

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

利用信令資料推估用戶常住與工作位置之研究  
The Study on Estimating User's Home and Work Location from  
Mobile Phone Data

研 究 生：王日隆(Wang, Jih-Lung)

指 導 教 授：王晉元(Wang, Jin-Yuan)

中華民國一一一年 七 月

July 2022

利用信令資料推估用戶常住與工作位置之研究  
The Study on Estimating User's Home and Work Location from  
Mobile Phone Data

研究生：王日隆

Student : Jih-Lung Wang

指導教授：王晉元 博士

Advisor : Dr. Jin-Yuan Wang

國立陽明交通大學  
運輸與物流管理學系  
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

July 2022

Taiwan, Republic of China

中華民國一一一年七月

## 誌謝

本篇論文得以順利完成，首先要感謝我的恩師 王晉元老師。兩年前考進交大運管所，很幸運的進入了老師的實驗室接受訓練，並開啟了我的兩年研究生活。在研究期間中，先是擁有自己的想法，僅是研究的開端。而學到如何克服研究過程中的種種困難才是最重要的。每道困難的問題都必須自行思考與尋找解決方法來去處理。這過程可能很累、很焦慮甚至是煩躁，但卻成為訓練我的磨刀石，成就了我的耐力與韌性。有時會在研究過程中不小心迷路，謝謝老師適時的出現並指引前進的方向。有時會很想主觀地表達自己的論述，謝謝老師教導凡事都需要根據與驗證，使我往後的思路更加周全、論述更加堅固。最後想謝謝老師訓練我成為獨立思考的研究生而不在只是聽命行事。最後無論在課業與研究上，或是身教與言教上，老師不辭辛勞的諄諄教誨，對我影響極其深遠且銘記於心。在此致上最深的敬意與謝意。

感謝逢甲大學蘇昭銘老師與陽明交通大學閻姿慧老師在口試時給予寶貴的建議與指導、感謝系上黃家耀老師在論文進度審查期間撥冗時間細閱並給予學生寶貴的建議，使本篇論文更臻完備。

感謝我在碩班碰到的每一位同學。感謝玟蘭好姊妹，每當研究碰到困難或是疑惑一定會先與妳討論！也祝妳在教育心理的路上發光發熱！順利畢業！但記得要適時放鬆讓自己休息噢！感謝盛泳、祐慈和耿峰從大學到研究所這幾年學習路上的陪伴！希望以後還能夠定期旅遊，一起出去玩耍！感謝同窗一起做計劃的子賢，當我學習程式語言的啟蒙導師之一！感謝語芳這兩年一起在 meeting 前焦慮 meeting 後吃鬆餅，做研究做到很煩或不想寫程式時就跑去打擾妳一下也真是辛苦妳了～要怪就要怪子賢提早畢業！感謝盧 lab 的朝瑋與瑋珊，無聊的時候還可以一起聊天或出去買吃的！感謝吃飯糰荏晴、宜芳還有李文一起煩惱學餐與 uber 要叫啥。感謝黃桂一起出差的時，耐心回答我的好奇心，也細心講解出雷達、GoPro 還有路口細節對於運輸安全的必要與重要性！感謝晉瑋讓新竹整個好玩起來，根本就是新竹移動美食地圖，也讓我無聊就可以去姚 lab 串門子、看風景然後出門買飲料。感謝宸誌、偉軒、大成與宜臻一起在姚 lab 耍廢、聊天、看電影，進行一系列紓壓的活動！雖然我們都不是姚 lab 的。感謝于容出差好夥伴，台灣在地美食專家、Google 專業評論員，以後工作要加油知道嗎哈哈！感謝憲驊修課好戰友，啟發式到運經的路上還好有你一起打拼！感謝書豪、斯涵、凱韻和欣好一起分擔助教業務！

最後要感謝我的父母全心全意支持我，讓我無後顧之憂順利完成學業，還有感謝我的承婷，碩班生活有妳陪伴真的很幸運！可以互相訴苦或談心，放假的時候一起出去散散心吃好吃的！以後在新竹也要一起加油！要感謝的人真的太多，又不能流於形式僅謝天，再次誠心感謝陪我走過碩班生活的所有人。

王日隆 謹誌

2022 年 7 月 於新竹交大

# 利用信令資料推估用戶常住與工作位置之研究

學生：王日隆

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

本研究目的為推估居民的常住與工作等重要位置。由於居民的移動方式與活動位置會根據常住與工作等重要位置的分佈，在時間上表現出高度的順序與規律性，進而決定了城市的結構與居民活動方式，因此推估重要位置這類訊息在許多應用領域都很受用且重要，其中包括都市規劃、交通運輸等相關分析。以往了解居民的重要位置與移動方式通常是透過家戶旅行調查來收集居民的旅運資料。但家戶旅行調查資料收集成本高，並且收集的樣本數不足，常使分析結果造成偏誤。因此，近年來學者常使用信令資料進行許多交通運輸領域相關研究，例如：人類移動模式判斷與起訖點推估等等。但在利用信令資料進行分析前，須先優化訊號飄移等問題，避免訊號飄移問題影響後續推估用戶常住與工作位置之結果。

本研究主要貢獻主要有三點：首先，本研究自行設計航向變化法(Bearing Difference Method)針對訊號飄移問題進行優化。其次，本研究設計與改良一系列判斷用戶常住與工作位置流程，用以推估出更精準的常住與工作位置。最後，透過實驗設計找出預測績效最佳的模式進行推估用戶常住與工作位置。經資料實測與模式驗證後，證明航向變化法確實有效提升預測位置精準度。使用本研究完整的資料清洗流程，並直接預測用戶常住與工作位置的模式最為精準。27 位用戶的資料實測結果中，預測與實際常住位置平均誤差為 148 公尺；預測與實際工作位置平均誤差為 269 公尺。顯示本研究之成果適合運用於交通運輸領域中。

關鍵詞： 交通運輸、信令資料、訊號飄移、航向變化法

# **The Study on Estimating User's Home and Work Location from Mobile Phone Data**

Student: Jih-Lung Wang

Advisor: Dr. Jin-Yuan Wang

Department of Transportation and Logistics Management

National Yang Ming Chiao Tung University

## **Abstract**

The purpose of this study is to estimate significant locations such as home and workplace. Due to the distribution of these significant location, roughly defines the spatial structure and activity patterns of urban area. The movement pattern and significant location of residents exhibit pretty much regularity in terms of sequence and time. Estimated significant location information is useful and important in many applications, including urban planning, transportation and other related research. There are three main contributions of this study :

1. This study designed Bearing Difference Method to optimize phenomenon of signal oscillation.
2. This study designs and improves a series of processes for estimating the user's home and workplace to get more accurate.
3. The model with the best predictive performance is extracted through the experimental design to facilitate the estimation of the user's home and workplace.

According to the actual test and model verification results of the mobile phone data, it is indicated that the Bearing Difference Method can effectively improve the accuracy of the predicted location. In addition, using the model designed in this study is the most accurate. Among the 27 users, the average error between the estimated and the actual home location is 148 meters. and the average error between the estimated and actual workplace is 269 meters. It shows that the results of this study are suitable for application in the field of transportation.

**Keywords:** transportation, mobile phone data, signal oscillation, bearing difference method

# 目錄

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.....	1
1.3 研究範圍.....	2
1.4 研究流程與架構.....	3
第二章 文獻回顧.....	5
2.1 信令資料.....	5
2.2 資料清洗.....	6
2.2.1 信令資料不合理移動速度定義.....	6
2.2.2 信令資料訊號飄移判別與處理.....	6
2.3 旅次產生方法.....	8
2.3.1 旅次停留定義.....	8
2.3.2 旅次停留點判斷方法.....	9
2.3.3 常住和工作地點判斷.....	10
2.4 小結.....	11
第三章 模式建構.....	13
3.1 資料清洗.....	13
3.1.1 基於速度判斷與刪除法.....	13
3.1.2 基於模式判斷與修正法.....	16
3.2 移動與停留狀態判斷.....	22
3.3 常住與工作地點判斷.....	24
3.3.1 常住位置判斷.....	24
3.3.2 工作位置判斷.....	27
3.4 實驗設計.....	28
3.4.1 四種模式.....	28
3.4.2 四種比較情境.....	29
3.5 小結.....	30
第四章 研究結果.....	31
4.1 資料收集內容.....	31
4.2 資料清洗結果.....	31
4.3 移動與停留狀態判斷結果.....	35
4.4 常住與工作位置判斷結果.....	36

4.4.1 常住位置預測結果 .....	37
4.4.2 工作位置預測結果 .....	42
4.4.3 航向變化法預測績效與模式驗證 .....	49
4.4.4 移動停留狀態判斷流程預測績效與模式驗證 .....	53
4.5 小結 .....	55
第五章 結論與建議 .....	58
5.1 結論 .....	58
5.2 建議 .....	59
參考文獻 .....	60

# 圖目錄

圖 1.4.1 研究流程圖 .....	3
圖 2.2.2.1 利用 K-means 集群方法修正增量集群方法前後比較.....	8
圖 2.3.1.1 不同距離集群門檻值下的不同位置數與集群停留時間關係.....	9
圖 2.3.2.1 將多個停留點集群示意圖 .....	10
圖 2.4.1 來回事件與疑似來回事件比較圖 .....	11
圖 3.1.1 兩階段資料清洗流程 .....	13
圖 3.1.1.1 基於速度判斷法之邏輯 .....	15
圖 3.1.1.2 速度判斷與刪除方法示意圖 .....	15
圖 3.1.2.1 來回事件與疑似來回事件的方向變化 .....	19
圖 3.1.2.2 移動方向的變化角度出現次數直方圖 .....	19
圖 3.1.2.3 疑似來回事件發生區域示意圖 .....	20
圖 3.2.1 移動與停留狀態判斷流程 .....	22
圖 3.2.2 (左圖)顧凱嫻(2021)與(右圖)本研究之涵蓋範圍概念 .....	23
圖 3.2.3 移動與停留狀態判斷流程範例示意圖 .....	23
圖 3.3.1.1 用戶每日類型分類流程圖 .....	24
圖 3.3.1.2 常住網格性質篩選示意圖 .....	26
圖 3.3.2.1 推估工作網格流程 .....	28
圖 3.4.1.1 模式 1-1 與模式 1-2 .....	29
圖 3.4.1.2 模式 2-1 與模式 2-2 .....	29
圖 4.2.1 資料清洗流程圖 .....	32
圖 4.2.2.1 號用戶跳點規則修正示意圖 .....	33
圖 4.2.3.1 號用戶航向變化法修正示意圖 .....	33
圖 4.2.4.1 號用戶時窗法修正示意圖 .....	34
圖 4.2.5.1 號用戶資料清洗前後差異圖 .....	34
圖 4.3.1.1 號用戶移動狀態之信令點與移動軌跡 .....	36
圖 4.3.2.1 號用戶停留狀態之信令點分佈 .....	36
圖 4.4.1 常住與工作位置預測流程圖 .....	36
圖 4.6.1.1 號用戶某日實際常住與工作位置附近之信令資料分布圖 .....	57

## 表目錄

表 2.2.1.1 各文獻對於不合理移動速度定義總整理 .....	6
表 2.2.2.1 各文獻針對訊號飄移之處理方法整理 .....	8
表 3.1.2.1 跳點規則範例 .....	17
表 3.1.2.2 跳點規則錯誤修正式意圖 .....	18
表 3.1.2.3 移動方向計算範例 .....	20
表 3.1.2.4 移動方向的變化角度計算範例 .....	20
表 3.1.2.5 航向變化法步驟三與步驟四之範例 .....	21
表 3.3.1.1 示意用戶每日分類示意 .....	25
表 3.3.1.2 示意推估用戶的常住候選網格 .....	25
表 3.3.1.3 推估每日候選常住網格性質 .....	26
表 3.3.1.4 近似比對推估最終常住代表網格 .....	27
表 4.1.1 前五筆原始信令資料 .....	31
表 4.2.1 各流程中資料清洗筆數 .....	32
表 4.3.1 移動與停留狀態篩選結果 .....	35
表 4.4.1 四種模式概述表 .....	36
表 4.4.2 四種情境概述表 .....	37
表 4.4.1.1 四種模式預測與實際常住位置誤差結果 .....	37
表 4.4.1.2 模式 1-1 預測常住位置結果 .....	38
表 4.4.1.3 模式 1-2 預測常住位置結果 .....	39
表 4.4.1.4 模式 2-1 預測常住位置結果 .....	40
表 4.4.1.5 模式 2-2 預測常住位置結果 .....	41
表 4.4.2.1 四種模式預測與實際工作位置誤差比較(單位:公尺) .....	43
表 4.4.2.2 有問題用戶之工作候選網格近似比對結果表 .....	43
表 4.4.2.3 有問題用戶之前三順位工作候選網格出現天數結果表 .....	44
表 4.4.2.4 預測與實際整體位置最大、最小與平均誤差值 .....	44
表 4.4.2.5 模式 1-1 預測工作位置結果 .....	45
表 4.4.2.6 模式 1-2 預測工作位置結果 .....	45
表 4.4.2.7 模式 2-1 預測工作位置結果 .....	46
表 4.4.2.8 模式 2-2 預測工作位置結果 .....	47
表 4.4.3.1 比較情境 1 之有無航向變化法預測常住位置統計結果表 .....	50
表 4.4.3.2 比較情境 1 之有無航向變化法預測工作位置統計結果表 .....	50
表 4.4.3.3 比較情境 1 之有無航向變化法預測整體位置統計結果表 .....	51
表 4.4.3.4 比較情境 1 預測績效與模式驗證結果表 .....	51
表 4.4.3.5 比較情境 2 之有無航向變化法預測常住位置統計結果表 .....	51
表 4.4.3.6 比較情境 2 之有無航向變化法預測工作位置統計結果表 .....	52

表 4.4.3.7 比較情境 2 之有無航向變化法預測整體位置統計結果表 .....	52
表 4.4.3.8 比較情境 2 預測績效與模式驗證結果表 .....	53
表 4.4.4.1 比較情境 3 之有無移動與停留狀態判斷流程預測常住位置統計結果表 .....	53
表 4.4.4.2 比較情境 3 之有無移動與停留狀態判斷流程預測整體位置統計結果表 .....	54
表 4.4.4.3 比較情境 3 預測績效與模式驗證結果表 .....	54
表 4.4.4.4 比較情境 4 之有無移動與停留狀態判斷流程預測常住位置統計結果表 .....	54
表 4.4.4.5 比較情境 4 之有無移動與停留狀態判斷流程預測整體位置統計結果表 .....	55
表 4.4.4.6 比較情境 4 預測績效與模式驗證結果表 .....	55

# 第一章 緒論

## 1.1 研究背景與動機

在一天當中，居民多數時間會待在常住、學校或工作等少數幾個重要的位置。他們的移動方式與活動位置也會在時間上表現出高度的順序與規律性。如同 Schneider 等人(2013)所述，大多數旅次都是城市或某區域內的旅行、通勤或娛樂購物等日常生活的短途旅次，而這些短途旅次會呈現規律性。而長時間與長距離的旅次對於居民日常生活而言是不常發生的，例如：國際航班與城際間的移動。因此，在一都市中重要位置的分佈與居民們集體的流動性大致決定了城市的結構與其活動方式，使重要位置這類訊息在許多應用領域都很受用，包括都市規劃、交通運輸與商業地選址 (Tongsinoot 等人，2017)。

以往了解居民的重要位置與移動方式通常是透過家戶旅行調查來收集居民或其家戶成員之旅運資料，並做進一步分析。然而，家戶旅行調查需要投入較高的金錢與人力成本進行樣本收集，使此份資料更新頻率非常緩慢，並且收集的樣本數不足，因此還需進一步放大才能做後續相關研究的分析，但經過放大後的樣本會導致分析過程與結果容易產生偏差，使得預估與實際的結果仍有一大段差距。近年來智慧型手機與基地台日漸普及，被動生成的手機資料(以下簡稱為信令資料)。信令資料記錄了每一用戶使用手機(收發訊息或與網路串聯)當下之時間與位置經緯度，使其運用層面相當廣泛。因此近年來學者常應用信令資料於交通領域許多研究，例如：判斷人類移動模式(Jiang 等人，2017)、旅行需求估計(Toole 等人，2015)、旅行時間預測(Lai 等人，2016)與起訖點推估(Alexander 等人，2015)等。信令資料在交通領域如此被廣泛運用的原因是基於以下優點：1. 資料收集容易且成本較家戶旅行調查資料低。2. 資料量龐大。3. 資料帶有時空性質。

根據國家通訊傳播委員會的「109年通訊市場調查結果報告」指出，台灣16歲以上民眾有使用手機的比例高達95.4%，代表台灣大多數民眾高度依賴使用手機的現況，使信令資料適合運用於了解居民重要位置分布與移動方式。但是，即便信令資料可用於廣泛的研究與分析，仍須先優化信令資料之缺陷「訊號飄移」，才能使後續分析結果更具有參考價值。訊號飄移的成因是基於眾多因素導致訊號時強時弱，使信令資料點在短時間內於兩至多個位置間頻繁切換。因此本研究著重於優化訊號飄移的問題，並提出新的資料清洗方法「航向變化法」用以優化訊號飄移問題，並加強資料清洗效果。其次透過改良一系列判斷與分析流程，用以預測出用戶常住與工作等重要位置，使得結果在交通運輸領域的應用更為可靠。

## 1.2 研究目的

根據研究背景與動機的探討，將研究目的整理為下列兩點：

1. 本研究欲深入探討信令資料訊號飄移之問題與特性，並針對問題的現象進行優化，使信令資料後續運用在交通運輸領域甚至在其他領域上的應用更加可靠。
2. 本研究欲透過手機的信令資料，經過一系列判斷與分析流程，預測出用戶的常住與工作位置。

### 1.3 研究範圍

本研究所使用的信令資料產生時機與資料內容如下：

1. 本研究所使用的信令資料是基於兩種情況被動產生，首先是基於受測者使用手機的情況下產生，例如使用者打電話、收發訊息與使用網路服務，只要手機與基地台互動就會被動產生當下時間與位置經緯度；其次，若使用者長時間未使用手機，導致手機與基地台無互動，只要手機有在正常開機的情況下，基地台會固定一段時間，如一小時，呼叫手機回傳當下時間與位置經緯度。
2. 本研究使用信令資料的內容包含：使用者匿名化 ID、時間以及所在位置經緯度。

## 1.4 研究流程與架構

本研究欲預測用戶常住與工作位置所建立之研究流程如圖 1.4.1 所示，共分為 6 個步驟。

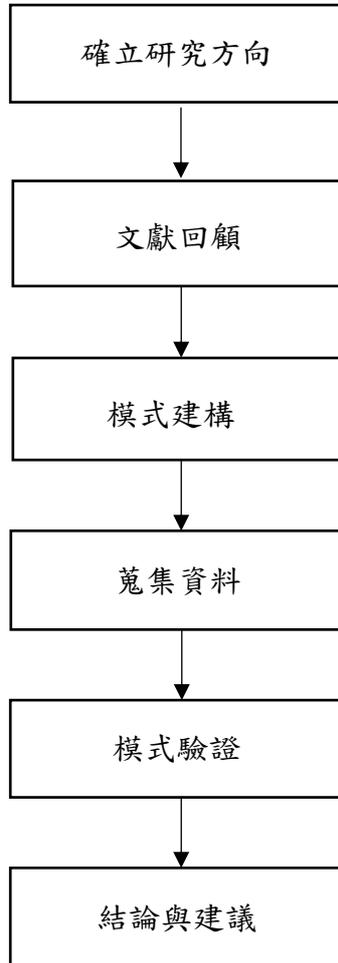


圖 1.4.1 研究流程圖

### 1. 確立研究方向

首先基於研究背景與動機來定義問題，了解信令資料應用於交通運輸領域的價值與需克服的困難。

### 2. 文獻回顧

本研究透過文獻回顧的方式回顧過往文獻的作法、成果與未能處理的問題。首先回顧信令資料處理與清洗的流程與方式，其中包括訊號飄移資料的判斷與處理。其次回顧過往文獻如何判定用戶的停留或移動狀態，最後回顧如何預測用戶常住與工作位置。

### 3. 模式建構

本研究依序講解資料清洗、移動停留狀態判斷與常住及工作位置判斷等三個流程。首先根據文獻回顧了解信令資料飄移的特性以及過往文獻未能處理的問題，並在資料清洗流程中加入本研究自行設計的航向變化法對訊號飄移問題進行優

化。信令資料清洗結束後進行信令資料的移動與停留狀態判斷。最後將停留狀態的信令資料帶入用戶常住與工作位置之判斷，並輸出預測常住與工作位置的結果以及預測與實際之常住與工作位置的誤差距離。本研究根據前述流程分別設計出四種不同模式進行實驗設計，用以探討本研究自行設計的航向變化法是否能有效優化訊號飄移問題並提升預測用戶常住與工作位置精準度。

#### 4. 蒐集資料

資料來源為計畫中所招募共 27 位主要生活於桃園市的居民，身分包含學生、上班族以及退休人士等，收集其於受測期間的 C 公司信令資料。受測者除同意提供信令資料外，亦提供其實際常住與工作位置做為本研究預測的常住與工作位置的標準答案。

#### 5. 模式驗證

透過統計檢定來檢驗本研究自行設計的航向變化法是否能夠有效優化信令資料訊號飄移之問題並提升預測位置精準度。以及檢驗移動停留狀態判斷流程所萃取停留狀態信令資料是否能夠有效提升預測位置之精準度。

#### 6. 結論與建議

針對本研究推估常住與工作位置與受測者實際提供的常住與工作位置比較與統計驗證結果說明本研究設計之精準度與適用情境。此外，在此部分說明目前常住與工作位置判斷模式可改進的方向與範疇，以及信令資料所面臨的限制與挑戰，提供未來信令資料應用於交通運輸領域中後續研究參考。

## 第二章 文獻回顧

由於本研究旨在優化信令資料之問題並預測出更準確的用戶常住與工作位置，因此除了需要了解旅次定義之外，信令資料本身缺點與問題處理方式也需要透過相關文獻去了解，如資料缺漏與訊號飄移等。本研究文獻回顧主要分為信令資料、資料清洗、旅次定義與產生方法、停留點判斷與常住與工作地點判斷。

### 2.1 信令資料

信令資料為被動產生的資料，資料原先為電信公司用於計費目的而產生的資料(Kujala 等人，2016)。近年來，信令資料被廣泛運用於各個領域，而交通領域也因此份資料的應用受益良多。由於用戶使用手機頻率或基地台固定一段時間呼叫手機回傳信息，每一筆信令資料皆帶有時間與空間的資訊，時間資訊是手機回傳信號當下的時間，而空間資訊通常為基地台的位置或是經過三角定位後的位置(Alexander 等人，2015)，因此透過分析用戶的移動距離與時間變化，可推估出使用者的軌跡與停留點。

信令資料在時間與空間上皆帶有不確定性(Wang and Chen, 2018)，這帶給分析者一項很大的挑戰。在時間不確定上，由於用戶不會無時無刻都在使用行動裝置，因此只要行動裝置未與基地台互動就不會持續產生資料，而要等待基地台過了一段時間主動呼叫行動裝置，行動裝置才會回饋並產生資料。根據用戶使用行動裝置的情況不同，造成有些用戶的信令資料非常多，但有些用戶的信令資料卻非常稀少的現象，因此文獻常以「時間稀疏性」來描述信令資料在時間上缺漏的現象(Hasan 等人，2017)。

在空間不確定性上，主要分為三大問題。首先是用戶位置預測的準確性會因為基地台設置的密度疏密而有所不同，如 Airsage 公司信令資料所記載用戶位置平均誤差落在 300 公尺，但在市區通常誤差僅在 50 至 100 公尺之間(Hard 等人，2016)。再者，不同電信公司會因應其需求不同，使得信令資料的空間單位記載形式亦不相同，有些傳統的信令資料其位置資訊是以基地台經緯度為代表，誤差可從數百公尺至數公里(Järv 等人，2014)，在空間上的單位會較大。而有些信令資料的位置資訊是經由三角定位計算出的經緯度為代表，其誤差可縮小至 200 公尺至 300 公尺(Alexander 等人，2015)，空間單位會較小。最後，除了空間單位記載形式不同會影響研究者推估用戶真實位置之精準度，信令資料本身也帶有一項重大問題，其為「訊號飄移」，訊號飄移帶來最明顯的問題為信令資料記載的空間資訊在二至多個位置之間反覆移動，但這些資訊不能反映用戶的實際移動狀況(Lee and Hou, 2006)。其主要原因為基地台接受行動裝置訊號會因為空間條件而被影響，如用戶周遭的建築物密度與天線類型等(Caceres 等人，2012)，另外，用戶們使用頻率也會影響基地台接收訊息，若當下某基地台附近用戶們使用手機頻

率高，導致訊號會被其他基地台接收，平攤該基地台負載容量，因此使訊號位置有跳動的現象(Ahas 等人，2010)。

## 2.2 資料清洗

過往文獻有透過偵測與修正不合理移動速度之信令資料的方法進行資料清洗。另外，過往文獻也透過偵測信令序列呈現在二至多個位置之間反覆移動的訊號飄移狀態進行資料清洗。以下 2.2.1 與 2.2.2 小節分別回顧過往文獻如何透過偵測不合理移動速度或反覆移動之現象進行優化信令資料的問題。

### 2.2.1 信令資料不合理移動速度定義

Iovan (2013)說明由於相鄰基地台為了平衡各基地台訊號的負載容量以及各基地台信號強度所導致的訊號飄移現象會產生信令點不合理的移動速度，因此兩連續信令點所計算出的移動速度大於 200km/hr，則會被視為飄移的點，隨後進行刪除。

Ma 等人(2013)提出了應用於信令資料清洗的模糊邏輯的專家系統(expert system,ES)，其中訂定若信令點之間移動速度超過 160.9 km/hr，則會被視為飄移的點，若有發現信令點有來回事件發生，則會將其刪除與合併。

Kujala 等人(2016)在估計兩地之間的旅行時間前，先將移動速度大於 100km/hr 的信令點視為不合理移動速度的信令點，並予以汰除。

Hasan 等人(2017)再 Dhaka 市透過信令資料做旅行時間計算時提出信令點間移動速度若大於 40km/hr 即為不合理移動速度，由於 Dhaka 市交通基礎設施規劃不善，長年飽受交通壅擠之苦，市區平均速度約 3km/hr~40km/hr 之間，且市民多為走路上班，因此速度上限訂定的如此之低。最後表 2.2.1.1 為各文獻對於不合理移動速度定義總整理。

表 2.2.1.1 各文獻對於不合理移動速度定義總整理

文獻	不合理移動速度定義
Iovan (2013)	信令點間移動速度大於 200km/hr
Ma 等人(2013)	信令點間移動速度大於 160.9km/hr
Kujala 等人(2016)	信令點間移動速度大於 100km/hr
Hasan 等人(2017)	信令點間移動速度大於 40km/hr

### 2.2.2 信令資料訊號飄移判別與處理

Schlaich 等人(2010)提出利用跳點規則修正訊號飄移的問題，該研究表示判別用戶軌跡至少需要三個不同位置，因此，一個資料區間須由三個不同位置組成，若在此資料區間中發現這三個不同位置之間有多次來會跳動的形況發生，則會視為有訊號飄移，判別公式如下：

$$jumpiness_{factor} = \frac{\text{資料區間中信令點跳動次數}}{\text{不同位置總數}} > 2$$

舉例說明，假設有一信令資料經過時間排序形成一個信令序列，此序列中含有四個不同位置 ABCD，將不同位置總數設定為 3，意即資料區間是由三個不同位置所組成。信令序列為 ABCABABCABCCBD，根據三個不同位置依時間先後順序形成一資料區間 ABCABABCABCCB，可以發現 D 位置這筆資料已超出資料區間所設定三個不同位置總數，因此不屬於該資料區間所涵蓋資料。位置間的跳動如 A 位置跳往 B 位置為一次跳動，因此可以發現資料區間中信令點跳動次數為 11 次，11 次位置間跳動除以 3 個不同位置總數，因此  $jumpiness_{factor}$  約等於 3.67，而超過 2 就代表該資料區間為飄移區間，則對飄移區間進行修正，修正方式以出現次數最多的位置為代表，因此飄移區間修正過後為 BBBBBBBBBBBBBB。

Wu 等人(2014)提出了 DECREASE algorithm 對信令資料清洗的演算法，共有四步驟，第一步為偵測(detect)，依序找出三種明顯有訊號震盪的情境以及一種疑似訊號飄移的情境，由於前一步驟偵測出大量疑似訊號飄移的事件，透過擴展(extend)與檢查(check)步驟來找出證據使這些疑似訊號飄移事件成立，確定飄移事件後，由最後移除(remove)步驟來將飄移事件移除。

Wang (2014)在研究中提出兩步驟處理訊號飄移的問題，第一步先處理位置不確定性，某一信令序列： $L = \{l_1, \dots, l_n\}$ ，以  $l_1$  為初始集群的質心並計算與下筆信令點的距離差距是否在 1 km 內，假設  $l_1$  與  $l_2$  距離差距小於 1 km，那麼集群這兩點並更新質心，更新後的質心在與  $l_3$  計算距離差距；若  $l_1$  與  $l_2$  距離差距大於 1 km，那麼  $l_1$  集群自成一類， $l_2$  與後續的信令點  $l_3$  計算距離差距，若在距離門檻值內，那麼集群這兩點並更新質心，更新後的質心在與  $l_4$  計算距離差距，以各集群質心代表集群內信令點的位置。第二步處理訊號飄移，首先檢測信令序列中是否有來回事件且來回切換次數超過 3 次，如  $\{A_1, B_2, A_3, B_4, A_5\}$ ，更新震盪序列位置：多數決來決定序列位置，A 出現 3 次；B 出現 2 次，修正結果以 A 為代表位置，如  $\{A_1, A_2, A_3, A_4, A_5\}$ 。

Wang & Chen (2018)透過四步驟處理訊號飄移所帶來位置不確定性的問題。第一步將信令資料依時間排序過後，透過增量集群方法(Incremental clustering method)將時間與位置相近的信令點集群，代表各集群內的信令點位置都是非常相近，以質心代表該集群所有信令點位置；第二步，計算集群內最後一筆與第一筆信令點的時間差，時間差若大於 5 分鐘，則該集群被視為停留點或活動地點；第三步，在此步驟反覆測試增量集群方法中集群半徑之參數；第四步，由於增量集群方法因信令序列時間限制，會導致相近的位置有多個集群，因此透過 K-means 集群方法將位置相近的集群合併，如圖 2.2.2.1。各集群位置確定後將集群內信令點位置修正為該集群質心之位置，隨後透過時窗分析法來修正訊號飄移的問題，設定分析時窗為 5 分鐘，若偵測到該分析時窗內有不合理的來回事件發

生，如 A-B-A，則會將信令序列的位置修正為 A-A-A，以時窗內出現頻率最高的位置為代表。

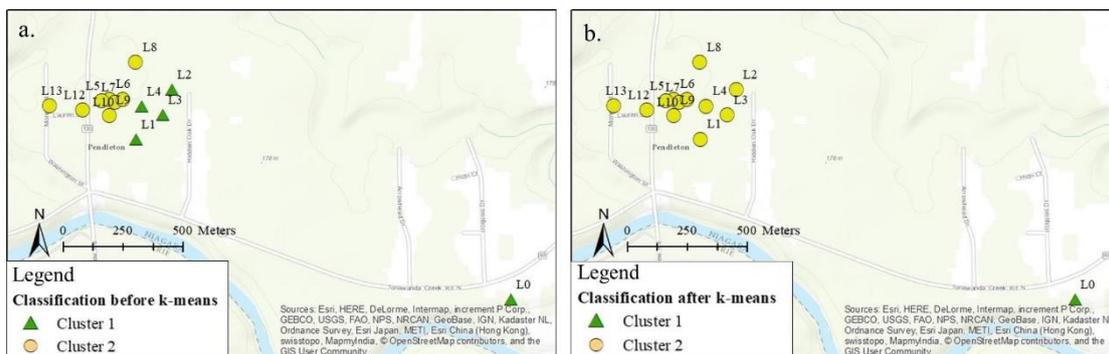


圖 2.2.2.1 利用 K-means 集群方法修正增量集群方法前後比較

顧凱嫻(2021)結合基於速度與模式兩種方法來進行資料清洗，首先基於速度模式判斷法主要參考 Ma 等人(2013)與 Iovan 等人(2013)的方法將不合理移動速度的信令點進行刪除。其次透過模式判斷法，主要結合 Schlaich 等人(2010)的跳點規則與 Wang and Chen(2018)的時窗法進行偵測訊號飄移並將其修正。最後表 2.2.2.1 為各文獻針對訊號飄移之處理方法總整理。

表 2.2.2.1 各文獻針對訊號飄移之處理方法整理

文獻	方法
Schlaich 等人(2010)	跳點規則(Jumpiness rule)
Wu 等人(2014)	DECRE algorithm
Wang (2014)	修正來回事件
Wang & Chen (2018)	時窗分析法(Window-based method)
顧凱嫻(2021)	基於速度與模式判斷法

## 2.3 旅次產生方法

旅次包含目的、停留點、運輸工具與路線等，本研究透過了解過往文獻如何定義旅次，以便於後續區分信令資料移動與停留狀態以及推估用戶常住與工作位置，因此，了解旅次定義對本研究來說是非常重要的一環，後續分為三個小節來回顧過去文獻如何透過信令資料產生旅次。首先，回顧各文獻對於旅次停留定義，參考如何為後續利用空間集群判別停留點與起訖點的分析設定，其次回顧文獻應用在旅次停留點判斷的各種方法，最後當確定停留點後，回顧文獻中如何判定起訖點性質，其中最重要為常住和工作地點的判斷。

### 2.3.1 旅次停留定義

Schlaic 等人(2010)於德國西南部的研究中提出，旅次的構成至少需有三個以上不同地點組成的軌跡，而用戶須待在同一地點超過 60 分鐘，才會將該區域視

為旅次的起點或迄點。

Wang (2014) 提出，用戶需在同一地點停留 10 分鐘以上，該點才會被視為停留點，而考量到信令資料有訊號飄移的問題，因此在研究中分別設定兩停留點間移動距離標準為 500 公尺、1000 公尺與 1500 公尺，最後在本篇研究中考量到位置估計誤差，最終選擇 1000 公尺作為移動距離標準，若用戶在 A 點停留超過 10 分鐘，並移動至 1 公里外的 B 點且在 B 點停留超過 10 分鐘，則將用戶從 A 點移動至 B 點視為一旅次。

Çolak 等人(2015) 以 300 公尺作為旅次移動的定義，且定義停留 10 分鐘代表用戶在該位置有停留。在此篇研究之定義運用判斷波士頓居們的旅次效果良好。

Wang & Chen (2018) 的研究中定義用戶在該地點停留 5 分鐘以上，該地點才會被視為停留點，而兩地點間移動距離門檻是經由增量集群方法反覆測試得出，如下圖 2.3.1.1 表示， $n_{AL}$  表示每人每天訪問不同位置的平均數量、 $R_c$  表示集群的半徑。首先可以發現當  $R_c$  越小， $n_{AL}$  就會越高；但隨著  $R_c$  增加， $n_{AL}$  就會遞減，漸漸收斂找到常見的活動位置，但隨後  $R_c$  持續增加， $n_{AL}$  開始遞增，代表集群開始納入較遠的位置，因此集群停留時間(Duration per cluster)也不斷攀升，因此在本篇文獻最終選擇 1000 公尺當作停留點集群的範圍，兩點需移動超過 1000 公尺才會被視為一旅次。

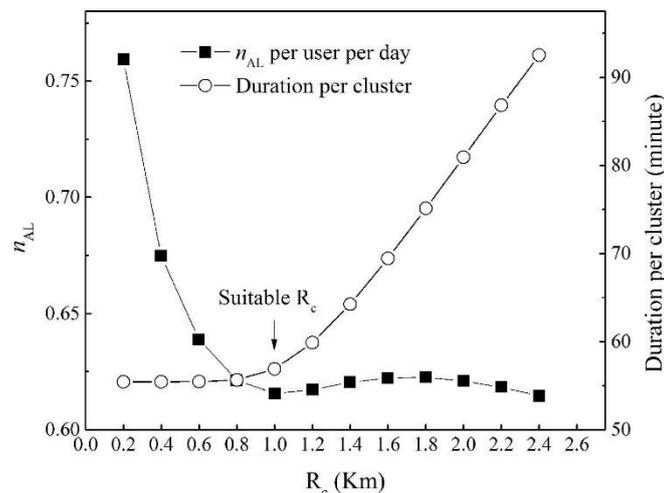


圖 2.3.1.1 不同距離集群門檻值下的不同位置數與集群停留時間關係

### 2.3.2 旅次停留點判斷方法

停留點的判斷對於後續推估用戶的常住或工作地點是至關重要的，若將兩連續信令點間視為一旅次，則會造成大量的短途旅次，並且信令資料訊號飄移的問題造成用戶實際上沒有移動，但資料顯示用戶有移動的情況。造成用戶停留位置不確定性的現象，導致後續推估用戶常住與工作位置時有嚴重的偏誤。

Toole 等人(2015)為了找出停留點，首先將各個用戶的信令資料依時間進行排序成為信令序列， $P = \{P_1(x_1, y_1, t_1), \dots, P_n(x_n, y_n, t_n)\}$ ， $P$  為信令序列、 $x$  為經度、 $y$  為緯度、 $t$  為時間；隨後計算連續兩信令點是否小於距離門檻值，若小

於距離門檻值，則將兩信令點納入候選點資料集( $S$ )， $S = \{P_s(x_s, y_s, t_s), \dots, P_t(x_t, y_t, t_t)\}$ ，候選點資料集中有許多可能會被視為停留點的信令資料，隨後在候選點資料集中計算連續兩信令點是否大於時間門檻值，若大於時間門檻值，則將該信令點的位置視為停留點。

Çolak 等人(2015)研究中將連續信令點距離差距小於 300 公尺且時間差距大於 10 分鐘，則將這些信令點的位置集群，以質心代表位置。但每個用戶因每天可能會有不同移動軌跡，因此可能會在某一小區域有位置相近的停留點 ( $s_2, s_3$ )，而這些停留點可能是指同一個位置，因此找出鄰近停留點的質心( $r_1$ )來代表最終的停留點， $r_1$  也代表停留區域(stay region)，如圖 2.3.2.1 所示。

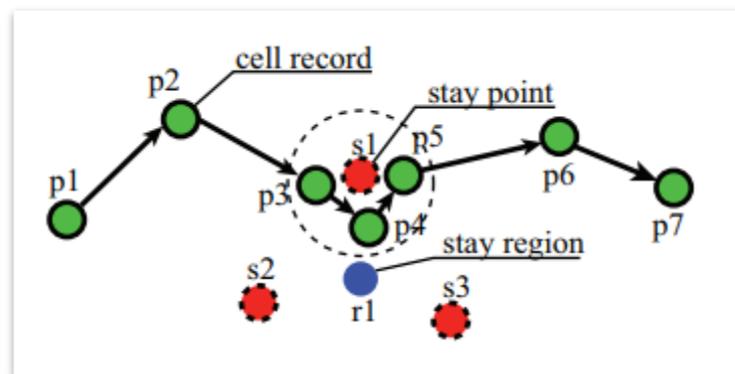


圖 2.3.2.1 將多個停留點集群示意圖

### 2.3.3 常住和工作地點判斷

Ma 等人(2013)透過觀察一整個月內，分析時段區間的方法判別用戶所在位置為常住或工作。會將晚間 8 點至翌日早晨 7 點之間，若 20 天以上都待在同一地點則視為常住位置；而工作地點則是早晨 7 點後至晚間 8 點之間最常發生的位置，若 20 天以上都待在同一地點則視為工作位置。

Toole 等人(2015)，以分析時段區間的方法判別用戶所在位置性質為常住、工作或其他。會將晚間 8 點至翌日早晨 7 點之間，最常發生的停留點位置識別為常住地點；而工作地點則是早晨 7 點後至晚間 8 點之間最常發生的位置；而其他地點是經歷過常住與工作地點篩選與識別後，未分類之地點在常住地點 500 公尺內或該地點每周訪問頻率不超過 1 次，則會被視為其他地點。

Çolak 等人(2015)，以分析時段區間的方法判別用戶所在位置性質為常住、工作或其他。會將晚間 7 點至翌日早晨 8 點之間，最常發生的停留點位置識別為常住地點；而工作地點則是早晨 8 點後至晚間 7 點之間最常發生的位置，其他地點使代表未標記常住或工作地點之位置。

交通部運輸研究所(2018)，以每日 0 點至 6 點為分析時窗來推估用戶的常住網格，並以用戶於每日夜間使用手機情況分為三類，分別為「第一類：無手機使用行為，且無被動接收訊息之用戶」，若用戶該日被分類為第一類，則會取每日第一段出現次數連續為 1 的網格進行近似比對並篩選最頻繁出現的網格為常住網格；「第二類：無手機使用行為，但有被動接收訊息之用戶」，該日為第二類

則會只取每日凌晨兩點至五點最頻繁出現的網格為常住網格；「第三類：開啟飛航模式或關機之用戶」，該日為第三類用戶則會由每日第一個無代表網格的時窗往前或往後篩選常住網格。

顧凱嫻(2021)，參考交通部運輸研究所(2018)並進行改良，在此研究預測用戶常住位置時，首先會將用戶分為一至三類，而第二類用戶是指夜間時段每小時資料超過三筆以上的用戶，若 14 日內有超過 10 日分類為第一類或第三類，則不參考用戶該日為第二類的資料來推估常住網格，因其該日夜間有可能外出，不適合用做預測。另外工作網格的推估則是建議使用 DBSCAN 分析方法。

## 2.4 小結

首先，在資料清洗部分本研究觀察到過往文獻通常使用一種方法來優化訊號飄移其中之一的問題現象。顧凱嫻(2021)將上述資料清洗方法結合來做資料清洗，但仍有問題未被處理。在信令資料的文獻回顧部分提到空間單位的大小也可能影響到誤差的大小，因此衍伸出前述資料清洗未能處理的問題，例如 Wang & Chen (2018)欲透過時窗法修正來回事件，但將其方法運用於空間單位較小的信令資料則可能使得成效不彰，其原因為空間單位較小的資料所引發的問題不僅是來回事事件，也可能發生本研究所提出的「疑似來回事事件」的現象。疑似來回事事件的現象為信令序列從 A 點跳往 B 點再跳往 C 點，而 A 點與 C 點是相近的點，這樣的現象使來回事事件變得不容易偵測，觀察下圖 2.4.1 來回事件與疑似來回事事件的差異比較圖。因此，本研究發現仍有優化信令資料問題的空間。

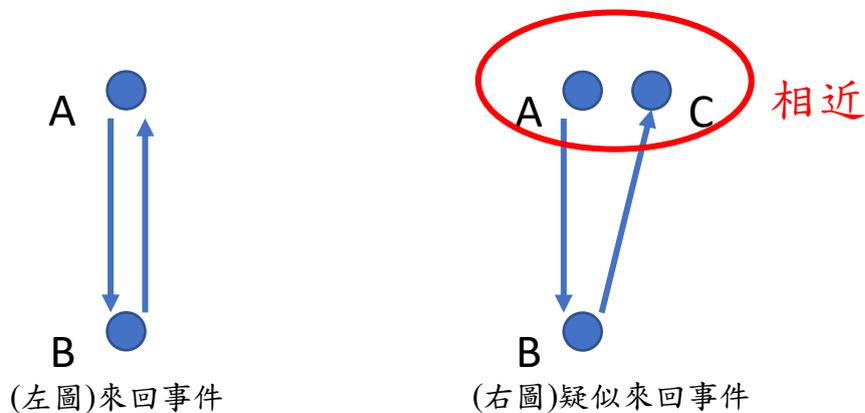


圖 2.4.1 來回事件與疑似來回事件比較圖

過往文獻為了推估用戶的旅次目的或旅次起訖表，往往會先找出用戶的停留點，但找出停留點仍不足以辨別這些停留點的性質，如用戶的常住位置或是工作位置，這使得旅次目的或旅次起訖表難以推估並建立。因此，為了推估用戶的常住與工作位置，過往文獻通常會設定一個分析時段來推估用戶所在位置為常住或工作位置，例如凌晨 12 點至早晨 6 點最常出現的位置為用戶之常住位置。但顧凱嫻(2021)指出其推估用戶常住位置時，不會參考在凌晨 12 點至早晨 6 點出現過多資料的天數，其認為用戶可能於夜間活動外出，因此不適合用於推估用戶常住位置，但本研究認為可能需要一個判別流程來決定每日的資料是否取用，而不

是統一不採用。透過文獻回顧了解過往文獻如何針對某項問題處理的方法與其所碰到的困難之處，並將研究缺口找出，透過後續模式建構針對前述問題進行改良與設計出解決方案，使研究成果更為優異。

## 第三章 模式建構

本研究目的為透過設計更優良的資料清洗方法用以優化信令資料飄移問題，並透過信令資料研擬出一系列演算流程預測用戶常住與工作位置，以利後續交通運輸領域應用。本章先分別講述資料清洗的流程、移動與停留狀態判斷流程、常住與工作位置之判斷流程。另外，分別講述實驗設計中四種模式之差異與目的。

### 3.1 資料清洗

在介紹資料清洗前，在本研究中先定義兩個名詞。首先，信令序列是指將相同用戶的信令資料，經日期與時間排序整理，稱之為信令序列。其次，信令點是指信令序列中的某一筆信令資料。訊號飄移帶來最明顯的問題為信令資料記載的空間資訊在二至多個位置之間反覆移動，但這些資訊不能反映用戶的實際移動狀況(Lee and Hou, 2006)。由此可知，訊號飄移會使用戶看似在短時間內反覆移動且移動數百公尺至數公里的距離，實際為不合理的移動現象。因此，本研究參考顧凱姝(2021)的兩階段資料清洗流程，對訊號飄移問題進行優化。階段一為基於速度的判斷與刪除方法，刪除不合理移動速度的資料；階段二為基於模式判斷與修正階段，依序包含 3 種方法為跳點規則、航向變化法及及時窗法。其中航向變化法為本研究自行設計的資料清洗方法，目的為加強優化訊號飄移問題。圖 3.1.1 為兩階段資料清洗流程。

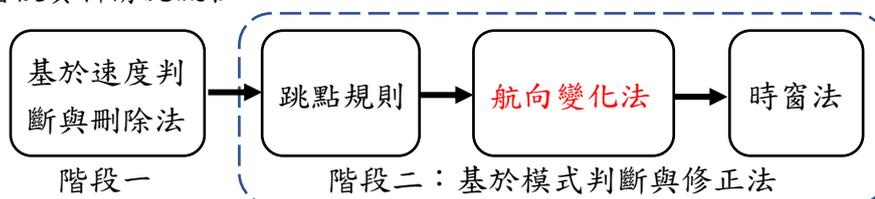


圖 3.1.1 兩階段資料清洗流程

#### 3.1.1 基於速度判斷與刪除法

階段一基於速度判斷與刪除法的目標為偵測信令序列中不合理移動速度的信令點，並將其刪除，主要參考 Ma 等人(2013)、Iovan 等人(2013)與顧凱姝(2021)的方法，判斷邏輯如下圖 3.1.1.1。本研究考量到用戶可能使用不同運具移動，應要訂定不同的不合理移動速度。台灣高鐵最高營運速度為 300km/hr，但基於安全理由，目前台灣高鐵平均行駛速度約 260km/hr，若兩連續信令點位置在高鐵沿線 500 公尺內，則信令點移動速度不可超過門檻值 300km/hr，若移動速度超過門檻值，則判定飄移，隨後進行刪除，移動速度未超過門檻值則標記非飄移；其餘信令點移動速度不可超過門檻值 150km/hr，若移動速度超過門檻值，亦判定為飄移，隨後進行刪除，移動速度未超過門檻值亦標記為非飄移。本研究發現若只刪除一次不合理移動的信令點，卻未檢查刪除過後的信令序列中，是否依然含有不

合理移動速度的點，則會影響到後續起訖點預測結果。因此當每一位用戶的信令點皆判斷完畢，保留非飄移的信令點，進入下一迭代。下一迭代開始再重新判斷其是否有飄移，直至無任何信令點被標記為飄移狀態，則停止此流程。

舉例說明如圖 3.1.1.2，首先為迭代一，在信令序列  $A \rightarrow B \rightarrow C$  中，A 點至 B 點的速度為 560km/hr，明顯為不合理移動速度情境，隨後計算 B 點往 C 點的移動速度為 18km/hr，在迭代一結束前，刪除標記為飄移的 B 點。迭代二開始時，重新計算 A 點往 C 點的移動速度，發現 A 點至 C 點的移動速度為 530km/hr，因此將 C 點標記為飄移的點，在迭代結束時進行刪除。事實上，發生訊號飄移時 B 點與 C 點位置相近，若未重新計算連續兩信令點移動速度，那麼 C 點的亦常飄移現象就不會被偵測出，並且投入後續的分析當中。因此，需檢查信令序列是否還有不合理移動速度的信令點，才能進行下個流程。

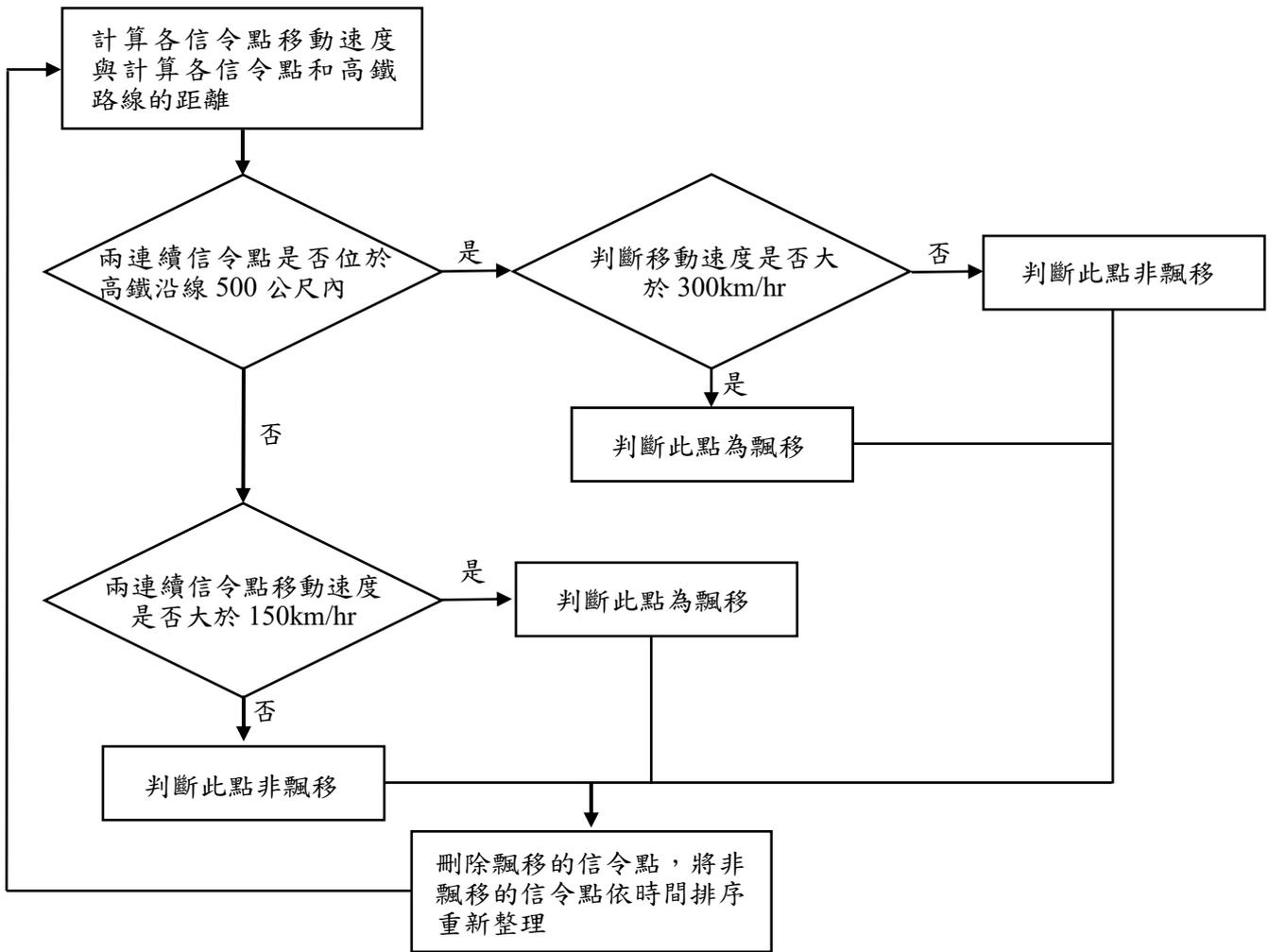


圖 3.1.1.1 基於速度判斷法之邏輯

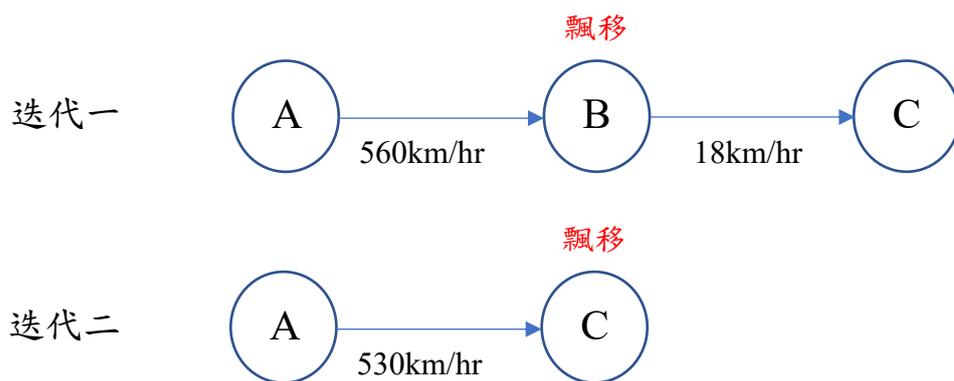


圖 3.1.1.2 速度判斷與刪除方法示意圖

### 3.1.2 基於模式判斷與修正法

階段二為基於模式判斷與修正方法主要目的是偵測信令資料是否在不同位置之間頻繁的切換，並予以修正。本研究參考 Schlaich 等人(2010)之跳點規則與 Wang and Chen(2018)所提出的時窗法，並結合本研究自行設計的航向變化法來進行階段二的資料清洗。

#### 1. 跳點規則：

本研究會將原始信令資料 50 公尺\*50 公尺網格替換為較大尺寸 100 公尺\*100 公尺之網格主要基於兩個原因。首先，原始信令資料空間單位是 50 公尺\*50 公尺的網格，這會導致訊號位置在網格之間的跳動過於頻繁；若將原始網格尺寸合併為較大空間單位的網格，例如 500 公尺\*500 公尺的網格，可能會增加預測用戶位置的誤差，使其在精準度的提升方面會有所限制。再者，交通部運輸研究所(2018)指出，若網格範圍過小可能會導致信令資料飄移之問題變得嚴重或無法克服。後續會於本節一併講解跳點規則操作方式以及過小的網格空間單位會造成的問題。另外，本研究希望分析結果能運用於交通運輸當中，而交通運輸領運分析課題之一的運輸規劃常使用交通分析區域(traffic analysis zone, TAZ)作為旅次起訖點的空間單位，TAZ 是基於人口數所劃分的，通常 3000 人組成一個 TAZ，因此 TAZ 的大小各不相同，但每一 TAZ 面積絕大部分超過 100 公尺\*100 公尺網格的面積，因此 100 公尺的誤差對於運輸規劃相關分析影響不大，甚至可以運用於都市規劃等其他相關分析。

階段二為模式判斷與修正方法，首先會以跳點規則進行資料清洗，參考 Schlaich 等人(2010)的公式以及不同網格總數設定為 4，若資料區間中網格切換次數大於等於 8 次， $jumpiness_{factor}$  大於等於 2 的話就會將資料區間判斷為飄移區間並進行修正，修正方法是以飄移區間內出現最多次的網格為修正代表。舉例說明，從表 3.1.2.1 可以發現由四個不同網格 36877、36878、36865、36864 所組成一資料區間(藍色方框)，資料區間內網格切換次數為 10 次(紅色橫線代表網格切換一次)，可計算出  $jumpiness_{factor}$  等於 2.5，因此，該資料區間即被視為飄移區間，隨後以出現次數最多的網格(36877 網格)修正該飄移區間的位置。

$$jumpiness_{factor} = \frac{\text{資料區間中網格切換次數}}{\text{不同網格總數}} \geq 2$$

表 3.1.2.1 跳點規則範例

日期	時	分	網格	修正後的網格
20200306	21	2	45888	45888
20200306	21	8	45888	45888
20200306	23	6	36877	36877
20200306	23	7	36878	36877
20200306	23	8	36877	36877
20200306	23	9	36877	36877
20200306	23	10	36865	36877
20200306	23	11	36864	36877
20200306	23	12	36878	36877
20200306	23	13	36877	36877
20200306	23	14	36877	36877
20200306	23	15	36865	36877
20200306	23	17	36864	36877
20200306	23	18	36877	36877
20200306	23	21	36878	36877
20200307	5	3	35641	35641
20200307	5	4	35641	35641

Schlaich 等人(2010)的研究中，建立資料區間目的是要檢測區間內位置是否有頻繁切換的現象發生，但其並無對資料區間設定時間限制，會造成資料區間內，信令序列橫跨的時間長度或長或短，因此會造成一些錯誤修正的疑慮，也就是前述提到過小的網格空間單會會造成的問題。以下表 3.1.2.2 為例，表中網格單位為 50 公尺\*50 公尺。用來建立資料區間的不同網格總數設定為 4，因此資料區間會由 A、B、C 與 D 四個不同位置組成(藍色方框)。資料區間內網格切換次數為 15 次，因此 $jumpiness_{factor}$  等於 3.75，因此，該資料區間即被視為飄移區間，隨後以出現次數最多的網格(A 網格)修正該飄移區間的位置。此時間問題開始浮現，實際上，A 與 B 兩位置相近且十分靠近該用戶常住位置，而 C 與 D 兩位置相近且十分靠近公司位置。資料區間無考慮時間限制，因此將此四個位置建立為一個資料區間，其實此一區間隱含著常住位置往工作位置的旅次，然而未到常住位置與工作位置預測流程前，已將此段資料區間的位置修正為出現次數最多的 A 網格，導致在資料清洗階段就已經錯誤修正該資料區間的網格，勢必會影響到後續狀與工作位置預測的結果。因此，若將網格合併為 100 公尺\*100 公尺網格會優化此類問題，因此 A 與 B 兩點會合併為 A 點；C 與 D 兩點會合併為 C 點，這使他們過於頻繁的位置切換穩定下來，避免錯誤修正的疑慮。

表 3.1.2.2 跳點規則錯誤修正式意圖

日期	小時	分鐘	網格	修正後的網格
20200306	6	33	E	E
20200306	6	34	E	E
20200306	7	6	A	A
20200306	7	7	B	A
20200306	7	8	A	A
20200306	7	9	B	A
20200306	7	10	A	A
20200306	7	11	B	A
20200306	7	12	A	A
20200306	7	13	B	A
20200306	7	14	A	A
20200306	7	15	A	A
20200306	8	59	C	A
20200306	9	1	D	A
20200306	9	2	C	A
20200306	9	3	D	A
20200306	9	4	C	A
20200306	9	5	D	A
20200306	12	1	H	H
20200306	12	2	H	H

## 2. 航向變化法：

過往文獻如 Wang (2014)與 Wang & Chen(2018)皆透過資料清洗流程優化訊號飄移中的來回事件。來回事件是指 A 點往 B 點又返回 A 點，如圖 3.1.2.1 之左圖，最明顯的特徵是移動方向(又稱為航向)的變化角度會成 $180^\circ$ ；由於本研究所使用的 100 公尺\*100 公尺的網格皆小於 Wang (2014)與 Wang & Chen(2018)這兩篇文獻所使用的資料空間單位，這導致本研究若只有使用時窗法偵測來回事件進行資料清洗，會忽略其他訊號飄移所產生之問題，如疑似來回事件。而疑似來回事件是指 A 點往 B 點再往 C 點，但 A 點與 C 點是彼此相近的點，如圖 3.1.2.1 之右圖。其移動方向變化角度根據本研究統計，如下圖 3.1.2.2 所示，我們觀察疑似來回事件的方向變化角度介於 $180^\circ \pm 20^\circ$ 之間居多(紅色標住處)。後續移動方向的變化角度介於 $180^\circ \pm 20^\circ$ 之間，就將其視為疑似來回事件。因此，本研究自行設計航向變化法加強優化訊號飄移中疑似來回事件之問題。

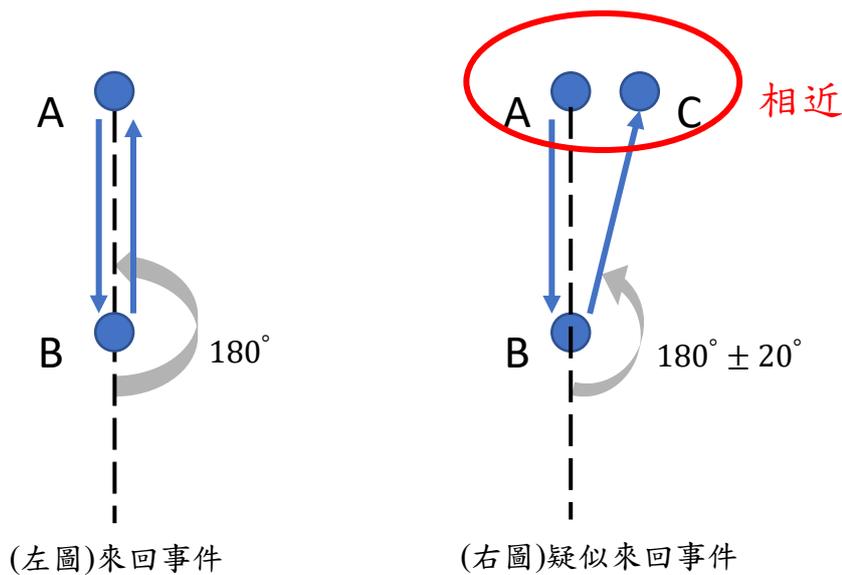


圖 3.1.2.1 來回事件與疑似來回事件的方向變化

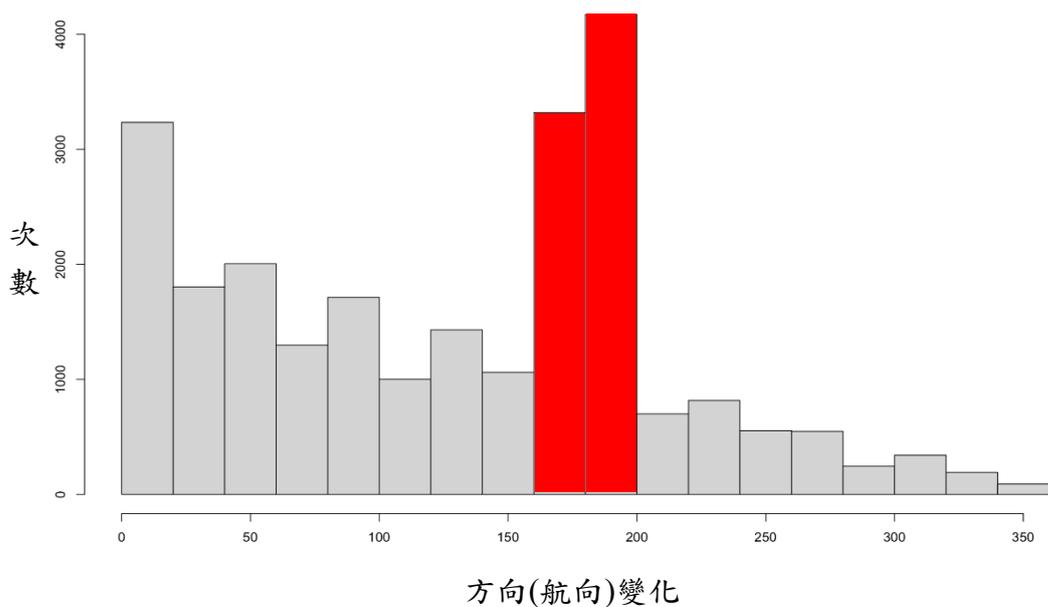


圖 3.1.2.2 移動方向的變化角度出現次數直方圖

航向變化法的操作步驟如下：

- 步驟一：計算每一信令點移動的方向，以角度表示。
- 步驟二：計算兩兩連續信令點移動方向的變化，以角度表示。
- 步驟三：設定不同網格總數並建立資料區間。
- 步驟四：檢測每一資料區間是否有兩次以上(含兩次)來回事件發生。

首先，步驟一要計算每一個信令點移動的方向。舉例說明，假設要計算 A 點的方向，實際上是計算 A 點所在位置經緯度往下一點 B 點所在位置經緯度的移動方向。因此，移動方向是以 A 點往 B 點的方向作為基準，並以角度來表示，

0°往北方、90°往東方、180°往南方、270°往西方。如下表 3.2.2.3 藍色方框處，若 A 點往 B 點的移動方向為89°，代表 A 點大致上朝東方向往 B 點移動。

表 3.1.2.3 移動方向計算範例

日期	時	分	經度	緯度	網格	方向 (角度)
20200308	19	29	121.3165255	25.00711561	A	89
20200308	19	30	121.3235255	25.00711561	B	270
...	...	...	...	...	...	...

當我們計算完所有信令點的移動方向時，接著進入第二步驟。第二步驟計算兩兩連續信令點移動方向的變化。舉例說明方向變化值計算方式，假設[A,B,C]三點組成一信令序列。計算方向變化角度則是將 A 往 B 的方向（角度）減去 B 往 C 的方向（角度）取絕對值。觀察下表 3.1.2.4，方向變化值為藍色方框之欄位，以知 A 往 B 是89°、B 往 C 是270°。兩者相減取絕對值為181°。隨後可觀察圖 3.1.2.3，只要方向變化值經過計算後落在 $180^\circ \pm 20^\circ$ 內(淡橘色區域)，即被識別為疑似來回事件的發生。因此，A 至 C 的181°方向變化屬於疑似來回事件。

表 3.1.2.4 移動方向的变化角度計算範例

日期	時	分	經度	緯度	網格	方向(角 度)	方向變 化值(角 度)
20200308	19	29	121.3165255	25.00711561	A	89	181
20200308	19	30	121.3235255	25.00711561	B	270	181
20200308	19	31	121.3225255	25.00711561	C	89	181
20200308	19	32	121.3235255	25.00711561	B	270	39
20200308	19	33	121.3225255	25.00711561	C	309	20
...	...	...	...	...	...	...	...

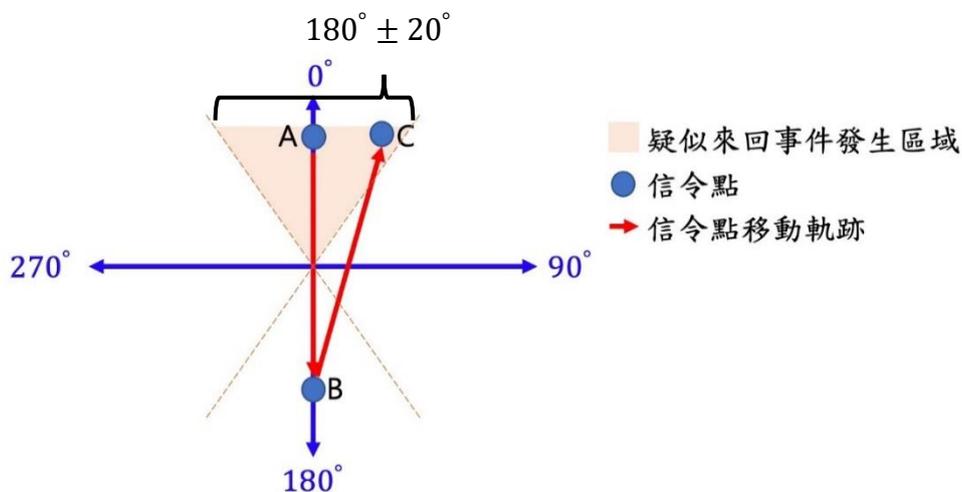


圖 3.1.2.3 疑似來回事件發生區域示意圖

計算完移動方向變化後，透過步驟三建立資料區間，建立資料區間方法與跳點規則相同。而步驟四的目標在於檢測每一資料區間是否有兩次以上(含兩次)疑似來回事件發生，若資料區間超過兩次疑似來回事件的發生，就會將該資料區間視為飄移區間，隨後以出現次數最多的網格修正該飄移區間的位置。觀察下表 3.1.2.5，資料區間由〔A,B,C,D〕四個位置組成(藍色方框處)。隨後偵測資料區件內，共有三次疑似來回事件發生，方向變化值介於 $180^{\circ} \pm 20^{\circ}$ 之間(紅色標住處)，因此將其視為飄移區間並進行修正，最後以出現次數最多的網格(A 網格)修正該飄移區間的位置。

表 3.1.2.5 航向變化法步驟三與步驟四之範例

日期	時	分	經度	緯度	網格	方向	方向變化值	修正後的網格
20200308	19	23	121.3155255	25.00711561	E	89	無變化	E
20200308	19	24	121.3165255	25.00711561	A	無移動	無變化	A
20200308	19	25	121.3165255	25.00711561	A	無移動	無變化	A
20200308	19	26	121.3165255	25.00711561	A	無移動	無變化	A
20200308	19	27	121.3165255	25.00711561	A	無移動	無變化	A
20200308	19	28	121.3165255	25.00711561	A	無移動	無變化	A
20200308	19	29	121.3165255	25.00711561	A	89	181	A
20200308	19	30	121.3235255	25.00711561	B	270	181	A
20200308	19	31	121.3225255	25.00711561	C	89	181	A
20200308	19	32	121.3235255	25.00711561	B	270	39	A
20200308	19	33	121.3225255	25.00711561	C	309	無變化	A
20200308	19	50	121.3185255	25.01011561	D	無移動	無變化	A
20200308	19	51	121.3185255	25.01011561	D	180	148	A
20200308	20	7	121.3185255	25.00111561	G	328	239	G
20200308	20	8	121.3145255	25.00711561	H	89	180	H
...	...	...	...	...	...	...	...	...

### 3. 時窗分析法：

階段二最後方法為時窗分析法，以 5 分鐘為一分析時窗，若在分析時窗內發現來回事件，則將該時窗內信令點網格進行修正。舉例說明，若在五分鐘內發現網格跳動呈現為 A-B-A，則將此序列修正為 A-A-A，以此類推。

### 3.2 移動與停留狀態判斷

移動與停留狀態判斷流程主要參考運輸研究所(2018)與顧凱嫻(2021)年之研究，此流程主要有兩個目的。首先，本研究目的在於預測出用戶常住與工作位置，供後續運輸規劃分析參考。但在預測位置前，若無區分信令資料移動或停留狀態則會造成過多雜訊輸入至預測位置流程之中，最後可能影響到預測結果。因此，在此流程主要用於區分信令點為移動與停留狀態，隨後保留停留狀態之信令點進入最後用戶的常住與工作位置判斷流程，圖 3.2.1 為移動與停留狀態判斷之流程。接下來介紹此流程概念。

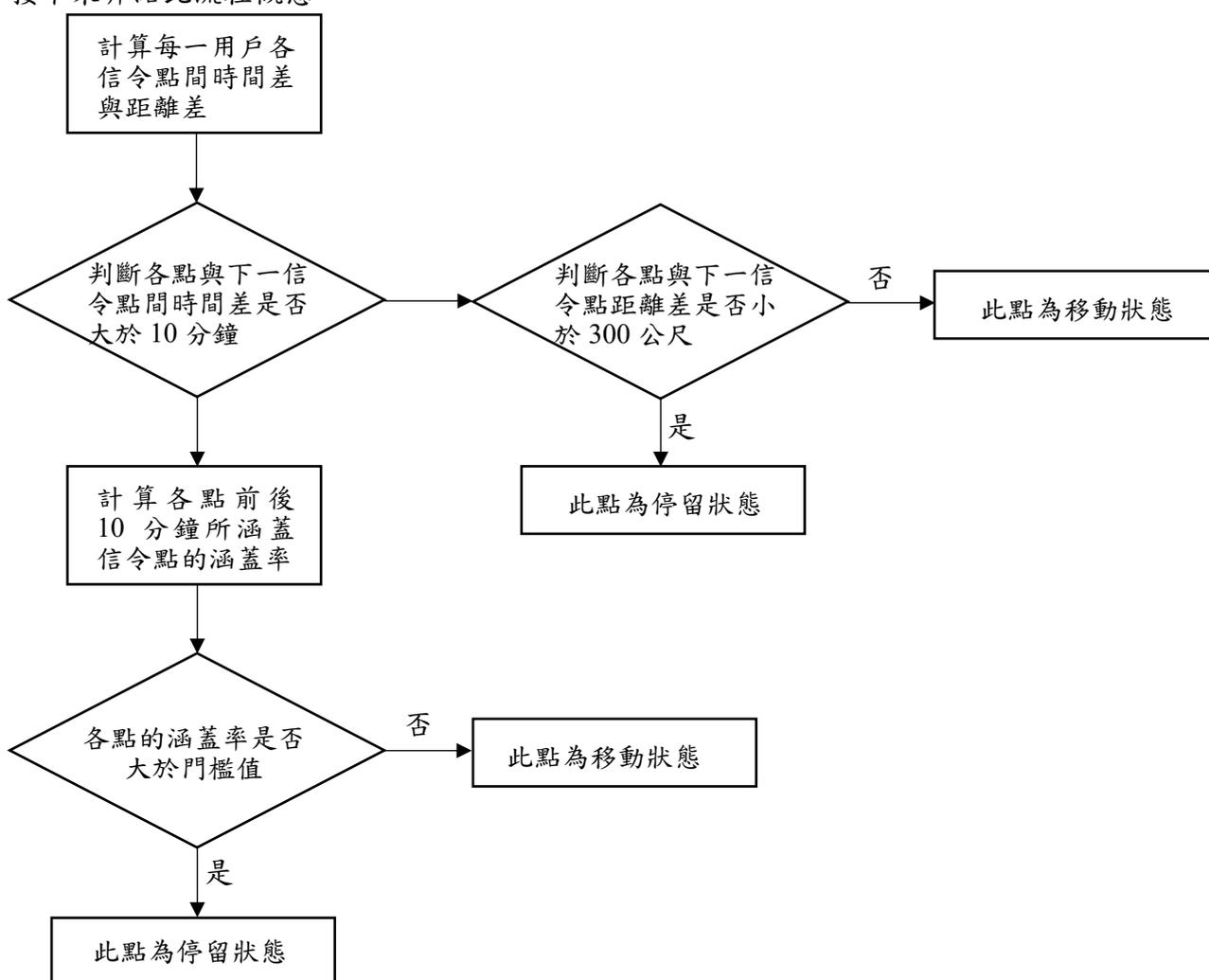


圖 3.2.1 移動與停留狀態判斷流程

如圖 3.2.1，移動與停留狀態判斷流程中，信令點之時間差分別以大於 10 分鐘與小於 10 分鐘來分辨信令點的移動與停留狀態。假設欲判斷 A 點是否為停留狀態，但其與下一點時間差大於 10 分鐘，但距離差在 300 公尺涵蓋範圍內，則該點即識別為停留狀態，反之為移動狀態；若時間差小於 10 分鐘，則會先計算 A 點前後 10 分鐘內之信令點是否落於以 A 點為中心所圍出涵蓋範圍內，並計算

涵蓋率，若涵蓋率大於等於 0.5，則將 A 點識別為停留狀態。移動與停留狀態判斷流程中的時間差與涵蓋率是參考顧凱嫻(2021)的研究所設定。另外，顧凱嫻(2021)在研究中所使用的網格空間單位為 500 公尺\*500 公尺，涵蓋範圍距離定義為 710 公尺(一環，黃色區域)與 1420 公尺(二環，藍色區域)，如圖 3.2.2 左圖。而本研究所使用的網格空間單位為 100 公尺\*100 公尺，由於空間單位小於 500 公尺\*500 公尺之網格，因此涵蓋範圍距離設定為較為保守的 300 公尺(二環，橘色區域)，如圖 3.2.2 右圖。

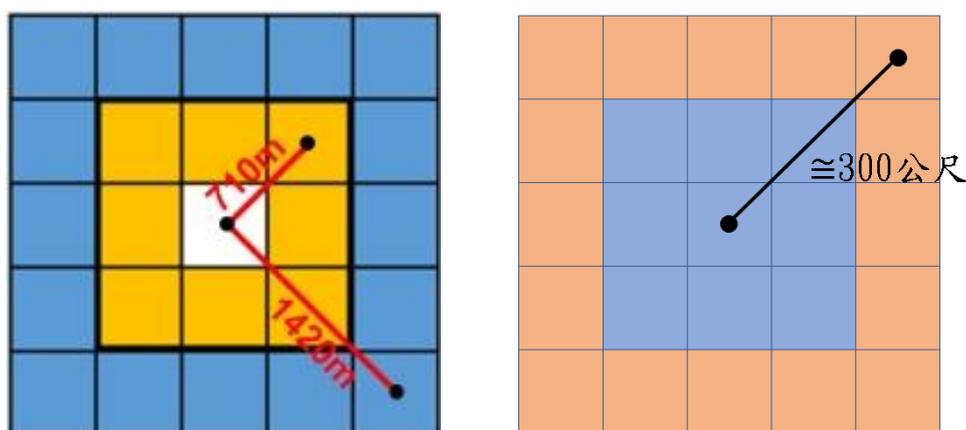


圖 3.2.2 (左圖)顧凱嫻(2021)與(右圖)本研究之涵蓋範圍概念

如圖 3.2.3，舉例說明移動與停留狀態判斷流程的操作方式。欲判斷紅色點 A 的狀態為何，假設 A 點前後 10 分鐘涵蓋共 15 個信令點，其中有 11 個信令點落入涵蓋範圍內，涵蓋率計算方式為  $\frac{11}{15} \cong 0.73$ ，則 A 點涵蓋率為 0.73 高於 0.5 門檻值，則將 A 點視為停留狀態，反之，涵蓋率小於 0.5 則視為移動狀態。若 A 點前後十分鐘內無其他信令點，則直接與其下一點進行距離差比較，發現 A 點與其下一點距離差為 286 公尺，在 300 公尺距離門檻值之內，因此將 A 點視為停留狀態，反之，距離大於 300 公尺則視為移動狀態。

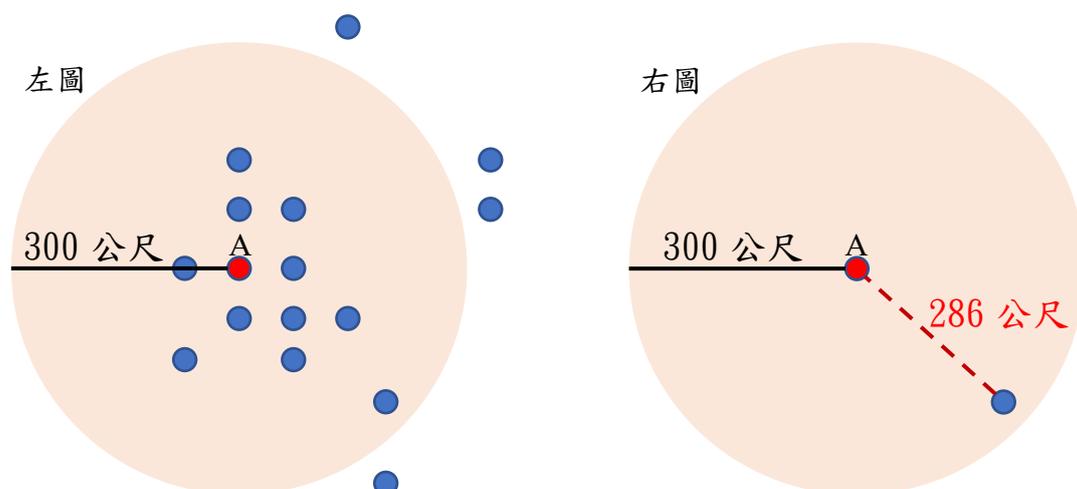


圖 3.2.3 移動與停留狀態判斷流程範例示意圖

### 3.3 常住與工作地點判斷

#### 3.3.1 常住位置判斷

參考 Toole 等人(2015)與 Çolak 等人(2015)提出在分析時窗內，以出現最頻繁的網格為常住位置代表。由於研究區域的不同，本研究主要參考交通部運輸研究所(2018)定義的合理就寢時間(凌晨 0 點至早晨 6 點)為常住網格篩選流程之分析時窗，其中也會針對常住網格篩選流程進行改良。本研究判斷常住網格的流程分為四步驟。第一步驟為區分用戶每日使用手機的類別，將用戶每日在夜間的使用型態分為兩類來推估常住網格。「第一類：正常開機」與「第二類：開啟飛航模式或關機」，分類方式以 0~6 時之間，每一小時內資料須超過三筆信令資料才可推估該小時常住候選網格，並以多數決方式決定該小時的常住候選網格。若該小時內無超過三筆資料，則無法推估該小時的候選網格，隨後只要分析時段中任三個小時內每一小時資料筆數不超過三筆，代表該日用戶可能開啟飛航模式或關機，即被視為第二類，反之皆為第一類，圖 3.3.1.1 為用戶分類篩選流程圖。

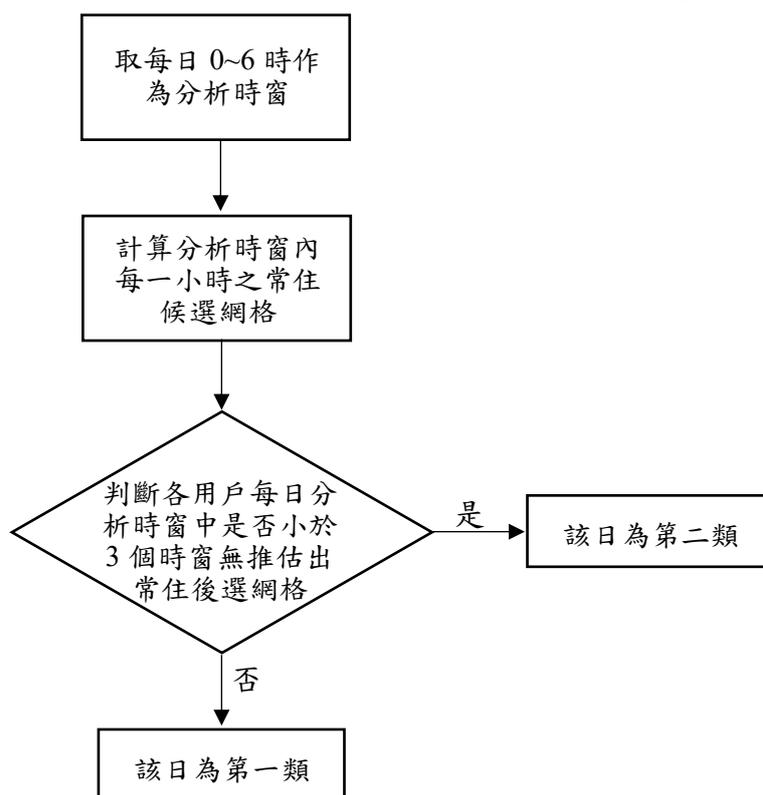


圖 3.3.1.1 用戶每日類型分類流程圖

以下表 3.3.1.1 為例，觀察如何實際區分用戶每日類別。觀察表 3.3.1.1 之左表，0~6 時的分析時窗中，僅有一個時窗內無超過 3 筆資料用以推估常住候選網格，因此標記為“-”，則 1 號用戶在 3 月 6 號則被分類為第一類。表 3.3.1.1 之右

表為 1 號用戶 3 月 7 號時的資料，分析時段內無任何代表小時的候選網格被推派出，已超過三個時窗無代表小時的候選網格資料，因此該日被分類為第二類。

表 3.3.1.1 示意用戶每日分類示意

第一類				第二類			
用戶	日期	小時	網格	用戶	日期	小時	網格
1	20200306	0	83925	1	20200306	23	84423
		1	-			20200307	0
		2	84922		1		-
		3	84423		2		-
		4	84424		3		-
		5	84423		4		-
		6	84423		5		-
		6	-				

第二步驟為判斷用戶每日常住候選網格，若該日類別被分為第一類，已知 0~6 時每一時窗的常住候選網格，隨後會以多數決推派該日的常住候選網格。若為第二類，則會往昨日最後一個有出現資料的時窗推估該日的常住候選網格。下表 3.3.1.2 為推估用戶每日常住候選網格，首先如上表 3.3.1.1 左表可以發現 84423 為出現頻率最高之網格，出現次數共三次，因此推估 03/06 該日常住候選網格為 84423，如表 3.3.1.2 之左表；而 3.3.1.1 右表呈現第二類用戶的情況，0 至 6 點皆無推派出該小時候選網格，則須往昨日前一個有資料的時段推估其常住候選格，因此回到 03/06 23 時推估常住候選網格為 84423，記為該用戶 03/07 之常住候選網格，如表 3.3.1.2 右表。

表 3.3.1.2 示意推估用戶的常住候選網格

第一類				第二類			
用戶	日期	類型	網格	用戶	日期	類型	網格
1	20200306	Type1	84423	1	20200307	Type2	84423

文獻回顧提到顧凱嫻(2021)指出其研究不考慮夜間出現過的資料的天數進行推估用戶常住位置。但本研究認為其做法也會刪除有用的資訊，應該要加入一個判別步驟決定資料是否用於後續預測，而不是直接刪除。第三步驟主要目的是判斷每位用戶每日常住候選網格性質，由於有些用戶於夜間外出或在外遊玩，因此可透過此流程篩選並汰除不具參考價值的每日常住候選網格。此步驟會逐筆判斷每日常住候選網格是否與其他日的常住候選網格距離在 500 公尺內，設定 500 公尺為距離門檻值的原因是基于本步驟僅用於汰除離多數常住候選網格過遠的網格，下一步驟會有更細緻的篩選，更小的距離門檻值。因此，若欲判斷的

該日候選網格與五成以上其餘天數的候選網格相近在 500 公尺內，則該日會被標註為”Home”，反之則為”Other”(代表該日的常住候選網格與多數的常住候選網格較不相近)，將每日常住候選網格標註完畢後，將標註為”Other”的常住候選網格淘汰。第三步驟範例如下表 3.3.1.3 與圖 3.3.1.2 所示，經過計算後，發現 03/09 所推估出常住候選網格 84922 與其他天常住候選網格相距甚遠，超過 500 公尺，因此標註為”Other”(紅色標記處)，會於進入第四步驟前刪除。反之，彼此相近的常住候選網格則標記為”Home”，保留至第四步驟進行最終常住網格的篩選。

表 3.3.1.3 推估每日候選常住網格性質

用戶	日期	網格	標記
1	20200302	84423	Home
	20200303	84424	Home
	20200304	83925	Home
	20200305	84424	Home
	20200306	84423	Home
	20200307	84423	Home
	20200308	84424	Home
	20200309	84922	Other
	20200310	84423	Home
	20200311	84423	Home
	20200312	84423	Home
	20200313	84423	Home
	20200314	84423	Home
	20200315	83925	Home
	20200316	84423	Home

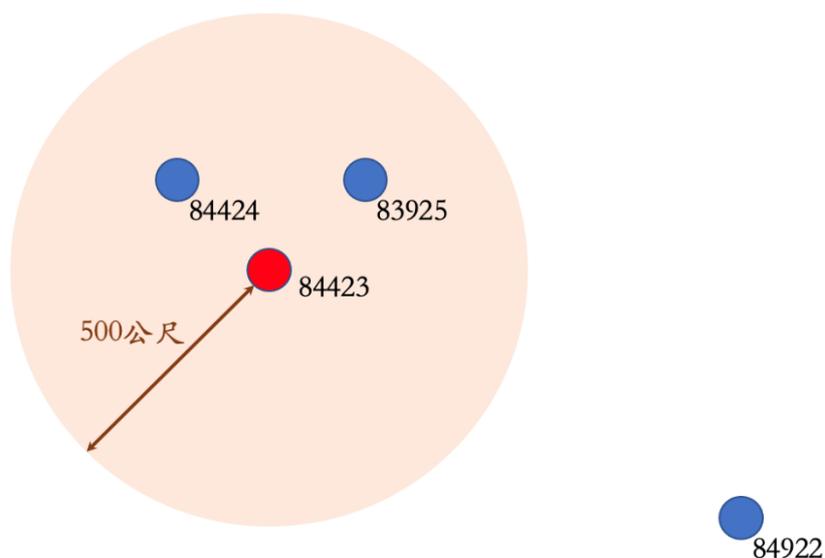


圖 3.3.1.2 常住網格性質篩選示意圖

第四步驟利用近似比對推估用戶的最終常住代表網格，近似比對方法分為兩階段的計算，階段一為依序以每日常住候選網格作為中心網格，計算共有幾天(包括網格中心)的常住候選網格與中心網格在距離門檻值內，為了更精準推估常住網格所在位置，因此門檻值設定為 200 公尺。階段二隨即將階段一相同網格計數數量進行加總，最後推估加總次數最多的網格為該用戶最終常住代表網格。以下表 3.3.1.4 為例，首先開始計算階段每日候選網格的出現天數，84423 網格與其他網格皆在距離門檻值 200 公尺內，因此包含自己為中心網格計算 1 次出現次數，加上其他 13 日出現次數，共為 14 日。而 84424 與 83925 兩常住候選網格皆只包含 84423，但不包含彼此，因此他們出現天數為 9(與 84423 在 200 公尺內)加上 1(自己為中心網格)，等於 10。階段二直接將相同網格在階段一出現的次數進行加總即可，家總結果為 84423 出現共 126 次、84424 出現共 30 次與 83925 出現共 20 次。最後將 84423 推估為此用戶的最終常住網格。

表 3.3.1.4 近似比對推估最終常住代表網格

用戶	日期	網格	標註	階段一	用戶	網格	階段二
1	20200302	84423	Home	14	1	84423	126
	20200303	84424	Home	10		84424	30
	20200304	83925	Home	10		83925	20
	20200305	84424	Home	10			
	20200306	84423	Home	14			
	20200307	84423	Home	14			
	20200308	84424	Home	10			
	20200310	84423	Home	14			
	20200311	84423	Home	14			
	20200312	84423	Home	14			
	20200313	84423	Home	14			
	20200314	84423	Home	14			
	20200315	83925	Home	10			
	20200316	84423	Home	14			

### 3.3.2 工作位置判斷

由於本研究主要目的在於探討加入航向變化法進入資料清洗流程是否能減少預測用戶常住與工作位置的誤差，因此在工作位置推估方法與常住位置推估方法相同，以利對於後續分析進行驗證與比較。推估工作位置方法僅更改分析時窗，本研究參考顧凱嫻(2021)所設定早上 8 點至晚間 8 點的分析時窗，每日工作候選網格以分析時窗內出現最頻繁的網格為工作位置代表，隨後利用近似比對推估最終工作網格為該用戶工作位置。由於就學與工作的性質相似，如長時間固定停留

於一位置，因此在本研究中工作網格包括就學網格。圖 3.3.2.1 為推估工作位置流程。

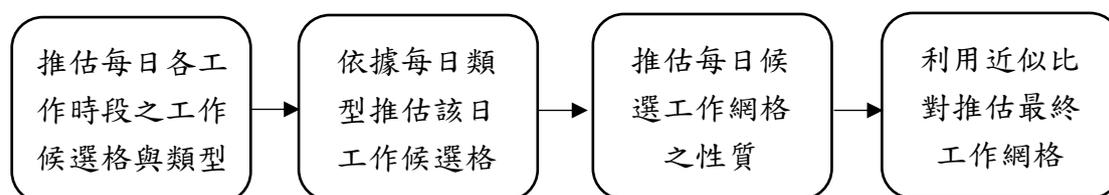


圖 3.3.2.1 推估工作網格流程

## 3.4 實驗設計

### 3.4.1 四種模式

本研究實驗設計中包含四個模式，分別為模式 1-1、模式 1-2、模式 2-1 以及模式 2-2，如圖 3.4.1.1 與圖 3.4.1.2。分為四種模式主要基於兩個原因。首先，為了檢驗加入本研究自行設計的航向變化法是否能夠有效優化信令資料訊號飄移之問題。其次檢驗改良後的整體預測用戶常住與工作位置流程，希望能夠提升預測用戶常住與工作位置之精準度。

模式 1-1 與模式 1-2 僅包含兩流程，資料清洗流程結束後直接進行用戶的常住與工作位置判斷流程，如圖 3.4.1.1。模式 1-1 與模式 1-2 主要區別在於模式 1-1 的資料清洗流程中有加入航向變化法，而模式 1-2 則無加入航向變化法。模式 2-1 與模式 2-2 為三流程，分別為資料清洗、移動停留狀態判斷與常住與工作位置判斷，如圖 3.4.1.2。模式 2-1 與模式 2-2 主要區別在於模式 2-1 的資料清洗流程中有加入航向變化法，而模式 2-2 則無加入航向變化法。

模式 1(兩個模式 1-1&1-2)與模式 2(兩個模式 2-1&2-2)，主要差異在於模式 2 多加入移動停留狀態判斷流程，主要原因在於資料清洗流程與移動停留狀態判斷流程皆可能有助於減少預測用戶常住與工作位置之誤差，為了能夠清楚了解在資料清洗流程中加入航向變化法是否有效優化訊號飄移問題，因此做此區分。過往文獻如 Ma 等人(2013)、Toole 等人(2015)與 Çolak 等人(2015)通常僅透過模式 1-2 預測用戶常住與工作位置。另外，顧凱嫻(2021)是透過模式 2-2 預測用戶常住與工作位置。本研究考量到在資料清洗流程中加入航向變化法優化訊號飄移問題以及用戶實際情況中有移動與停留狀態，因此模式 2-1 為本研究理想的模型(圖 3.4.1.2 紅色方框處)。

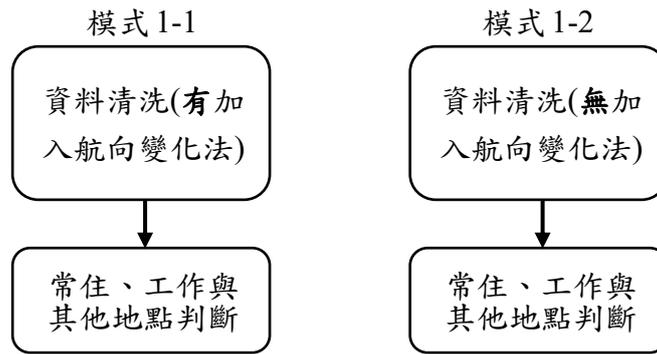


圖 3.4.1.1 模式 1-1 與模式 1-2

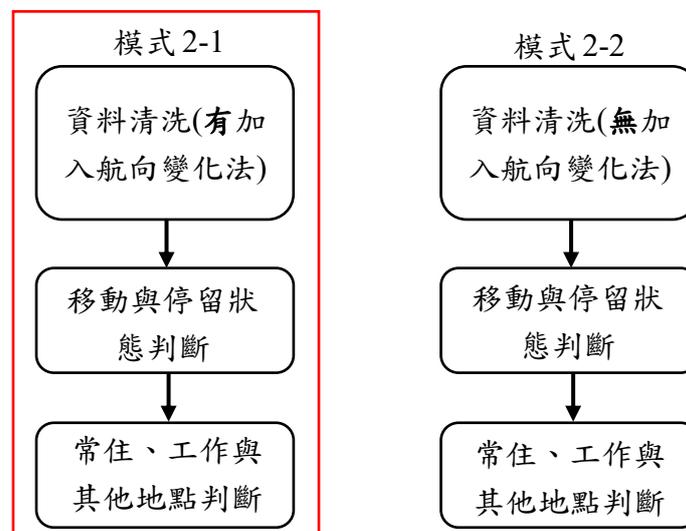


圖 3.4.1.2 模式 2-1 與模式 2-2

### 3.4.2 四種比較情境

前述提及航向變化法與移動停留狀態判斷流程皆可能提升預測用戶的常住與工作位置之精準度。為了要清楚了解加入兩方法前後是否能够有效減少預測用戶的常住與工作位置誤差，因此設計四種比較情境用於評估與驗證航向變化法以及移動停留判斷流程是否能夠有效減少預測用戶位置之誤差，隨後分別透過三種統計檢定方法進行航向變化法與移動停留狀態判斷流程預測績效之模式驗證。下述為四種比較情境詳細內容：

- 比較情境 1: 模式 1-1 與模式 1-2 兩模式皆無加入移動停留狀態判斷流程下，比較有加入航向變化法之模式 1-1 預測績效是否優於無加入航向變化法之預測績效。意即探討模式 1-1 預測結果誤差是否小於模式 1-2。後續比較情境以此類推。
- 比較情境 2: 模式 2-1 與模式 2-2 兩模式皆有加入移動停留狀態判斷流程下，比較有加入航向變化法之模式 2-1 預測績效是否優於無加入航向變化法之模式 2-2 預測績效。

- 比較情境 3：模式 1-2 與模式 2-2 兩模式在資料清洗流程中皆無加入航向變化法下，比較有加入移動停留狀態判斷流程之模式 2-2 預測績效是否優於無加入移動停留狀態判斷流程之模式 1-2 預測績效。
- 比較情境 4：模式 1-1 與模式 2-1 兩模式在資料清洗流程中皆有加入航向變化法下，比較有加入移動停留狀態判斷流程之模式 2-1 預測績效是否優於無加入移動停留狀態判斷流程之模式 1-1 預測績效。

### 3.5 小結

本研究旨在優化訊號飄移問題並求出更精準的用戶常住與工作位置。首先透過一系列資料清洗流程處理訊號飄移，在此流程中航向變化法為本研究提出用於加強優化訊號飄移中疑似來回事件之現象。隨後透過移動與停留狀態判斷區分每筆信令資料的性質，將停留狀態的信令資料輸入用戶的常住與工作位置判斷。而在最後用戶的常住與工作位置判斷流程中，本研究加入判斷每日候選網格之性質，欲提升後續用戶位置判斷結果之品質。最後分別將四種模式的預測與實際位置誤差結果透過四種比較情境探討航向變化法與移動停留狀態判斷流程之預測績效。

## 第四章 研究結果

本章呈現本研究透過 27 位用戶的信令資料進行實際測試，在資料清洗流程與移動停留狀態判斷流程中，透過視覺化與流程結果來檢視是否有達到預期的成果。最後討論四種模式預測與實際位置誤差結果，以及四種比較情境所評估航向變化法與移動停留狀態判斷流程之預測績效。

### 4.1 資料收集內容

本研究之研究區域為桃園市，收集的資料包含桃園市 27 位用戶基本資料與用戶 14 日內的信令資料。收集用戶基本資料的欄位包含姓名、性別、職業類別、聯絡方式、手機型號、常使用運具、實際常住地址以及工作或學校地址。用戶 14 日內的信令資料是由 C 公司提供，原始信令資料如表 4.1.1，資料內容依序包含匿名化用戶、日期、時間、網格 ID 與經緯度，其中位置經緯度是以 50 公尺\*50 公尺的網格中心點為代表。非業務組 14 日信令資料收集期間為西元 2020 年 3 月 2 日至西元 2020 年 3 月 16 日；業務組 14 日信令資料收集期間為 2020 年 3 月 16 日至 2020 年 4 月 1 日。其中有 20 位用戶為非業務組(一般用戶)，其特性是有固定常住與工作位置，資料總比數共 119420 筆；7 位用戶為業務組(物流士)，由於業務上需要不斷送貨或是移動，其特性為工作位置不固定，資料總比數共 62462 筆。將業務組與非業務組信令資料加總後，筆數共 181882 筆信令資料。本研究後續會參考用戶所提供的實際常住與工作位置，進行研究結果探討與模式驗證。

表 4.1.1 前五筆原始信令資料

用戶	日期	小時	分鐘	網格 ID	經度	緯度
3	20200302	0	4	376151068	121.2578	24.9431
5	20200303	0	25	374646059	121.2374	24.9391
6	20200302	22	50	375543060	121.2424	24.9531
1	20200302	0	7	377455052	121.292	24.99
7	20200303	0	3	375249031	121.2118	24.9761

### 4.2 資料清洗結果

進行資料清洗前，先將同一位用戶信令資料依照其時間進行排序成信令序列，隨後才進入資料清洗階段。此流程主要是偵測訊號飄移的信令資料並將其刪除或修正，使其能減少後續預測用戶常住與工作位置的誤差。資料清洗流程如下圖 4.2.1 所示：

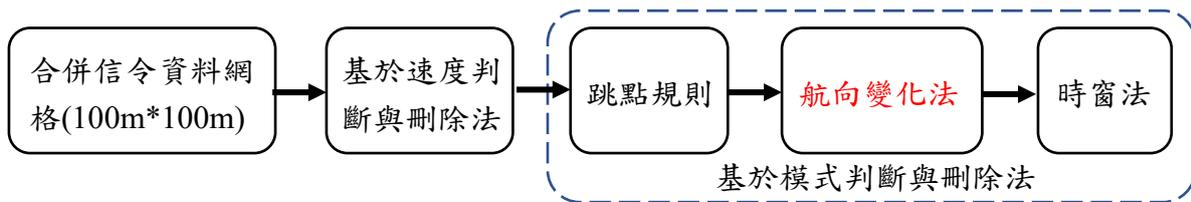


圖 4.2.1 資料清洗流程圖

表 4.2.1 各流程中資料清洗筆數

步驟依序	刪除或修正的資料筆數
基於速度判斷法	刪除：297 筆，約佔整體資料 0.16%
跳點規則	修正：19189 筆，約佔整體資料 11%
航向變化法	修正：29950 筆，約佔整體資料 17%
時窗法	修正：599 筆，約佔整體資料 0.3%

依序觀察各流程資料清洗狀況，以表 4.2.1 呈現兩階段資料清洗流程中每一方法所刪除或修正的資料筆數。首先，從表 4.2.1 可以發現利用速度判斷法僅刪除約整體資料 0.16%，其原因為訊號飄移所造成不正常移動速度在網格尺寸較小的資料中難以被偵測出來，因為訊號皆在較近的距離飄移與跳動，並且本研究所使用資料最小的時間單位為分鐘，因此利用時間與距離所推估的移動速度會被每分鐘時間間距所模糊，使用戶移動速度看似都在合理移動速度門檻內，而不被刪除。隨後進入基於模式判斷法的第一階段跳點規則，跳點規則主要用於修正一個過於頻繁跳動的資料區間，共修正整體資料 11%。跳點規則修正前後效果差異如圖 4.2.2。後續圖 4.2.2 至 4.2.5 中出現綠、黃與橘三色圓圈分別代表在該位置或該區域信令資料出現次數，出現次數少於 10 為綠色、出現次數介於 11 次至 99 次為黃色、出現次數超過 100 次以上則為橘色。圖 4.2.2 之左圖為跳點規則修正前，可以發現黃色圓圈內數字為信令點在該位置出現次數，藍色軌跡為信令點相互跳動所連成之軌跡，紅色線條為該網格中心位置與 1 號用戶實際常住位置之距離，橘色座標則為 1 號用戶實際常住位置，信令序列於四個網格中心間頻繁切換。圖 4.2.2 之右圖則為跳點規則修正後之結果，可以發現修正結果為最靠近實際常住位置之 A 網格。

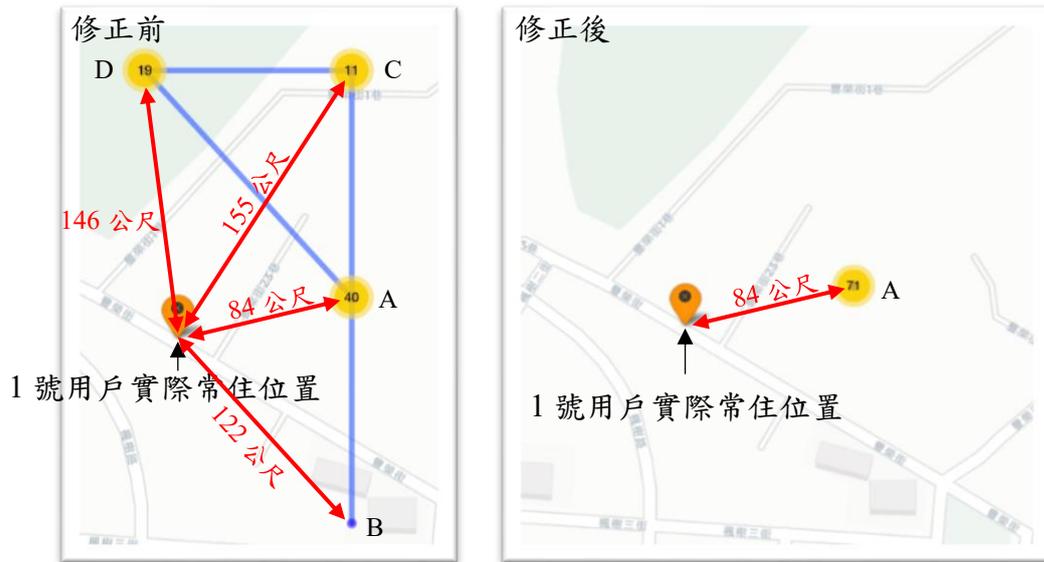


圖 4.2.2 1 號用戶跳點規則修正示意圖

接續為本研究所提出的「航向變化法」，其主要修正疑似來回事件的訊號飄移問題，從表 4.2.1 可觀察出航向變化法共修正整體資料 17%。而從圖 4.2.3 可觀察出航向變化法修正先後的的效果差異。先觀察圖 4.2.3 之左圖，在航向變化法修正前，除了 A 網格，其餘網格都與 1 號用戶實際常住位置相距超過 100 公尺。觀察圖 4.2.3 之右圖，經過航向變化法修正後，飄移至 C、D 與 E 的網格都已修正至離 1 號用戶實際常住位置最近之 A 網格。

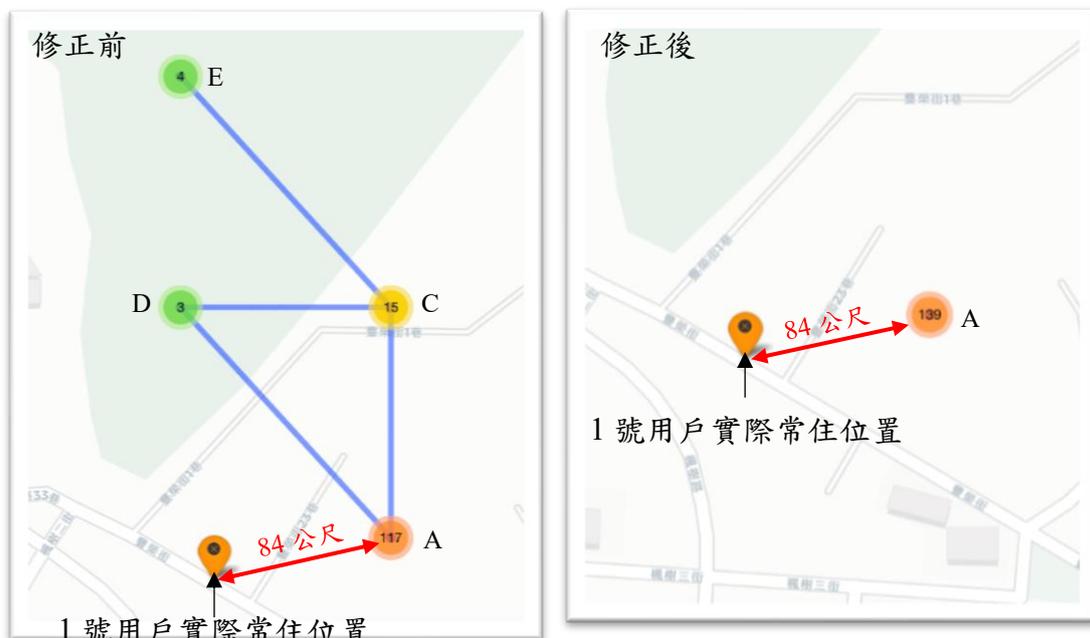


圖 4.2.3 1 號用戶航向變化法修正示意圖

模式判斷與修正方法的最後階段為時窗法，時窗法主要修正在五分鐘內發生的來回事件，由於在網格尺寸較小的資料中難以偵測到來回事件，因此僅修正整

體資料 0.3%。而從圖 4.2.4 可觀察出時窗法修正前後的效果差異。先觀察圖 4.2.4 之左圖，五分鐘內網格位置變化為 G 點-F 點-G 點的來回事件。右圖為時窗法修正後之結果，將 F 點拉回與用戶實際工作位置較近之網格 G。



圖 4.2.4 1 號用戶時窗法修正示意圖

圖 4.2.5 為 1 號用戶一日資料清洗前後差異圖，圖中黃、綠、橘三色圓圈表示圓圈附近區域信令資料出現次數。可從上圖 4.2.2 至 4.2.4 觀察得知本研究使用的網格空間單位較小，因此經過修正前後的網格變化都較細微。綜觀 1 號用戶一日的資料清洗結果可以發現用戶的移動軌跡較平滑，並且可以觀察到圖 4.2.5 左圖紅色圓圈為嚴重飄移的信令點，而經過修正後，圖 4.2.5 之右圖顯示飄移的信令點都已歸位。



圖 4.2.5 1 號用戶資料清洗前後差異圖

### 4.3 移動與停留狀態判斷結果

移動與停留狀態判斷流程主要目的為區分每一筆信令資料移動或是停留狀態，隨後將停留狀態的信令資料用於判斷用戶的常住與工作位置，藉此提高預測用戶位置精準度。如圖 4.3.1 呈現 1 號用戶 3 月 5 日之移動狀態信令點，綠色線條為移動狀態性質的信令點連結而成的移動軌跡，藍色地標點分別為實際常住與工作地點位置，因此觀察出軌跡的兩端連結著用戶常住與工作位置。圖 4.3.2 為 1 號用戶 3 月 5 日之停留狀態信令點，紅色點為停留狀態之信令點，黃與橘兩色圓圈數字為該用戶被識別停留在該區域的信令點出現次數，藍色地標點為實際常住或工作地點位置。因此觀察出停留狀態信令點大部分都分佈在實際的常住位置與工作位置附近，停留狀態信令點在實際常住位置附近的約為 330 筆、實際工作地點附近約 280 筆。

整體用戶移動與停留狀態的資料比例經過計算後，如表 4.3.1 移動與停留狀態篩選結果。從表中可以發現業務組(物流士)因其工作性質，在工作期間常常移動交付貨物等等，在本流程分析結果發現其移動狀態資料比率確實比非業務組高上許多(業務組：26% > 非業務組：9%)，也應證此演算流程的合理性。最後總和業務組與非業務組移動狀態資料為 26799 筆，占整體資料 15%；停留狀態資料為 154786 筆，占整體資料 85%。

表 4.3.1 移動與停留狀態篩選結果

	移動狀態筆數	停留比率	停留狀態筆數	移動比率
非業務組	10907	約 91%	108359	約 9%
業務組(物流士)	15892	約 74%	46427	約 26%
整體	26799	約 15%	154786	約 85%



圖 4.3.1 1 號用戶移動狀態之信令點與移動軌跡



圖 4.3.2 1 號用戶停留狀態之信令點分佈

#### 4.4 常住與工作位置判斷結果

本研究透過四種模式預測出用戶常住與工作位置，並計算預測與實際位置誤差結果，表 4.4.1 為四種模式之概述表。隨後將四種模式預測與實際位置誤差結果分別輸入至四種比較情境，用以評估與驗證有無航向變化法與移動停留狀態判斷流程預測用戶常住與工作位置之績效，表 4.4.2 為四種比較情境之概述表。圖 4.4.1 呈現模式 1(指模式 1-1 與模式 1-2)與模式 2(指模式 2-1 與模式 2-2)主要區別，模式 1-1 與模式 1-2 為藍色箭頭，皆有資料清洗流程以及常住與工作位置預測流程；模式 2-1 與模式 2-2 為黑色箭頭，皆有三流程。

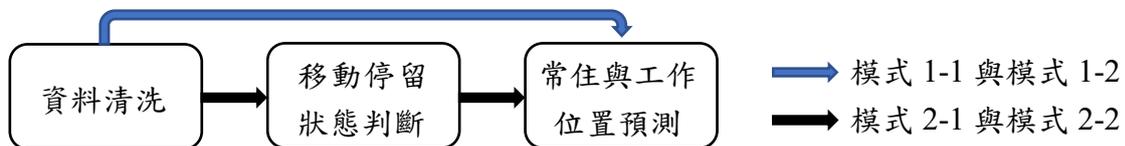


圖 4.4.1 常住與工作位置預測流程圖

表 4.4.1 四種模式概述表

模式 1	模式 1-1	資料清洗(有航向變化法)後直接進行用戶常住與工作位置預測
	模式 1-2	資料清洗(無航向變化法)後直接進行用戶常住與工作位置預測

模式 2	模式 2-1	資料清洗(有航向變化法)後，先進行資料的移動與停留狀態判斷，才進行用戶常住與工作位置預測
	模式 2-2	資料清洗(無航向變化法)後，先進行資料的移動與停留狀態判斷，才進行用戶常住與工作位置預測

表 4.4.2 四種情境概述表

目標	前提	模式驗證
航向變化法預測 績效	兩模式皆無加入移動停留判斷流程 (比較情境 1)	<ul style="list-style-type: none"> <li>若群體成常態分佈，使用成對樣本 T 檢定進行驗證</li> <li>若群體不成常態分佈，使用符號檢定與魏克生符號檢定進行驗證</li> </ul>
	兩模式皆有加入移動停留判斷流程 (比較情境 2)	
移動停留狀態判斷 流程預測績效	兩模式皆無加入航向變化法 (比較情境 3)	
	兩模式皆有加入航向變化法 (比較情境 4)	

#### 4.4.1 常住位置預測結果

將模式 1-1、模式 1-2、模式 2-1 與模式 2-2 預測與實際的常住位置誤差進行比較，整理成表 4.4.1.1 用以清楚觀察出四種模式預測的效果差異。舉例說明，模式 1-1 預測 1 號用戶常住位置與實際常住位置誤差為 87 公尺。後續表 4.4.1.2 至表 4.4.1.5 為模式 1-1、1-2、2-1 與 2-2 為四種模式預測與實際常住位置結果誤差表，表中包含匿名化的用戶 ID、預測網格 ID、預測網格之經緯度、實際位置經緯度以及預測與實際位置誤差(公尺)。從表 4.4.1.1 可以觀察出模式 1-1 預測與實際常住位置平均誤差為 148 公尺，比其他三模式預測與實際常住位置誤差結果都還低。另外，四種模式預測與實際常住位置誤差結果皆在 500 公尺內。

表 4.4.1.1 四種模式預測與實際常住位置誤差結果

用戶	模式 1-1	模式 1-2	模式 2-1	模式 2-2
1	87	87	87	87
2	492	492	492	492
3	137	137	137	137
4	5	5	5	5
5	217	407	207	207
6	492	492	492	492
7	31	31	31	31
8	140	140	140	140
9	62	62	62	62

10	109	109	109	109
11	61	196	172	172
12	61	61	61	61
13	131	131	131	131
14	48	48	48	48
15	441	441	441	441
16	37	37	37	37
17	93	93	93	93
18	37	37	37	37
19	406	406	406	406
20	216	216	216	216
21	30	30	30	30
22	38	148	38	148
23	35	35	35	35
24	117	117	117	117
25	98	98	98	98
26	134	134	134	134
27	245	245	245	245
最大誤差	492	492	492	492
最小誤差	5	5	5	5
平均誤差	148	164	152	156

表 4.4.1.2 模式 1-1 預測常住位置結果

用戶	預測的 常住網 格 ID	預測的常 住緯度	預測的常 住經度	實際的常 住緯度	實際的常 住經度	距離 誤差 (公尺)
1	58632	25.00611561	121.3475255	25.00594335	121.3466806	87
2	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
3	67067	24.98911561	121.3165255	24.98788222	121.3166427	137
4	89405	24.94411561	121.2445255	24.94407862	121.244504	5
5	89408	24.94411561	121.2475255	24.94344837	121.2455099	217
6	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
7	87926	24.94711561	121.2595255	24.94704439	121.2592261	31
8	72001	24.97911561	121.2705255	24.98027052	121.2699546	140
9	94335	24.93411561	121.1945255	24.93430517	121.1951019	62
10	66047	24.99111561	121.2925255	24.99017726	121.292206	109
11	92385	24.93811561	121.2365255	24.93782021	121.2360124	61
12	92385	24.93811561	121.2365255	24.93782021	121.2360124	61

13	62584	24.99811561	121.3155255	24.9990245	121.3146944	131
14	83920	24.95511561	121.2375255	24.95533049	121.2379433	48
15	89917	24.94311561	121.2585255	24.94704442	121.2592258	441
16	78469	24.96611561	121.2645255	24.96645215	121.2645597	37
17	106275	24.91011561	121.1825255	24.90927897	121.1824085	93
18	96864	24.92911561	121.2335255	24.92891544	121.2338193	37
19	73436	24.97611561	121.2115255	24.97381128	121.2146548	406
20	64562	24.99411561	121.3015255	24.99420074	121.299387	216
21	76448	24.97011561	121.2355255	24.97034779	121.2353799	30
22	82898	24.95711561	121.2115255	24.95745406	121.2114913	38
23	105739	24.91111561	121.1445255	24.9112752	121.1448276	35
24	89409	24.94411561	121.2485255	24.94385396	121.2496477	117
25	62017	24.99911561	121.2465255	24.99982129	121.2459427	98
26	64993	24.99311561	121.2345255	24.99255273	121.2333452	134
27	144645	24.83311561	121.2065255	24.8353219	121.2067324	245

表 4.4.1.3 模式 1-2 預測常住位置結果

用戶	預測的 常住網 格 ID	預測的常 住緯度	預測的常 住經度	實際的常 住緯度	實際的常 住經度	距離 誤差 (公尺)
1	58632	25.00611561	121.3475255	25.00594335	121.3466806	87
2	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
3	67067	24.98911561	121.3165255	24.98788222	121.3166427	137
4	89405	24.94411561	121.2445255	24.94407862	121.244504	5
5	89908	24.94311561	121.2495255	24.94344837	121.2455099	407
6	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
7	87926	24.94711561	121.2595255	24.94704439	121.2592261	31
8	72001	24.97911561	121.2705255	24.98027052	121.2699546	140
9	94335	24.93411561	121.1945255	24.93430517	121.1951019	62
10	66047	24.99111561	121.2925255	24.99017726	121.292206	109
11	93381	24.93611561	121.2365255	24.93782021	121.2360124	196
12	92385	24.93811561	121.2365255	24.93782021	121.2360124	61
13	62584	24.99811561	121.3155255	24.9990245	121.3146944	131
14	83920	24.95511561	121.2375255	24.95533049	121.2379433	48
15	89917	24.94311561	121.2585255	24.94704442	121.2592258	441
16	78469	24.96611561	121.2645255	24.96645215	121.2645597	37
17	106275	24.91011561	121.1825255	24.90927897	121.1824085	93
18	96864	24.92911561	121.2335255	24.92891544	121.2338193	37

19	73436	24.97611561	121.2115255	24.97381128	121.2146548	406
20	64562	24.99411561	121.3015255	24.99420074	121.299387	216
21	76448	24.97011561	121.2355255	24.97034779	121.2353799	30
22	83396	24.95611561	121.2115255	24.95745406	121.2114913	148
23	105739	24.91111561	121.1445255	24.9112752	121.1448276	35
24	89409	24.94411561	121.2485255	24.94385396	121.2496477	117
25	62017	24.99911561	121.2465255	24.99982129	121.2459427	98
26	64993	24.99311561	121.2345255	24.99255273	121.2333452	134
27	144645	24.83311561	121.2065255	24.8353219	121.2067324	245

表 4.4.1.4 模式 2-1 預測常住位置結果

用戶	預測的 常住網 格 ID	預測的常 住緯度	預測的常 住經度	實際的常 住緯度	實際的常 住經度	距離 誤差 (公尺)
1	58632	25.00611561	121.3475255	25.00594335	121.3466806	87
2	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
3	67067	24.98911561	121.3165255	24.98788222	121.3166427	137
4	89405	24.94411561	121.2445255	24.94407862	121.244504	5
5	89906	24.94311561	121.2475255	24.94344837	121.2455099	207
6	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
7	87926	24.94711561	121.2595255	24.94704439	121.2592261	31
8	72001	24.97911561	121.2705255	24.98027052	121.2699546	140
9	94335	24.93411561	121.1945255	24.93430517	121.1951019	62
10	66047	24.99111561	121.2925255	24.99017726	121.292206	109
11	92884	24.93711561	121.2375255	24.93782021	121.2360124	172
12	92385	24.93811561	121.2365255	24.93782021	121.2360124	61
13	62584	24.99811561	121.3155255	24.9990245	121.3146944	131
14	83920	24.95511561	121.2375255	24.95533049	121.2379433	48
15	89917	24.94311561	121.2585255	24.94704442	121.2592258	441
16	78469	24.96611561	121.2645255	24.96645215	121.2645597	37
17	106275	24.91011561	121.1825255	24.90927897	121.1824085	93
18	96864	24.92911561	121.2335255	24.92891544	121.2338193	37
19	73436	24.97611561	121.2115255	24.97381128	121.2146548	406
20	64562	24.99411561	121.3015255	24.99420074	121.299387	216
21	76448	24.97011561	121.2355255	24.97034779	121.2353799	30
22	82898	24.95711561	121.2115255	24.95745406	121.2114913	38
23	105739	24.91111561	121.1445255	24.9112752	121.1448276	35

24	89409	24.94411561	121.2485255	24.94385396	121.2496477	117
25	62017	24.99911561	121.2465255	24.99982129	121.2459427	98
26	64993	24.99311561	121.2345255	24.99255273	121.2333452	134
27	144645	24.83311561	121.2065255	24.8353219	121.2067324	245

表 4.4.1.5 模式 2-2 預測常住位置結果

用戶	預測的 常住網 格 ID	預測的常 住緯度	預測的常 住經度	實際的常 住緯度	實際的常 住經度	距離 誤差 (公尺)
1	58632	25.00611561	121.3475255	25.00594335	121.3466806	87
2	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
3	67067	24.98911561	121.3165255	24.98788222	121.3166427	137
4	89405	24.94411561	121.2445255	24.94407862	121.244504	5
5	89906	24.94311561	121.2475255	24.94344837	121.2455099	207
6	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
7	87926	24.94711561	121.2595255	24.94704439	121.2592261	31
8	72001	24.97911561	121.2705255	24.98027052	121.2699546	140
9	94335	24.93411561	121.1945255	24.93430517	121.1951019	62
10	66047	24.99111561	121.2925255	24.99017726	121.292206	109
11	92884	24.93711561	121.2375255	24.93782021	121.2360124	172
12	92385	24.93811561	121.2365255	24.93782021	121.2360124	61
13	62584	24.99811561	121.3155255	24.9990245	121.3146944	131
14	83920	24.95511561	121.2375255	24.95533049	121.2379433	48
15	89917	24.94311561	121.2585255	24.94704442	121.2592258	441
16	78469	24.96611561	121.2645255	24.96645215	121.2645597	37
17	106275	24.91011561	121.1825255	24.90927897	121.1824085	93
18	96864	24.92911561	121.2335255	24.92891544	121.2338193	37
19	73436	24.97611561	121.2115255	24.97381128	121.2146548	406
20	64562	24.99411561	121.3015255	24.99420074	121.299387	216
21	76448	24.97011561	121.2355255	24.97034779	121.2353799	30
22	83396	24.95611561	121.2115255	24.95745406	121.2114913	148
23	105739	24.91111561	121.1445255	24.9112752	121.1448276	35
24	89409	24.94411561	121.2485255	24.94385396	121.2496477	117
25	62017	24.99911561	121.2465255	24.99982129	121.2459427	98
26	64993	24.99311561	121.2345255	24.99255273	121.2333452	134
27	144645	24.83311561	121.2065255	24.8353219	121.2067324	245

## 4.4.2 工作位置預測結果

將模式 1-1、模式 1-2、模式 2-1 與模式 2-2 預測與實際的工作位置誤差進行比較，整理成表 4.4.2.1 用以清楚觀察出四種模式預測的效果差異。舉例說明，模式 1-1 預測 1 號用戶工作位置與實際工作位置誤差為 152 公尺。後續表 4.4.2.5 至表 4.4.2.8 為模式 1-1、1-2、2-1 與 2-2 為四種模式預測與實際工作位置誤差表，表中包含預測網格 ID、預測網格之經緯度、實際位置經緯度以及預測與實際工作位置誤差(公尺)。

首先，觀察表 4.4.2.1 紅色方框處，21 號至 27 號用戶為物流士，由於物流士不斷移動的工作性質導致無法準確預測物流士之工作位置，使物流士預測與實際工作位置的誤差皆超過 1000 公尺。因為訊號飄移為系統性的錯誤，非人為所致。若將物流士工作位置預測結果用以探討有無加入航向變化法或是移動停留狀態判斷流程是否能夠有效降低預測誤差可能不太合適，因此後續對於預測工作位置的比較情境與模式驗證皆使用 1 號至 20 號一般用戶預測與實際工作位置誤差結果進行探討。

另外，從表 4.4.2.1 發現四種模式預測工作位置的結果中，除了上述 7 位物流士的預測與實際工作位置誤差超過 1000 公尺，一般用戶的 8 號、17 號與 20 號的預測與實際工作位置誤差距離也超過 1000 公尺。進一步探討高誤差預測結果的成因，觀察表 4.4.2.2 為 8 號、17 號與 20 號三位一般用戶近似比對結果表，表中預測結果是取用模式 2-1 預測與實際工作位置流程誤差結果，僅列出現頻率前三高之網格 ID，以 8 號用戶為例，58568 網格出現 27 次位列最高，因此 8 號用戶最終工作代表網格為 58568。從表 4.4.2.2 中發現 8 號用戶前三順位工作候選網格皆與真實工作位置誤差約為 2200 公尺上下，可能受該區域基地台較稀疏或受地勢影響等系統性問題，使多數信令點的網格皆與真實工作位置距離的較遠，導致本研究無法準確預測此用戶工作位置。從表 4.4.2.2 中可以觀察到 17 號用戶第二順位的工作候選網格為 64567，與真實工作位置僅誤差 345 公尺；且從表中發現 20 號用戶之第三順位的工作候選網格 47129 與真實工作位置誤差距離約為 831 公尺。也就是說，17 號與 20 號用戶的前三候選網格中發現有誤差極高的候選網格，也有誤差低於 1000 公尺的網格。隨後觀察表 4.4.2.3 探討為何離真實工作位置較近的網格無法成為最終工作位置代表網格的原因，發現網格 106275 與網格 65557 分別出現本研究所設定工作分析天數 10 日中的 5 天與 3 天，這也代表該用戶在這些天內多數時段用戶是待在這些位置中，因此本研究才將其判斷為工作位置，但其實這兩個位置分別為前一階段預測用戶的最終常住位置相同，而我們預測的常住位置也與實際常住位置相近，因此可推估此誤判位置的原因為用戶排休日所導致，使本研究最後誤判其工作位置。若收集信令資料的資料期間可延伸至一個月甚至更長，則可避免此誤判問題。

最後，基於上述原因，表 4.4.2.1 四種模式預測與實際工作位置最大誤差、最小誤差與平均誤差計算不考慮 17 號、20 號與物流士之預測與實際工作位置誤差

結果(灰色字體)。模式 1-1 之預測與實際工作位置平均誤差為 269 公尺，其結果與模式 2-1 相同，代表模式 1-1 與模式 2-1 預測效果相當。無論是模式 1-1 或是模式 2-1，加入航向變化法之後，除了 8 號用戶預測與實際工作位置誤差結果超過 1000 公尺，其餘誤差皆在 500 公尺內。

表 4.4.2.1 四種模式預測與實際工作位置誤差比較(單位:公尺)

用戶	模式 1-1	模式 1-2	模式 2-1	模式 2-2
1	152	152	152	152
2	98	492	98	492
3	131	131	131	131
4	100	100	100	100
5	74	106	74	106
6	415	684	415	684
7	253	253	253	253
8	2209	2301	2209	2301
9	49	49	49	49
10	33	33	33	33
11	172	172	172	172
12	172	172	172	172
13	277	277	277	277
14	48	141	48	141
15	473	1259	473	1259
16	30	30	30	30
17	15557	15557	15557	15557
18	37	37	37	37
19	124	124	124	124
20	4885	4885	4885	4885
21	1848	1848	1848	1848
22	4512	4512	4512	4512
23	12872	12872	12872	12872
24	4636	4636	4636	4636
25	1556	1556	1556	1556
26	1149	1149	1149	1149
27	17273	17273	17273	17273
最大誤差	2909	2301	2909	2301
最小誤差	30	30	30	30
平均誤差	269	362	269	362

表 4.4.2.2 有問題用戶之工作候選網格近似比對結果表

用戶	網格 ID	近似比對次數	與真實工作位置 誤差距離(公尺)
8	58568	27	2209
8	58569	21	2179
8	59065	12	2301
17	106275	30	15557
17	64567	3	345
17	60580	2	1290
20	65557	18	4885
20	65556	12	4887
20	47129	3	831

表 4.4.2.3 有問題用戶之前三順位工作候選網格出現天數結果表

用戶	網格 ID	出現總天數
8	58568	3
8	58569	3
8	59065	2
17	106275	5
17	64567	1
17	60580	1
20	65557	3
20	65556	2
20	47129	1

透過表 4.4.1.1 四種模式預測與實際常住位置誤差結果與表 4.4.2.1 四種模式預測與實際工作位置誤差結果(除 17 號、20 號與物流士用戶預測結果)，計算用戶預測與實際整體(常住與工作)位置最大誤差、最小誤差與平均誤差，如表 4.4.2.4。從表 4.4.2.4 中可以發現模式 1-1 平均誤差最小。另外，模式 1-1 與模式 2-1 的最大誤差值與最小誤差值相同。

表 4.4.2.4 預測與實際整體位置最大、最小與平均誤差值

	模式 1-1	模式 1-2	模式 2-1	模式 2-2
最大誤差(公尺)	2209	2301	2209	2301
最小誤差(公尺)	5	5	5	5
平均誤差(公尺)	197	243	199	239

表 4.4.2.5 模式 1-1 預測工作位置結果

用戶	預測的常住網格 ID	預測的常住緯度	預測的常住經度	實際的常住緯度	實際的常住經度	距離誤差 (公尺)
1	48626	25.02611561	121.3015255	25.02745214	121.3012023	152
2	81931	24.95911561	121.2405255	24.95825539	121.2407674	98
3	67988	24.98711561	121.2415255	24.98621026	121.2423647	131
4	70051	24.98311561	121.3125255	24.98222116	121.3123847	100
5	89406	24.94411561	121.2455255	24.94344837	121.2455099	74
6	83921	24.95511561	121.2385255	24.95825539	121.2407674	415
7	65065	24.99311561	121.3065255	24.99138198	121.3081616	253
8	58568	25.00611561	121.2835255	24.98721498	121.2905122	2209
9	21172	25.08111561	121.2375255	25.08114079	121.2370452	49
10	66545	24.99011561	121.2925255	24.99017726	121.292206	33
11	92884	24.93711561	121.2375255	24.93782021	121.2360124	172
12	92884	24.93711561	121.2375255	24.93782021	121.2360124	172
13	68533	24.98611561	121.2885255	24.98744365	121.2862056	277
14	83920	24.95511561	121.2375255	24.95533049	121.2379433	48
15	9241	25.10511561	121.2585255	25.10909251	121.2602316	473
16	77946	24.96711561	121.2395255	24.9673855	121.2395067	30
17	106275	24.91011561	121.1825255	24.99138213	121.3081622	15557
18	96864	24.92911561	121.2335255	24.92891782	121.2338187	37
19	77405	24.96811561	121.1965255	24.96816515	121.1953009	124
20	65557	24.99211561	121.3005255	25.03620918	121.3012132	4885
21	76448	24.97011561	121.2355255	24.98559	121.24236	1848
22	83396	24.95611561	121.2115255	24.98559	121.24236	4512
23	105739	24.91111561	121.1445255	24.98559	121.24236	12872
24	89409	24.94411561	121.2485255	24.98559	121.24236	4636
25	62017	24.99911561	121.2465255	24.98559	121.24236	1556
26	64993	24.99311561	121.2345255	24.98559	121.24236	1149
27	144645	24.83311561	121.2065255	24.98559	121.24236	17273

表 4.4.2.6 模式 1-2 預測工作位置結果

用戶	預測的常住網格 ID	預測的常住緯度	預測的常住經度	實際的常住緯度	實際的常住經度	距離誤差 (公尺)
----	------------	---------	---------	---------	---------	-----------

1	48626	25.026115607	121.301525461	25.02745214	121.3012023	152
2	84423	24.954115607	121.242525461	24.95825539	121.2407674	492
3	67988	24.987115607	121.241525461	24.98621026	121.2423647	131
4	70051	24.983115607	121.312525461	24.98222116	121.3123847	100
5	89903	24.943115607	121.244525461	24.94344837	121.2455099	106
6	85418	24.952115607	121.241525461	24.95825539	121.2407674	684
7	65065	24.993115607	121.306525461	24.99138198	121.3081616	253
8	72000	24.979115607	121.269525461	24.98721498	121.2905122	2301
9	21172	25.081115607	121.237525461	25.08114079	121.2370452	49
10	66545	24.990115607	121.292525461	24.99017726	121.292206	33
11	92884	24.937115607	121.237525461	24.93782021	121.2360124	172
12	92884	24.937115607	121.237525461	24.93782021	121.2360124	172
13	68533	24.986115607	121.288525461	24.98744365	121.2862056	277
14	84418	24.954115607	121.237525461	24.95533049	121.2379433	141
15	8259	25.107115607	121.272525461	25.10909251	121.2602316	1259
16	77946	24.967115607	121.239525461	24.9673855	121.2395067	30
17	106275	24.910115607	121.182525461	24.99138213	121.3081622	15557
18	96864	24.929115607	121.233525461	24.92891782	121.2338187	37
19	77405	24.968115607	121.196525461	24.96816515	121.1953009	124
20	65557	24.992115607	121.300525461	25.03620918	121.3012132	4885
21	76448	24.97011561	121.2355255	24.98559	121.24236	1848
22	83396	24.95611561	121.2115255	24.98559	121.24236	4512
23	105739	24.91111561	121.1445255	24.98559	121.24236	12872
24	89409	24.94411561	121.2485255	24.98559	121.24236	4636
25	62017	24.99911561	121.2465255	24.98559	121.24236	1556
26	64993	24.99311561	121.2345255	24.98559	121.24236	1149
27	144645	24.83311561	121.2065255	24.98559	121.24236	17273

表 4.4.2.7 模式 2-1 預測工作位置結果

用戶	預測的 常住網 格 ID	預測的常 住緯度	預測的常 住經度	實際的常 住緯度	實際的常 住經度	距離誤差 (公尺)
1	48626	25.02611561	121.3015255	25.02745214	121.3012023	152
2	81931	24.95911561	121.2405255	24.95825539	121.2407674	98
3	67988	24.98711561	121.2415255	24.98621026	121.2423647	131
4	70051	24.98311561	121.3125255	24.98222116	121.3123847	100
5	89406	24.94411561	121.2455255	24.94344837	121.2455099	74
6	83921	24.95511561	121.2385255	24.95825539	121.2407674	415

7	65065	24.99311561	121.3065255	24.99138198	121.3081616	253
8	58568	25.00611561	121.2835255	24.98721498	121.2905122	2209
9	21172	25.08111561	121.2375255	25.08114079	121.2370452	49
10	66545	24.99011561	121.2925255	24.99017726	121.292206	33
11	92884	24.93711561	121.2375255	24.93782021	121.2360124	172
12	92884	24.93711561	121.2375255	24.93782021	121.2360124	172
13	68533	24.98611561	121.2885255	24.98744365	121.2862056	277
14	83920	24.95511561	121.2375255	24.95533049	121.2379433	48
15	9241	25.10511561	121.2585255	25.10909251	121.2602316	473
16	77946	24.96711561	121.2395255	24.9673855	121.2395067	30
17	106275	24.91011561	121.1825255	24.99138213	121.3081622	15557
18	96864	24.92911561	121.2335255	24.92891782	121.2338187	37
19	77405	24.96811561	121.1965255	24.96816515	121.1953009	124
20	65557	24.99211561	121.3005255	25.03620918	121.3012132	4885
21	76448	24.97011561	121.2355255	24.98559	121.24236	1848
22	83396	24.95611561	121.2115255	24.98559	121.24236	4512
23	105739	24.91111561	121.1445255	24.98559	121.24236	12872
24	89409	24.94411561	121.2485255	24.98559	121.24236	4636
25	62017	24.99911561	121.2465255	24.98559	121.24236	1556
26	64993	24.99311561	121.2345255	24.98559	121.24236	1149
27	144645	24.83311561	121.2065255	24.98559	121.24236	17273

表 4.4.2.8 模式 2-2 預測工作位置結果

用戶	預測的 常住網 格 ID	預測的常住 緯度	預測的常住經 度	實際的常住 緯度	實際的常 住經度	距離 誤差 (公 尺)
1	48626	25.02611561	121.3015255	25.02745214	121.3012023	152
2	84423	24.95411561	121.2425255	24.95825539	121.2407674	492
3	67988	24.98711561	121.2415255	24.98621026	121.2423647	131
4	70051	24.98311561	121.3125255	24.98222116	121.3123847	100
5	89903	24.94311561	121.2445255	24.94344837	121.2455099	106
6	85418	24.95211561	121.2415255	24.95825539	121.2407674	684
7	65065	24.99311561	121.3065255	24.99138198	121.3081616	253
8	72000	24.97911561	121.2695255	24.98721498	121.2905122	2301
9	21172	25.08111561	121.2375255	25.08114079	121.2370452	49
10	66545	24.99011561	121.2925255	24.99017726	121.292206	33
11	92884	24.93711561	121.2375255	24.93782021	121.2360124	172

12	92884	24.93711561	121.2375255	24.93782021	121.2360124	172
13	68533	24.98611561	121.2885255	24.98744365	121.2862056	277
14	84418	24.95411561	121.2375255	24.95533049	121.2379433	141
15	8259	25.10711561	121.2725255	25.10909251	121.2602316	1259
16	77946	24.96711561	121.2395255	24.9673855	121.2395067	30
17	106275	24.91011561	121.1825255	24.99138213	121.3081622	15557
18	96864	24.92911561	121.2335255	24.92891782	121.2338187	37
19	77405	24.96811561	121.1965255	24.96816515	121.1953009	124
20	65557	24.99211561	121.3005255	25.03620918	121.3012132	4885
21	76448	24.97011561	121.2355255	24.98559	121.24236	1848
22	83396	24.95611561	121.2115255	24.98559	121.24236	4512
23	105739	24.91111561	121.1445255	24.98559	121.24236	12872
24	89409	24.94411561	121.2485255	24.98559	121.24236	4636
25	62017	24.99911561	121.2465255	24.98559	121.24236	1556
26	64993	24.99311561	121.2345255	24.98559	121.24236	1149
27	144645	24.83311561	121.2065255	24.98559	121.24236	17273

#### 4.4.3 航向變化法預測績效與模式驗證

透過成對樣本 T 檢定、符號檢定(Sign Test)與魏克生符號檢定(Wilcoxon Signed Ranked Test)三種統計檢定方法來分別檢驗比較情境 1 與比較情境 2 加入航向變化法後，是否能有效減少預測用戶位置之誤差(公尺)；另外，也檢驗比較情境 3 與比較情境 4 加入移動與停留狀態判斷流程後，是否能有效減少預測用戶位置之誤差。位置指用戶的常住或工作位置。

- 成對樣本 T 檢定之假設：

若群體分佈呈常態分佈，則使用成對樣本 T 檢定，以下為成對樣本 T 檢定之假設。舉例說明：加入某流程或方法下，探討模式 A 預測與實際位置平均誤差是否比模式 B 預測與實際位置平均誤差小。若 P-value > 0.05 則不拒絕虛無假設，反之，若 P-value ≤ 0.05 則拒絕虛無假設。

$\mu_u$ ：模式 A 之預測與實際位置的平均誤差

$\mu_s$ ：模式 B 之預測與實際位置的平均誤差

$$\begin{cases} \text{虛無假說 } H_0 : \mu_u = \mu_s \\ \text{對立假說 } H_1 : \mu_u < \mu_s \end{cases}$$

- 符號檢定之假設與魏克生符號檢定之假設：

若群體不呈現常態分佈，本研究會採用符號檢定與魏克生符號檢定兩種無母數方法進行統計檢定，由於魏克生符號檢定為改良後的符號檢定，因此兩方法假說檢定之假設相同。舉例說明：加入某流程或方法下，探討模式 A 預測與實際位置誤差是否比模式 B 預測與實際位置誤差小。若 P-value > 0.05 則不拒絕虛無假設，反之，若 P-value ≤ 0.05 則拒絕虛無假設。

$\eta_1$ ：模式 A 之預測與實際位置的誤差

$\eta_2$ ：模式 B 之預測與實際位置的誤差

$\eta_D = \eta_1 - \eta_2$ ：成對樣本之間的差

$$\begin{cases} \text{虛無假說 } H_0 : \eta_D = 0 \\ \text{對立假說 } H_1 : \eta_D < 0 \end{cases}$$

- 比較情境 1: 模式 1-1 與模式 1-2 兩模式皆無加入移動停留狀態判斷流程下，比較有加入航向變化法之模式 1-1 預測績效是否優於無加入航向變化法之預測績效。位置分別為常住、工作與整體位置。

##### (1) 常住位置

從表 4.4.1.1 可以觀察出在兩模式比較下，有加入航向變化法的模式 1-1 經過預測後，有 3 位用戶的精準度優於模式 1-2 之預測結果，如 5 號、11 號與 22 號等用戶誤差分別改善(減少)約 190 公尺、135 公尺與 110 公尺，沒有其他用戶預測誤差比模式 1-2 高。3 位用戶中，改善最多誤差為 190 公尺、改善最小誤差為 110 公尺，平均誤差改善 145 公尺。結果顯示在預測常住位置時，模式 1-1 略

優於模式 1-2。

將表 4.4.1.1 中，整體 27 位用戶在模式 1-1 與模式 1-2 的預測常住位置誤差結果進行三種統計方法檢定並輸出表 4.5.1.1，驗證模式 1-1 在加入航向變化法下，預測與實際常住位置誤差是否比模式 1-2 預測與實際常住位置誤差小。從表 4.4.3.1 可以發現成對樣本 T 檢定不符合常態分配，而其餘比較情境驗證中，成對樣本 T 檢定皆不符合常態分配，因此後續皆不參考其檢定結果。另外從表 4.4.3.1 可以發現符號檢定與魏克生符號檢定結果皆不拒絕虛無假設，表示在預測常住位置時，模式 1-1 在資料清洗流程中加入航向變化法後，仍無明顯減少預測常住位置誤差。

**表 4.4.3.1 比較情境 1 之有無航向變化法預測常住位置統計結果表**

	p-value	$\alpha = 0.05$	備註
<del>成對樣本 T 檢定</del>	<del>0.04586</del>	<del>顯著</del>	<del>不符合常態分配</del>
符號檢定	0.125	不顯著	無
魏克生符號檢定	0.09072	不顯著	無

(2) 工作位置

從表 4.4.2.1 可以觀察出在兩模式比較下，有加入航向變化法的模式 1-1 經過預測後，有 6 位用戶的精準度優於模式 1-2 預測結果，如 2 號、5 號、6 號、8 號、14 號與 15 號等用戶誤差約減少分別為 394 公尺、32 公尺、269 公尺、92 公尺、93 公尺與 786 公尺，且沒有其他用戶誤差比模式 1-2 預測結果的誤差還高。6 位用戶中，誤差改善最多為 786 公尺、誤差改善最小為 32 公尺，平均誤差改善約為 278 公尺。結果顯示在預測工作位置時，模式 1-1 優於模式 1-2。

將表 4.4.2.1 中，整體 20 位一般用戶在模式 1-1 與模式 1-2 的預測工作位置誤差結果進行三種統計方法檢定並輸出表 4.4.3.2，驗證模式 1-1 在加入航向變化法下，預測與實際工作位置誤差是否比模式 1-2 預測與實際工作位置誤差小。觀察表 4.4.3.2 可以發現成對樣本 T 檢定不符合常態分配，因此不參考其檢定結果。另外發現符號檢定與魏克生符號檢定結果皆拒絕虛無假設，表示在預測工作位置時，有加入航向變化法的模式 1-1 能夠顯著減少整體一般用戶預測與實際工作位置誤差的效果。

**表 4.4.3.2 比較情境 1 之有無航向變化法預測工作位置統計結果表**

	p-value	$\alpha = 0.05$
符號檢定	0.01563	顯著
魏克生符號檢定	0.01802	顯著

(3) 整體位置

將表 4.4.1.1 中整體 27 位用戶在模式 1-1 與模式 1-2 的預測常住位置誤差結果以及表 4.4.2.1 中整體 20 位一般用戶在模式 1-1 與模式 1-2 的預測工作位置誤差結果結合，可以觀察出在兩模式比較下，有加入航向變化法的模式 1-1 經過預測後，共有 9 個位置預測誤差得以改善。9 個位置中，誤差改善最多為 786 公尺、

誤差改善最小為 32 公尺，平均誤差改善約為 233 公尺。結果顯示在預測整體位置時，模式 1-1 優於模式 1-2。

將表 4.4.1.1 中整體 27 位用戶在模式 1-1 與模式 1-2 的預測常住位置誤差結果以及表 4.4.2.1 中整體 20 位一般用戶在模式 1-1 與模式 1-2 的預測工作位置誤差結果結合，進行三種統計方法檢定並輸出表 4.4.3.3，驗證模式 1-1 在加入航向變化法下，預測與實際整體位置誤差是否比模式 1-2 預測與實際整體位置誤差小。觀察表 4.4.3.3 可以發現符號檢定與魏克生符號檢定結果皆拒絕虛無假設。代表在資料清洗流程中加入航向變化法的模式 1-1 能夠明顯減少預測用戶整體(常住與工作)位置的誤差。綜合上述比較結果，整理至表 4.4.3.4。

**表 4.4.3.3 比較情境 1 之有無航向變化法預測整體位置統計結果表**

	p-value	$\alpha = 0.05$
符號檢定	0.001953	顯著
魏克生符號檢定	0.004576	顯著

**表 4.4.3.4 比較情境 1 預測績效與模式驗證結果表**

預測位置	樣本數	誤差有改善之樣本數	改善最多誤差(公尺)	改善最少誤差(公尺)	平均改善誤差(公尺)
常住位置	27	3	190	110	145
模式驗證	符號檢定：不顯著、魏克生符號檢定：不顯著				
工作位置	20	6	786	32	278
模式驗證	符號檢定：顯著、魏克生符號檢定：顯著				
整體位置	47	9	786	32	233
模式驗證	符號檢定：顯著、魏克生符號檢定：顯著				

- 比較情境 2: 模式 2-1 與模式 2-2 兩模式皆有加入移動停留狀態判斷流程下，比較有加入航向變化法之模式 2-1 預測績效是否優於無加入航向變化法之模式 2-2 預測績效。

(1) 常住位置

從表 4.4.1.1 可以觀察出在兩模式比較下，有加入航向變化法的模式 2-1 經過預測後，僅 22 號用戶誤差改善 11 公尺。沒有其他用戶因為加入航向變化法使誤差增加。結果顯示在預測常住位置時，模式 2-1 略優於模式 2-2。

將表 4.4.1.1 中，整體 27 位用戶在模式 2-1 與模式 2-2 的預測常住位置誤差結果進行三種統計方法檢定並輸出表 4.4.3.5，驗證模式 2-1 在加入航向變化法下，預測與實際常住位置誤差是否比模式 2-2 預測與實際常住位置誤差小。從表 4.4.3.5 可以發現符號檢定與魏克生符號檢定結果皆不拒絕虛無假設，表示在預測常住位置時，模式 2-1 雖然在資料清洗流程中加入航向變化法用以優化訊號飄移問題，但無明顯減少整體用戶預測常住位置之誤差。

**表 4.4.3.5 比較情境 2 之有無航向變化法預測常住位置統計結果表**

	p-value	$\alpha = 0.05$
符號檢定	0.221	不顯著
魏克生符號檢定	0.06355	不顯著

### (2) 工作位置

從表 4.4.2.1 可以觀察出在兩模式比較下，有加入航向變化法的模式 2-1 經過預測後，有 6 位用戶的精準度優於模式 2-2 預測結果，如 2 號、5 號、6 號、8 號、14 號與 15 號等用戶誤差約減少分別為 394 公尺、32 公尺、269 公尺、92 公尺、93 公尺與 786 公尺，沒有其他用戶預測誤差比模式 2-2 預測結果的誤差還高。6 位用戶中，誤差改善最多為 786 公尺、誤差改善最小為 32 公尺，平均誤差改善約為 278 公尺。結果顯示在預測工作位置時，模式 2-1 優於模式 2-2。

將表 4.4.2.1 中，整體 20 位一般用戶在模式 2-1 與模式 2-2 的預測工作位置誤差結果進行三種統計方法檢定並輸出表 4.4.3.6，驗證模式 2-1 在加入航向變化法下，預測與實際工作位置誤差是否比模式 2-2 預測與實際工作位置誤差小。觀察表 4.4.3.6 可以發現符號檢定與魏克生符號檢定結果皆拒絕虛無假設，表示在預測工作位置時，有加入航向變化法的模式 2-1 能夠顯著減少整體一般用戶預測與實際工作位置誤差的效果。

**表 4.4.3.6 比較情境 2 之有無航向變化法預測工作位置統計結果表**

	p-value	$\alpha = 0.05$
符號檢定	0.01563	顯著
魏克生符號檢定	0.01802	顯著

### (3) 整體位置

將表 4.4.1.1 中整體 27 位用戶在模式 2-1 與模式 2-2 的預測常住位置誤差結果以及表 4.4.2.1 中整體 20 位一般用戶在模式 2-1 與模式 2-2 的預測工作位置誤差結果結合，可以觀察出在兩模式比較下，有加入航向變化法的模式 1-1 經過預測後，共有 7 個位置預測誤差得以改善。誤差改善最多為 786 公尺、誤差改善最小為 11 公尺，平均誤差改善約為 240 公尺。沒有其他位置因為加入航向變化法使誤差增加。結果顯示在預測整體位置時，模式 2-1 優於模式 2-2。

將表 4.4.1.1 中整體 27 位用戶在模式 2-1 與模式 2-2 的預測常住位置誤差結果以及表 4.4.2.1 中整體 20 位一般用戶在模式 2-1 與模式 2-2 的預測工作位置誤差結果結合，進行三種統計方法檢定並輸出表 4.4.3.7，驗證模式 2-1 在加入航向變化法下，預測與實際整體位置誤差是否比模式 2-2 預測與實際整體位置誤差小。觀察表 4.4.3.7 可以發現符號檢定與魏克生符號檢定結果皆拒絕虛無假設。代表在資料清洗流程中加入航向變化法的模式 2-1 能夠明顯減少預測用戶整體常住與工作位置的誤差。綜合上述比較結果，整理至表 4.4.3.8。

**表 4.4.3.7 比較情境 2 之有無航向變化法預測整體位置統計結果表**

	p-value	$\alpha = 0.05$
--	---------	-----------------

符號檢定	0.04007	顯著
魏克生符號檢定	0.002807	顯著

表 4.4.3.8 比較情境 2 預測績效與模式驗證結果表

預測位置	樣本數	誤差有改善之樣本數	改善最多誤差(公尺)	改善最少誤差(公尺)	平均改善誤差(公尺)
常住位置	27	1	11	-	-
模式驗證	符號檢定：不顯著、魏克生符號檢定：不顯著				
工作位置	20	6	786	32	278
模式驗證	符號檢定：顯著、魏克生符號檢定：顯著				
整體位置	47	7	786	11	240
模式驗證	符號檢定：顯著、魏克生符號檢定：顯著				

#### 4.4.4 移動停留狀態判斷流程預測績效與模式驗證

- 比較情境 3：模式 1-2 與模式 2-2 兩模式在資料清洗流程中皆無加入航向變化法下，比較有加入移動停留狀態判斷流程之模式 2-2 預測績效是否優於無加入移動停留狀態判斷流程之模式 1-2 預測績效。

##### (1) 常住位置

從表 4.4.1.1 可以觀察出在兩模式比較下，有加入移動與停留狀態判斷流程的模式 2-2 經過預測後，5 號與 11 號用戶誤差分別改善 200 公尺與 24 公尺，沒有其他用戶預測誤差比模式 1-2 高。2 位用戶平均誤差改善為 112 公尺。結果顯示在預測常住位置時，模式 2-2 略優於模式 1-2。

將表 4.4.1.1 中模式 2-2 與模式 1-2 的預測與實際常住位置結果結合進行三種統計方法檢定並輸出表 4.4.4.1，驗證模式 2-2 在加入移動與停留狀態判斷流程下，預測與實際常住位置誤差是否比模式 1-2 預測與實際常住位置誤差小。觀察表 4.4.4.1 可以發現符號檢定與魏克生符號檢定結果皆不拒絕虛無假設，表示在預測常住位置時，有加入移動停留狀態判斷流程的模式 2-2 沒有顯著減少整體用戶預測與實際常住位置誤差的效果。

表 4.4.4.1 比較情境 3 之有無移動與停留狀態判斷流程預測常住位置統計結果表

	p-value	$\alpha = 0.05$
符號檢定	0.25	不顯著
魏克生符號檢定	0.1855	不顯著

##### (2) 工作位置

從表 4.4.2.1 可以觀察出在兩模式比較下，模式 2-2 與模式 1-2 預測與實際工作位置誤差結果完全相同。代表加入移動停留判斷流程沒有明顯減少預測工作位置之誤差。結果顯示在預測工作位置時，模式 2-2 與模式 1-2 預測效果相當。

##### (3) 整體位置

將表 4.4.1.1 中整體 27 位用戶在模式 1-2 與模式 2-2 的預測常住位置誤差結

果以及表 4.4.2.1 中整體 20 位一般用戶在模式 1-2 與模式 2-2 的預測工作位置誤差結果結合，可以觀察出在兩模式比較下，有加入航向變化法的模式 2-2 經過預測後，共有 2 個位置預測誤差得以改善，分別改善 200 公尺與 24 公尺。平均誤差改善為 112 公尺。結果顯示在預測整體位置時，模式 2-2 略優於模式 1-2。

將表 4.4.1.1 與表 4.4.2.1 中的模式 2-2 與模式 1-2 結合進行三種統計方法檢定並輸出表 4.4.4.2，驗證模式 2-2 在加入移動與停留狀態判斷流程下，預測與實際常住位置誤差是否比模式 1-2 預測與實際整體位置誤差小。觀察表 4.4.4.2 可以發現符號檢定與魏克生符號檢定結果皆不拒絕虛無假設，表示在預測常住位置時，有加入移動停留狀態判斷流程的模式 2-2 沒有顯著減少整體用戶預測與實際常住與工作位置誤差的效果。綜合上述比較結果，整理至表 4.4.4.3。

**表 4.4.4.2 比較情境 3 之有無移動與停留狀態判斷流程預測整體位置統計結果表**

	p-value	$\alpha = 0.05$
符號檢定	0.25	不顯著
魏克生符號檢定	0.1855	不顯著

**表 4.4.4.3 比較情境 3 預測績效與模式驗證結果表**

預測位置	樣本數	誤差有改善之樣本數	改善最多誤差(公尺)	改善最少誤差(公尺)	平均改善誤差(公尺)
常住位置	27	2	200	24	112
模式驗證	符號檢定：不顯著、魏克生符號檢定：不顯著				
工作位置	20	0	兩模式預測結果完全相同		
模式驗證	兩模式無差異				
整體位置	47	2	200	24	112
模式驗證	符號檢定：不顯著、魏克生符號檢定：不顯著				

- 比較情境 4：模式 1-1 與模式 2-1 兩模式在資料清洗流程中皆有加入航向變化法下，比較有加入移動停留狀態判斷流程之模式 2-1 預測績效是否優於無加入移動停留狀態判斷流程之模式 1-1 預測績效。

(1) 常住位置

從表 4.4.1.1 可以觀察出在兩模式比較下，有加入移動與停留狀態判斷流程的模式 2-1 經過預測後，5 號用戶誤差改善 10 公尺；但 11 號用戶誤差增加 111 公尺。結果顯示在預測常住位置時，模式 1-1 略優於模式 2-1。

將表 4.4.1.1 中模式 2-1 與模式 1-1 的預測與實際常住位置結果結合進行三種統計方法檢定並輸出表 4.4.4.4，驗證模式 2-1 在加入移動與停留狀態判斷流程下，預測與實際常住位置誤差是否比模式 1-1 預測與實際常住位置誤差小。觀察表 4.4.4.4 可以發現符號檢定與魏克生符號檢定結果皆不拒絕虛無假設，表示在預測常住位置時，有加入移動停留狀態判斷流程的模式 2-1 沒有顯著減少整體用戶預測與實際常住位置誤差的效果。

**表 4.4.4.4 比較情境 4 之有無移動與停留狀態判斷流程預測常住位置統計結果表**

	p-value	$\alpha = 0.05$
符號檢定	0.75	不顯著
魏克生符號檢定	0.8145	不顯著

### (2) 工作位置

從表 4.4.2.1 可以觀察出在兩模式比較下，模式 2-1 與模式 1-1 預測與實際工作位置誤差結果完全相同。代表加入移動停留判斷流程沒有明顯減少預測工作位置之誤差。結果顯示在預測工作位置時，模式 2-1 與模式 1-1 預測能力無差異。

### (3) 整體位置

將表 4.4.1.1 中整體 27 位用戶在模式 1-1 與模式 2-1 的預測常住位置誤差結果以及表 4.4.2.1 中整體 20 位一般用戶在模式 1-1 與模式 2-1 的預測工作位置誤差結果結合，可以觀察出在兩模式比較下，有加入航向變化法的模式 2-1 經過預測後，有 1 個位置預測誤差得以改善 10 公尺；但另有 1 個位置誤差增加 111 公尺。結果顯示在預測整體位置時，模式 1-1 略優於模式 2-1。

將表 4.4.1.1 中模式 2-1 與模式 1-1 的預測與實際常住位置結果結合進行三種統計方法檢定並輸出表 4.5.2.4，驗證模式 2-1 在加入移動與停留狀態判斷流程下，預測與實際常住位置誤差是否比模式 1-1 預測與實際常住位置誤差小。觀察表 4.4.4.5 可以發現符號檢定與魏克生符號檢定結果皆不拒絕虛無假設，表示在預測常住位置時，有加入移動停留狀態判斷流程的模式 2-1 沒有顯著減少整體用戶預測與實際常住位置誤差的效果。綜合上述比較結果，整理至表 4.4.4.6。

表 4.4.4.5 比較情境 4 之有無移動與停留狀態判斷流程預測整體位置統計結果表

	p-value	$\alpha = 0.05$
符號檢定	0.75	不顯著
魏克生符號檢定	0.8145	不顯著

表 4.4.4.6 比較情境 4 預測績效與模式驗證結果表

預測位置	樣本數	誤差有改善之樣本數	改善誤差 (公尺)	誤差有變差之樣本數	增加誤差 (公尺)
常住位置	27	1	10	1	111
模式驗證	符號檢定：不顯著、魏克生符號檢定：不顯著				
工作位置	20	兩模式預測結果完全相同			
模式驗證	兩模式無差異				
整體位置	47	1	10	1	111
模式驗證	符號檢定：不顯著、魏克生符號檢定：不顯著				

## 4.5 小結

首先，從 4.4.3 小節中比較情境 1 與比較情境 2 可以觀察出航向變化法在預

測常住位置上，有少部分用戶誤差得以改善；而在預測工作或整體位置上，有較多用戶誤差得到改善。比較情境 1 與比較情境 2 經過模式驗證後，發現航向變化法之預測績效有明確顯示降低預測與實際整體位置誤差，尤其在降低預測工作位置誤差方面；但在預測常住位置方面，航向變化法之預測績效沒有明顯減少預測與實際常住位置誤差效果。

進一步探討加入航向變化法之模式對於預測用戶常住位置誤差沒有得到明顯減少的原因。先觀察圖 4.5.1 之右圖，可以發現預測用戶常住位置時所取用到的信令資料位置分布，最遠的信令點位置僅離實際常住位置 360 公尺左右，信令資料之標記位置雖有分散的現象，但這些位置皆分佈在用戶實際常住位置附近。其原因為通常用戶實際常住位置的面積並不大，用戶帶著手機活動的空間有限，無法在家中移動太遠。相同的，用戶在其實際常住位置中產生信令資料的位置不會離實際常住位置太遠，因此本研究在後續預測用戶常住位置時，預測常住位置誤差皆在 500 公尺內，平均誤差約為 148 公尺。基於上述原因，導致資料清洗流程中加入航向變化法後，可能對於減少預測與實際位置誤差的空間並不多。相反的，從圖 4.5.2 左圖中可以觀察出預測 1 號用戶工作位置時所取用信令資料分布的現象，其信令資料紀錄的位置分佈較散，從比較情境 1 與比較情境 2 皆可看出加入航向變化法後，皆有 6 位一般用戶預測與實際工作位置誤差得以改善，改善最多誤差高達 786 公尺。而在 4.4.3 小節模式驗證中明確顯示加入航向變化法後，用戶預測與實際整體位置誤差得到顯著的減少，尤其在預測工作位置方面得到明顯降低誤差的結果。從表 4.4.2.1 可以得知有部分一般用戶在模式 1-2 之預測與實際工作位置誤差超過 500 公尺，代表可以改善誤差距離的空間很大。因為與預測常住位置相比，預測工作位置誤差較大的原因可能為工作位置通常比住常住大，例如在工業區或科學園區，且用戶在工作中可能會帶著手機在公司或工廠內活動，這類的活動會使用戶信令資料所標記位置的出現範圍比常住位置中信令資料所標記位置的出現範圍大上許多也較複雜，導致本研究預測工作位置誤差會比預測常住位置的誤差還高，但較差的條件反而賦予航向變化法有了充足的改善誤差的空間。綜合上述實測與驗證結果後做出以下兩點解釋：

- (1) 大多數用戶的實際常住位置面積並不大，並且在實際常住位置中，用戶活動的空間有限，導致被動產生的信令資料位置標記不會離實際常住位置太遠，因此無論使用何種模式預測常住位置誤差結果皆在 500 公尺內，使本研究加入航向變化法改善預測常住位置誤差的空間不大。
- (2) 一般來說，用戶的工作位置面積較大，例如位於工業區或科學園區，工作時可能會在工作位置內部移動，導致被動產生的信令資料位置標記有些可能會離實際位置較遠，但這樣較差的現象也提供航向變化法減少預測用戶工作位置誤差的空間，因此有較多的用戶在加入航向變戶法後，預測與實際的工作位置誤差得以改善。

另外，從比較情境 3 與比較情境 4 中可以發現加入移動停留判斷流程之預測績效皆無法明確顯示減少預測與實際常住、工作與整體位置誤差。從表 4.4.1.1 與

表 4.4.2.1 中可以發現預測與實際平均誤差中，模式 1-1 平均誤差略低於模式 2-1 之平均誤差。因此本研究認為未來在預測用戶位置時，可考慮不加入移動與停留狀態判斷流程。



圖 4.6.1 1 號用戶某日實際常住與工作位置附近之信令資料分布圖

## 第五章 結論與建議

本研究之主要目標為優化訊號飄移問題並預測用戶常住與工作位置應用於後續都市規劃、交通運輸等相關分析中。經過資料實測與模式驗證後，首先討論在資料清洗流程中加入航向變化法後，對於優化訊號飄移問題的效果以及適用情境。其次，討論本研究對於改良預測常住與工作位置之一系列流程後，預測用戶常住與工作位置誤差減少之成效。因此，以下結論為本研究之研究成果，並提出對未來相關研究之建議。

### 5.1 結論

1. 本研究在資料清洗流程中，加入自行設計的航向變化法能夠有效優化訊號飄移問題，使研究成果更為可靠。本研究中以預測用戶的常住與工作位置做為實測，經過統計檢定結果發現加入航向變化法後，確實能減少預測整體位置的誤差，尤其在預測工作位置方面。
2. 航向變化法無明顯改善預測用戶常住位置誤差之原因為大多數用戶的實際常住位置面積並不大，並且在實際常住位置中用戶活動的空間有限，導致信令資料位置標記不會離實際常住位置太遠，因此在本研究資料實測中無論使用何種模式預測常住位置誤差結果皆在 500 公尺內，使本研究加入航向變化法改善預測常住位置誤差的空間不大。
3. 本研究與過往文獻的資料清洗流程相比下，發現本研究資料清洗流程優化訊號飄移現象的成效優於過往文獻之流程。從資料實測結果中可以清楚得知模式 1-1 優於模式 1-2；模式 2-1 優於模式 2-2。只要資料清洗流程中有包含航向變化法就能使預測結果更為精準。
4. 資料實測結果之精準度顯示本研究一系列預測常住與工作位置流程確實能準確得出用戶的常住、學校與工作等重要位置，以利後續應用於交通運輸領域相關當中，甚至是都市規劃等其他領域。傳統運輸規劃所使用的交通分析區域等空間單位皆大於 500 公尺。而本研究在預測與實際常住位置的誤差都在 500 公尺以內，平均誤差為 148 公尺。多數用戶預測與實際工作位置的誤差都在 500 公尺內，平均誤差為 269 公尺。
5. 本研究在用戶常住與工作位置判斷邏輯中進行改良。在流程中加入位置性質的判斷流程，目的是在推派最終的常住或工作網格前，找出並標記每日可能非常住或工作候選網格，使其不影響最終常住或工作網格的預測。
6. 從研究結果中可以發現移動停留狀態判斷流程可以合理的區分出用戶信令資料移動狀態與停留狀態。一般用戶所有信令資料中，有超過九成以上資料被判斷為停留狀態，而物流士所有信令資料中，有接近三成資料被判斷為移動狀態。
7. 使用本研究的資料清洗流程(跳點規則、航向變化法與時窗法)，並直接進行

預測位置的模式 1-1 預測績效最佳。

## 5.2 建議

1. 由於本研究僅收集 14 日中 27 位用戶之信令資料，並以用戶所提供實際的常住與工作位置進行驗證，以樣本數量來說略顯不足。本研究皆以少量樣本測試得出資料清洗演算法、停留與移動狀態判斷邏輯與常住與工作地點判斷邏輯之最適參數。建議未來可蒐集更多的樣本進行調教個流程所需之參數，使得出的參數更具參考價值。
2. 本研究有少數用戶(17 號與 20 號用戶)在預測與實際工作位置誤差有超過兩公里的情況，深入探討後發現問題原因為資料收集的天數過短，導致用戶在一般週間有排休的情況，依本研究邏輯會誤將其識別為上班日，並進行工作位置的預測導致影響預測最終工作位置的結果。建議未來可延長資料收集期間即可避免此問題。
3. 本研究發現移動停留狀態判斷流程無法明顯提升預測用戶位置之精準度，建議未來在預測位置時可將此流程刪除。
4. 建議加入其他資料與設計不同的判斷邏輯來辨別特殊情況的用戶，使後續分析結果更符合實際情況。例如可以透過資料移動的比例高低來推估用戶的工作性質是否與物流士相符。
5. 由於航向變化法在偵測不同網格形成資料區間時並無考慮時間限制，可能會使得部分資料區間中的第一筆資料與最後一筆資料時差過大，會造成模式判斷與修正法錯誤修正信令位置的疑慮。因此，建議後續可針對如航向變化法取用資料區間的方法進行改良。

## 參考文獻

- [1] 交通部運輸研究所，應用旅次特性大數據精進公共運輸服務計畫，2018年，高雄，台灣。
- [2] 台灣世曦工程顧問股份有限公司，應用行動信令資料於運輸規劃之分析程序與運算邏輯，2019年，台灣。
- [3] 顧凱嫻，「使用信令資料判斷旅運起訖點」，國立陽明交通大學，碩士論文，2021年。
- [4] Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., & González, M. C. (2015). Analyzing cell phone location data for urban travel: current methods, limitations, and opportunities. *Transportation Research Record*, 2526(1), 126-135.
- [5] Caceres, N., Romero, L. M., Benitez, F. G., & del Castillo, J. M. (2012). Traffic flow estimation models using cellular phone data. *IEEE Transactions on Intelligent Transportation Systems*, 13(3), 1430-1441.
- [6] Hard, E., Chigoy, B., Songchitruksa, P., Farnsworth, S., Borchardt, D. W., & Green, L. (2016). Synopsis of new methods and technologies to collect Origin-Destination (OD) data (No. FHWA-HEP-16-083). *United States Federal Highway Administration*.
- [7] Iovan, C., Olteanu-Raimond, A. M., Couronné, T., & Smoreda, Z. (2013). Moving and calling: Mobile phone data quality measurements and spatiotemporal uncertainty in human mobility studies. In *Geographic information science at the heart of Europe* (pp. 247-265). Springer, Cham.
- [8] Jameson L., Serdar Colak, Bradley Sturt, Lauren P., Alexander Evsukoff, Marta C. (2015). The path most traveled: Travel demand estimation using big data resources. *Transportation Research Part C: Emerging Technologies*, 58, 162-177.
- [9] Järv, O., Ahas, R., & Witlox, F. (2014). Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies*, 38, 122-135.
- [10] Lee, J. K., & Hou, J. C. (2006, May). Modeling steady-state and transient behaviors of user mobility: formulation, analysis, and application. *In*

*Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing* (pp. 85-96).

- [11] Lauren Alexander, Shang Jiang, Mikel Murga, Marta C. Gonzalez (2015). Origin-destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*, 58, 240-250.
- [12] Md. Mahedi Hasan, & Mohammed Eunus Ali (2017). Estimating Travel Time of Dhaka City from Mobile Phone Call Detail Records *In Proceedings of ICTD '17. ACM*, New York, NY, USA, 11 pages.
- [13] Ma, J., Li, H., Yuan, F., & Bauer, T. (2013). Deriving operational origin-destination matrices from large scale mobile phone data. *International Journal of Transportation Science and Technology*, 2(3), 183-204.
- [14] Rainer Kujala, Talayeh Aledavood & Jari Saramäki (2016). Estimation and monitoring of city-to-city travel times using call detail records. *EPJ Data Science*, 5:6. DOI 10.1140/epjds/s13688-016-0067-3.
- [15] Shan Jiang, Joseph Ferreira, Jr., and Marta C. Gonzalez (2017). Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE TRANSPORTATIONS ON BIG DATA* , VOL.3, NO.2.
- [16] Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, 10(84), 20130246.
- [17] Schlaich, J., Otterstätter, T., & Friedrich, M. (2010, January). Generating trajectories from mobile phone data. In *Proceedings of the 89th annual meeting compendium of papers, transportation research board of the national academies*.
- [18] Tongsinoot, L., & Muangsin, V. (2017, December). Exploring home and work locations in a city from mobile phone data. In *2017 IEEE 19th International Conference on High Performance Computing and Communications; IEEE 15th International Conference on Smart City; IEEE 3rd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)* (pp. 123-129). IEEE.
- [19] Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A., & González, M. C. (2015). The path most traveled: Travel demand estimation using big

data resources. *Transportation Research Part C: Emerging Technologies*, 58, 162-177.

- [20] Wei-Kuang Lai & Ting-Huan Kuo (2016). Vehicle position and speed estimation based on cellular network signal for urban roads. *International Journal of Geo-Information*.
- [21] Wang, M. (2014). Understanding activity location choice with mobile phone data (Doctoral dissertation).
- [22] Wang, F., & Chen, C. (2018). On data processing required to derive mobility patterns from passively-generated mobile phone data. *Transportation Research Part C: Emerging Technologies*, 87, 58-74.
- [23] Wu, W., Wang, Y., Gomes, J. B., Anh, D. T., Antonatos, S., Xue, M., ... & Shi-Nash, A. (2014, July). Oscillation resolution for mobile phone cellular tower data to enable mobility modelling. In *2014 IEEE 15th International Conference on Mobile Data Management* (Vol. 1, pp. 321-328). IEEE.