

# 應用手機信令資料預測觀光景點拜訪人數

## APPLYING CELLULAR-BASED VEHICLE PROBE DATA TO PREDICT NUMBER OF VISITORS AT ATTRACTIONS

盧宗成 Chung-Cheng Lu<sup>1</sup>

梁竣凱 Jyun-Kai Liang<sup>2</sup>

曲 平 Ping Cyu<sup>3</sup>

王晉元 Jin-Yuan Wang<sup>4</sup>

吳東凌 Tung-Ling Wu<sup>5</sup>

陳翔捷 Siang-Jie Chen<sup>6</sup>

(109年5月1日收稿，109年7月28日第一次修改，  
109年8月31日第二次修改，110年5月17日定稿)

### 摘 要

了解遊客在觀光景點間的移動行為與景點的拜訪人次，對於改善觀光地區公共運輸服務與觀光管理當局之資源配置具有極大助益。隨著科技發展的日新月異，以及人們使用手機行動上網的普及率提升，在人口移動預測上，手機信令資料有著樣本數量、涵蓋範圍廣且蒐集成本相對低廉的優點。手機信令資料分析可以有效建構使用者的時空軌跡，進而得到使用者在景點間潛在的移動型態。本研究建立一套有系統的景點拜訪人數預測方法，首先透過分析使用者的信令資料建立旅次鏈與旅次起迄矩陣，作為景點間轉移矩陣推估的基礎，再利用馬可夫鏈預測每小時觀光景點拜訪人

- 
1. 國立陽明交通大學運輸與物流管理系教授 (聯絡地址：100 臺北市中正區忠孝西路 1 段 118 號 4 樓，電話：02-23494960，傳真：02-23494953，E-mail：jasonccclu@nctu.edu.tw)。
  2. 玄奘大學應用心理學系助理教授。
  3. 國立陽明交通大學運輸與物流管理學系碩士。
  4. 國立陽明交通大學運輸與物流管理學系副教授。
  5. 交通部運輸研究所運輸資訊組組長。
  6. 交通部運輸研究所運輸資訊組副研究員。

次。本研究以花蓮縣作為研究場域，以交通部觀光局推薦的 59 個主要觀光景點為對象，透過由電信公司取得觀光客的信令資料，評估所提出預測方法的績效。結果顯示預測之平均絕對百分比誤差約為 20%，屬於實務上可接受的範圍，表示本研究之預測方法具有不錯的成效。

**關鍵詞：** 信令資料、旅次鏈、旅次起迄矩陣、馬可夫鏈、觀光景點拜訪人次

## ABSTRACT

*Understanding the movement of tourists between attractions and the number of visitors at attractions is helpful for the authorities to improve the public transportation services in scenic areas and to allocate resources. With the rapid development of technology and the increasing popularity of people using mobile phones to access the Internet, cellular-based vehicle probe (CVP) data has the advantages of larger sample size, broader coverage and lower collection cost in human mobility prediction. Users' spatial-temporal trajectories can be effectively constructed by analyzing CVP data. The potential movement pattern of users can be extracted from those trajectories. This study develops a systematic approach to predict the number of visitors at attractions. Firstly, trip chains and origin-destination (OD) matrix are constructed by analyzing users' CVP data. The OD matrix is used as the basis for estimating the transition matrix between attractions. Then, a Markov Chain model is established to predict the number of tourists at each attraction in each hour. The proposed method is applied to predict the numbers of tourists at 59 major attractions, recommended by the Tourism Bureau, in Hualien county. The CVP data is provided by a major telecom in Taiwan. The mean absolute percentage error of the prediction results is about 20%, which is practically acceptable. The evaluation results indicate that the proposed method has a good prediction performance.*

**Key Words :** *cellular-based vehicle probe data, trip chain, origin-destination (OD) matrix, Markov Chain, number of visitors at attractions.*

## 一、緒 論

近年來觀光產業廣受世界各國所重視，對於創造就業機會及賺取外匯收入上具有極大效益。世界觀光組織 (United Nations World Tourism Organization, UNWTO) 統計 2018 年國際遊客人數成長了 6% 達到 14 億人次，同時也預測 2019 年的國際遊客人數將再成長 3% 到 4%，顯示全球觀光呈現長期穩定成長的趨勢。臺灣由於地理位置特殊，擁有豐富的觀光資源，發展觀光為政府促進經濟發展之重點產業，除了吸引外國人士來臺觀光外，並讓

國人樂於留在國內度假。其中，花蓮縣位於臺灣東部，開發較西部緩慢，保留了優美、無汙染的自然景觀與多元的族群文化特色，是許多人國內旅遊的第一選擇。花蓮的旅遊業相當興盛且觀光景點眾多，若能得知一天中各個時段景點內之拜訪人數，有助於景點內資源配置發揮最大效益；而政府或業者透過任何時間各景點的人數多寡，可了解國人的旅遊需求，進而把效益提高，觀察資源配置在觀光地區是否有不足的情形，例如：景點間公共運輸服務路線能否滿足遊客需求等，並針對資源配置不足的地方進行改善。

隨著大數據分析技術的發展及手機行動上網的普及率不斷提升，行動通訊資料有著較低的蒐集成本、樣本數大、更新快、範圍廣的優點。藉由使用者與電信公司基地台互動，可產生使用者活動時的紀錄，能以較直接的方式取得民眾的移動資訊。將手機的行動通訊資料作為資料蒐集的來源，藉由此資訊可間接得知使用者的時空軌跡、並分析使用者的旅運行為，在實務上十分具有參考價值。本研究利用手機信令資料 (cellular-based vehicle probe, CVP)，建立一有系統的預測方法，以馬可夫鏈模式 (Markov Chain Model) 預測各時段之景點拜訪人次，提供觀光景點各時段拜訪人數之變化，了解旅遊需求，輔助政府機關之政策規劃或觀光景點間之公共運輸服務，同時可作為業者在景點內設施規劃與景點內部人力資源之配置參考，達到促進觀光產業發展之目的。

本研究以交通部觀光局推薦的 59 個花蓮主要觀光景點為研究場域，以前往花蓮旅遊之觀光客為研究對象，使用國內某大電信公司所提供的手機信令資料，其數據內容包含使用者去個資之識別碼、日期、時間與所在的網格 ID，由使用者的歷史信令資料推估旅次起迄矩陣，計算各時段遊客在各景點間移動的轉移機率，接續以馬可夫鏈模式預測未來各時段景點拜訪人數，以掌握一天中各時段景點拜訪人次的變化，並針對預測結果計算準確率。由於信令資料可能因信號飄移導致資料缺漏而難以判斷使用者的移動軌跡，因此，本研究取得之信令資料已由該電信公司進行資料清洗與插補處理，以減少上述情形發生。

本文之架構說明如下：第二節為相關研究回顧，包含利用手機信令資料推估旅運行為與大眾運輸服務改善之應用，及應用馬可夫鏈預測使用者移動行為之相關研究；第三節為本研究所建構的分析架構與演算流程，包含旅次切割與起訖矩陣產生、轉移矩陣推估及各時段各景點人數推估；第四節為旅次數數據分析結果；最後一節則為研究結果與討論及未來研究建議。

## 二、文獻回顧

### 2.1 手機信令資料對旅運需求之推估

隨著手機使用行為及行動上網的普及，信令資料成為一種分析使用者特性的資料來源，本小節將針對信令資料應用於民眾旅運需求之分析進行回顧。

Ma et al.<sup>[1]</sup> 以美國沙加緬度市 (Sacramento) 的兩條主要高速公路為研究對象，利用

手機信令資料提出起迄矩陣推估流程。該研究以模擬軟體進行交通量指派模擬，得到相關係數為 0.88 至 0.92，顯示推估結果與實際情況符合程度高。Iqbal et al.<sup>[2]</sup> 以孟加拉的達卡市 (Dhaka) 為研究區域，利用電信公司提供的手機信令資料推估旅次起迄矩陣，該研究透過與路網疊圖、篩選機制、使用交通模擬結果估計之放大因子，將信令資料轉換為路網節點對路網節點的起迄矩陣，以模擬軟體進行交通量指派，其方均根百分誤差 (Root Mean Square Percent Error) 為 13.59%。Li et al.<sup>[3]</sup> 提出一個三階段程序，利用手機信令資料分析用戶的空間互動 (spatial interaction) 情形，其以上海三個社區為研究區域，利用頻繁樣式 (frequency pattern) 探勘技術，推估出該三個社區居民的旅次起迄分布。Alexander et al.<sup>[4]</sup> 以美國波士頓為研究區域，提出推估用戶每天不同旅次目的的起迄矩陣方法，以一天內連續使用的兩點為每個用戶構建行程，乘以人口普查的放大因子、除以觀察天數推估每日起迄量，再與美國運輸規劃人口普查調查相比，兩者之相關係數接近於 1。Qian et al.<sup>[5]</sup> 提出一套手機信令資料處理程序，用以識別並分析旅客及其在城市中的旅遊時空分布情形，其程序首先利用 DBSCAN (Density-based spatial clustering of applications with noise) 演算法辨識出旅客，後續根據互動強度 (strength of interaction)、互動對稱性 (symmetry of interaction) 及網路結構 (structure of network) 建立三個 S 維度 (s-dimension) 指標，並以上海市為分析範圍，辨識出 6.44% 的手機用戶為旅客，並發現節假日、高鐵線路和城市位置對遊客時空分佈有顯著影響。邱裕鈞和謝志偉<sup>[6]</sup> 透過蒐集手機信令資料，結合地圖匹配演算法及最短路徑演算法，提出一套即時估計道路交通資訊演算法。該研究以 GVP (GPS-based vehicle probe) 交通資訊作為比較的基準值，結果顯示，所提出的演算法推估結果之平均絕對誤差百分比 (Mean Absolute Percentage Error, MAPE) 值為 32%，且道路資訊覆蓋率可達 86%。洪琮博<sup>[7]</sup> 利用透過串接不同來源之資料，引入擴增係數的概念，透過加乘結果回推實際人數，並探討影響擴增係數的因素。在 9 種推估結果中，在晨峰方面，以擴增係數因時而異且設定 2 條屏柵線者表現最佳，其整體準確率 MAPE 值為 117.2%。

## 2.2 手機信令資料應用於大眾運輸之服務改善

由於手機信令資料具有高覆蓋率、更新頻率迅速、樣本數大、以及成本低的優點，可以進行更深入的了解個體行為及民眾潛在需求，對大眾運輸之服務改善十分有幫助，本小節將探討手機信令資料應用於大眾運輸服務改善之相關研究。

Demissie et al.<sup>[8]</sup> 透過手機信令資料估計大眾運輸服務之需求，並提出新增或是延長公車路線的建議。研究中整合距離 500 公尺以內的基地台，表示為車站候選點，視不同預算限制的情境，選擇需要新增的站點，由此分析結果可以看出需要增加班次之路線、可以延駛之路線，以及是否需要增開新的服務路線。Ying et al.<sup>[9]</sup> 利用手機信令資料作為雨天的公車排班基礎，研究中將場站有大量乘客分為兩種情況，一是乘客在等待特定路線，擁擠站點較集中，這時只要增加公車服務頻率，二是乘客不是在等待特定路線，擁擠站點較分散，這就可以利用增加接駁公車來解決。王晉元等人<sup>[10]</sup> 結合手機信令資料及「公共運輸整合資訊流通服務平台 (PTX)」所擷取的公車動態資訊，經由時空軌跡比對的方式發展

出不同運具之判斷邏輯。該研究於高雄市及花蓮縣進行實測，其結果顯示，運具判讀的正確率在 0.75~0.8 之間，屬於可接受的範圍。再根據前述運具選擇判讀結果，產製四種旅運起迄表格分析公共運輸可能存在的缺口。該研究同時利用使用者的信令資料推測使用者之常住區域，不僅較傳統的戶籍資料更能掌握民眾實際居住分布情形，並能根據上述成果作進一步之公共運輸需求分析。王以萱<sup>[11]</sup>以手機信令資料為基礎，推估使用者的旅次鏈資料，考量總成本限制，設計使用者總搭乘人數最大化之公車服務路網及行駛班距。研究中以宜蘭縣觀光景點為實例驗證地區，共設計四條公車路線銜接主要觀光景點及轉運站，每條路線都依據旅次鏈需求有不同之班次設計，以符合旅客的實際需求。施冠毅<sup>[12]</sup>利用手機信令資料建立用戶的時空軌跡，依據用戶特徵將用戶進行分群，並挑選具有重現性需求之用戶，再從以上用戶的旅次鏈中，以循序探勘的方式找出高頻的需求樣態。該研究以電信公司所提供之高雄市隨機取樣信令資料進行實測分析，所找出的高頻樣式結果主要集中於市區，代表此處可能會有較多的民眾聚集停留，而且是經常性的需求，並非偶發性的需求，根據此資訊可以提供公部門作為規劃公車路線之參考。

### 2.3 應用馬可夫鏈預測使用者移動行為

本研究使用馬可夫鏈模式 (Markov Chain Model) 為預測方法，馬可夫鏈模式是一種具有狀態的隨機過程 (Stochastic process)，模式假設狀態之間的轉移機率 (Transition probability) 不會受到狀態以外的因素所影響，例如：時間、先前狀態等，並以此轉移機率來推測未來事物的分布狀態 (Norris<sup>[13]</sup>)。馬可夫鏈模式的分析方法已被廣泛運用在許多不同領域，例如：管理決策分析、醫療追蹤、土地利用等領域。一階的馬可夫鏈僅能描述前一個狀態的變化率，若給定當前狀態及前 N-1 個狀態，且未來狀態與先前狀態獨立，則稱為 N 階馬可夫鏈 (N-order Markov Chain)，高階的馬可夫鏈可降低預測的不確定性，因此有較佳的預測力。

Gambs, Killijian & Cortez<sup>[14]</sup>以馬可夫鏈模式考慮使用者前 n 個拜訪過的興趣點 (Points of interest, POIs)，預測使用者的下一個位置。該研究以三組 GPS 資料進行實測評估，分別計算準確度與預測力，結果顯示當 n = 2 時，準確度與預測力大約落在 70%到 95%，若再選擇 n > 2 以上的模式對於準確度與預測力沒有明顯的改善。Ashbrook & Starner<sup>[15]</sup>透過 GPS 定位蒐集使用者的位置資訊，再利用馬可夫鏈模式預測使用者的位置。研究結果顯示該演算法對於不同研究對象具有一致性。Xia et al.<sup>[16]</sup>利用馬可夫鏈模式對遊客的時空移動型態進行分析，並以澳洲自然公園 9 個觀光景點進行實徵研究，透過問卷的方式蒐集遊客在景點間移動行為的資訊。該研究結果顯示資料與模式有良好的配適度；然而，當景點拜訪數量 n 愈大時，模式的配適程度會降低，因此該模式適用於行程數量不要太多的情況。Mathivaruni & Vaidehi<sup>[17]</sup>設計一以預測使用者活動為基礎的移動模式，考慮使用者過去的移動型態、依照時間的規律性進行預測，給定前 n 天的資訊以預測第 n+1 天的移動，並以馬可夫鏈模式預測第 n+1 天的下一個活動。該研究以印度清奈使用者為研究對象，以 hit ratio 為績效指標，可達到 82.6%的準確度。Jiang et al.<sup>[18]</sup>透過蒐集使用者目前位置資料

及相關的歷史移動資料，以馬可夫鏈模式預測使用者未來的移動行為。該研究蒐集 182 位使用者五年的 GPS 軌跡資料，透過訊息熵 (Information Entropy) 方法分別計算一階及二階馬可夫鏈之轉移矩陣的平均訊息熵，結果顯示，K 階馬可夫鏈模式可以降低預測較多的不確定性，得到較準確的結果。

綜上所述，目前已有不少研究應用手機信令資料在交通旅運需求之分析與推估，而且皆有不錯的成效，表示以手機信令資料作為交通旅運需求之分析與推估之資料來源是可行且效果不錯的方法。然而目前較少針對觀光旅運需求之研究，因此本研究希望以觀光景點為研究對象，了解前往旅遊的使用者旅運需求的所在地點以及時間。本研究主要參考 Xia et al.<sup>[16]</sup> 之研究，透過離散時間的馬可夫鏈模式 (Discrete Time Markov Chain Model) 建立一個預測觀光景點拜訪人數的方法。

### 三、研究方法

#### 3.1 研究架構

本研究利用手機信令資料，並以交通部觀光局推薦之花蓮縣 59 個主要觀光景點為對象，篩選於景點中移動的使用者。首先針對上述使用者找出使用者整日的時空軌跡，計算每位使用者於景點的停留時間，進行旅次鏈分析。接著以特定停留時間作為旅次切割之門檻值，產製使用者一天內各時段之起迄矩陣。透過各時段起迄矩陣內記錄之各景點移動旅次數，作為推估馬可夫鏈模式中轉移機率矩陣之依據，轉移機率矩陣用以描述使用者在未來時段向各個景點移動之機率。因此只要取得當前景點內的使用者人數，即可利用馬可夫鏈模式之轉移機率矩陣預測未來對應時段中的景點拜訪人數。本研究之轉移機率矩陣將透過多天歷史資料訓練，使轉移機率矩陣趨近於穩定狀態，使各景點間之轉移機率較不會因時間改變而呈現不同的狀態分布。

由於本研究設計之各時段景點人數預測方法，並未於封閉系統內與特定限制條件下進行。也就是說，系統內總人數將會因時段的不同而有所改變。由於每位使用者之旅次鏈開始時間與結束時間不盡相同，若僅透過馬可夫鏈模式之預測方法，將無法完整地考慮到因系統總人數變化所帶來的預測誤差。因此，本研究透過於馬可夫鏈模式預測方法後加入一「誤差項」的方式給予補償，用以描述因系統總人數變化造成之預測誤差，提升整體預測準確率。本研究景點人數預測架構可表示如圖 1 所示，其中各步驟將於後續章節中詳細說明。

本研究之研究方法與流程可分為兩階段，第一階段為訓練階段，信令資料在蒐集時已經過資料插補與清洗步驟，降低信令資料缺漏及修正信號飄移問題。應用信令資料分析用戶旅運行為時，最理想狀況為每一分鐘都有一筆信令資料，以便完整定義出使用者的停留或移動行為，故若無連續資料則進行資料清洗與插補。其過程如圖 2 所示，首先找出個別的代表點 G\_ID (欄位 1) 來代表此停留群組，而每個信令資料皆有變化百分比的數值，該

數值表示此 G\_ID 前後 10 分鐘內其他信令資料落在其 5\*5 宮格內比例，所以將每個群組取出變化百分比最高的 G\_ID 代表每個群組，並將同一群組內的信令資料 G\_ID 以其取代之，作為該停留群組的代表網格；如圖 2 欄位 4 所示，百分比最高為 85772062 的 94%，作為代表點取代群組內其他信令資料，最後，再於同一群組內檢視信令資料之時間間距，以相同 G\_ID 插補出每一分鐘的資料 (13:50-13:54, 欄位 2、3)，如此完成靜態停留清洗與插補作業。接著利用歷史資料進行旅次鏈分析、旅次切割及馬可夫鏈模式之轉移機率矩陣推估。

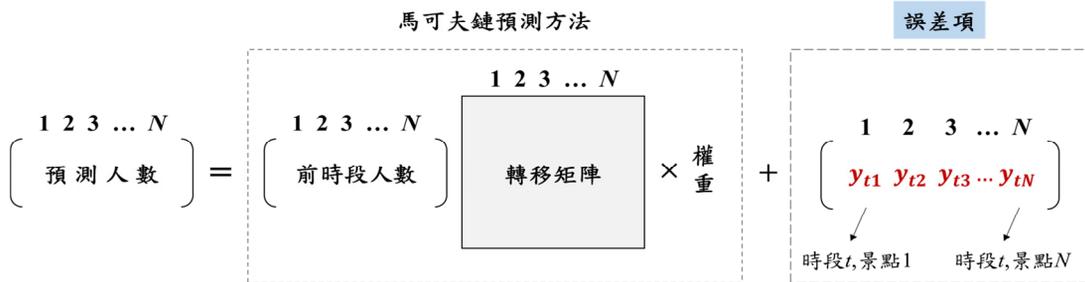


圖 1 各景點人數預測架構

|          |                |           |    |          |                |             |    |          |                |            |
|----------|----------------|-----------|----|----------|----------------|-------------|----|----------|----------------|------------|
| 85772063 | 2018/6/6 13:43 | 0.625     |    | 85772062 | 2018/6/6 13:43 | 0.625       |    | 85772062 | 2018/6/6 13:43 | 0.625      |
| 85772064 | 2018/6/6 13:44 | 0.6875    |    | 85772062 | 2018/6/6 13:44 | 0.6875      |    | 85772062 | 2018/6/6 13:44 | 0.6875     |
| 85772065 | 2018/6/6 13:45 | 0.7058824 |    | 85772062 | 2018/6/6 13:45 | 0.705882353 |    | 85772062 | 2018/6/6 13:45 | 0.70588235 |
| 85772063 | 2018/6/6 13:46 | 0.7222222 |    | 85772062 | 2018/6/6 13:46 | 0.722222222 |    | 85772062 | 2018/6/6 13:46 | 0.72222222 |
| 85772063 | 2018/6/6 13:47 | 0.7777778 |    | 85772062 | 2018/6/6 13:47 | 0.77777778  |    | 85772062 | 2018/6/6 13:47 | 0.77777778 |
| 85772068 | 2018/6/6 13:48 | 0.8333333 |    | 85772062 | 2018/6/6 13:48 | 0.833333333 |    | 85772062 | 2018/6/6 13:48 | 0.83333333 |
| 85772063 | 2018/6/6 13:49 | 0.8823529 |    | 85772062 | 2018/6/6 13:49 | 0.882352941 |    | 85772062 | 2018/6/6 13:49 | 0.88235294 |
| 85772062 | 2018/6/6 13:55 | 0.9411765 |    | 85772062 | 2018/6/6 13:55 | 0.941176471 |    | 85772062 | 2018/6/6 13:50 | 插補         |
| 85772062 | 2018/6/6 13:56 | 0.8823529 |    | 85772062 | 2018/6/6 13:56 | 0.882352941 |    | 85772062 | 2018/6/6 13:51 | 插補         |
| 85772062 | 2018/6/6 13:57 | 0.8235294 |    | 85772062 | 2018/6/6 13:57 | 0.823529412 |    | 85772062 | 2018/6/6 13:52 | 插補         |
| 85772062 | 2018/6/6 13:58 | 0.7647059 |    | 85772062 | 2018/6/6 13:58 | 0.764705882 |    | 85772062 | 2018/6/6 13:53 | 插補         |
| 85772063 | 2018/6/6 14:00 | 0.6470588 |    | 85772062 | 2018/6/6 14:00 | 0.647058824 |    | 85772062 | 2018/6/6 13:54 | 插補         |
| 85772067 | 2018/6/6 14:01 | 0.6111111 | 清洗 | 85772062 | 2018/6/6 14:01 | 0.611111111 | 插補 | 85772062 | 2018/6/6 13:55 | 0.94117647 |
| 85772063 | 2018/6/6 14:02 | 0.5789474 |    | 85772062 | 2018/6/6 14:02 | 0.578947368 |    | 85772062 | 2018/6/6 13:56 | 0.88235294 |
| 85772063 | 2018/6/6 14:03 | 0.55      |    | 85772062 | 2018/6/6 14:03 | 0.55        |    | 85772062 | 2018/6/6 13:57 | 0.82352941 |
| 85772089 | 2018/6/6 14:04 | 0.5       |    | 85772062 | 2018/6/6 14:04 | 0.5         |    | 85772062 | 2018/6/6 13:58 | 0.76470588 |
| 85758015 | 2018/6/6 14:05 | 0.15      |    |          |                |             |    | 85772062 | 2018/6/6 14:00 | 0.64705882 |
| 85758015 | 2018/6/6 14:05 | 0.15      |    |          |                |             |    | 85772062 | 2018/6/6 14:01 | 0.61111111 |
| 85758015 | 2018/6/6 14:06 | 0.15      |    |          |                |             |    | 85772062 | 2018/6/6 14:02 | 0.57894737 |
| 85496057 | 2018/6/6 14:07 | 0.15      |    |          |                |             |    | 85772062 | 2018/6/6 14:03 | 0.55       |
| 85496057 | 2018/6/6 14:08 | 0.1       |    |          |                |             |    | 85772062 | 2018/6/6 14:04 | 0.5        |
| 85487014 | 2018/6/6 14:09 | 0.1       |    |          |                |             |    |          |                |            |

圖 2 停留資料清洗插補

第二階段為預測階段，則利用馬可夫鏈模式以及誤差項估計進行各個時段景點拜訪人數之推估與預測，最後針對上述預測結果之準確性進行驗證，主要研究架構如圖 3 所示：

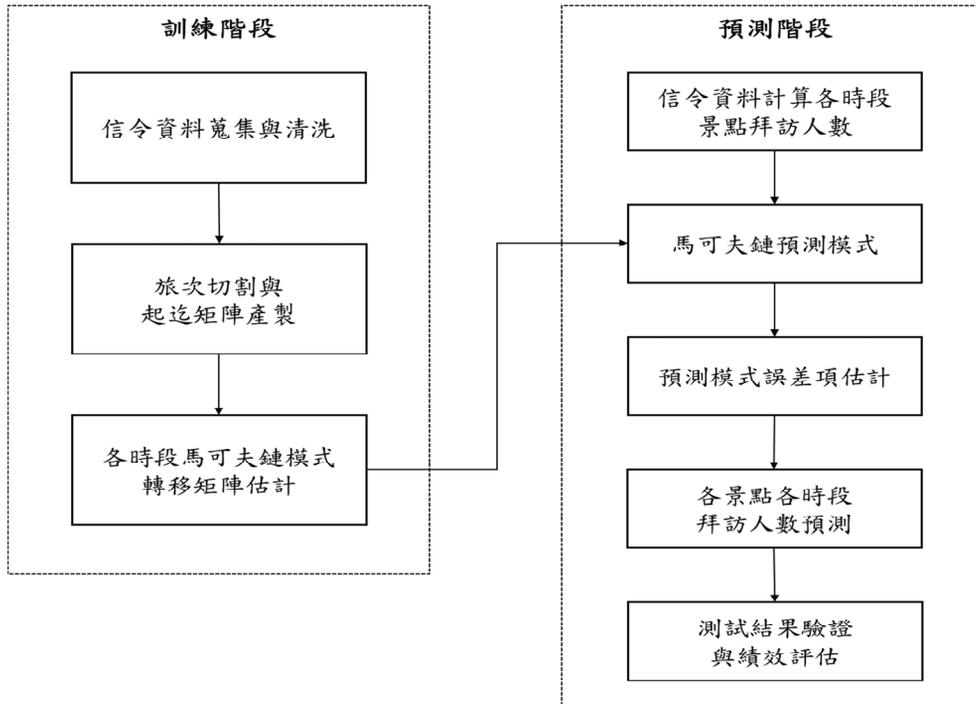


圖 3 研究流程

### 3.2 旅次切割與起迄矩陣產製

本研究的資料來源為國內某大電信公司所提供的去個資信令資料，其原始資料已經過處理，扣除屬於當地居民的使用者，同時使用者中可能包含本國遊客及外國遊客。信令資料的時間單位為每分鐘一筆，空間單位為將研究場域切割為許多 50 公尺\*50 公尺的網格分區，並給予一網格 ID 命名，信令資料範例如表 1 所示。其中第一欄為使用者去個資之識別碼，為電信公司對用戶門號進行去個資處理，以避免產生個資外洩的問題。第二欄與第三欄為時間單位，代表使用者使用手機與基地台互動時所記錄的日期與時間。第四欄為空間單位，代表使用者使用手機與基地台互動時所在的網格。以表 1 中的第一筆資料為例，該名去個資使用者 ID 前五碼為 4d99d 的使用者，在時間 2018 年 5 月 7 日的 18 時 50 分位於網格 ID 為 69325100 的位置。

本研究設計之旅次起迄矩陣產製邏輯如圖 4 所示。由於觀光景點範圍較大，通常涵蓋數十個網格，僅透過 50 公尺\*50 公尺的網格記錄使用者移動資訊，容易因範圍過於精細造成判斷上的困難。因此，必須先建立各景點與網格分區之間的對應關係，將網格聚合成較大範圍的單位，篩選出重要性較高的地點。本研究將各景點區所對應的網格群組，視為同一個 POI (Points of Interest, 興趣點)，並給予其相同的名稱，若某網格群組屬於相同的 POI，代表該網格群組位於同一個景點中。網格分區與 POI 之對應關係範例如表 2 所示，每個觀光景點均視為一個 POI，並分別以 POI ID 表示之。

表 1 信令資料格式範例

| 去個資使用者 ID  | 日期       | 時間           | 網格 ID    |
|--|----------|--------------|----------|
| 4d99d1ebfb91a05575baf2ec3873bedda08<br>bd32e435ca42d98ec294d998abad7 | 2018/5/7 | 18 : 50 : 00 | 69325100 |
| 4d99d1ebfb91a05575baf2ec3873bedda08<br>bd32e435ca42d98ec294d998abad7 | 2018/5/7 | 18 : 51 : 00 | 69325100 |
| 4d99d1ebfb91a05575baf2ec3873bedda08<br>bd32e435ca42d98ec294d998abad7 | 2018/5/7 | 18 : 52 : 00 | 69325100 |
| 7e05c5eade0783464ccee3da64301c68f9e<br>f232ac359da405ac299016de8a0ef | 2018/5/7 | 05 : 49 : 00 | 69328010 |
| 761d8fd1f7643d17ae5106aa2df21255105<br>b75e8b77ab80753d4e3c4be10e463 | 2018/5/7 | 05 : 50 : 00 | 69328045 |

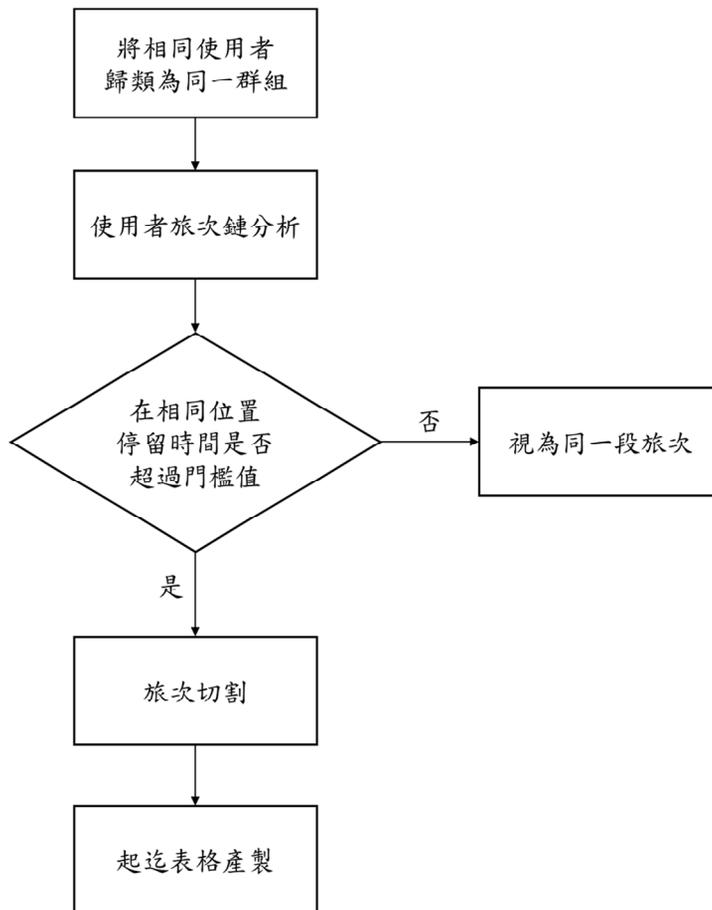


圖 4 旅次起迄矩陣產製流程

表 2 網格 ID 與景點 POI ID 對照表範例

| POI 名稱       | POI ID    | 網格 ID     | POI 名稱         | POI ID    | 網格 ID     |
|--------------|-----------|-----------|----------------|-----------|-----------|
| 秀姑巒溪<br>泛舟中心 | 310001711 | 195241044 | 花蓮文化創<br>意產業園區 | 310001275 | 254233058 |
|              |           | 195241045 |                |           | 254233059 |
|              |           | 195241046 |                |           | 254233060 |
|              |           | 195241054 |                |           | 254233068 |
|              |           | 195241055 |                |           | 254233069 |
|              |           | 195241056 |                |           | 254233070 |
|              |           | 195241064 |                |           | 254233078 |
|              |           | 195241065 |                |           | 254233079 |
|              |           | 195241066 |                |           | 254233080 |

1. 將相同使用者歸類為同一群組：依據電信公司提供的信令資料，將每位使用者的時空軌跡資料，以使用者 ID 進行分群，將相同使用者的資料歸類為同一群組，以利進行後續分析。其範例如表 3 所示，表中為使用者 A 一天中停留過的 POI，及停留 POI 時所記錄的時間。

表 3 以相同使用者 ID 分群範例

| 使用者 ID | POI  | 時間      | POI  | 時間      |
|--------|------|---------|------|---------|
| 使用者 A  | 景點 1 | 10 : 30 | 景點 2 | 12 : 04 |
|        | 景點 1 | 10 : 31 | 景點 2 | 12 : 05 |
|        | 景點 1 | 10 : 32 | 景點 2 | 12 : 06 |
|        | 景點 1 | 10 : 33 | 景點 2 | 12 : 07 |
|        | 景點 1 | 10 : 34 | 景點 2 | 12 : 08 |
|        | 景點 1 | 10 : 35 | 景點 2 | 12 : 09 |
|        | 景點 1 | 10 : 36 | 景點 2 | 12 : 10 |
|        | 景點 1 | 10 : 37 | 景點 3 | 13 : 20 |
|        | 景點 1 | 10 : 38 | 景點 3 | 13 : 21 |
|        | 景點 1 | 10 : 39 | 景點 3 | 13 : 22 |
|        | 景點 1 | 10 : 40 | 景點 3 | 13 : 23 |
|        | 景點 1 | 10 : 41 | 景點 3 | 13 : 24 |
|        | 景點 1 | 10 : 42 | 景點 3 | 13 : 25 |

2. 使用者旅次鏈分析：將每一個具有相同使用者 ID 的群組以時間做排序，得到每位使用者一天中之時空軌跡，同時透過「時間」欄位內所記錄在各 POI 的時間，計算使用者在該 POI 的「停留時間」。使用者一天中出現在各 POI 的第一筆資料視為開始時間，最後一筆資料視為結束時間，兩者時間經相減後，得到使用者在各 POI 的「停留時間」。整理後可取得使用者一天之旅次鏈，以及在各 POI 的停留情況。範例如表 4 所示，承接表 3 之使用者 A 資料，該名使用者於 10：30 至 10：42 停留於景點 1，停留時間為 12 分鐘，於 12：04 至 12：10 停留於景點 2，停留時間為 6 分鐘，於 13：20 至 13：25 停留於景點 3，停留時間為 5 分鐘，可得該名使用者一天之旅次鏈。

表 4 使用者旅次鏈範例

| 使用者 ID | POI  | 開始時間  | 結束時間  | 停留時間  |
|--------|------|-------|-------|-------|
| 使用者 A  | 景點 1 | 10：30 | 10：42 | 12 分鐘 |
| 使用者 A  | 景點 2 | 12：04 | 12：10 | 6 分鐘  |
| 使用者 A  | 景點 3 | 13：20 | 13：25 | 5 分鐘  |

3. 判斷在相同位置的停留時間是否超過門檻值：由於信令資料特性，發生信號飄移的地點通常停留時間極短，透過刪去低於門檻值的地點，較能判斷使用者是否確實有在當地觀光。停留時間之門檻值可根據經驗法則及敏感度分析，加上考慮遊客在景點旅遊時，使用手機上網的時間少，回傳的資料筆數不多，因此門檻值不會設定太高。

如表 5 所示，表中為使用者 B 之旅次鏈及停留時間範例，假設門檻值設定為「五分鐘」，依照其旅次鏈中在各 POI 的停留時間，可看出該使用者於景點 3、景點 5 之停留時間未達門檻值所設定的五分鐘，因此將會被判定為未在該 POI 停留，經篩選後，使用者 B 一天中的旅次鏈所停留的地點分別為景點 1、景點 2、景點 4。根據電信公司的經驗，本研究將門檻值設定為「1 分鐘」，以減少忽略使用者因回傳的資料不足而無法被判斷為停留的情況，導致遺失某些使用者的移動資訊。

表 5 使用者旅次鏈及停留時間範例

| 使用者 ID | POI  | 開始時間  | 結束時間  | 停留時間  |
|--------|------|-------|-------|-------|
| 使用者 B  | 景點 1 | 9：10  | 9：30  | 20 分鐘 |
| 使用者 B  | 景點 2 | 10：30 | 10：40 | 10 分鐘 |
| 使用者 B  | 景點 3 | 13：20 | 13：23 | 3 分鐘  |
| 使用者 B  | 景點 4 | 15：00 | 15：05 | 5 分鐘  |
| 使用者 B  | 景點 5 | 17：10 | 17：12 | 2 分鐘  |

4. 進行旅次切割：當在 POI 的停留時間大於所設定的門檻值，則視為一個停留區間，因此一個停留區間共可切割出前後兩段旅次。以表 5 使用者 B 所停留之前兩個 POI 為例，9：10 至 9：30 為一個停留區間，10：30 至 10：40 為下一個停留區間，因此兩個停留區間之間可切割出一個旅次，如圖 5 所示。每個停留區間的前後分別銜接不同的旅次，代表開始停留點及最後停留點，分別視為一段旅次的迄點與起點。



圖 5 旅次切割示意圖

圖 5 中每段旅次的「出發時間」以前一段停留區間中，最後一筆資料的時間為準。以表 5 之使用者 B 信令資料為例，從景點 1 出發時間可表示為上午 9 點 30 分，每段旅次的「抵達時間」以下一段停留區間中，第一筆資料的時間為準，因此其抵達景點 2 的時間為上午 10 點 30 分。當使用者在相同位置停留時間超過門檻值，則視為使用者有在該景點停留觀光，而下一個停留點則代表使用者下一個前往拜訪的觀光景點。如表 5 所示，該使用者所停留之 POI 景點 1 可視為一段旅次的起點，而下一個停留點為 POI 景點 2，則視為該段旅次的迄點。由上述邏輯即可將使用者一天中之旅次鏈切割為景點至景點間的起迄對，如表 6 所示。

表 6 使用者旅次切割範例

| 使用者 ID | 出發時間  | 抵達時間  | 起點   | 迄點   |
|--------|-------|-------|------|------|
| 使用者 B  | 9：30  | 10：30 | 景點 1 | 景點 2 |
| 使用者 B  | 10：40 | 15：00 | 景點 2 | 景點 4 |

5. 起迄矩陣產製：為達到即時地預測各景點拜訪人數變化，本研究以每小時為單位分別產製起迄表格，例如：時間欄位中若為上午 10 點 13 分，本研究僅記錄至小時單位，表示為「時段 10」。將前一步驟之每位使用者起迄對，根據出發時間及抵達時間填入對應時段的起迄表格，如表 6 所示，使用者 B 於時段 9 由景點 1 出發，於時段 10 抵達景點 2，此段旅次將填入「時段 9 出發、時段 10 抵達」的起迄表格中，起迄表格範例如表 7 所示。

表 7 起迄表格範例

| 時段 9<br> <br>時段 10 | 使用者 ID | 出發時間 | 抵達時間 | 起點   | 迄點   |
|--------------------|--------|------|------|------|------|
|                    | 使用者 A  | 9    | 10   | 景點 1 | 景點 3 |
|                    | 使用者 B  | 9    | 10   | 景點 1 | 景點 2 |
|                    | 使用者 C  | 9    | 10   | 景點 2 | 景點 3 |

### 3.3 轉移矩陣推估方法

馬可夫鏈模式具有「無記憶性」，在狀態轉移的過程中，下一個狀態只受到當前狀態的影響，與過去狀態彼此獨立，定義如下：

隨機過程  $\{X_t\}_{t \in T}$  為一組隨機變數集合，其中  $T = \{0, 1, 2, \dots\}$  表示該過程之時間序列。狀態空間  $S$  則表示  $X_t$  可產生之所有可能結果。馬可夫鏈模式具有以下性質：

$$P(X_{t+1} = j | X_t = i, X_{t-1} = i_{t-1}, \dots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = j | X_t = i) \quad (1)$$

$$P(X_{t+1} = j | X_t = i) = p_{ij}(t) \quad (2)$$

其中  $i, j \in S$ ， $p_{ij}(t)$  稱為馬可夫鏈一階轉移機率，滿足：

對於所有  $i, j \in S$ ， $p_{ij}(t) \geq 0$

$$\sum_{j \in S} p_{ij}(t) = 1$$

馬可夫鏈一階轉移機率可表示為矩陣型態，稱為一階轉移機率矩陣  $P$ ：

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (3)$$

透過初始狀態與轉移矩陣的運算，即可描述未來一連串事件之狀態轉移現象。本研究利用離散時間馬可夫鏈模式，以每小時為一時間段，透過時空軌跡資料的蒐集，整理出各景點初始人數並推估景點間的轉移矩陣，推測未來各時段遊客的移動狀況，轉移矩陣推估流程如圖 6 所示。

1. 考慮各出發時間與相對應抵達時間之組合：由於使用者有著不同的旅運行為，途中可能前往其他非 POI 的地點，或是使用不同運具前往目的地等因素，使得抵達時間不相同。為了記錄在景點間移動時各種出發時間及抵達時間的組合，以時空網路方式呈現如圖 7。本研究於轉移矩陣增加一個「時間維度」，記錄在景點間所有出發時間及抵達時間的組合，稱之為「多維度轉移矩陣」，以計算各時段移動之使用者人數並得到其在景點間之轉移機率。例如在圖 7 中， $f_{ij}(0,1)$  表示於  $t=0$  由景點  $i$  出發前往景點  $j$ ，於  $t=1$  抵達的使用者人數； $f_{ij}(0,2)$  表示於  $t=0$  由景點  $i$  出發前往景點  $j$ ，於  $t=2$  抵達的使用者人數； $f_{ij}(0,3)$  表示於  $t=0$  由景點  $i$  出發前往景點  $j$ ，於  $t=3$  抵達的使用者人數。

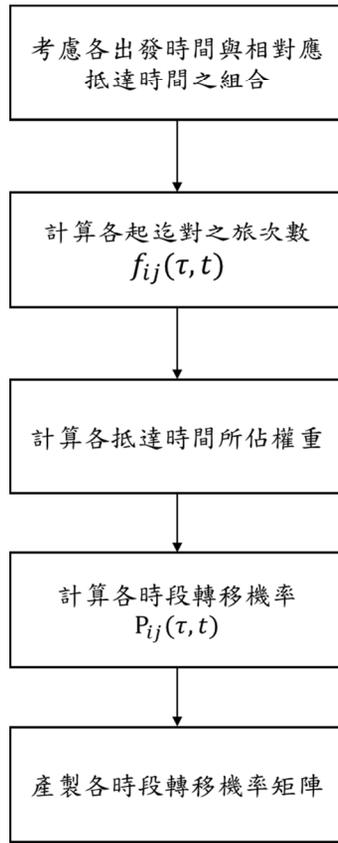


圖 6 多維度轉移矩陣推估流程

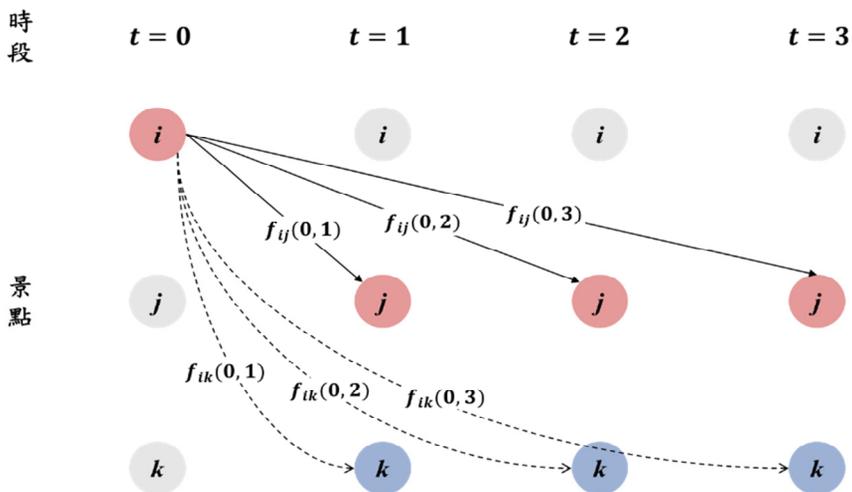


圖 7 以時空網路方式呈現使用者之移動

2. 計算表格內各起迄對之旅次數：針對旅次起迄矩陣結果，進一步計算各時段組合矩陣內各起迄對之旅次數，旅次起迄矩陣之行與列均表示景點個數，每列代表使用者出發景點，每行則代表使用者會抵達的景點，表 8 列出旅次起迄矩陣範例，矩陣內元素用來記錄各景點間移動的旅次數。

表 8 旅次起迄矩陣結果範例

| 景點 | 0 | 1 | 2 | 3 | ... | 57 | 58 |
|----|---|---|---|---|-----|----|----|
| 0  | 0 | 2 | 0 | 0 | ... | 0  | 0  |
| 1  | 0 | 0 | 0 | 0 | ... | 0  | 0  |
| 2  | 0 | 0 | 0 | 0 | ... | 0  | 0  |
| 3  | 0 | 0 | 0 | 0 | ... | 0  | 0  |
| ⋮  | ⋮ | ⋮ | ⋮ | ⋮ | ⋮   | ⋮  | ⋮  |
| 57 | 3 | 0 | 0 | 0 | ... | 0  | 0  |
| 58 | 0 | 0 | 0 | 0 | ... | 0  | 0  |

3. 計算各抵達時間所佔權重：由於有著同一群使用者即使在同一時間離開同一個景點  $i$ ，前往同一個景點  $j$ ，但被觀察到抵達景點  $j$  的時間也不同的情況。Almanaa et al. [19] 將馬可夫鏈模式應用於公共自行車租借站之最佳初始狀態預測的研究中，於目標式內針對轉移矩陣加入每小時分配之權重，以預測每小時的自行車需求量，解決自行車共享系統的供需失衡問題。參考上述研究，本研究考慮使用者各種可能的旅行時間，分配各時段所佔權重，計算在同一時間離開同一個景點  $i$  會在後續各時段抵達景點  $j$  的比例，以降低整體誤差。由  $\tau$  時間出發、 $t$  時間抵達的權重  $W(\tau, t)$  計算方法為：

$$W(\tau, t) = \frac{f(\tau, t)}{\sum_{t=\tau}^T f(\tau, t)} \quad (4)$$

上式表示出發時間  $\tau$ 、抵達時間  $t$  的權重為：由  $\tau$  時間出發、 $t$  時間抵達的人數  $f(\tau, t)$  與相同出發時間  $\tau$ 、所有可能時間抵達的人數總和  $\sum_{t=\tau}^T f(\tau, t)$  之比值。

4. 計算各時段轉移機率：在各景點間的轉移機率  $P_{ij}(\tau, t)$ ，表示使用者由景點  $i$  轉移至景點  $j$ ，並於  $\tau$  時間出發、 $t$  時間抵達的機率。其中  $f_{ij}(\tau, t)$  表示  $\tau$  時段出發、 $t$  時間抵達之起迄對  $i$ 、 $j$  之旅次數。轉移機率  $P_{ij}(\tau, t)$  的計算方法為：

$$P_{ij}(\tau, t) = \frac{f_{ij}(\tau, t)}{\sum_{j \in S} f_{ij}(\tau, t)} \quad (5)$$

上式表示在相同的出發、抵達時間組合中，分子為由  $i$  景點出發、抵達  $j$  景點的人數，分母為由  $i$  景點出發、抵達其他各景點的人數總和。因此對於相同出發時間  $\tau$ 、抵達時間  $t$  而言， $\sum_{j \in S} P_{ij}(\tau, t) = 1$ 。

5. 產製各時段轉移機率矩陣：將各景點轉移機率計算結果填入轉移矩陣對應的位置中，多維度轉移矩陣示意圖如圖 8 所示，其中第一個矩陣內之轉移機率  $P_{ij}(\tau, \tau + 1)$  表示於  $\tau$  時間時，由景點  $i$  出發前往景點  $j$ ，並於  $\tau + 1$  時間抵達的機率，轉移機率  $P_{ij}(\tau, \tau + 2)$  則表示於  $\tau$  時間時，由景點  $i$  出發前往景點  $j$ ，並於  $\tau + 2$  時間抵達的機率，依此類推，考慮所有可能的抵達時間。

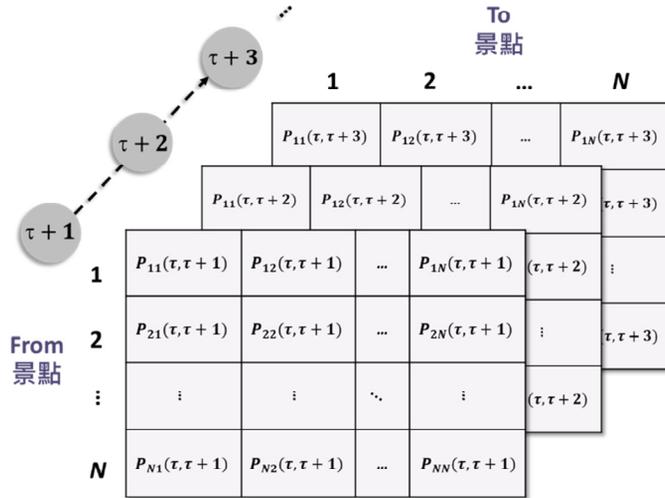


圖 8 多維度轉移矩陣示意圖

### 3.4 各時段景點人數預測

#### 1. 馬可夫鏈模式

本研究之目的為預測各景點在各時段之拜訪人數，因此將以使用者之「抵達時間」為基礎，向前推算使用者之出發時間。景點  $j$  在  $t$  時段時的拜訪人數為：

$$\begin{cases} C_j(t) = \sum_{\tau=0}^{t-1} \sum_{i=1}^N C_i(\tau) \times P_{ij}(\tau, t) \times W(\tau, t), t \geq 1 \\ C_j(t) = \overline{C_j}(t), t = 0 \end{cases} \quad (6)$$

其中，

$\tau$  為出發時段，

$t$  為抵達時段，

$i, j$  為景點， $i, j \in S$

$N$  為景點總數，

$C_j(t)$  為景點  $j$  在  $t$  時段時的拜訪人數，

$C_i(\tau)$  為景點  $i$  在  $\tau$  時段時的拜訪人數，

$\overline{C_j(t)}$ ,  $j = 1, 2, \dots, N$  為在  $t = 0$  時之各景點起始人數分布，本研究取各景點歷史資料人數之平均作為該值，

$P_{ij}(\tau, t)$  為在  $\tau$  到  $t$  時段內，分別由  $i$  到  $j$  之轉移機率，

$W(\tau, t)$  為  $\tau$  到  $t$  時段所分配之權重。

首先以歷史資料各景點人數平均作為初始狀態，接者考慮當前時段以前所有可能的時段，分別加入對應的轉移矩陣與權重進行，將各時段結果加總，得到當前時段之各景點拜訪人數預測結果，以上方法可呈現如表 9 所示。

表 9 各時段景點拜訪人數計算方法

| 時段      | 出發/抵達                 | 計算過程   | 加總結果       |
|---------|-----------------------|--|------------|
| $t = 0$ | 由歷史資料取得： $t = 0$ 起始人數 |  |            |
| $t = 1$ | 0~1                   | $t = 0$ 起始人數 $\times$ 轉移矩陣 (0,1) $\times$ 權重 (0,1) | $t = 1$ 人數 |
| $t = 2$ | 0~2                   | $t = 0$ 起始人數 $\times$ 轉移矩陣 (0,2) $\times$ 權重 (0,2) | $t = 2$ 人數 |
|         | 1~2                   | $t = 1$ 人數 $\times$ 轉移矩陣 (1,2) $\times$ 權重 (1,2)   |            |
| $t = 3$ | 0~3                   | $t = 0$ 起始人數 $\times$ 轉移矩陣 (0,3) $\times$ 權重 (0,3) | $t = 3$ 人數 |
|         | 1~3                   | $t = 1$ 人數 $\times$ 轉移矩陣 (1,3) $\times$ 權重 (1,3)   |            |
|         | 2~3                   | $t = 2$ 人數 $\times$ 轉移矩陣 (2,3) $\times$ 權重 (2,3)   |            |
| ...     | ...                   | ...  | ...        |

## 2. 誤差項估計

由前步驟介紹之透過馬可夫鏈模式預測過程並未於封閉系統內進行，使得各時段系統總人數的變化將會對預測結果造成一定程度影響。因此本研究提出結合馬可夫鏈模式預測結果與「誤差項」的方式，對原預測結果給予補償。以向量表示誤差項方法如 (7) 式所示：

$$\begin{aligned}
 & [C_1(t) \quad C_2(t) \quad \dots \quad C_N(t)] \\
 & = \left( [C_1(\tau) \quad C_2(\tau) \quad \dots \quad C_N(\tau)] \begin{bmatrix} P_{11}(\tau, t) & \dots & P_{1N}(\tau, t) \\ \vdots & \ddots & \vdots \\ P_{N1}(\tau, t) & \dots & P_{NN}(\tau, t) \end{bmatrix} \times W(\tau, t) \right) \\
 & \quad + [y_1(t) \quad y_2(t) \quad \dots \quad y_N(t)]
 \end{aligned} \tag{7}$$

其中，

$\tau$  為出發時段，

$t$  為抵達時段，

$i, j$  為景點， $i, j \in S$

$N$  為景點總數，

$C_j(t)$  為景點  $j$  在  $t$  時段時的拜訪人數，

$C_i(\tau)$  為景點  $i$  在  $\tau$  時段時的拜訪人數，

$P_{ij}(\tau, t)$  為在  $\tau$  到  $t$  時段內，分別由  $i$  到  $j$  之轉移機率，

$W(\tau, t)$  為  $\tau$  到  $t$  時段所分配之權重，

$y_j(t)$  為景點  $j$  在  $t$  時段時的誤差項。

上式小括號內為原馬可夫鏈模式預測方法之結果，於原預測方法後方加入一誤差項  $y_j(t)$ ，得到在  $t$  時段時的景點拜訪人數。式 (7) 經移項後可得：

$$\begin{aligned} & [y_1(t) \quad y_2(t) \quad \cdots \quad y_N(t)] \\ & = [C_1(t) \quad C_2(t) \quad \cdots \quad C_N(t)] \\ & - \left( [C_1(\tau) \quad C_2(\tau) \quad \cdots \quad C_N(\tau)] \begin{bmatrix} P_{11}(\tau, t) & \cdots & P_{1N}(\tau, t) \\ \vdots & \ddots & \vdots \\ P_{N1}(\tau, t) & \cdots & P_{NN}(\tau, t) \end{bmatrix} \times W(\tau, t) \right) \end{aligned} \quad (8)$$

即誤差項可由「景點拜訪人數」與原先「馬可夫鏈模式預測值」相減取得。在誤差項的估測上，認為一天中各時段總人數變化與產生之誤差應有規律性，所以本研究透過簡單線性迴歸模型 (Simple Linear Regression) 完成，目的為找出應變數  $y$  與自變數  $x$  間的關係式，以推論與預測感興趣的應變數  $y$ ，其中應變數  $y$  為時段  $t$  實際人數和 (馬可夫鏈模式) 預測人數差距，自變數  $x$  為時段  $t$  各景點人數變化。藉由自變數  $x$  考慮每個時段各景點的人數變化，以描述系統總人數的增減所帶來的誤差，以補償馬可夫鏈模式的不足。由於馬可夫鏈模式具有「無記憶性」，當前狀態僅與前一狀態有關，因此本研究針對當前時段與前一時段各景點人數的差異作為自變數  $x$ ，以描述其與應變數  $y$  所代表的誤差項兩者之間的關聯。誤差項之迴歸方程式可表示為：

$$y_d(t, i) = \alpha(t, i)x_d(t, i) + \beta(t, i) \quad (9)$$

其中，

$t$  為時段，

$i$  為景點，

$d$  為樣本觀察值，

$y_d(t, i)$  為第  $d$  天觀察值中， $t$  時段中景點  $i$  預測人數與實際人數之差距，

$x_d(t, i)$  為第  $d$  天觀察值中， $t$  時段中景點  $i$  的人數變化，

$\alpha(t, i)$  為  $t$  時段中景點  $i$  之迴歸方程式斜率，

$\beta(t, i)$  為  $t$  時段中景點  $i$  之迴歸方程式截距。

式 (9) 中，針對各時段各景點均會求出一條迴歸方程式，以推估各個景點人數隨時間變化與系統總人數的增減所產生的誤差，舉例來說，若取  $T$  個時段、 $N$  個景點為預測對象，則共有  $T \times N$  條迴歸方程式。在迴歸方程式參數推估上，將歷史資料擁有的天數中，每天所對應的時段  $t$  與景點  $i$  作為一個觀察值  $d$  代入，以推估迴歸方程式的斜率  $\alpha(t, i)$  與截距  $\beta(t, i)$ 。並藉由計算當前時段與前一時段各個景點人數的變化代入自變數  $x$  中，以推估應變數  $y$  所代表的誤差項。最後將估計得之誤差項與原預測人數相加，對於系統總人數增減造成之誤差，進行預測結果上的調整，得到最終預測結果。

綜觀來說，影響總人數變化的因素可能有天氣狀況、假別、特殊活動及政府觀光政策等等，都會直接影響各景點的拜訪人數，而本研究的誤差項受限於資料數量僅透過簡單線性迴歸模型估計，若能將以上相關因素納入模型，進而以多元迴歸分析 (Multiple Regression Analysis) 描述誤差項，相信可更準確的評估各時段且不同景點預測人數。

## 四、研究成果與應用

本研究透過蒐集拜訪花蓮縣之使用者信令資料，分析使用者在觀光景點間的移動型態。在交通部運輸研究所與電信公司協助下，共取得 6 天、三個週末之信令資料，分別為 2018 年 5 月 19 日星期六、2018 年 5 月 20 日星期日、2018 年 5 月 26 日星期六、2018 年 5 月 27 日星期日、2018 年 6 月 2 日星期六、2018 年 6 月 3 日星期日，各天使用者人數統計如表 10 所示。

表 10 信令資料各天使用者人數統計

| 日期       | 使用者人數  | 經過 POI 使用者人數 | 停留景點使用者人數 |
|----------|--------|--------------|-----------|
| 5/19 (六) | 18,607 | 10,738       | 3,520     |
| 5/20 (日) | 16,911 | 7,281        | 2,057     |
| 5/26 (六) | 20,005 | 11,648       | 4,225     |
| 5/27 (日) | 19,944 | 11,239       | 4,274     |
| 6/2 (六)  | 18,136 | 9,789        | 3,385     |
| 6/3 (日)  | 18,104 | 9,731        | 3,480     |

### 4.1 模式驗證

本研究將多天資料加以彙整，以獲得較適用於各天的轉移矩陣穩定狀態。在驗證流程上，使用交叉驗證 (cross validation) 方法，將 6 天資料，每次取 5 天不同組合之訓練資料及 1 天之測試資料，共重複上述步驟，如圖 9 所示，並分別計算 6 組訓練資料預測之準確率。

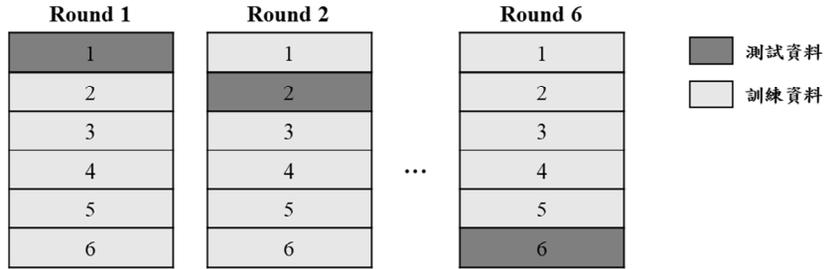


圖 9 交叉驗證示意圖

本研究之轉移矩陣驗證流程如圖 10 所示：

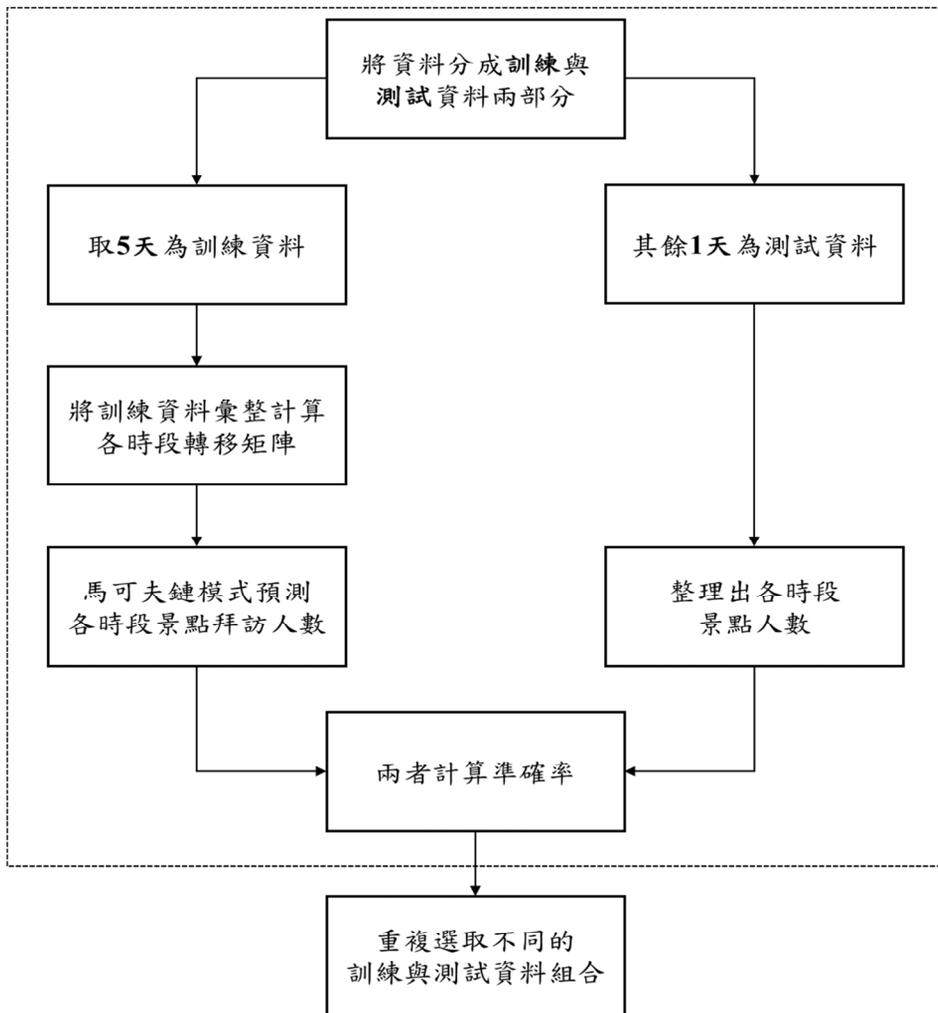


圖 10 多維度轉移矩陣訓練流程

1. 將資料分成訓練資料與測試資料兩部分：本研究共取得花蓮縣使用者 6 天之信令資料，將其中 5 天作為訓練資料，並以剩餘的 1 天作為測試資料。
2. 將資料彙整計算各時段轉移矩陣：將 5 天訓練資料進行彙整，以達到較趨近於穩定狀態之轉移矩陣，以轉移矩陣推估邏輯產製各時段各景點間的多維度轉移矩陣。
3. 整理出各時段景點人數：將剩餘 1 天的測試資料，根據信令資料紀錄之時間欄位與使用者所在的景點，整理出各時段各景點內的人數實際值，以作為後續驗證訓練結果之標準答案。
4. 馬可夫鏈模式與誤差項預測各時段景點拜訪人數：經由 6 天信令資料作為歷史資料，計算各景點平均人數，作為各景點的起始人數，透過由訓練資料推估之多維度轉移矩陣與權重，得到各時段景點人數之馬可夫鏈預測結果。再透過誤差項方法，分別以 6 天歷史資料作為樣本觀察值，得到各時段景點人數預測結果。
5. 計算準確率：由前一步驟之馬可夫鏈模式與誤差項景點拜訪人數預測結果與由測試資料整理之各時段景點人數標準答案，計算準確率。本研究以「平均絕對百分比誤差」(Mean Absolute Percentage Error, MAPE) 與「平均絕對誤差」(Mean Absolute Error, MAE) 作為預測模式準確率之衡量指標。MAPE 值為相對數值，能較客觀獲得預測值與實際值間之差異程度，計算方法為：

$$MAPE = \frac{\sum \frac{|\text{實際值}_i - \text{預測值}_i|}{\text{實際值}_i}}{n} \times 100\%$$

一般而言，MAPE 值若愈接近 0 表示估計效果愈佳。Lewis (1982) 曾為 MAPE 值訂定相關評估標準，若 MAPE 小於 10%，為高準確預測；10%~20%，為優良的預測；20%~50%，為合理的預測；大於 50%，為不準確的預測。MAE 值則將各次測量的誤差取絕對值後再求平均值，能反映預測誤差之實際情況，MAE 值之計算方法為：

$$MAE = \frac{\sum |\text{實際值}_i - \text{預測值}_i|}{n}$$

6. 選取不同訓練資料組合：由交叉驗證方法，每次選取不同之 5 天訓練資料組合，與剩餘的 1 天測試資料驗證模式，計算準確率。

## 4.2 預測結果

考慮大部分觀光景點營業時間為白天，本研究選取每日的「上午 10 點」至「下午 5 點」共 8 個時段作為預測時段，透過馬可夫鏈模式與誤差項方法進行預測。如表 11 所示，表中列出時段 14 之景點人數預測過程，並代入實際數值作為範例。 $t = 0$  起始人數矩陣由歷史資料平均人數取得，其餘  $t = 1 \sim t = 13$  各時段人數則透過測試資料歸納出各景點人

數，乘以所對應時段之轉移矩陣與權重，由欲預測之時段 14 向前推算，計算出由各時段出發、於時段 14 抵達各景點的人數，將上述各時段結果加總即可得到時段 14 各景點人數預測結果，而各景點人數預測結果如表 12 所示。

表 11 時段 14 景點人數預測範例

| 時段       | 出發/抵達 | 計算過程   | 加總             |
|----------|-------|--|----------------|
| $t = 14$ | 0~14  | $t = 0$ 起始人數 $\times$ 轉移矩陣 $(0, 14) \times$ 權重 $(0, 14)$<br>$[89.34 \quad 53.61 \quad \dots \quad 3.61] \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \times 0.08$<br>$= [7.15 \quad 1.36 \quad \dots \quad 0]$     | $t = 14$<br>人數 |
|          | 1~14  | $t = 1$ 人數 $\times$ 轉移矩陣 $(1, 14) \times$ 權重 $(1, 14)$<br>$[14 \quad 1 \quad \dots \quad 0] \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \times 0.03 = [0.42 \quad 0 \quad \dots \quad 0]$                         |                |
|          | ⋮     | ⋮  |                |
|          | 13~14 | $t = 13$ 人數 $\times$ 轉移矩陣 $(13, 14) \times$ 權重 $(13, 14)$<br>$[208 \quad 131 \quad \dots \quad 9] \begin{bmatrix} 0.02 & 0.2 & \dots & 0 \\ 0.02 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \times 0.31$<br>$= [25.88 \quad 37.88 \quad \dots \quad 0]$ |                |

表 12 各景點人數預測結果 (5 月 19 日時段 14)

| 景點 | 預測人數   | 景點 | 預測人數  | ... | 景點 | 預測人數  |
|----|--------|----|-------|-----|----|-------|
| 0  | 60.09  | 12 | 1.93  | ... | 48 | 36.29 |
| 1  | 143.57 | 13 | 0.00  |     | 49 | 10.37 |
| 2  | 4.79   | 14 | 4.29  |     | 50 | 43.82 |
| 3  | 0.00   | 15 | 21.90 |     | 51 | 0.40  |
| 4  | 11.05  | 16 | 49.63 |     | 52 | 22.84 |
| 5  | 3.57   | 17 | 11.23 |     | 53 | 18.27 |
| 6  | 159.08 | 18 | 3.31  |     | 54 | 20.08 |
| 7  | 0.00   | 19 | 9.58  |     | 55 | 23.85 |
| 8  | 1.93   | 20 | 20.36 |     | 56 | 3.82  |
| 9  | 134.65 | 21 | 0.00  |     | 57 | 41.81 |
| 10 | 22.66  | 22 | 95.67 |     | 58 | 6.13  |
| 11 | 0.00   | 23 | 17.60 |     |    |       |

而由馬可夫鏈模式之轉移機率矩陣推估結果中，可以觀察到民眾旅運行為的型態。以 5 月 19 日為測試資料的結果為例子說明（訓練資料為 5 月 20 日、5 月 26 日、5 月 27 日、6 月 2 日、6 月 3 日），取出時段 14-15 以視覺化方式呈現各景點間的轉移機率，如圖 11 所示。圖的最右邊表示轉移機率，從機率低 (0.0，深藍色) ~ 機率高 (1.0，淺黃色)，圖中表示於該時段由景點  $i$  (縱軸，0~58) 轉移至景點  $j$  (橫軸，0~58) 的轉移機率大小，顏色愈淺的區塊表示由景點  $i$  轉移至景點  $j$  的機率愈高。

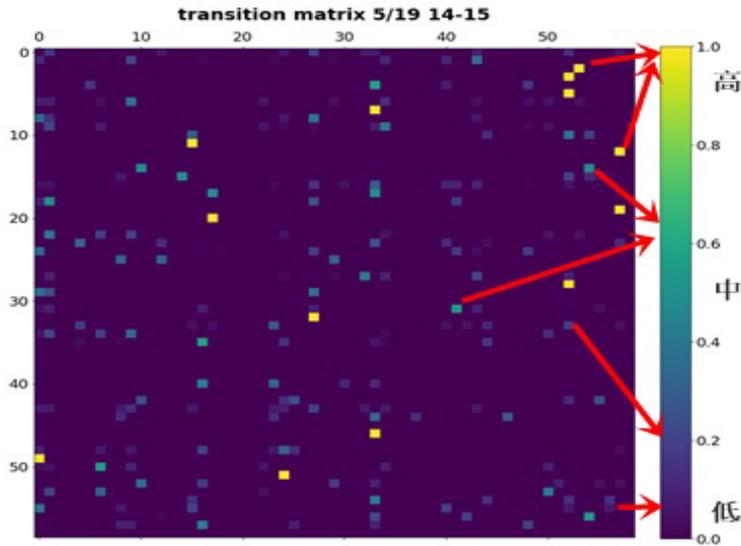


圖 11 時段 14-15 各景點間轉移機率視覺化結果

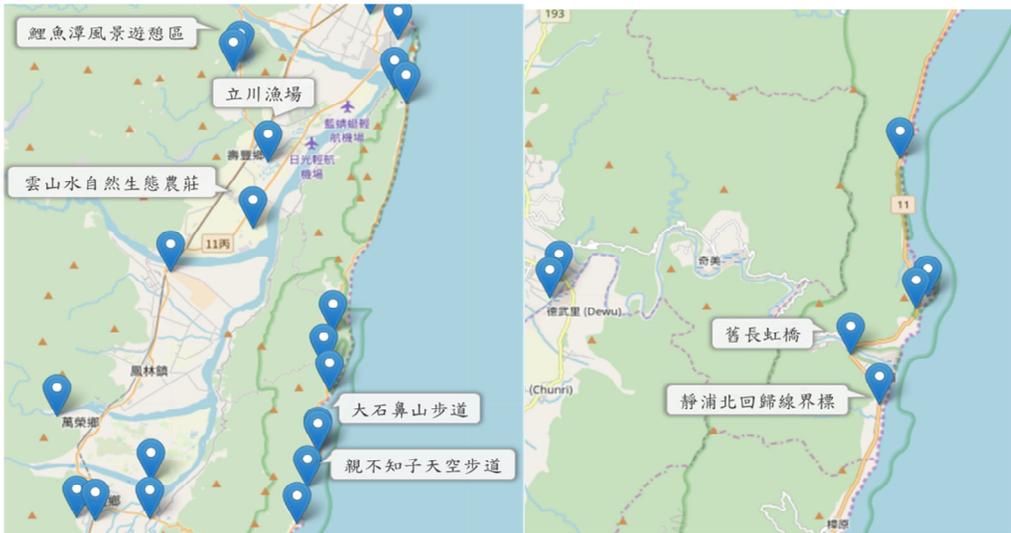
經交叉驗證後，景點間轉移機率較高的起迄組合有：「雲山水自然生態農莊 (景點 40) - 立川漁場 (景點 16)」、「鯉魚潭風景遊憩區 (景點 57) - 立川漁場 (景點 16)」、「舊長虹橋 (景點 56) - 靜浦北回歸線界標 (景點 54)」、「大石鼻山步道 (景點 3) - 親不知子天空步道 (景點 52)」，以及太魯閣國家公園沿線景點間，如「天祥 (景點 6)、岳王亭 (景點 22)、九曲洞隧道 (景點 2)、錐麓古道 (景點 53)、燕子口步道 (景點 50)、砂卡礑步道 (景點 34)、太魯閣國家公園遊客中心 (景點 9)」。

將上述景點於地圖上標示其相對位置，如圖 12 所示。可以發現轉移機率較高的兩個起迄景點通常位於相鄰區域，例如：圖 12 (a) - 太魯閣國家公園沿線，景點有：天祥 (景點 6)、岳王亭 (景點 22)、九曲洞隧道 (景點 2)、錐麓古道 (景點 53)、燕子口步道 (景點 50)、砂卡礑步道 (景點 34)、太魯閣國家公園遊客中心 (景點 9)。圖 12 (b) - 臺灣好行縱谷花蓮線與海岸公路沿線，景點有雲山水自然生態農莊 (景點 40)、立川漁場 (景點 16)、鯉魚潭風景遊憩區 (景點 57)；大石鼻山步道 (景點 3)、親不知子天空步道 (景點 52)；舊長虹橋 (景點 56)、靜浦北回歸線界標 (景點 54)。表示遊客在花蓮旅遊時選擇景點具有明顯的區域之分，推測可能為區域間交通便利性低，遊客通常只選擇在同一區域內之景點觀光。

後續將推估得到的誤差項與前步驟馬可夫鏈預測結果加總，計算最終之預測人數，並透過由測試資料整理之標準答案進行準確率計算，表 13 列出以 5 月 26 日作為測試資料之時段 12 部分景點結果。表中「結果」欄位由「原預測值」與「誤差項」兩欄位加總，透過「實際值」計算 MAPE 值；表中「原 MAPE 值」欄位，則僅以原預測值未新增誤差項的情況下，與實際值計算之 MAPE 值。從 MAPE 值可看出，本研究所提出方法的預測結果屬於高準確至合理的預測範圍（小於 10%，為高準確預測；10%~20%，為優良的預測；20%~50%，為合理的預測），經由新增誤差項對原預測結果進行補償後，MAPE 值均有明顯降低的趨勢，表示對於提升預測準確率具相當成效。



(a)太魯閣國家公園



(b)臺灣好行縱谷花蓮線與海岸公路沿線

圖 12 轉移機率較高之景點位置分布

表 13 部分景點準確率計算結果 (5 月 26 日時段 12)

| 景點      | 原預測值   | 誤差項    | 結果     | 實際值 | MAPE 值 | 原 MAPE 值 |
|---------|--------|--------|--------|-----|--------|----------|
| 花蓮文創園區  | 75.93  | 113.13 | 189.06 | 225 | 15.97  | 66.25    |
| 七星潭風景區  | 26.52  | 80.59  | 107.11 | 146 | 26.64  | 81.83    |
| 天祥風景區   | 185.44 | 108.40 | 293.85 | 269 | 9.24   | 31.06    |
| 太平洋公園   | 20.61  | 44.36  | 64.98  | 80  | 18.78  | 74.23    |
| 太魯閣國家公園 | 172.26 | 54.07  | 226.34 | 293 | 22.75  | 41.21    |
| 石梯坪     | 22.83  | 30.21  | 53.04  | 70  | 24.23  | 67.38    |
| 立川漁場    | 92.46  | 13.09  | 105.54 | 154 | 31.47  | 39.96    |

### 4.3 績效評估

本研究以 6 天資料進行交叉驗證，分別計算各天測試資料之景點內預測人數與實際人數之 MAPE 值與 MAE 值。由於取得之信令資料同時包含週六及週日，而遊客前的旅遊型態也會因是週六或週日而有不同，相較於週日，週六之行程通常較多且較緊湊，因此相對來說，在週六的使用者數量較多、能篩選出的 POI 資料較多，計算之轉移矩陣也會較為穩定。因此本研究選取三個週六分別作為測試資料，表 14 列出三天各時段之 MAPE 值與 MAE 值，並針對各天誤差取平均值。

表 14 三天測試資料之 MAPE 值與 MAE 值

| 測試日期 | 5 月 19 日 |         | 5 月 26 日 |         | 6 月 2 日  |         |
|------|----------|---------|----------|---------|----------|---------|
|      | MAPE (%) | MAE (人) | MAPE (%) | MAE (人) | MAPE (%) | MAE (人) |
| 10   | 29.18    | 7.79    | 28.20    | 7.95    | 27.92    | 8.79    |
| 11   | 22.16    | 23.96   | 28.47    | 23.00   | 32.67    | 21.84   |
| 12   | 23.70    | 12.64   | 32.48    | 15.55   | 35.44    | 16.29   |
| 13   | 28.19    | 25.07   | 27.39    | 28.41   | 32.92    | 21.63   |
| 14   | 25.80    | 16.27   | 19.77    | 19.25   | 43.09    | 16.74   |
| 15   | 24.71    | 21.77   | 24.34    | 30.34   | 31.20    | 24.66   |
| 16   | 33.70    | 17.98   | 24.60    | 22.80   | 32.10    | 16.47   |
| 17   | 46.25    | 23.26   | 29.55    | 30.26   | 37.77    | 20.50   |
| 平均   | 29.21    | 18.59   | 26.85    | 22.20   | 34.14    | 18.37   |

由表 14 可知，三天之平均 MAPE 值分別為 29.21%、26.85%、34.14%，屬於「合理的預測範圍」。而透過 MAE 值觀察實際情況誤差，顯示這三天中各景點之預測人數與實際人數的差異平均約為 18.59 人、20.20 人、18.37 人。由各時段來看，MAPE 值最佳的時段大約落在時段 14 到時段 15 的範圍，推測原因為下午時段是大多數景點內拜訪人次最多的時候，能蒐集到之使用者移動資料較其他時段充足，接近早晨或是傍晚則拜訪人次較少。將 5 月 19 日時段 15 之實際值與預測值以視覺化方式呈現，如圖 13 所示，各景點人數之實際值與預測值大致呈現相同趨勢；並以 5 月 19 日景點人數預測結果為例，各時段各景點之 MAPE 值等級如表 15 所示。

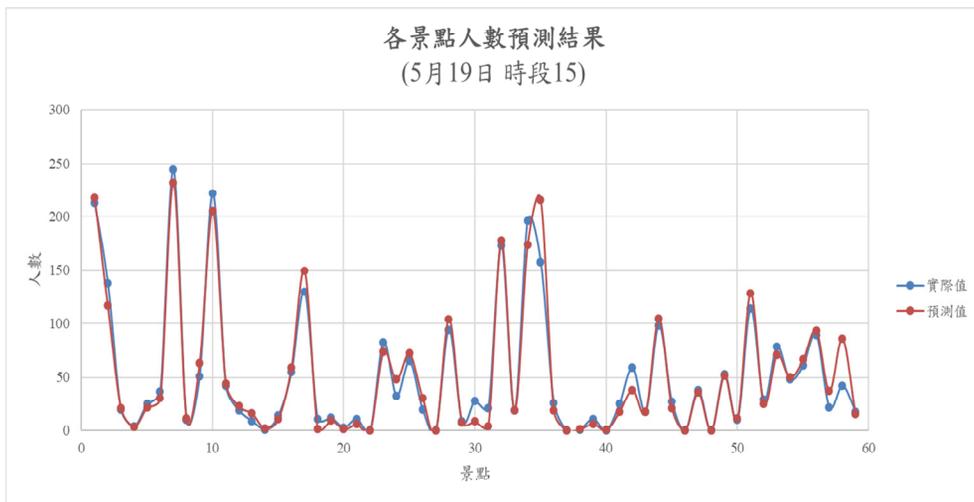


圖 13 5 月 19 日時段 15 實際值與預測值視覺化結果

表 15 5 月 19 日各景點各時段 MAPE 等級

| 編號 | 景點         | 時段 |    |    |    |    |    |    |    |  |
|----|------------|----|----|----|----|----|----|----|----|--|
|    |            | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |  |
| 0  | 花蓮文化創意產業園區 | 1  | 2  | 1  | 1  | 1  | 2  | 3  | 3  |  |
| 1  | 七星潭風景區     | 1  | 2  | 2  | 3  | 2  | 3  | 3  | 1  |  |
| 2  | 九曲洞隧道      | 3  | 3  | 3  | 3  | 1  | 1  | 1  | 2  |  |
| 3  | 大石鼻山步道     | 2  | 1  | 1  | 2  | 1  | 1  | 1  | 1  |  |
| 4  | 大農大富平地森林園區 | 3  | 2  | 2  | 3  | 2  | 1  | 3  | 3  |  |
| 5  | 六十石山       | 2  | 2  | 1  | 2  | 2  | 3  | 1  | 1  |  |
| 6  | 天祥風景區      | 1  | 2  | 1  | 1  | 1  | 3  | 3  | 1  |  |
| 7  | 太巴壠文化園區    | 1  | 3  | 2  | 3  | 2  | 1  | 2  | 1  |  |
| 8  | 太平洋公園      | 1  | 1  | 1  | 1  | 3  | 3  | 3  | 1  |  |

表 15 5 月 19 日各景點各時段 MAPE 等級 (續 1)

| 編號 | 景點          | 時段 |    |    |    |    |    |    |    |
|----|-------------|----|----|----|----|----|----|----|----|
|    |             | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 9  | 太魯閣國家公園     | 2  | 3  | 2  | 1  | 1  | 2  | 1  | 3  |
| 10 | 月洞遊憩區       | 2  | 2  | 2  | 1  | 1  | 3  | 2  | 1  |
| 11 | 牛山風景區       | 3  | 3  | 3  | 3  | 3  | 3  | 3  | 3  |
| 12 | 台開心農場       | 1  | 3  | 3  | 3  | 3  | 3  | 1  | 2  |
| 13 | 玉里神社遺址      | 3  | 3  | 3  | 3  | 3  | 1  | 3  | 3  |
| 14 | 石門遊憩區       | 2  | 3  | 3  | 2  | 3  | 1  | 3  | 3  |
| 15 | 石梯坪遊客中心     | 1  | 1  | 2  | 3  | 1  | 3  | 2  | 2  |
| 16 | 立川漁場        | 2  | 2  | 1  | 1  | 2  | 1  | 3  | 2  |
| 17 | 吉蒸牧場        | 3  | 2  | 3  | 1  | 3  | 1  | 1  | 3  |
| 18 | 向日廣場        | 3  | 2  | 1  | 3  | 2  | 1  | 3  | 3  |
| 19 | 池南國家森林遊樂區   | 1  | 3  | 3  | 3  | 3  | 1  | 3  | 1  |
| 20 | 秀姑巒溪泛舟中心    | 3  | 3  | 3  | 2  | 3  | 1  | 1  | 3  |
| 21 | 赤科山         | 1  | 1  | 2  | 3  | 1  | 1  | 1  | 3  |
| 22 | 岳王亭         | 3  | 3  | 1  | 2  | 2  | 1  | 1  | 2  |
| 23 | 林田山林業文化園區   | 3  | 3  | 3  | 3  | 3  | 2  | 3  | 3  |
| 24 | 松園別館        | 2  | 1  | 1  | 1  | 2  | 3  | 1  | 2  |
| 25 | 知卡宣綠森林親水公園  | 3  | 3  | 3  | 3  | 3  | 3  | 1  | 3  |
| 26 | 芭崎遊憩區瞭望台    | 3  | 1  | 1  | 3  | 1  | 1  | 3  | 3  |
| 27 | 花蓮文化中心石雕公園  | 3  | 2  | 2  | 2  | 2  | 3  | 2  | 1  |
| 28 | 花蓮東大門國際觀光夜市 | 3  | 3  | 1  | 3  | 2  | 2  | 2  | 3  |
| 29 | 花蓮美侖大飯店     | 3  | 3  | 2  | 1  | 3  | 2  | 3  | 3  |
| 30 | 花蓮遠雄悅來大飯店   | 3  | 3  | 3  | 3  | 3  | 1  | 3  | 2  |
| 31 | 花蓮遠雄海洋公園    | 3  | 3  | 1  | 1  | 1  | 3  | 1  | 3  |
| 32 | 花蓮翰品酒店      | 3  | 1  | 3  | 3  | 1  | 2  | 1  | 3  |
| 33 | 花蓮觀光糖廠      | 1  | 1  | 1  | 1  | 2  | 3  | 3  | 3  |
| 34 | 砂卡礑步道       | 2  | 2  | 3  | 1  | 3  | 3  | 1  | 2  |
| 35 | 馬太鞍溼地生態館    | 3  | 3  | 3  | 2  | 3  | 3  | 1  | 3  |
| 36 | 曼波海灘        | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  |
| 37 | 崇德礫灘        | 3  | 2  | 3  | 1  | 3  | 1  | 2  | 1  |
| 38 | 掃叭石柱 (舞鶴石柱) | 2  | 3  | 3  | 3  | 3  | 1  | 3  | 1  |

表 15 5 月 19 日各景點各時段 MAPE 等級 (續 2)

| 編號 | 景點         | 時段 |    |    |    |    |    |    |    |
|----|------------|----|----|----|----|----|----|----|----|
|    |            | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 39 | 富源國家森林遊樂區  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 3  |
| 40 | 雲山水自然生態農莊  | 2  | 1  | 3  | 2  | 3  | 1  | 3  | 3  |
| 41 | 新光兆豐休閒農場   | 3  | 2  | 3  | 2  | 3  | 1  | 3  | 3  |
| 42 | 新社梯田       | 1  | 2  | 3  | 2  | 1  | 1  | 3  | 2  |
| 43 | 煙波大飯店      | 2  | 1  | 1  | 1  | 1  | 2  | 3  | 3  |
| 44 | 瑞穗牧場       | 3  | 1  | 2  | 3  | 3  | 3  | 3  | 3  |
| 45 | 翡翠谷        | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  |
| 46 | 舞鶴觀光茶園     | 3  | 1  | 2  | 3  | 1  | 1  | 3  | 2  |
| 47 | 鳳林環保科技園區   | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  |
| 48 | 慶修院        | 3  | 2  | 1  | 2  | 1  | 2  | 3  | 3  |
| 49 | 慕谷慕魚遊客中心   | 3  | 2  | 1  | 3  | 2  | 1  | 1  | 3  |
| 50 | 燕子口步道      | 3  | 2  | 2  | 1  | 2  | 3  | 1  | 3  |
| 51 | 蕃薯寮休息區     | 3  | 1  | 1  | 2  | 2  | 3  | 3  | 3  |
| 52 | 親不知子天空步道   | 3  | 3  | 2  | 1  | 1  | 3  | 3  | 3  |
| 53 | 錐麓古道       | 1  | 1  | 2  | 2  | 1  | 3  | 1  | 3  |
| 54 | 靜浦北回歸線界標   | 1  | 2  | 3  | 1  | 1  | 3  | 3  | 2  |
| 55 | 磯崎海水浴場     | 3  | 1  | 3  | 3  | 1  | 3  | 3  | 1  |
| 56 | 舊長虹橋       | 2  | 3  | 3  | 3  | 3  | 2  | 3  | 3  |
| 57 | 鯉魚潭風景遊憩區   | 3  | 3  | 3  | 3  | 3  | 3  | 3  | 3  |
| 58 | 羅山管理站暨遊客中心 | 1  | 3  | 3  | 2  | 2  | 2  | 3  | 2  |

註：1：MAPE<10%。2：10%<MAPE<20%。3：MAPE>20%。

#### 4.4 成果應用探討

本研究所提出的觀光景點人數預測方法與成果，於實務上有下列方向可供主管機關於規劃旅遊景點間交通運輸管理措施時參考：

1. 藉由即時的手機信令資料，透過本研究之分析程序所獲得的景點人數推估資料，可讓主管機關隨時間不同，進行公共運輸班次彈性調整，例如：於特殊活動或連假時，針對觀光景點內較非常態性出現的人潮，進行公共運輸服務之排班調整或臨時加開熱門景點間的接駁專車，以符合遊客的實際需求。
2. 由動態的即時景點間移動人數推估，主管機關可預先知道該景點的熱門時段，針對該景

點周邊的交通道路實施對應的管制措施，例如於熱門時段實施調撥車道，增加進入景點區的道路交通容量供給；或者於靠近景點周遭實施私人車輛進入管制措施，進入景點區的民眾皆須搭乘接駁車，降低大量私人運具湧入對景點環境的衝擊。

- 藉由動態的人數預估，主管機關可針對該景點發布「擁擠度」指標預測。例如自 5 月起，臺北捷運公司公布「車廂擁擠度即時告知功能」功能，透過「臺北捷運 GO」App 及板南線月台電視，旅客可輕鬆掌握板南線進站列車內的人潮狀況，將車廂擁擠程度分為舒適（綠色）、普通（黃色）、略多（橘色）及人多（紅色）等 4 個等級，供旅客查詢（臺北大眾捷運股份有限公司<sup>[20]</sup>）。於後新冠肺炎時期，降低病毒傳染的運輸管理方法將逐漸受到重視，此資訊將可供欲前往該景點的民眾先行了解各時段景點的擁擠程度，自行評估是否改變行程先行前往人數較少的景點參訪。
- 景點的經營業者，也可經由本研究之研究成果，彈性配置各時段景點內的觀光資源。例如於熱門時段增加景點內的餐飲存貨數量，以滿足旅客的需求。

## 五、結論與建議

### 5.1 結論

過去研究中大多透過問卷蒐集或追蹤少數對象的 GPS 資料以預測使用者未來前往的地點，而本研究的貢獻主要在於透過蒐集信令資料，建立一套資料分析的流程，捕捉大量使用者的移動軌跡，使推估之轉移機率矩陣趨近於穩定，較不容易因時間不同而有所改變或受到某些離群值影響預測結果。本研究利用電信公司取得之花蓮縣使用者之手機信令資料，包含六天、分別為三個週末，目的為建立一有系統的預測方法，預測各時段在花蓮縣 59 個觀光景點地拜訪人次，提供觀光景點各時段拜訪人數變化，輔助觀光景點間之大眾運輸服務與觀光景點管理單位之資源配置。

研究中首先篩選在花蓮縣 59 個觀光景點旅遊的使用者，分析使用者在景點間之移動行為，透過找出使用者於一天中的旅次鏈，並產製旅次起迄矩陣，作為馬可夫鏈模式中推估轉移機率矩陣之基礎。在馬可夫鏈模式中，藉由蒐集多天資料，推估在景點間移動之轉移機率矩陣，以每小時為一個時間區段，預測每小時內各景點的拜訪人數。同時考慮到每位使用者旅運行為的差異與信令資料特性，本研究考慮不同的「出發—抵達時間」組合，於轉移機率矩陣增加一「時間維度」，在預測上透過各時段組合累加的方式計算某時段的拜訪人數。以 5 月 19 日、5 月 26 日、6 月 2 日分別作為測試日期時，三天之各時段平均 MAPE 值分別為 29.21%、26.85%、34.14%，屬於合理的預測範圍。

在模式驗證上，由於本研究取得之信令資料同時包含週六與週日，在測試資料選擇上，準確率會因為使用週六或週日的資料而有所差異。一般而言，遊客於週六通常會停留較多景點、旅遊時間持續較長，且晚上會於當地過夜。而週日則因為是假期最後一天需要

返家，遊客行程通常到下午即結束，停留的景點數量通常較少、時間也較短。因此在本研究中，週六所能篩選到的遊客移動型態較多且轉移矩陣較穩定，相較於週日的資料，以週六作為測試資料時，能取得較良好的準確率。於本研究同時段中，各景點實際人數的多寡變異大，假設誤差皆是 20 人的情況下，其實際值的大小差異會影響 MAPE 值大小而難以評斷，各時段的 MAPE 值計算上，可能受到某些實際人數極少的景點影響而提高整體 MAPE 值。

本研究提出之預測方法，主要透過蒐集信令資料的方式得到使用者的移動資訊，若以其他資料來源蒐集使用者的時空軌跡資料也能套用相同的方法進行景點人數預測，例如：GPS 定位資料、Wi-Fi 無線網路資料等等，且能得到與信令資料相同的成效。在實際應用上，透過取得足夠天數的歷史資料，推估各景點間的轉移矩陣以及各時段個別景點的誤差項，並藉由即時獲得的景點人數資料進行運算，以預測未來時段的各景點拜訪人數。

## 5.2 研究限制與建議

由於信令資料的特性，使得本研究在資料分析上有所限制，後續研究可參考以下之建議，將得以提高預測的準確度。(1) 以 50 公尺\*50 公尺網格作為記錄使用者位置的單位，空間定位十分狹小，使得在使用者的信令資料上，可能出現每分鐘在鄰近地區來回飄移的現象，造成無法正確取得使用者的停留時間。而信號飄移是一個難以量化且預測的問題，透過資料清洗與插補處理仍無法完全避免，因此在未來的研究中，建議可以將信令資料的空間定位進一步放大，例如：改為 500 公尺\*500 公尺網格，藉此避免遺失許多使用者的停留時間資訊；後續研究亦可對於網格的分區大小進行敏感度分析。(2) 使用者處於待機或使用 Wi-Fi 時將無信令資料回傳，而在無手機使用行為的情況下，信令資料每小時僅回傳一筆，使得本研究在旅次鏈分析步驟中，將使用者出現在該景點第一筆資料的時間欄位作為開始停留時間，而在該景點最後一筆資料的時間欄位作為結束時間的判斷，將會受限於信令資料的特性，使得停留時間的判斷不一定正確。

受限於取得信令資料天數過少，使得本研究之馬可夫鏈模式預測方法僅使用 5 天訓練資料推估轉移機率矩陣，參考過去的研究，大多以一個月以上的資料進行模式訓練，皆能得到不錯的效果。建議在後續的研究中，在取得足夠訓練資料的情況下，針對特定群體使用者進行追蹤，使馬可夫鏈模式於封閉系統中進行，以減少後續透過迴歸模型估測誤差項可能額外增加之誤差，改善預測結果。舉例來說，研究對象可以為一天中以同時段同地點作為旅次鏈起點的遊客，追蹤該群遊客一天中在景點間的移動行為。

同時，本研究受研究資源限制，所能取得的信令資料僅源自於單一電信公司，其用戶市占率約為 32%，故無法完整代表各觀光景點的總人數；再加以考量各景點的旅客並非全部擁有手機，如部分銀髮族、小孩等，故此些族群人數無法由手機信令資料獲得，建議後續研究可擴大實際觀光景點的人數收集管道，例如交通部觀光局的主要觀光休憩區遊客人數統計資料（交通部觀光局觀光統計資料庫<sup>[21]</sup>）、各觀光景點的售票數據乃至於觀光景點的

錄影資料等，並進一步洽詢國內各大電信公司，取得更完整的手機信令資料，透過資料融合技術將各來源的資料進行清洗及彙整，以此做為模式預測能力分析時的實際值，將更能作為檢驗模式預測能力的標準。

建議未來研究可與其他預測方法進行比較，例如：可以透過高階馬可夫鏈模式 (High-order Markov Chain) 進行進一步的人數推估。以二階馬可夫鏈模式馬可夫鏈 (Second-order Markov Chain) 為例，將考慮使用者前兩個拜訪過的景點產製轉移機率矩陣，過去研究結果顯示，高階馬可夫鏈模式相較於一階馬可夫鏈模式具有較高的預測力。另外，尚可以透過隱馬可夫模式 (Hidden Markov Model, HMM) 考慮某些無法被觀察到的影響因素作為隱藏狀態，透過觀察到的狀態來得到隱藏狀態的一些資訊，提供更廣泛的應用。

## 參考文獻

1. Ma, J., Li, H., Yuan, F. and Bauer, T. “Deriving Operational Origin-Destination Matrices From Large Scale Mobile Phone Data”, *International Journal of Transportation Science and Technology*, Vol.2, No.3, 2013, pp.183-204.
2. Iqbal, M. S., Choudhury, C. F., Wang, P. and González, M. C. “Development of Origin–Destination Matrices Using Mobile Phone Call Data”, *Transportation Research Part C : Emerging Technologies*, Vol.40, 2014, pp.63-74.
3. Li, W., Cheng, X., Duan, Z., Yang, D., & Guo, G. “A Framework For Spatial Interaction Analysis Based On Large-Scale Mobile Phone Data”, *Computational intelligence and neuroscience*, Vol.2014, 2014, pp.1-11.
4. Alexander, L., Jiang, S., Murga, M. and González, M. C. “Origin–Destination Trips By Purpose And Time Of Day Inferred From Mobile Phone Data”, *Transportation Research Part C : Emerging Technologies*, Vol.58, 2015, pp.240-250.
5. C. Qian, W. Li, D. Yang, B. Ran and F. Li. “Measuring Spatial Distribution of Tourist Flows Based on Cellular Signalling Data : A Case Study of Shanghai,” 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, IEEE, 2019, pp. 2584-2590.
6. 邱裕鈞、謝志偉，「應用行動通訊資料於道路交通資訊之蒐集與分析」，中華民國運輸學會 105 年年會暨學術論文國際研討會論文集，民國 105 年。
7. 洪琮博，「利用手機信令推估旅運起迄矩陣」，交通大學運輸與物流管理學系碩士論文，民國 106 年。
8. Demissie, M. G., Phithakkitnukoon, S., Sukhvibul, T., Antunes, F., Gomes, R. and Bento, C. “Inferring Passenger Travel Demand To Improve Urban Mobility In Developing Countries Using Cell Phone Data : A Case Study Of Senegal”, *IEEE Transactions on Intelligent Transportation Systems*, Vol.17, No.9, 2016, pp. 2466-2478.
9. Ying, L., Baorui, H. and Yimin, L. “Bus Scheduling Feasibility Study of Rainy Day Based on the Mobile Phone Signal Data.” 2015 International Conference on Intelligent

- Transportation, Big Data & Smart City (ICITBS), Halong Bay, Vietnam, 2015, pp.192-195.
10. 交通部運輸研究所，**旅運時空資料分析與公共運輸服務應用發展計畫**，民國 107 年。
  11. 王以萱，「以行動信令資料推估觀光需求為基礎之公車服務最佳化模式」，交通大學運輸與物流管理學系碩士論文，民國 107 年。
  12. 施冠毅，「應用信令資料於用戶重現性需求之分析」，交通大學運輸與物流管理學系碩士論文，民國 107 年。
  13. Norris, J. R., *Markov chains*, Cambridge university press, Cambridge, 1998.
  14. Gambs, S., Killijian, M. O. and del Prado Cortez, M. N. “Next Place Prediction using Mobility Markov Chains”, MPM '12 : Proceedings of the First Workshop on Measurement, Privacy, and Mobility, 2012, pp.1-6.
  15. Ashbrook, D. and Starner, T. “Using GPS To Learn Significant Locations And Predict Movement Across Multiple Users”, *Personal and Ubiquitous Computing*, Vol.7, No.5, 2003, pp.275-286.
  16. Xia, J. C., Zeepongsekul, P. and Arrowsmith, C. “Modelling Spatio-Temporal Movement of Tourists Using Finite Markov Chains”, *Mathematics and Computers in Simulation*, Vol.79, No.5, 2009, pp.1544-1553.
  17. Mathivaruni, R.V. and Vaidehi V. “An Activity Based Mobility Prediction Strategy Using Markov Modeling for Wireless Networks”, Proceedings of the World Congress on Engineering and Computer Science, 2008.
  18. Jiang, J., Pan, C., Liu, H. and Yang, G. “Predicting Human Mobility based on location data modeled by Markov chains”, In Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), 2016 Fourth International Conference, IEEE, 2016, pp.145-151.
  19. Almannaa, M., Elhenawy, M. and Rakha, H. “Identifying Optimum Bike Station Initial Conditions using Markov Chains Modeling”, *Transport Findings*. <https://doi.org/10.32866/6801>, 2019.
  20. 臺北大眾捷運股份有限公司，<https://www.metro.taipei/>，2020。
  21. 交通部觀光局觀光統計資料庫，<https://stat.taiwan.net.tw/>，2020。