩定度較高。

$$RMSE = \sqrt{\frac{\sum_{i=1}^S (x(i)-y(i))^2}{S}} \quad \text{式 11}$$

S : 預測數

$X(i)$: 第 i 次預測之真實值

$y(i)$: 第 i 次預測之預測值

3. 平均絕對誤差 (Mean Absolute Error, MAE)

平均絕對誤差公式如式 12 所示，越低代表模式預測效果越好。

$$MAE = \frac{1}{S} \sum_{i=1}^S |x(i) - y(i)| \quad \text{式 12}$$

S : 預測數

$X(i)$: 第 i 次預測之真實值

$y(i)$: 第 i 次預測之預測值

3.5.5 演算範例

為了更清楚說明模式預測流程，本研究接續 3.4 小節提到的假設情境進行說明，預測流程如圖 17 所示。

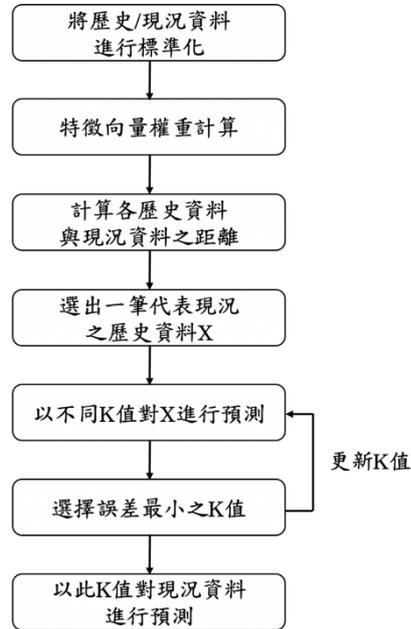


圖 17 預測流程

首先可得到標準化後現況資料(如表 17)、標準化後歷史資料庫(如表 18)，與依照式 8 計算各特徵向量與旅行時間之相關係數，取絕對值後得到特徵向量權重(如表 19)。

表 17 演算範例之現況資料

時間	VD1 速率	VD1 流量	VD1 佔有率	VD2 速率	VD2 流量	VD2 佔有率	VD3 速率	VD3 流量	VD3 佔有率
6/1 00:30	0.64	0.75	0.2	0.58	0.83	0.75	0.44	0.71	0.7

表 18 演算範例之歷史資料

時間	VD1 速率	VD1 流量	VD1 佔有率	VD2 速率	VD2 流量	VD2 佔有率	VD3 速率	VD3 流量	VD3 佔有率	旅行 時間
6/1 00:00	0.27	0.69	0.6	0.5	1	0.75	0	0.76	0.8	120
6/1 00:05	0.45	1	1	0	0.97	1	0.22	1	1.0	140
6/1 00:10	0	0.38	0.5	0.75	0.49	0.75	0.56	0.18	0.2	95
6/1 00:15	0.91	0.25	0.3	0.92	0.21	0.25	1	0	0	80
6/1 00:20	0.82	0.06	0.1	0.83	0	0	0.89	0.29	0.3	85
6/1 00:25	1	0	0	1	0.49	0.13	0.67	0.12	0.1	70

表 19 演算範例之特徵向量權重

	VD1 速率	VD1 流量	VD1 佔有率	VD2 速率	VD2 流量	VD2 佔有率	VD3 速率	VD3 流量	VD3 佔有率
相關係數	-0.59	0.98	0.95	-0.98	0.81	0.88	-0.81	0.95	0.96
權重	0.59	0.98	0.95	0.98	0.81	0.88	0.81	0.95	0.96

獲取上述資料後，即可計算各歷史資料與現況資料之距離，本研究以 6/1 00:00 為例，其與現況資料 6/1 00:30 之距離為所比對的第一個特徵向量，VD1 速率差值 $|0.27-0.64|$ 乘以 VD1 速率之權重 0.59 可得 0.2183，加上所比對的第二個特徵向量，VD1 流量之差值 $|0.69-0.75|$ 乘以 VD1 流量之權重 0.98 可得 0.0588，以此方式加總至所比對的第九個特徵向量，VD3 佔有率之差值 $|0.8-0.7|$ 乘以 VD3 佔有率之權重 0.96 可得 0.096，再除以所有進行比對的特徵向量其權重總和 $(0.59+0.98+\dots+0.96)$ ，即可求出歷史資料 00:00 與現況資料 00:30 之距離 1.3845。以此距離計算方法求得 00:00 至 00:25 六筆歷史資料與現況資料之距離（如表 20）。

表 20 歷史資料與現況資料之距離

歷史資料	與現況資料之距離
6/1 00:00	1.3845
6/1 00:05	2.7598
6/1 00:10	2.5410
6/1 00:15	3.8047
6/1 00:20	3.5821
6/1 00:25	3.6852

接著再依據表 20 選出一筆距離最小的歷史資料 X，即為 6/1 00:00 這筆資料，並依相同的距離計算方式算出歷史資料庫其餘資料與 X 之距離（如表 21），並找出與 X 最相近的 K-1、K、K+1 筆資料，在此我們假設初始 K 值為 2，因此分別選出最相近的 1、2、3 筆資料（表 21 灰底處），接著分別以 K=1、K=2、K=3 來預測 X 的旅行時間，利用 Rank-based weight 給予不同樣本權重（如表 22），K=1 時預測 X 之旅行時間即為 140 乘以 1，K=2 時即為 140 乘以 0.8 加上 95 乘以 0.2，K=3 時即為 140 乘以 0.64 加上 95 乘以 0.29 加上 85 乘以 0.07。

表 21 歷史資料與 X 之距離

歷史資料	與現況資料之距離
6/1 00:05	2.1182
6/1 00:10	2.8053
6/1 00:15	4.8772
6/1 00:20	4.8446
6/1 00:25	4.9476

表 22 不同 K 值下之預測

歷史資料	旅行時間	距離排序	K=1 權重	K=2 權重	K=3 權重
6/1 00:05	140	1	1	0.8	0.64
6/1 00:10	95	2	0	0.2	0.29
6/1 00:20	85	3	0	0	0.07

預測 X 旅行時間之結果與誤差如表 23 所示，誤差最小之 K 值為 3，我們需更新初始 K 值（由 2 更新至 3，下次預測將選擇 3-1、3、3+1 筆資料），且以 K=3 來對現況資料（6/1 00:25）進行預測（如表 24），選擇與現況資料最相近的 3 筆資料來進行預測，為 120 乘以 0.64 加上 95 乘以 0.29 加上 140 乘以 0.07 等於 114.15。

表 23 不同 K 值下之預測誤差

X 實際旅行時間	K=1 時預測	K=2 時預測	K=3 時預測
120	140	131	123.1
預測誤差	20	11	3.1

表 24 K=3 預測現況資料旅行時間

歷史資料	與現況資料之距離	旅行時間	距離排序	K=3 權重
6/1 00:00	1.3845	120	1	0.64
6/1 00:05	2.7598	140	3	0.07
6/1 00:10	2.5410	95	2	0.29
6/1 00:15	3.8047	80	6	0
6/1 00:20	3.5821	85	4	0
6/1 00:25	3.6852	70	5	0

第四章 實測分析

4.1 實測時空間範圍

本研究選定實測空間範圍為國一南向台南系統交流道至鼎金系統交流道。根據高公局提供的車輛偵測器經緯度資料及政府資料開放平臺提供的國道計費門架座標經緯度資料，於地圖上標示出範圍內包含的車輛偵測器與 ETC 門架分佈（如圖 18 所示）。共 79 個車輛偵測器（以藍色圖釘標示所在位置）及 11 個 ETC 門架（以深紅色圖釘標示所在位置）。資料蒐集時間範圍則為 2020 年 6 月 1 日至 2020 年 10 月 1 日共計 123 天，每五分鐘蒐集一筆資料，共蒐集 35424 筆。

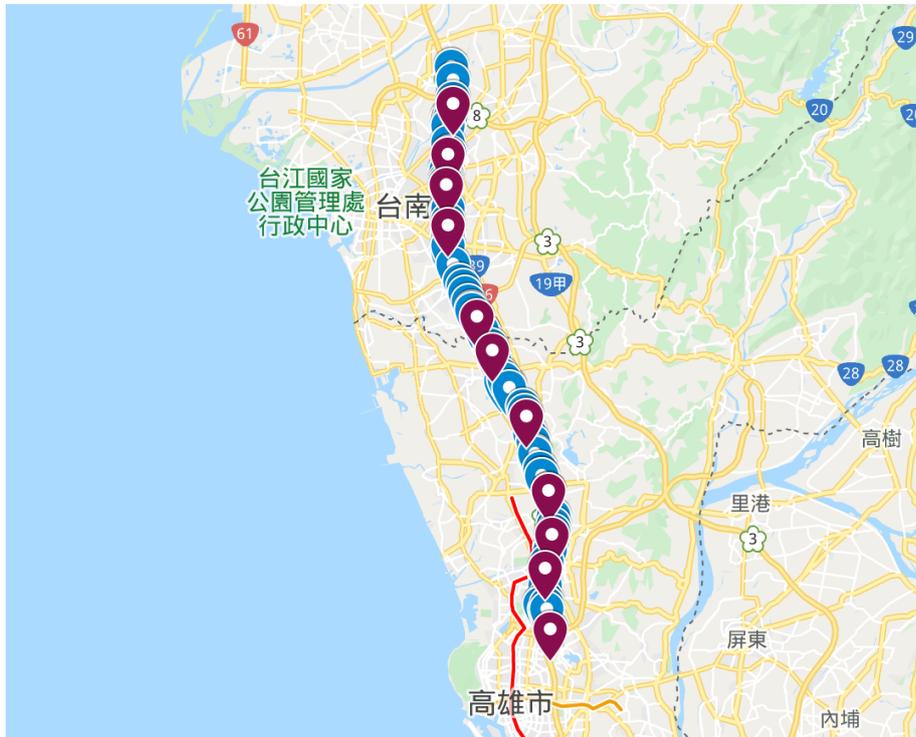


圖 18 預測之空間範圍

4.2 資料處理

依 VD 資料處理流程與 ETC 資料處理流程，需先將各 VD、ETC 缺漏比例、異常比例未達門檻值之 VD、ETC 進行刪除，而由於 ETC 資料之高完整性，無任何缺漏、異常比例達門檻值之 ETC，因此本小節僅列出 VD 資料部分。

4.2.1 遺失值比例

本研究對遺失值的定義為原始檔案無任何資料，數值異常之情況不包含在遺失值範圍內。各車輛偵測器遺失值比例如圖 19 所示，遺失值比例大於 20% 的車輛偵測器共有七個（如表 25 所示），因此這七個 VD 將不以加入預測，進行刪除。



圖 19 各 VD 遺失率

表 25 嚴重遺失的車輛偵測器

VD代碼	車道數	應有筆數	實際筆數	遺失率
VD-N1-S-317.600-M-RS	4	141696	13433	90.52%
VD-N1-S-318.550-M-Loop	無任何資料			100%
VD-N1-S-327.650-N-Loop	3	106272	85011	20.01%
VD-N1-S-329.875-M-RS	無任何資料			100%
VD-N1-S-343.695-M-Loop	3	106272	47107	55.67%
VD-N1-S-348.615-M-Loop	無任何資料			100%
VD-N1-S-361.000-M-Loop	4	141696	112300	20.74%

4.2.2 異常值比例

本研究對於獲得之原始資料（即速率、流量、佔有率）一旦為負值，即定義為異常值，參考交通部運研所對 VD 一分鐘動態資訊連續相同資料不合理判斷條件，訂定流量、速率、佔有率若連續兩筆為零，第二筆資料將被視為異常值，扣掉上述遺失值過高之車輛偵測器，剩餘 72 個車輛偵測器中異常值比例大於 20% 之 VD 如表 26 所示，本研究以佔有率、速率、流量三者之中異常值筆數最高者進行異常值佔比判斷（如表 26 灰底處），共七個，同樣不列入預測，進行刪除，空間範圍內剩餘需比對的 VD 為 65 個。

表 26 異常值佔比高的車輛偵測器

VD 代碼	應有筆數	佔有率 異常筆數	速率 異常筆數	流量 異常筆數	異常值 佔比
VD-N1-S-334.450-M-RS	35424	12148	11939	2656	39.2%
VD-N1-S-336.450-M-RS		8970	8832	834	25.3%
VD-N1-S-342-O-NW-2-Loop		10906	11274	10922	31.8%
VD-N1-S-345.620-M-RS		9947	9801	851	28.1%
VD-N1-S-349-O-NW-1-Loop		10855	11107	10879	31.4%
VD-N1-S-351.590-M-Loop		23208	24445	23281	69.0%
VD-N1-S-351.600-M-RS		15846	15692	4278	44.7%

4.2.3 VD 速率資料差補

本研究先行找出離台南系統交流道最近之 ETC 門架（編號 01F3185S）與離鼎金系統交流道最近之 ETC 門架（編號 01F3640S），並蒐集 ETC 資料獲得兩門架範圍內各門架間（橘色圈圈）之旅行時間（如圖 20）。

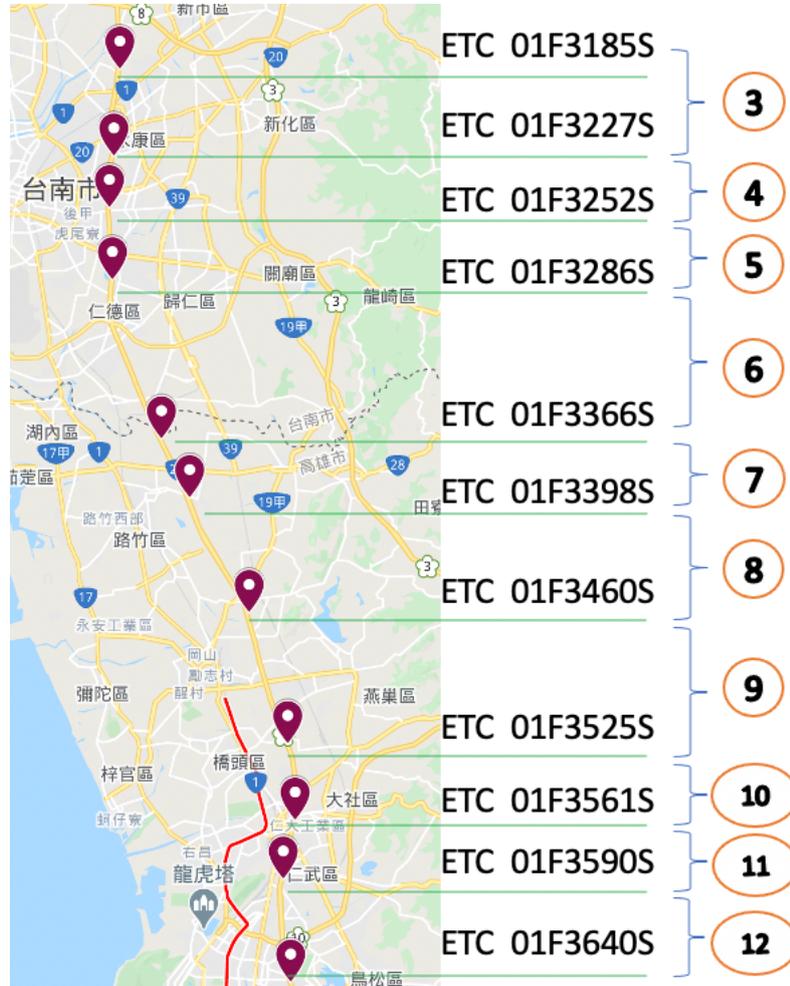


圖 20 ETC 資料獲取範圍

接著針對前後端與交流道之距離進行差補，編號 01F3185S 與台南系統交流道相對位置如圖 21 所示，VD-N1-315.910-N-Loop 位於台南系統交流道上，需將此 VD 至 01F3185S 門架間旅行時間補上，以 VD 位置做切割，需填補之路段有路段 1、路段 2，而路段 1 旅行時間計算方式為兩相鄰之 VD（VD-N1-315.910-N-Loop 與 VD-N1-S-317.017-M-Loop）速率取平均，除上路段 1 長度，路段 2 旅行時間為 VD-N1-S-317.017-M-Loop 與 VD-N1-S-319.825-N-Loop 速率取平均，除上路段 2 長度，VD-N1-S-318.550-M-Loop 本身遺失率達 100%，無法取得速率資料，而 VD-N1-S-319-O-NW-1-Loop 為上下匝道之 VD，並非位於主線上，會造成速率大幅低估情形，也不適合作為填補，因而使用 VD-N1-S-317.017-M-Loop 與 VD-N1-S-319.825-N-Loop 速率取平均，除上路段 2 長度。

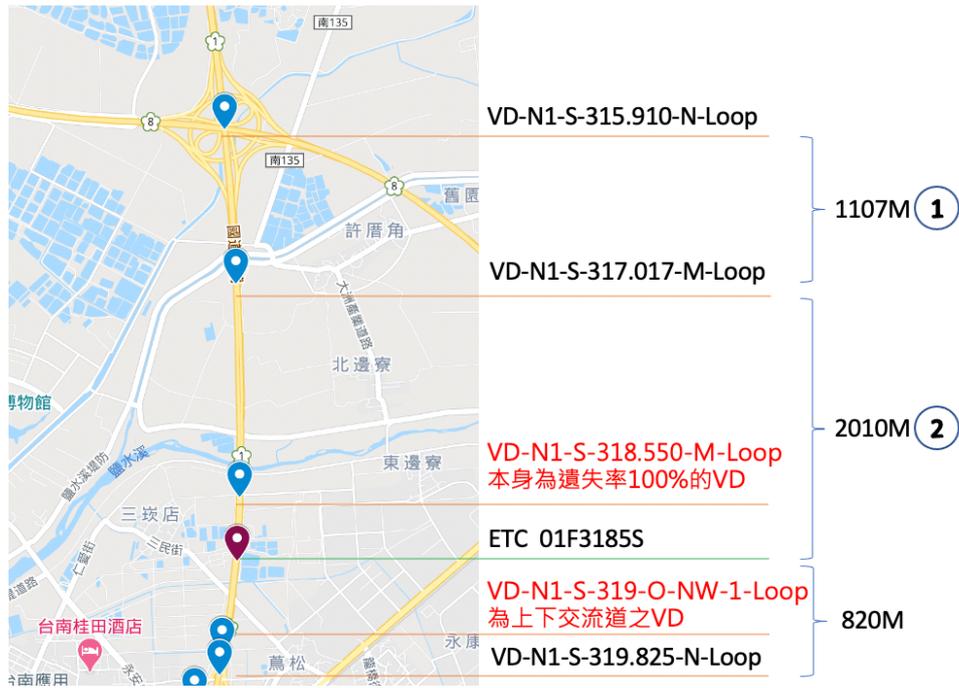


圖 21 前端旅行時間差補

鼎金系統交流道與距離最近之 ETC 門架 (編號 01F3640S) 相對位置如圖 22 所示，需將路段 12 扣除多算的部分 (即位於鼎金系統交流道上之 VD-N1-S-363.000-N-Loop 至 01F3640S 旅行時間，可以 VD 位置切割為路段 13 至路段 16)，路段 13 旅行時間為 VD-N1-S-363.000-N-Loop 與 VD-N1-S-363.500-M-Loop 之平均速率除以路段 13 長度，路段 14 旅行時間為 VD-N1-S-363.500-M-Loop 與 VD-N1-S-363.800-M-Loop 之平均速率除以路段 14 長度，路段 15 旅行時間為 VD-N1-S-363.800-M-Loop 與 VD-N1-S-364.300-M-Loop 之平均速率除以路段 15 長度，路段 16 旅行時間為 VD-N1-S-364.300-M-Loop 與 VD-N1-S-364.800-M-Loop 之平均速率除以路段 16 長度。



圖 22 後端旅行時間差補

接著將原先多算的路段 12 扣除路段 13 至路段 16 之旅行時間，得到新的路段 12 之旅行時間，再將路段 1 至新的路段 12 旅行時間進行延遲加總，即可求得最終旅行時間。

4.3 實測結果

本研究以 2020 年 6 月 1 日至 2020 年 9 月 26 日共 118 天資料進行歷史資料庫建置，而由於旅行時間於離峰時段起伏並不大，對其進行預測較無效益，因此本研究將以尖峰時段進行實測，而考量平日相較假日有更明顯的尖離峰，因此選取較多的平日尖峰時段進行實測分析，平日與假日比例為 3:1（如表 27），每次預測為兩小時，將預測此時間點由台南系統交流道行駛至鼎金系統交流道的旅行時間。

表 27 實測選定時段

實測日期	實測時間	備註
2020/9/27 星期日	10:00 至 12:00	上午尖峰
2020/9/27 星期日	15:00 至 17:00	下午尖峰
2020/9/28 星期一	06:30 至 08:30	上午尖峰
2020/9/28 星期一	17:30 至 19:30	下午尖峰
2020/9/29 星期二	06:30 至 08:30	上午尖峰
2020/9/29 星期二	17:30 至 19:30	下午尖峰
2020/9/30 星期三	06:30 至 08:30	上午尖峰
2020/9/30 星期三	17:30 至 19:30	下午尖峰

首先在初始 K 值的選擇上，本研究於 9/29 星期二上午尖峰進行不同 K 值下之預測，將 K 值設定為 1、5、10、15 及 20，其誤差指標如表 28 所示。可看到在三種指標 K=5 表現皆為最佳，由此可知若以固定 k 值進行預測，K=5 為最佳設定，因此本研究之變動 k-NN 法初始 k 值設定為 5，並考量不同的變動 i 值（k 值每次變動上下限）進行預測，由於 K=1 與 K=10 誤差皆較 K=5 來得高，因此我們不希望 K 值在變動過程中小至 1 或大至 10，而是於 2 至 9 間進行變動，因此設定區間為 2-9。

表 28 不同 k 值下之誤差指標

K 值設定	K=1	K=5 (最佳)	K=10	K=15	K=20
MAPE(%)	0.97%	0.61%	0.63%	0.69%	0.70%
RMSE(Sec)	16.2	10.2	10.5	11.5	11.7
MAE(Sec)	19.1	13.0	13.7	14.7	15.4

接著本研究針對不同的特徵向量組合於八個測試時段進行預測，分析本研究與過往使用單一或少數文獻相比，使用哪些特徵向量做搭配能夠有效提升預測準確度。首先設定五種不同的特徵向量組合（如表 29 所示）。包含使用所蒐集的 12

種特徵向量預測、只使用流量預測、只使用速率進行預測、只使用佔有率進行預測及只使用大型車佔比進行預測。

表 29 特徵向量搭配

特徵向量組合	選取之特徵向量
組合一	流量、速率、佔有率、大型車佔比、五分鐘速率變化、五分鐘流量變化、十分鐘速率變化、十分鐘流量變化、十五分鐘速率變化、十五分鐘流量變化、二十分鐘速率變化、二十分鐘流量變化
組合二	流量
組合三	速率
組合四	佔有率
組合五	大型車佔比

整體預測情形如表 30 所示，可看到本研究所使用的 12 種特徵向量在 MAPE 與 MAE 表現皆為最佳，而以速率進行比對則在 RMSE 表現最佳，而其中有一現象為在四個主要資訊（流量、速率、佔有率、大型車佔比）中，使用速率、佔有率進行預測有著較高的準確率（MAPE 約 2-3%，MAE 約 65-75，RMSE 多在 120-150），而使用流量、大型車佔比進行預測有較低的準確率（MAPE 約為 4-5%，MAE 都在 100 以上，RMSE 約在 140-170），因此判斷在本研究使用的 12 種特徵向量中，進一步選取表現較佳的速率、佔有率進行特徵向量重組（如表 31），包含速率、5-20 分鐘速率趨勢及佔有率。

表 30 整體預測情形

選取組合(估計預測值)	MAPE	MAE	RMSE
12 種全使用 (平均法)	2.54% (最佳)	62.3 (最佳)	122.3
12 種全使用 (Rank)	2.74%	66.4	131.9
流量(平均法)	4.94%	108.3	210.8
流量(Rank)	5.34%	117.0	222.5
速率(平均法)	2.80%	67.3	120.5 (最佳)
速率(Rank)	2.87%	68.8	127.7
佔有率(平均法)	3.01%	73.6	145.0
佔有率(Rank)	3.16%	77.3	152.1
大型車佔比(平均法)	4.45%	105.2	167.6
大型車佔比(Rank)	4.61%	107.5	177.9

表 31 特徵向量重組

特徵向量組合	選取之特徵向量
組合六	速率、佔有率
組合七	速率、佔有率、五分鐘速率變化、十分鐘速率變化、十五分鐘速率變化、二十分鐘速率變化

預測情形如表 32 所示，可觀察進一步選擇表現良好之特徵向量重組後的兩種新組合，與原先表現最佳的 12 種特徵向量、速率進行比較，在 MAPE 的部分，由原先 2.54% 放大至 2.58%，MAE 則由 62.3 降低至 60.9，RMSE 則由原先 120.5 降低至 115.7。而在特徵向量搭配組合選擇上，準確度較高的組合為組合一、組合六，而考量到模式計算速度與資料取得方便性，本研究選擇以組合六（速率+佔有率）進行下一步分析，並探討在此特徵向量組合下，以本研究所改良之變動 k 值是否能提升準確度。

表 32 重組後整體預測情形

選取組合(估計預測值)	MAPE	MAE	RMSE
速率+佔有率 (平均法)	2.58% (並列)	60.9 (最佳)	115.7 (最佳)
速率+佔有率 (Rank)	2.81%	66.7	128.2
速率+佔有率+速率趨勢 (平均法)	2.58% (並列)	63.9	117.7
速率+佔有率+速率趨勢 (Rank)	2.77%	68.0	125.5

選擇組合六為最佳特徵向量搭配組合後，接著本研究以組合六針對固定 k 值與本研究設計之變動 k 值進行分析，由於先前測試固定 k 值時發現 k = 1 與 k = 10 表現均未比 k = 5 好，因此將 k 值設定區間定為 2 至 9，k 值小於 1 或大於 9 時，皆以上一次預測之 k 值進行取代。

結果如表 33 所示。可看到本研究設計之變動 k 值在每次變動量為 2 時，採用平均法估計預測值之狀態下，於 MAPE 誤差指標具有最佳之預測效果，MAE、RMSE 誤差指標則些微增加，整體而言變動量 i=2 時優於變動量 i=1，但與固定 k 值之 k-NN 無明顯差異，相較於決定 k 值與否變動，選擇最佳特徵向量組合進行預測更為重要。

表 33 固定/變動 K 整體預測情形

固定/變動 k 值	估計預測方法	MAPE	MAE	RMSE
固定 K=5	平均法	2.58%	60.9 (最佳)	115.7 (最佳)
固定 K=5	Rank	2.81%	66.7	128.2
變動 i=1	平均法	2.61%	64.5	116.9
變動 i=1	Rank	2.84%	70.2	130.7
變動 i=2	平均法	2.57% (最佳)	63.4	118.9
變動 i=2	Rank	2.79%	68.9	127.7

而基於速率+佔有率組合下，在變動量 i 為 0（固定 K 值）至 i 為 2 之條件，分別針對兩種估計預測值方法，平均法與 Rank-based Weight 進行誤差分析，結果如表 34 所示，Rank-based Weight 相較於平均法誤差較高，並不能有效提升預測準確度，由於本研究所蒐集之歷史資料庫具有三萬多筆資料，而其中用於每次預測比對的資料也都有五千筆以上，距離現況資料最近之歷史資料（ $K=1$ ）與距離現況資料第五近之歷史資料（ $K=5$ ）在大數據的角度來看，或許無太大差異，然而，使用 Rank-based Weight 則會讓最近的歷史資料對預測之影響大大提升，當 $k=1$ 時往往又相較 $k=5$ 擁有較大誤差，這也是本研究認為 Rank-based Weight 不適合用於 k -NN 之原因。

表 34 不同估計預測值方法比較

估計預測值方法	MAPE	MAE	RMSE
平均法	2.59%	62.91	117.16
Rank-based Weight	2.81%	68.59	128.85

而基於速率+佔有率組合、且使用平均法與固定 K 值條件下，對本研究所提出之特徵向量權重計算方式進行誤差分析，與無設置任何特徵向量權重之情況（即所有比對之特徵向量權重皆為一）進行比較，分析結果如表 35 所示。可明顯看出本研究所提出特徵向量權重計算方式在各項誤差指標皆有顯著下降，MAPE 由 2.84% 下降至 2.58%，MAE 由 71.5 下降至 60.9，RMSE 由 138.5 下降至 115.7，代表此權重計算方式相較無設置任何權重而言，能有效提升預測準確率。

表 35 本研究所提之權重計算方式效果評估

特徵向量權重	MAPE	MAE	RMSE
權重皆為一	2.84%	71.5	138.5
依其與旅行時間相關係數之絕對值作為權重	2.58%	60.9	115.7

本研究所得出之最佳結果為使用速率加上佔有率之組合，且使用本研究所提出之特徵向量權重計算方式與使用平均法作為估計預測值方法， k 值方面，固定 k 與變動 k （變動量 $i=2$ ）效果相當，皆為最佳結果，本研究以變動量 $i=2$ 為例，其於八個實測時段之預測情況與 k 值變動情形如表 36 所示。

變動量欄位中，若該欄位顯示 i 代表維持既有之 k 值，一旦 k 值改變了（如本次預測 k 值相較於上次預測之 k 值增加 1），於變動量欄位則為「 $i+1$ 」，而一旦 k 值超過本研究設定區間，如 k 值於該次預測降至 1，此時會判斷 k 值過小，於變動量該欄位則為「過小」，並於 k 值之欄位維持上一次預測之 k 值，反之如 k 值提升至 9 以上，則會判斷 k 值過大，於變動量欄位則為「過大」，並同樣於 k 值之欄位維持上一次預測之 k 值。

表 36 本研究所得之最佳結果預測情形

條件	初始 k 值=5，變動量 i=2，平均法				
時間	真實旅行時間	預測值	誤差 (秒)	變動量	k 值
2020/9/27 上午尖峰					
10:00	1649.3	1645.5	3.7	i	5
10:05	1659.1	1649.7	9.4	i+2	7
10:10	1649.4	1641.8	7.6	i+1	8
10:15	1660.2	1654.1	6.1	i-2	6
10:20	1689.3	1666.9	22.4	i-2	4
10:25	1639.2	1647.5	8.3	i-2	2
10:30	1660.2	1669.5	9.4	i+1	3
10:35	1680.3	1663.6	16.7	過小	3
10:40	1665.2	1685.4	20.2	i+2	5
10:45	1665.1	1681.5	16.5	i+2	7
10:50	1674.2	1671.9	2.3	i+2	9
10:55	1702.3	1692.4	9.9	過大	9
11:00	1668.5	1676.6	8.1	i-2	7
11:05	1648.5	1672.6	24.1	i+2	9
11:10	1660.7	1671.5	10.8	過大	9
11:15	1667.4	1692.3	24.9	過大	9
11:20	1678.5	1677.5	1.0	過大	9
11:25	1663.6	1719.6	56.0	i-2	7
11:30	1653.7	1679.5	25.8	i+2	9
11:35	1655.5	1671.1	15.7	過大	9
11:40	1646.6	1662.3	15.7	i-2	7
11:45	1647.7	1660.1	12.4	i-2	5
11:50	1647.4	1647.0	0.4	i-1	4
11:55	1659.1	1641.4	17.7	i+2	6
2020/9/27 下午尖峰					
15:00	1768.0	1737.5	30.6	i	5
15:05	1781.3	1715.5	65.8	i-2	3
15:10	1749.6	1774.5	24.9	i+1	4
15:15	1795.6	1757.1	38.5	i-1	3
15:20	1731.5	1734.8	3.3	i	3
15:25	1776.0	1775.5	0.5	i+1	4
15:30	1776.5	1771.2	5.2	i-2	2

15:35	1798.5	1825.7	27.2	i+1	3
15:40	1792.0	1808.9	16.9	i-1	2
15:45	1776.2	1786.0	9.8	i+2	4
15:50	1776.9	1792.0	15.1	i+2	6
15:55	1755.1	1768.8	13.7	i+2	8
16:00	1764.8	1773.6	8.8	i-2	6
16:05	1813.5	1785.7	27.8	i-1	5
16:10	1809.1	1786.9	22.3	i+1	6
16:15	1860.4	1801.0	59.4	i+2	8
16:20	1859.9	1827.9	32.0	過大	8
16:25	1850.3	1822.1	28.2	i-2	6
16:30	1824.4	1847.3	22.9	i+2	8
16:35	1888.8	1930.5	41.8	過大	8
16:40	1889.6	2043.0	153.4	i-2	6
16:45	1910.6	1943.6	33.1	i+2	8
16:50	1931.0	1950.8	19.8	過大	8
16:55	1868.2	1938.1	69.9	i-2	6
2020 / 9 / 28 上午尖峰					
06:30	1554.9	1557.2	2.3	i	5
06:35	1554.3	1564.2	9.9	i+2	7
06:40	1566.4	1577.4	11.1	i-2	5
06:45	1573.4	1587.3	14.0	i	5
06:50	1591.2	1589.7	1.5	i-1	4
06:55	1611.3	1619.7	8.4	i-2	2
07:00	1633.9	1637.7	3.8	i+2	4
07:05	1644.2	1657.1	12.9	i-2	2
07:10	1647.8	1661.3	13.5	過小	2
07:15	1684.8	1685.0	0.2	i+2	4
07:20	1734.0	1700.6	33.4	i-2	2
07:25	1719.4	1714.3	5.1	i+2	4
07:30	1699.8	1709.7	9.9	i-2	2
07:35	1686.2	1688.9	2.7	i	2
07:40	1668.6	1693.9	25.3	i+2	4
07:45	1691.3	1682.1	9.2	i-1	3
07:50	1658.5	1676.8	18.2	i+2	5
07:55	1685.4	1702.7	17.3	i+2	7
08:00	1706.1	1699.4	6.7	i+2	9

08:05	1697.6	1695.7	2.0	i-2	7
08:10	1698.8	1704.6	5.8	i+2	9
08:15	1688.2	1726.7	38.5	i-2	7
08:20	1660.6	1684.3	23.7	i-2	5
08:25	1693.1	1704.1	11.0	i+2	7
2020 / 9 / 28 下午尖峰					
17:30	3202.2	2868.8	333.4	i	5
17:35	3219.8	3229.1	9.3	i-2	3
17:40	3203.2	3393.0	189.8	i	3
17:45	3250.4	3293.1	42.7	i+2	5
17:50	3188.7	2964.4	224.4	i+2	7
17:55	3155.8	2970.3	185.6	i-2	5
18:00	3127.1	2663.4	463.6	i+2	7
18:05	3109.1	2731.7	377.4	i-2	5
18:10	3134.8	2756.4	378.4	i+2	7
18:15	3173.5	2781.6	391.9	i-2	5
18:20	3068.0	2595.9	472.1	i-2	3
18:25	3048.6	2752.6	296.0	過小	3
18:30	2993.0	2501.9	491.1	過小	3
18:35	2802.5	2729.6	72.8	過小	3
18:40	2661.9	2571.0	90.9	i-1	2
18:45	2512.1	2471.1	41.0	i+2	4
18:50	2365.8	2342.3	23.5	i+2	6
18:55	2216.4	2264.7	48.3	i	6
19:00	2057.5	2196.7	139.2	i+2	8
19:05	1955.6	2041.4	85.8	過大	8
19:10	1875.5	1905.1	29.6	過大	8
19:15	1788.7	1842.9	54.2	i-2	6
19:20	1709.9	1767.9	58.1	i-2	4
19:25	1714.3	1877.4	163.2	i-2	2
2020 / 9 / 29 上午尖峰					
06:30	1552.0	1563.5	11.5	i	5
06:35	1566.9	1563.9	3.0	i+2	7
06:40	1564.4	1571.7	7.3	i+2	9
06:45	1576.6	1579.1	2.6	i-2	7
06:50	1592.4	1586.7	5.7	i-2	5
06:55	1609.1	1605.1	4.0	i-2	3

07:00	1619.5	1624.4	5.0	i-1	2
07:05	1621.4	1637.5	16.1	i	2
07:10	1652.6	1646.2	6.3	i+2	4
07:15	1687.8	1684.7	3.0	i-2	2
07:20	1692.5	1693.5	1.0	i+2	4
07:25	1706.6	1720.6	14.0	i-2	2
07:30	1722.2	1697.6	24.6	i	2
07:35	1704.9	1719.3	14.4	i	2
07:40	1718.2	1721.4	3.2	過小	2
07:45	1674.6	1695.9	21.3	i+2	4
07:50	1665.1	1695.1	30.0	i-2	2
07:55	1695.9	1705.5	9.5	過小	2
08:00	1678.8	1719.2	40.4	i+2	4
08:05	1735.1	1698.1	37.0	i-2	2
08:10	1713.8	1724.4	10.6	過小	2
08:15	1745.5	1731.7	13.8	過小	2
08:20	1759.0	1758.1	0.9	i	2
08:25	1725.2	1746.4	21.1	i+2	4
2020 / 9 / 29 下午尖峰					
17:30	2935.6	2738.3	197.3	i	5
17:35	2975.8	2894.8	81.0	i+2	7
17:40	2970.5	2892.6	77.9	i+2	9
17:45	2875.4	2878.3	2.8	i-2	7
17:50	2878.4	2930.9	52.5	i-2	5
17:55	2899.7	2788.6	111.1	i-2	3
18:00	2931.2	2992.2	61.0	i+2	5
18:05	2935.7	3001.0	65.3	i	5
18:10	2858.3	2895.7	37.4	i	5
18:15	2933.7	2940.3	6.6	i	5
18:20	2841.7	2806.3	35.4	i-2	3
18:25	2756.5	2791.0	34.5	i+2	5
18:30	2748.7	2996.8	248.2	i-2	3
18:35	2651.2	2845.4	194.2	i+2	5
18:40	2587.9	2822.3	234.4	i+2	7
18:45	2555.5	2819.0	263.5	i-2	5
18:50	2446.1	2494.8	48.8	i	5
18:55	2312.1	2669.6	357.5	i+2	7

19:00	2176.7	2434.1	257.4	i-2	5
19:05	2076.7	2031.3	45.3	i-1	4
19:10	2056.1	2030.7	25.3	i-2	2
19:15	1976.2	2079.9	103.8	i	2
19:20	1888.0	1861.8	26.2	i+2	4
19:25	1822.8	1868.2	45.3	i-2	2
2020 / 9 / 30 上午尖峰					
06:30	1556.2	1557.5	1.2	i	5
06:35	1555.4	1566.5	11.2	i+2	7
06:40	1563.7	1566.5	2.8	i-2	5
06:45	1577.0	1582.0	5.0	i-1	4
06:50	1585.1	1579.2	5.8	i-1	3
06:55	1603.3	1606.6	3.3	i+2	5
07:00	1619.9	1628.9	9.0	i+2	7
07:05	1611.7	1631.5	19.8	i-2	5
07:10	1645.4	1653.0	7.5	i+1	6
07:15	1667.5	1693.6	26.1	i-1	5
07:20	1702.0	1705.9	3.9	i+2	7
07:25	1690.8	1705.0	14.2	i-2	5
07:30	1742.9	1739.4	3.5	i-2	3
07:35	1752.3	1703.8	48.5	i-1	2
07:40	1720.6	1699.2	21.4	i+2	4
07:45	1690.8	1697.0	6.2	i+2	6
07:50	1682.4	1696.1	13.7	i+2	8
07:55	1654.0	1696.1	42.1	i-1	7
08:00	1681.3	1717.5	36.2	i-2	5
08:05	1690.6	1713.0	22.4	i+2	7
08:10	1693.7	1705.3	11.5	i+2	9
08:15	1678.7	1687.7	9.0	i-2	7
08:20	1720.2	1715.9	4.3	i+2	9
08:25	1715.3	1720.5	5.2	i-2	7
2020 / 9 / 30 下午尖峰					
17:30	2884.4	2691.7	192.6	i	5
17:35	2801.1	2621.7	179.4	i+2	7
17:40	2705.0	2495.4	209.5	i-2	5
17:45	2684.8	2480.4	204.4	i	5
17:50	2701.9	2713.5	11.6	i+1	6

17:55	2698.5	2649.6	48.8	i	6
18:00	2745.0	2642.3	102.7	i+2	8
18:05	2754.3	2777.3	23.0	i-1	7
18:10	2754.8	2781.6	26.8	i-2	5
18:15	2620.2	3007.2	387.0	i-2	3
18:20	2659.2	2976.0	316.8	過小	3
18:25	2614.1	2706.5	92.4	過小	3
18:30	2543.5	2420.0	123.5	i	3
18:35	2522.1	2277.2	244.9	i+1	4
18:40	2574.1	2351.9	222.1	i+2	6
18:45	2515.2	2350.4	164.9	i+2	8
18:50	2530.2	2552.0	21.8	i	8
18:55	2345.3	2338.7	6.6	i	8
19:00	2340.4	2442.6	102.1	過大	8
19:05	2349.4	2392.8	43.4	過大	8
19:10	2273.0	2272.1	0.8	i	8
19:15	2251.0	2278.0	27.0	i-2	6
19:20	2264.2	2263.4	0.8	i-2	4
19:25	2185.7	2279.6	93.9	i-2	2

4.4 統計分析

針對本研究所進行之改良，包含使用多元特徵向量、變動 k 值、特徵向量權重計算方法、不同的估計預測值方式所進行之預測，以使用與否或方法差異來進行統計分析，本研究利用 SPSS 軟體進行操作，所使用的統計方法為成對樣本 T 檢定，比較兩組相依樣本 (D1、D2) 之間 MAPE 指標之平均數差異，並於假說檢定方面進行以下假設：

- 虛無假說 H_0 : D_1 等同於 D_2
- 對立假說 H_1 : D_1 不等於 D_2

首先分析本研究所得出之最佳特徵向量組合 (速率、佔有率) 與使用單一特徵向量所進行之預測是否具顯著差異。比對項目如表 37 所示。

表 37 統計分析所比對之樣本 (一)

比對代號	選取之特徵向量	k 值	特徵向量權重	估計預測值方式
S	速率+佔有率	固定	與旅行時間 之相關係數	平均法
A	流量			
B	速率			
C	佔有率			

D	重車比			
---	-----	--	--	--

1. S 比對 A

總預測數為 192 次，成對樣本檢定中，雙尾顯著性 <0.001 ，其小於 0.05，拒絕 H_0 ，因此有顯著差異。

2. S 比對 B

總比對數為 192 次，成對樣本檢定中，雙尾顯著性為 0.124，大於 0.05，無法拒絕 H_0 ，因此無顯著差異。

3. S 比對 C

總預測數為 192 次，成對樣本檢定中，雙尾顯著性為 0.045，其小於 0.05，拒絕 H_0 ，因此有顯著差異。

4. S 比對 D

總比對數為 192 次，成對樣本檢定中，雙尾顯著性 <0.001 ，其小於 0.05，拒絕 H_0 ，因此有顯著差異。

接著分析變動 k 值，在每次變動量 $i=2$ 之情況下所進行之預測，與固定 k 值是否具顯著差異。比對項目如表 38 所示。

表 38 統計分析所比對之樣本 (二)

比對代號	k 值	選取之特徵向量	特徵向量權重	估計預測值方式
S	變動 k, $i=2$	速率+佔有率	與旅行時間	平均法
E	固定 k, $i=0$		之相關係數	

1. S 比對 E

總預測數為 192 次，成對樣本檢定中，雙尾顯著性為 0.931，大於 0.05，無法拒絕 H_0 ，因此無顯著差異。

接著分析使用平均法所進行之預測，與 Rank-based weight 法是否具顯著差異。比對項目如表 39 所示。

表 39 統計分析所比對之樣本 (三)

比對代號	估計預測值方式	k 值	選取之特徵向量	特徵向量權重
S	平均法	固定	速率+佔有率	與旅行時間 之相關係數
F	Rank-based weight			

1. S 比對 F

總預測數為 192 次，成對樣本檢定中，雙尾顯著性為 0.057，大於 0.05，無法拒絕 H_0 ，因此無顯著差異。

接著分析以各特徵向量間與旅行時間相關係數之絕對值作為特徵向量權重所進行之預測，與特徵向量權重皆為一之間是否具顯著差異。比對項目如表 40 所示。

表 40 統計分析所比對之樣本（四）

比對代號	特徵向量權重	選取之特徵向量	K 值	估計預測值方式
S	依其與旅行時間 相關係數絕對值	速率+佔有率	固定	平均法
G	皆為一			

1. S 比對 G

總預測數為 192 次，成對樣本檢定中，雙尾顯著性為 0.029，其小於 0.05，拒絕 H0，因此有顯著差異。

由以上統計檢定可作出以下結論，本研究所找出的最佳特徵向量組合為「速率+佔有率」，相較於使用流量、佔有率、重車比以上的「單一」特徵向量來說，是能有效提升模式預測準確度的，而相較於只使用速率作為特徵向量，僅在數學上具有意義，即意味證明其預測結果不會比較差，但不能做出足夠優於之結論。而本研究也發現不論 k 值變動與否，預測準確度上無明顯差異，至於估計預測值方法上，兩者之間雖無顯著差異，但其統計值其實十分接近臨界點（雙尾顯著性為 0.057），在可能的誤差之下，兩者間可能具備顯著差異，另外，本研究提出的特徵向量權重計算方法，相較於不去處理，即給予各特徵向量等值之權重來說，是能有效提升預測準確度的。

陽明交大
NYCU

第五章 結論與建議

本研究設計一變動 k-NN 法，並考量多個特徵向量與不同的估計預測值方式，且以各特徵向量與旅行時間之相關係數之絕對值作為其權重，來預測無號誌化道路旅行時間。資料蒐集上，使用了車輛偵測器 (VD) 與電子收費系統 (ETC) 兩種資料，且在旅行時間資訊推估上，使用更為合適之計算方式，藉由 ETC 資料計算門架間旅行時間，再透過 VD 速率資料差補，計算出更為實用之交流道間旅行時間。建構完歷史資料庫後，再透過實際分析證明模式之可行性，獲得最佳特徵向量組合，並再以該組合深入探討本研究所設計之變動 k-NN 法、不同的估計預測方式及特徵向量權重計算方式之成效。

與過往文獻以單一資料相比，使用多元特徵向量能明顯提升預測準確度，而 k 值變動與否則與準確度較無直接影響，考量到變動 k-NN 每預測一次須進行兩次距離計算，所耗費之時間為固定 k-NN 之兩倍，因此可先針對其餘部分進行改善，估計預測值方面，則建議以平均法進行預測，最後，若是考慮到多個特徵向量，本研究所提出之特徵向量權重計算方式則可有效提升準確度。

5.1 結論

1. 本研究利用車輛偵測器 (VD) 資料與電子收費系統 (ETC) 資料進行現況旅行時間預測，並在旅行時間計算上進行調整，使用 ETC 資料進行主路段計算，前後端再以 VD 速率資料差補，計算出更為實用之交流道間旅行時間。
2. 相較於以往文獻，蒐集更多特徵向量進行預測，藉由實測分析得知本研究所提出之特徵向量組合，在各項評估指標較原先使用單一特徵向量有改善之效果。
3. 本研究所提出之變動 k-NN 法，為將劉軒寧 (2016) 之方法進行調整，減少歷史資料庫數量並設立 k 值區間，以避免 k 值變動至 1，產生較大誤差之情況發生，且同樣證明於變動量 i 為 2 時，相較變動量為 1 有著較佳之效果，但與固定 k-NN 相比，並無顯著提升準確率之效果。
4. 使用兩種不同估計預測值方式進行預測，並發現平均法與 Rank-based Weight 相比，有著較好的準確率，Rank-based Weight 會提高距離最近之樣本權重，並降低其餘樣本權重，在資料量大的情況下，與現況資料進行比對後所選取之數個樣本，其與現況資料間相似度差異不大，因此過度放大最近樣本可能帶來更多的預測誤差。
5. 在考量多個特徵向量下，本研究所提出之特徵向量權重計算方式，相較於等值之權重，能有效提升預測準確率。
6. 對於本研究所進行的改良，其對於預測準確率重要性之排序應為選取最佳特徵向量組合為先，並使用平均法及以相關係數作為特徵向量權重，

再去考慮 k 值是否需變動。

5.2 建議

1. 本研究將歷史資料庫分為平日與假日，未來可納入更多情境，如連假、颱風天、濃霧等，且與事故資料進行結合，建立車禍發生時旅行時間預測。
2. 本研究對於模式估計預測值部分僅以平均法及等級排序法進行預測，建議可加入更多估計方法進行分析，如依照各歷史資料庫樣本與現況資料之距離倒數來估計；另在特徵向量權重上，可嘗試不同計算方式，並與本研究提出之方法進行比較。
3. 本研究共收集四個月之資料，建議未來可擴大蒐集時間範圍，觀察是否更加能提升準確性。
4. 可多加對於變動 k 值之機制進行研究與改善。

陽明交大
NYCU

參考文獻

- [1] Alghamdi, T., Elgazzar, K., Bayoumi, M., Sharaf, T., & Shah, S. (2019, June). Forecasting traffic congestion using ARIMA modeling. In 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC) (pp. 1227-1232). IEEE.
- [2] Billings, D., & Yang, J. S. (2006, October). Application of the ARIMA models to urban roadway travel time prediction-a case study. In 2006 IEEE International Conference on Systems, Man and Cybernetics (Vol. 3, pp. 2529-2534). IEEE.
- [3] Clark, S. (2003). Traffic prediction using multivariate nonparametric regression. *Journal of transportation engineering*, 129(2), 161-168.
- [4] Habtemichael, F. G., & Cetin, M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation research Part C: emerging technologies*, 66, 61-78.
- [5] Jang, J. (2019). Travel-time Prediction Using K-nearest Neighbor Method with Distance Metric of Correlation Coefficient. *The Open Transportation Journal*, 13(1).
- [6] Kwon, J., Coifman, B., & Bickel, P. (2000). Day-to-day travel-time trends and travel-time prediction from loop-detector data. *Transportation Research Record*, 1717(1), 120-129.
- [7] Long, K., Yao, W., Gu, J., Wu, W., & Han, L. D. (2019). Predicting freeway travel time using multiple-source heterogeneous data integration. *Applied Sciences*, 9(1), 104.
- [8] Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4), 818.
- [9] Oh, J. S., Jayakrishnan, R., & Recker, W. (2002). Section travel time estimation from point detection data.
- [10] Qiao, W., Haghani, A., & Hamedi, M. (2013). A nonparametric model for short-term travel time prediction using bluetooth data. *Journal of Intelligent Transportation Systems*, 17(2), 165-175.
- [11] Robinson, S., & Polak, J. W. (2005). Modeling urban link travel time with inductive loop detector data by using the k-NN method. *Transportation research record*, 1935(1), 47-56.
- [12] Van Hinsbergen, C. P., Van Lint, J. W., & Sanders, F. M. (2007). Short term traffic prediction models. In PROCEEDINGS OF THE 14TH WORLD CONGRESS ON INTELLIGENT TRANSPORT SYSTEMS (ITS), HELD BEIJING, OCTOBER 2007.
- [13] Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3-19.

- [14] Wu, J., Cai, Z., & Gao, Z. (2010, August). Dynamic K-Nearest-Neighbor with Distance and attribute weighted for classification. In 2010 International Conference on Electronics and Information Engineering (Vol. 1, pp. V1-356). IEEE.
- [15] 王晉元(2013)，高速公路中長程旅行時間預測模式之建立與應用，國道高速公路總局委託計畫。
- [16] 朱志杰, & 黃家耀. (2013). 使用車輛偵測器和自動車輛辨識之資料預測高速公路旅行時間 (Doctoral dissertation).
- [17] 邱逸彥, & 王晉元. (2013). 利用貝氏理論於旅行時間推估之研究 (Doctoral dissertation).
- [18] 林崇翔. (2018). 使用 GBRT 預測短期國道交通路況. 交通大學資訊科學與工程研究所學位論文, 1-26.
- [19] 陳志華, 謝欣翰, 高果, & 官大勝. (2017). 深度學習應用於交通旅行時間預測. 電工通訊季刊, 92-109.
- [20] 陳建旻, & 王晉元. (2008). 比較 k-NN 模式與時變係數模式對高速公路旅行時間預測之研究 (Doctoral dissertation).
- [21] 許添本. (2020). 專輯序言: 交通控制. 土木水利, 47(4), 3-4.
- [22] 黃冠維, 王晉元「利用卷積神經網路預測市區道路路況之研究」, 國立交通大學運輸與物流管理學系碩士論文, 民國一零七年。
- [23] 劉軒寧. (2016). 變動 k 值之 k-NN 法應用於高速公路旅行時間預測之研究. 交通大學運輸與物流管理學系學位論文, 1-32.
- [24] 蔡繼光, & 卓訓榮. (2008). 高速公路旅行時間預測-以 k-NN 法及分群方法探討 (Doctoral dissertation).