

手機信令資料探勘於改善觀光旅客公共 運輸服務之研究—以花蓮縣 臺灣好行路線為例

IMPROVING PUBLIC TRANSPORTATION SERVICES FOR
TOURISTS BY MINING CELLULAR-BASED VEHICLE PROBE
DATA – A CASE STUDY OF TAIWAN TOURIST SHUTTLE IN
HUALIEN COUNTY

王晉元 Jin-Yuan Wang¹
盧宗成 Chung-Cheng Lu²
李晟豪 Cheng-Hao Li³
陳其華 Chi-Hwa Chen⁴
吳東凌 Tung-Ling Wu⁵
陳翔捷 Siang-Jie Chen⁶

(107年12月19日收稿，108年3月4日第1次修改，108年6月11日定稿)

摘要

花蓮縣豐富的生態資源、多元的族群文化特色，每年皆吸引眾多的國內外觀光客前往旅遊，但是由於對外交通不便，仰賴臺鐵東部幹線提供有

-
1. 國立交通大學運輸與物流管理系副教授。
 2. 國立交通大學運輸與物流管理系教授，本文通訊作者（100 臺北市中正區忠孝西路 1 段 118 號 4 樓交通大學運輸與物流管理系；電話：+886-2-23494960；傳真：+886-2-23494953；E-mail：jasoncclu@nctu.edu.tw）。
 3. 國立交通大學運輸與物流管理系碩士研究生。
 4. 交通部運輸研究所運輸經營與管理組組長。
 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。

限的公共運輸服務，縣內公共運輸亦不發達，導致外來觀光客在花蓮旅遊時大多以自行開車或是租賃車輛的方式在縣內觀光。目前交通部觀光局提供兩條臺灣好行路線服務花蓮地區的遊客，然而這兩條路線是否能滿足多數前往花蓮地區旅遊觀光客的公共運輸需求是個值得探討的議題。為了解花蓮地區觀光旅客的潛在公共運輸需求，本研究利用資料探勘方法，分析花蓮地區觀光旅客的手機信令資料，比較潛在公共運輸旅客的觀光旅運需求樣態，以及現有兩條臺灣好行路線，找出服務缺口，並提出改善建議。從資料探勘結果發現，文化創意產業園區吸引相當多的遊客，但是臺灣好行路線並未經過這個景點，如果臺灣好行路線能夠繞駛，將能滿足超過八成以上潛在公共運輸旅客的觀光旅運需求。本研究結果可協助主管單位檢視當前的觀光公共運輸服務是否符合旅客的需求，進而對既有路線進行新增站點等相關改善措施。

關鍵詞： 信令資料；大數據；公共運輸；觀光管理；資料探勘

ABSTRACT

Characterized by its rich natural resources and diverse cultures of different ethnic groups, Hualien county has attracted many domestic and foreign tourists. However, most tourists travel by driving their own cars or rental cars in Hualien, because public transportation access to the county relies merely on the limited capacity provided by the Eastern Line of Taiwan Railway; the transit service inside the county is also inconvenient. Although the Tourism Bureau of MOTC provides two Taiwan Tourist Shuttle routes servicing tourists in Hualien, it remains a topic worth of investigation whether or not the public transportation demand of tourists in Hualien can be satisfied by these two routes. In order to explore tourists' potential demand of public transportation in Hualien, this study utilizes data mining techniques to analyze Cellular-based Vehicle Probe data of the tourists in Hualien. We compare the travel demand patterns of potential public transportation users with the two Taiwan Tourist Shuttle routes in Hualien. The service gaps are identified and improvement suggestions are proposed accordingly. The data mining results reveal that Hualien Cultural Creative Industries Park is a popular scenic spot, but it is not covered by the two Taiwan Tourist Shuttle routes. Had the routes included this spot, more than 80% of the potential public transportation demand of the tourists can be satisfied. The findings of this study can assist the authorities to examine whether or not existing tourist transit services are able to meet tourists' demands and to provide suggestions to improve the services, such as adding new stops to the existing routes.

Key Words: Cellular-based vehicle probe data; Big data; Public transportation; Tourism management; Data mining

一、前言

1.1 研究背景與動機

近年來，隨著生活型態的轉變，民眾的生活不僅僅只侷限於住家以及工作地或學校，而是更注重於休閒娛樂的生活，也因此使得觀光運輸的需求隨之增加。而在臺灣各個縣市的觀光發展中，又以有著豐富的人文景觀的花蓮縣尤其重要，在地理景點方面，花蓮境內有著特殊的斷崖景觀和峽谷地形；而在人文方面，也有著客家文化、原住民文化等多元的族群文化特色。也因為如此，花蓮縣深受國內外觀光客喜愛，也經常被觀光客拜訪。

花蓮縣政府觀光處統計，106 年度花蓮縣內十二大觀光遊憩區遊客人次合計達 1,027 萬人次，但花蓮縣不像臺灣西半部的各個城市之間有著臺鐵、高鐵以及國道客運等完善的公共運輸系統可以進行移動，而是僅能利用臺鐵的東部幹線聯外，因此外來旅客進入花蓮約有 51%自行開車，27%搭乘臺鐵，16%搭乘遊覽車。進入花蓮後，由於境內的公共運輸服務亦不發達，僅有 3 條主要的公車路線以及臺鐵進行服務，且公車方面的各個路線班次皆不算密集（1 個小時內僅有 1~2 班車），因此有 68%旅客在花蓮境內景點移動的交通工具選擇自小客車（55%）或機車（13%），17%搭乘遊覽車，利用（副）大眾運輸的比例僅 9%（計程車 5%，臺鐵 3%，公車 1%）。經調查旅客到花蓮後未搭乘公共運輸的原因包含：資訊不清楚、班次太少、站牌太少、路線太少及轉乘不便，顯見花蓮縣雖為觀光大縣，但其公共運輸設計對於觀光客而言仍非友善。旅客則反映若能提供公共運輸即時資訊、調整公共運輸路線及班次密度、可隨招隨停服務等，則會提升其搭乘意願。

在 2010 年時，交通部觀光局為了利於觀光客前往臺灣各處的景點旅遊，推出了「臺灣好行」的觀光巴士景點接駁服務，透過設計路線連結各地的臺鐵火車站以及觀光景點，並以公車的方式進行接駁，吸引民眾使用公共運輸系統於觀光景點之間移動。臺灣好行服務的收費方式有分為點對點的單趟旅次收費和一日之內無限次數搭乘的一日券兩種，前者為選定起迄點後，依照起迄點的距離收費；而後者則是選定日期及路線之後，在該日臺灣好行營運時間之內可無限次數在該路線的各個站點之間乘坐。

花蓮地區現有兩條臺灣好行路線進行營運，分別為縱谷花蓮線以及太魯閣線。縱谷花蓮線以臺鐵花蓮火車站為起點站，沿著花東縱谷行駛，一路停靠鯉魚潭、花蓮糖廠等景點，並以大農大富平地森林為區為終點站，共計 12 個站點，行車時間約 1 小時 42 分鐘；每日早上 8 點 40 分發首班車，下午 3 點 30 分發末班車，班車時間間隔 60 ~ 80 分鐘，一日共計 7 班次。太魯閣線則是以臺鐵花蓮火車站為起點站，往北行駛，行經七星潭、太魯閣等景點，並以天祥風景區為終點站，共計 10 個站點，行車時間約 1 小時 15 分鐘；每日早上 7 點發首班車，下午 3 點 10 分發末班車，班車時間間隔 40 ~ 90 分鐘，一日共計 9 班次。

目前縱谷花蓮線以及太魯閣線仍屬於營運虧損補貼路線，根據公路總局 104 年公路客運虧損補貼資料，花蓮縣公路客運虧損補貼共計 102,814 千元（30 條路線），其中縱谷花蓮線補貼 5,500 千元（佔 5.3%），太魯閣線補貼 4,901 千元（佔 4.8%），合計 10,401 千元，佔

所有補貼金額比例約 10.1%；而假日外來旅客到花蓮選擇臺灣好行當作交通工具的比例不到 1%，顯現其營運效率仍待提升。為了吸引民眾使用臺灣好行服務，除了在票價上做調整以及推出優惠套票等方案外，路線規劃與民眾旅遊行是否程貼近亦是非常重要的營運策略。本研究試從此兩條觀光巴士的服務路線與觀光客的實際旅運行為是否符合的角度切入，若觀光巴士服務路線與觀光客的旅運行為相似，表示觀光客欲前往的景點皆能被臺灣好行服務到，亦說明觀光客在出遊時可選擇利用公共運輸系統來進行移動；反之，若觀光巴士服務路線與觀光客的旅運行為不一致，表示多數觀光客在旅遊時較不會使用到臺灣好行的服務，可能造成資源的浪費。

為了得知觀光客的在各個觀光景點的旅運行為，需要蒐集與分析觀光客的旅運資料，傳統的旅運資料調查方式有電話訪問、問卷調查等，然而這些傳統方法可能需要大量的人力與時間。隨著行動通訊裝置的普及，蒐集與分析行動通訊資料亦成為得到民眾旅運資料的可行方法之一，相對於傳統問卷調查，手機基地台數據具有高覆蓋率、更新頻率迅速及效率高的優點；不僅如此，手機信令資料是手機使用者於使用 4G 網路時被基地站所紀錄的當下時空資訊，因此資料整體不會受到觀光用戶本身的人為因素影響，例如：刻意隱瞞行蹤、錯估旅行時間等 (Alexander 等人^[1])。然而，當手機使用者停止使用 4G 網路 (如：使用 Wi-Fi) 或是所在地無 4G 基地站服務時，信令資料便會有所漏失。另根據國家通訊傳播委員會 (NCC) 於民國 106 年 9 月發布的 106 年持有手機民眾數位機會調查報告^[2]，截至 2017 年 7 月，臺灣的 4G 用戶占了全體手機用戶的 73%，當中使用行動上網的用戶有 2270 萬戶，行動上網率達 79%。此現象代表手機上網的服務在臺灣民眾的生活占了不可或缺的地位，也表示將帶來極為可觀的行動通訊資料。綜上所述，雖然手機信令資料仍有其缺點，但依照目前臺灣的手機普及率以及 4G 用戶註冊比例，手機信令的資料品質可較傳統問卷調查與電話訪問為佳。

花蓮縣的面積廣達 4,629 平方公里，雖有享譽國內外之觀光景點，但卻分散廣大縣境內，也使得若欲使用傳統問卷調查方式至各景點進行調查，顯得費時又費人力；若使用電子票證資料，則因觀光旅客搭乘客運及觀光好行比例不高，票證資料相對侷限，其分析結果恐不具代表性，且電子票證係已搭乘之旅客資料，較難據以分析出潛在公共運輸使用者之旅次行為。為克服上述困難，本研究將以手機所產製的行動通訊資料，又稱手機信令 (cellular-based vehicle probe, CVP) 資料，作為分析花蓮縣觀光客旅運行為的基礎，探索花蓮地區觀光客潛在公共運輸需求用戶之旅運行為，透過資料科學方法中的關聯規則以及循序樣式探勘兩個技術找出常被觀光客拜訪的景點組合，再將結果與花蓮地區的兩條臺灣好行路線進行比對，以研擬出花蓮地區臺灣好行服務路線的相關建議。

1.2 研究目的與範圍

本研究之目的為利用關聯規則與循序樣式探勘兩個資料科學方法的技術，找出潛在公共運輸觀光需求用戶常去拜訪的景點組合，並將結果與現有之觀光公共運輸環境相比對，以探討是否存有服務缺口。透過大數據的技術，以及新興的資料型態，並希望能以此實務

數據的結果給予相對應之服務建議，而非單純採用經驗判斷或資料量相對較少的抽樣問卷調查等方式給出相關建議。且由於本研究之目的為探索觀光客的潛在公共運輸需求與觀光公共運輸的服務是否有落差，因此研究場域選在以觀光產業較為發達的花蓮地區。本研究僅考慮利用周末假期去花蓮進行兩天（星期六、日）或三天（星期五、六、日，或星期六、日、一）短期旅遊的觀光客，不考慮利用平常日之遊客與團客，因此取得的資料期間也選在有較多觀光旅次產生的周末假期，而非觀光客較少的平常日。

本篇論文後續章節內容簡要說明如下：第二小節為相關文獻回顧，第三小節說明本研究中所使用之相關資料，第四小節陳述研究流程與方法，第五小節為分析結果與討論，第六小節提出結論與建議。

二、文獻回顧

本研究回顧手機信令資料應用於改善公共運輸服務的文獻，以及手機信令資料應用於觀光旅運、或其他交通運輸領域之文獻。

交通部運輸研究所執行旅運時空資料分析與公共運輸服務應用發展計畫^[3]，該研究計畫利用手機信令資料與公共運輸動態資訊與班表，發展出產製分時分運具別之旅運起迄需求矩陣，並根據所產出之旅運起迄需求，找出公共運輸服務缺口，提出數種公共運輸服務改善策略，例如調整路線與發車頻率等。之外，該研究亦透過信令資料推估各地區常住人口數，以及利用資料探勘技術分析高雄市主要觀光景點間的關聯性。

Holleczeck 等人^[4]利用手機信令資料搭配捷運班表，發展出一個能及時回報捷運系統當時的路網情況的系統。此研究之場域選定在新加坡，而回報的資訊包含：月台等候人數、預期旅行時間，讓民眾可以根據當時各路線的服務狀況來避免民眾浪費太多時間於候車。

Demissie 等人^[5]將手機信令資料結合都市公共運輸資料，找出塞內爾加以及達卡（賽內爾加首都）是否存有公共運輸服務缺口，該研究以境內的公車路線作為研究對象。透過信令資料找出用戶經常活動的地點做為可能的起迄對，再比對當地的公車路線來找出是否存有服務缺口。

Ahas 等人^[6]利用手機信令資料分析各國觀光客於愛沙尼亞的相關資訊。該研究先篩選出非設籍於愛沙尼亞的用戶，並將他們定義為觀光客，再利用其信令資料分析觀光客於愛沙尼亞境內的時間分布、地理位置分布等資訊。

Trasarti 等人^[7]以用戶信令資料為基礎，透過循序樣式探勘 (sequential pattern mining, SPM) 的方法，找出一地區中具有關聯性的地點組合。該研究首先定義出搜索具有關聯性的地點間的最小支持度，再根據研究場域內人口稠密度的變化 (C-pattern)，配合循序樣式探勘的演算法找出具關連性的地點組合。

Mamei 與 Colonna^[8]將信令資料透過機器學習的方法為用戶進行分類，共有住戶、通勤人士、經過用戶、遊客、以及短途旅行者。最後將研究所完成的分類與政府部門公布的

旅遊人數比較後，發現兩者的結果有高度相關。

Holleczek 等人^[9] 利用手機信令資料結合電子票證資訊，分析新加坡私人運具與公共運輸運具的使用情形。此研究先利用手機信令資料得知起迄對之間的總旅次量，減去票證資料中同個起迄對之間的公共運輸旅次量，得到該起迄對之間的私人運具旅次量。結果顯示私人運具與公共運輸旅次量會因時間和空間的不同而相異。

Chen 等人^[10] 利用信令資料與興趣點資料，透過機器學習的方法找出觀光熱點區和觀光客的移動行為，研究場域為中國的海南地區。此研究先對興趣點進行整合，再透過循序樣式探勘找出在觀光客之間較常發生的移動樣態，最後再將結果進行進一步的分析。研究結果之一顯示，在三亞市中，亞龍灣是最熱門的景點，且觀光客幾乎會花上一天的時間在該處遊玩。

Bekhor 等人^[11] 利用手機信令資料，探討各個使用者長距離之旅次起迄，並提出將此結果應用在運輸規劃的建議。有別於傳統僅能取得旅運者在短距離移動的起迄，此研究是以更有效率的方法、耗費較低的成本來分析使用者長距離的旅次起迄。而研究結果與實際資料相比後也得到兩者結果有高度相關的結論。

Iqbal 等人^[12] 蒐集達卡市的手機信令資料，並透過此資料推估旅次起迄矩陣後，再利用電腦軟體進行交通量指派的推估。將推估結果與實際結果相互比對後顯示，兩結果的方均根誤差百分比僅為 13.8%。

從上述各個研究可看出手機信令資料可實際應用在交通運輸相關領域的研究，如：改善公共運輸系統、推估旅運者身分、觀光景點間的關連性分析等。同時，Mamei 等人^[8]、Bekhor 等人^[11] 和 Iqbal 等人^[12] 的研究，也說明了用信令資料所得到的交通運輸相關結果，與傳統或實際的調查資料所得出的結果有高度的相關。此外，Ahas 等人^[6] 提到手機信令資料在取得上，比傳統調查資料來得有效率且省成本。因此，本研究認為透過手機信令資料來探討公共運輸服務的改善是可行，且相較傳統調查方式來得要有效率的。

三、研究資料

本研究所取得之 CVP 資料為國內一家主要的電信業者所提供，時間為民國 107 年 5 月 19 到 20 號、5 月 26 到 27 號以及 6 月 2 號到 3 號這 3 個周末。此外，電信業者亦提供與每個信令資料的用戶資訊相對應之用戶居住地資料，以及各個網格中心點位置的網格中心點資料。而花蓮地區主要觀光景點（興趣點，points of interest）的相關資料，則是由國內某導航專業科技公司所提供之地理資訊圖資中擷取。以下針對上述 4 份資料進行詳細說明。

3.1 手機信令資料

手機信令資料 (cellular-based vehicle probe, CVP) 為手機與電信基地站產生互動時，

被基地站所記錄下來包含時間以及位置資訊的資料。當手機使用者透過 4G 網路從事上網活動，或是手機使用者因移動導致手機與連結的電信基地站有所改變時，電信基地站便會記錄該活動發生的時間以及該手機的所在位置，即為信令資料。此外，即便手機使用者長時間沒有讓手機產生與基地站互動之行為，如：長時間使用 Wi-Fi、關閉網路訊號並處於待機狀態等非關機的情況，電信公司仍會每小時主動式地蒐集各個用戶的行動裝置並產生信令資料紀錄。本研究取得電信業者的手機信令資料其定位精確度為 50*50 公尺為單位的「網格」，把手機使用者的位置資訊以網格的方式呈現。

表 1 為本研究所使用之信令資料格式。為保護手機用戶的隱私，第 1 個欄位並非直接顯示手機門號，而是以加密後之 64 位英數字混和的亂碼表示，同時為了方便研究進行，同一個門號會加密成同一組亂碼。第 2 個欄位為時間資訊，包括西元年、月、日、小時、分鐘以及秒鐘。第 3 個欄位則為位置資訊，並如同上一段落所提到的以網格的方式表示，而網格編碼為 9 為純數字所組成。表 1 的第 1 筆信令資料顯示用戶 4e4b4（為減少篇幅與方便表達，在本文中取使用者識別碼之前 5 碼代表該使用者）在西元 2018 年 5 月 19 日的早上 8 點 22 分時，在網格 214707015 使用手機進行上網的動作，即表示該名用戶在該時間身處在該網格的位置上。

表 1 手機信令資料格式

加密後之使用者識別碼	日期與時間	網格
4e4b496d3c4fc9fa888cd36743bbc7dc94 da492018dd41685e7dd0e02657d368	2018/05/19 08:22:00	214707015
ee3bdd30c1b93b9eb0f6a38488e5672a61 b7347a4adb48a06b3c8b500ea3a031	2018/05/19 08:24:00	213627013
4e4b496d3c4fc9fa888cd36743bbc7dc94 da492018dd41685e7dd0e02657d368	2018/05/19 08:25:00	254715036
4e4b496d3c4fc9fa888cd36743bbc7dc94 da492018dd41685e7dd0e02657d368	2018/05/19 08:32:00	242968003

3.2 用戶居住地資料

表 2 為用戶居住地資料表範例，本研究利用此份資料篩選出其居住地非為花蓮縣的使用者作為觀光旅運需求的分析對象。第 1 欄為加密後的使用者識別碼，和信令資料一樣為 64 位英數字混和之亂碼，但表 2 僅取每個用戶識別碼之前 5 碼表示。第 2 到第 4 欄為該用戶居住地的一、二級行政區、三級行政區以及四級行政區。表 2 第 1 筆資料即說明用戶 4e4b4 居住在花蓮縣吉安鄉東昌村。

表 2 用戶居住地資料範例

加密後之使用者識別碼	一、二級行政區	三級行政區	四級行政區
4e4b496d3c4fc9fa888cd36743bbc7dc94da492018dd41685e7dd0e02657d368	花蓮縣	吉安鄉	東昌村
ee3bdd30c1b93b9eb0f6a38488e5672a61b7347a4adb48a06b3c8b500ea3a031	高雄市	前鎮區	竹南里
11385848cc463de3e9338cbe377c59542b3b9c78fa51f6ee8ad5b60c1c9a3aaa	臺東縣	臺東市	岩灣里
8qs12b23747f1b60d5ae104ee1e155e4cb703d8c318a7abc52cfa51553c117406	花蓮縣	吉安鄉	東昌村

3.3 網格中心點資料

表 3 為網格中心點資料範例，第 1 欄為網格編碼，第 2 到第 4 欄為該網格所在的一級或二級、三級以及四級行政區，第 5 和第 6 欄則為該網格的中心點之經緯度座標。此份資料顯示每個網格的所在地，因此可直接藉由比對信令資料和此份資料「網格」欄位，來得知每一筆信令資料產生時，手機使用者當時的所在位置，舉例來說，表 1 的第 1 筆資料其網格欄位與表 3 的第 1 筆資料的網格欄位相同，表示該名用戶當時所處的位置在中心點為經度 121.260507、緯度 23.1281893 向外擴張 50*50 公尺的範圍內。除此之外，亦可透過經緯度得知每個網格之間的距離。

表 3 網格中心點資料範列表

網格	一、二級行政區	三級行政區	四級行政區	經度	緯度
214707015	花蓮縣	富里鄉	豐南村	121.260507	23.1281893
213627013	花蓮縣	富里鄉	豐南村	121.260507	23.12864083
254715036	花蓮縣	富里鄉	豐南村	121.260507	23.12909235
242968003	花蓮縣	富里鄉	豐南村	121.2609952	23.12502063

3.4 興趣點資料

本研究中所考慮之花蓮地區主要觀光景點（或稱為興趣點，points of interest）的選取方式，為交通部運輸研究所根據交通部觀光局之建議、Google Map 以及 TripAdvisor 兩個網站的遊客評論，選取花蓮境內前 60 名的觀光景點。選定觀光景點後，由國內某導航科技公司提供這些景點圖資與網格資料，將花蓮境內的 60 大景點以及 24 個臺鐵火車站，共

計 84 個興趣點轉換為網格的形式表示。表 4 為花蓮地區兩個興趣點的部分資料範例，資料的第 1 欄為興趣點的名稱，第 2 欄則為該興趣點的編號，編號為該科技公司內部之流水號，有 6 碼和 9 碼兩種，而每個興趣點只會對應到一個興趣點編號。第 3 個欄位為該興趣點所包含之網格，且由於一個興趣點的面積（大於 2500 平方公尺）大於一個網格的面積（250 平方公尺），因此一個興趣點會對應到多個網格，每個興趣點都有其對應之網格群。

表 4 興趣點資料範例

興趣點名稱	興趣點編號	網格
a-zone 花蓮文化創意產業園區	310001275	254233058
a-zone 花蓮文化創意產業園區	310001275	254233059
a-zone 花蓮文化創意產業園區	310001275	254233060
a-zone 花蓮文化創意產業園區	310001275	254233069
a-zone 花蓮文化創意產業園區	310001275	254233070
七星潭風景區	170995	259678001
七星潭風景區	170995	259678002
七星潭風景區	170995	259678011

四、研究流程與方法

本研究之研究流程如圖 1 所示。第 1 步為篩選觀光用戶，將 3 個周末的全部信令資料篩選出觀光客所產生的信令資料，也就是說將非觀光客所產生的信令資料予以剔除。第 2 步為建立觀光用戶的旅運資訊，將觀光客在哪些興趣點有停留的紀錄找出。第 3 步為找尋潛在公共運輸需求用戶，將觀光用戶中為全程使用私人運具的用戶先排除，並針對使用公共運輸運具（主要為火車）抵達花蓮的觀光客進行下一步分析。第 4 步則是將前一步驟所篩選出來的目標用戶之旅運資訊，透過關聯規則與循序樣式探勘兩種資料科學分析方法，找出經常被目標用戶拜訪的興趣點組合，並以該結果觀察目標用戶在興趣點之間移動。而最後一步驟為將前一步驟所得到的結果，作為潛在公共運輸需求觀光用戶在興趣點之間移動的依據，並比對臺灣好行目前在花蓮地區運行的兩條路線，以檢視是否存有服務缺口。

4.1 篩選觀光用戶

本研究以電信業者所提供信令資料用戶之居住地，作為判斷該用戶是否為觀光客之依據，將信令資料分為居住地在花蓮地區和居住地不在花蓮地區兩大類別；也就是說，當用戶的居住地是在花蓮地區，則該用戶會被視為是當地居民；反之，若用戶的居住地不在花蓮地區，則該用戶被視為是觀光客。

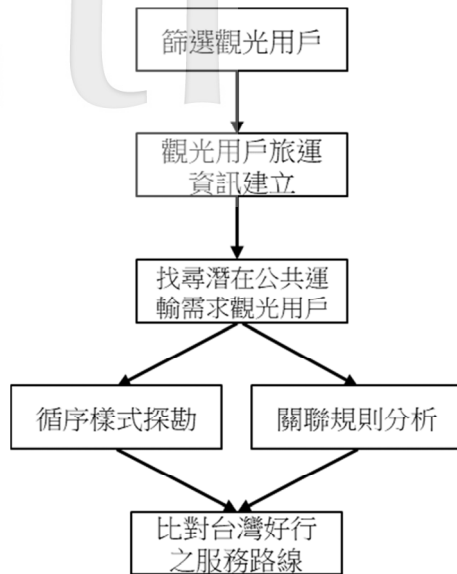


圖 1 研究流程

本研究比對信令資料 (3.1 小節) 與用戶居住地資料 (3.2 小節)，檢視每一筆信令資料的用戶其居住地是否為花蓮縣。以表 1 為例，第 1 筆信令資料的使用者識別碼和表 2 的第 1 位用戶相同，說明產生表 1 第 1 筆信令資料的用戶其居住地為花蓮縣，而產生第 2 筆信令資料的用戶，可對應到表 2 的第 2 位用戶，且該用戶之居住地並非在花蓮地區，所以此用戶所產生的信令資料，將會是本研究所需要的觀光用戶信令資料。在表 1 中，由於第 1、第 3 和第 4 筆的信令資料是由花蓮當地的居民所產生的，所以將這幾筆資料刪除，僅留下第 2 筆信令資料進入後續分析階段。表 2 的第 3 及第 4 筆資料所代表的用戶並非沒有信令資料的產生，而是由於表 1 只取該時段中的 4 筆信令資料作為範例，且為了說明本研究將居住於花蓮地區的用戶所產生的信令資料刪除，因此特別在 4 筆資料中取 3 筆由用戶 4e4b4 所產生的信令資料作為展示，而造成用戶 11385 和用戶 8qs12 在表 1 皆沒有信令資料紀錄。

4.2 觀光用戶旅運資訊建立

篩選完觀光客之後，接著對每位觀光客計算其個別在哪些興趣點有停留的行為，並依用戶在各個停留點發生的時間點來建立每位用戶的旅運資訊。本節將用戶旅運資訊的建立細分為 3 個子步驟，如圖 2 所示，分別為精準比對、容忍距離 200 公尺比對以及停留時間計算，以下對此 3 個步驟進行詳細說明。

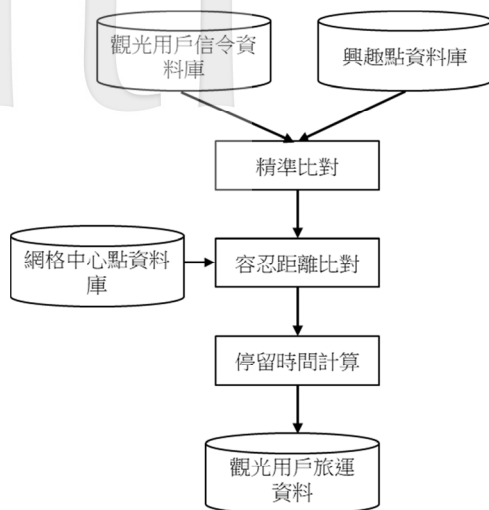


圖 2 觀光用戶旅運資訊建立流程圖

4.2.1 精準比對

此步驟為得知用戶的每一筆信令資料被記錄的地點是否在興趣點資料表格中，可透過比對觀光用戶信令資料與興趣點資料中的「網格」欄位來達成此目的。如圖 3 所示，圖左上方的表格為觀光用戶信令資料範例，而圖右上方的表格為興趣點資料範例，每個興趣點僅選擇一個所在網格作範例。圖 3 下方表格為比對結果，可發現信令資料表中的第 1 筆資料的網格與第 3 筆資料的網格，與興趣點資料的花蓮觀光糖廠網格一致，因此判定用戶 6609f 在 2018 年 5 月 19 號的早上 8 點 25 分以及早上 8 點 26 分在花蓮觀光糖廠；而信令資料第 5 筆資料的網格編號與興趣點資料中的花蓮火車站的網格編號一致，因此判定用戶 113fd 在 2018 年 5 月 19 號早上 8 點 30 分時在花蓮火車站。而信令資料第 2 筆及第 4 筆資

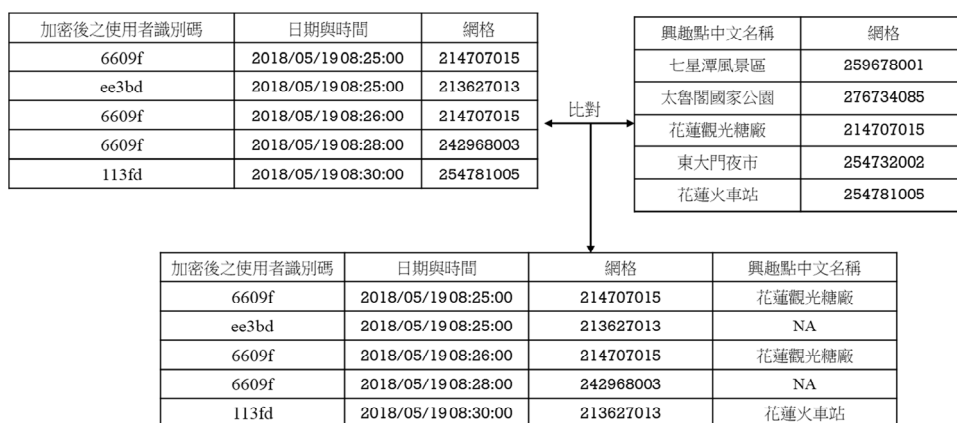


圖 3 精準比對示意圖

料，由於其網格編號並未比對到任一興趣點資料中的網格編號，因此判定該筆資料發生的位置並非於所選取之興趣點中，並以「NA」紀錄之，如圖 3 下方表格的第 2 及第 4 列。

4.2.2 容忍距離比對

由於手機與基地站之間的訊號可能會因天氣因素、建築物干擾等原因而受影響，進而導致信令資料出現漂移的現象；又或者是其他因素，如：手機使用者在觀光景點遊玩時不常使用手機、觀光景點不在電信業者基地站服務範圍內等，導致信令資料出現的位置並非能精準比對到觀光景點的所在位置，而是在觀光景點附近。為改善 4.2.1 小節所述精準比對方法之限制，本研究提出容忍距離比對的方法，將信令資料網格與觀光景點網格比對的範圍，擴大到觀光景點向外延伸某一給定長度（即容忍距離）的區域；也就是說，雖然某一筆信令資料網格與一個觀光景點網格不同，但該信令資料的網格與某一觀光景點的任一網格間之距離在給定之容忍距離內，則仍然判定該筆信令資料與該觀光景點之網格比對成功。本研究參考電信業者的建議，經由參數敏感度測試與評估，將容忍距離設定為 200 公尺。

容忍距離比對方法為利用第 3.3 小節提到的網格中心經緯度資料，得到未被比對到興趣點之信令資料的網格中心經緯度，再計算該中心點與各個興趣點的所在網格是否有距離小於 200 公尺者。以圖 3 為例作說明，首先針對圖 3 下方表格中興趣點名稱欄位顯示「NA」的每一筆信令資料（也就是精準比對失敗的信令資料）找出其網格之中心點經緯度，接著判斷其方圓 200 公尺以內是否有任一興趣點所包含之網格；若有，則為容忍距離比對成功，將所判斷到之興趣點填入該筆信令資料的興趣點名稱欄位；若否，則為容忍距離比對失敗，該筆信令資料之興趣點名稱欄位仍為「NA」。此外，若一筆信令資料網格的方圓 200 公尺內，同時比對到多個興趣點所在網格，則以距離最近之興趣點作為比對成功的結果。

4.2.3 停留時間計算

為了得知用戶信令資料出現在興趣點的紀錄是經過或是實際有停留行為發生，需計算用戶於各個興趣點停留的時間。計算的方法為透過 4.2.1 以及 4.2.2 小節得到的用戶信令資料網格與興趣點資料網格比對結果，先將其依照用戶、時間日期排序後，再檢視是否有連續被記錄在同一個興趣點的情況發生。若觀察到某用戶於同一興趣點有連續被記錄的情況，且該連續被記錄的情況持續時間超過預設門檻值，則判定用戶於該興趣點有停留的情況發生，並在記錄停留行為的同時，根據該停留點是用戶當周末的第幾個停留行為，而給予相對應的時間戳記。

以圖 4 為例，圖上方表格為用戶 6609f 在 2018 年 5 月 19 日上午 8 點 22 分到上午 9 點 10 分，與興趣點資料完成精準比對以及容忍距離 200 公尺比對的信令資料。如果定義停留時間須大於 5 分鐘才算是有興趣點停留行為發生，則該用戶在七星潭風景區從上午 8 點 24 分停留至上午 8 點 35 分，共計 11 分鐘，因此紀錄該用戶於七星潭風景區有停留行為。然而，雖然該用戶在煙波大飯店被記錄的時間有上午 9 點 01 分和上午 9 點 09 分，但

由於中間有一筆信令資料其位置並非被記錄於煙波大飯店，表示用戶並沒有在煙波大飯店有連續被記錄的情況，因此不可將煙波大飯店算成是該用戶的停留點。

對於用戶被記錄到的停留點將會以時間戳記來取代信令資料實際發生的時間。以圖 4 為例，用戶 6609f 在信令資料蒐集期間的第 1 個停留點為七星潭風景區，且停留時間為 11 分鐘，因此可看到圖 4 下方表格中用戶 6609f 在七星潭風景區停留的時間戳記紀錄為 1。倘若用戶 6609f 在後續的時間有其他的興趣點停留行為，則會依序給予相對應的時間戳記，以此類推。

加密後之使用者 識別碼	日期與時間	興趣點名稱
6609f	2015/05/19 08:22:00	NA
6609f	2015/05/19 08:24:00	七星潭風景區
6609f	2015/05/19 08:25:00	七星潭風景區
6609f	2015/05/19 08:35:00	七星潭風景區
6609f	2015/05/19 08:51:00	NA
6609f	2015/05/19 09:01:00	煙波大飯店
6609f	2015/05/19 09:03:00	NA
6609f	2015/05/19 09:09:00	煙波大飯店
6609f	2015/05/19 09:10:00	NA

↓

興趣點名稱	停留時間	時間戳記
七星潭風景區	11	1

圖 4 停留點計算示意圖

4.3 找尋潛在公共運輸需求觀光用戶

此小節說明如何從所篩選出的觀光用戶旅運資料中，再進一步找出可能會使用公共運輸於各個興趣點間移動的觀光用戶，本文中稱此類用戶為潛在公共運輸需求觀光用戶。外縣市或外籍觀光客到花蓮旅遊，除了少數搭乘飛機外，一般可以分為自行開車以及搭乘臺鐵東部幹線列車兩大類，由於花蓮最主要的聯外公共運輸為臺鐵，每到周末假期臺鐵東部幹線列車經常一票難求。自行開車的觀光客由於享有高度的彈性，極少在花蓮縣境內使用公共運輸；而搭乘臺鐵抵達花蓮的觀光客，其旅遊方式可以再分為租賃汽機車與使用當地

公共運輸兩類，因此本研究在找尋潛在公共運輸需求觀光用戶方面，主要是針對搭乘臺鐵抵達花蓮的觀光客。

如 1.2 小節所述，本研究僅考慮利用周末假期去花蓮進行兩天（星期六、日）或三天（星期五、六、日，或星期六、日、一）短期旅遊的觀光客，因此在找尋潛在公共運輸需求觀光用戶時，乃是透過檢查每個觀光用戶的停留點資訊，觀察是否該用戶在資料蒐集期間的第一個或最後一個停留點為臺鐵火車站，若是，則該用戶可能是搭乘臺鐵抵達花蓮（第一個停留點為臺鐵火車站），或是搭乘臺鐵離開花蓮（最後一個停留點為臺鐵火車站），可以視為潛在公共運輸需求觀光用戶。

表 5 為 3 名用戶的旅運資訊範例表。可看出用戶 6609f 在資料蒐集期間曾於七星潭風景區、花蓮火車站以及煙波大飯店有停留的行為；用戶 019cf 則曾於花蓮火車站、太魯閣國家公園以及太平洋公園有停留的行為；而用戶 8qs12 則曾於花蓮觀光糖廠以及富里火車站有停留的行為。根據以上之定義，由於用戶 019cf 以及用戶 8qs12 分別在資料蒐集期間的第一個或最後一個停留點為臺鐵火車站，因此此兩名用戶將被定義為潛在公共運輸需求觀光用戶。而用戶 6609f 雖曾在花蓮火車站有停留的紀錄，但該停留點並非用戶 6609f 的第一個或最後一個停留，因此用戶 6609f 並非本研究定義之潛在公共運輸需求觀光用戶。

表 5 觀光用戶旅運資訊範例表

加密後之使用者識別碼	興趣點名稱	停留時間	時間戳記
6609f	七星潭風景區	11	1
6609f	花蓮火車站	6	2
6609f	煙波大飯店	50	3
019cf	花蓮火車站	10	1
019cf	太魯閣國家公園	123	2
019cf	太平洋公園	17	3
8qs12	花蓮觀光糖廠	59	1
8qs12	富里火車站	8	2

4.4 關聯規則

關聯規則 (association rule, AR) 為資料探勘領域中常被使用的方法之一，用於找出資料庫中彼此之間具有關聯性的資料集。而關聯規則透過設定最小支持度、信賴度以及增益 3 個門檻值，來得到一個資料庫中互相具有關聯性的資料集。以表 6 的資料庫為例，為方便檢視，表 6 為將用戶旅運資料去掉時間戳記後，再以用戶為單位，將該用戶去過的各個興趣點整理成一個集合，即為第 2 欄位的興趣點集合，而表 6 顯示 4 名用戶以及該 4 名用戶在資料蒐集期間所停留之興趣點集合。此外，為了方便後續說明，表 6 將各個興趣點給

予相對應之英文代號，並在之後的說明以代號取代興趣點名稱，例如 A 代表砂卡礑步道，B 代表 a-zone 花蓮文化創意產業園區。

從表 6 可以得知用戶 ee3bd 在旅程中去了砂卡礑步道和 a-zone 花蓮文化創意產業園區兩個地點，而用戶 6609f 去了七星潭風景區、煙波大飯店和花蓮火車站 3 個地點，用戶 113fd 去了砂卡礑步道、花蓮火車站、煙波大飯店和七星潭風景區 4 個地點，以及用戶 00087 去了煙波大飯店和花蓮火車站兩個地點。而關聯規則首先透過最小支持度初步找出資料庫中彼此可能具有關聯性的資料，且支持度表示一個資料或資料集在資料庫中占有的比例，即 $\text{support}(A) = P(A)$ ，倘若某個資料（或資料集）的支持度大於最小支持度，則將該結果稱為高頻樣態。由於本研究利用關聯規則的目的為找出彼此之間具有關聯的興趣點組合，而不考慮單一資料所構成的高頻樣態，因此在最小支持度設定為 0.5 的情況下，可以找出如表 7 所列出的 12 個由 2 個或 2 個以上的資料所構成的高頻樣態。以表 7 的第 1 列為例， $\text{support}(C \Rightarrow D) = P(C \cap D)$ ，則可藉由表 6 看出旅程中會拜訪七星潭風景區和煙波大飯店的旅客數占全體觀光客的一半（在 4 名用戶中佔了 2 名），因此該樣態的支持度等於 0.5，符合高頻樣態之條件。

表 6 關聯規則示範資料庫

使用者	興趣點集合
ee3bd	砂卡礑步道 (A)、a-zone 花蓮文化創意產業園區 (B)
6609f	七星潭風景區 (C)、煙波大飯店 (D)、花蓮火車站 (E)
113fd	砂卡礑步道 (A)、花蓮火車站 (E)、煙波大飯店 (D)、七星潭風景區 (C)
00087	煙波大飯店 (D)、花蓮火車站 (E)

此時，為了檢視上述具有一定支持度的高頻樣態其規則是確實存在，抑或純粹只是發生的先導規則 (antecedent, left-hand-side, LHS) 太過普及，導致後繼規則 (consequent, right-hand-side, RHS) 的發生皆會成為高頻樣態，信賴度即成為一個衡量的重要指標。舉例來說，假設發現一個規則為「拜訪花蓮糖廠則會拜訪馬太鞍濕地生態館」，則可透過信賴度檢視該規則是否確實存在，抑或是該規則只是由於遊客到花蓮皆會拜訪花蓮糖廠，導致只要拜訪馬太鞍濕地生態館的遊客數超過一定比例，「若拜訪花蓮糖廠則會拜訪馬太鞍濕地生態館」的規則就會產生。因此，信賴度的計算方式為計算在發生先導規則的情況下 (拜訪花蓮糖廠)，發生後繼規則 (拜訪馬太鞍濕地生態館) 的機率為何，即 $\text{confidence}(\text{拜訪花蓮糖廠} \Rightarrow \text{太訪馬太鞍濕地生態館}) = P(\text{拜訪馬太鞍濕地生態館} | \text{拜訪花蓮糖廠})$ 。以表 7 的第 1 筆資料為例，在發生 D 的情況下，皆會有 C 與 D 事件同時發生的情況，因此該條件機率的結果為 1，也就是該規則的信賴度為 1。而以第 2 筆資料為例，在發生 D 的情況下，D 與 C 同時發生的機率僅有 0.667，因此其信賴度為 0.667。

最後，透過增益值可以讓研究人員得知，某樣態在用以預測結果時能比隨機發生的機

率好多少的程度。增益值的計算方式為 $P(\text{後繼規則}|\text{先導規則})/P(\text{後繼規則})$ ，也就是某樣態的信賴度與該樣態的後繼規則發生機率的比值。舉例來說，假設「拜訪花蓮糖廠則拜訪馬太鞍濕地生態館」規則的信賴度為 0.5，且在所有觀光客的停留點紀錄中，旅程中包含馬太鞍濕地生態館的比例佔 0.25，也就是說此規則的增益為 $0.5/0.25=2$ ，表示花蓮糖廠與馬太鞍濕地生態館是正關聯 (positive associated)，亦可表示花蓮糖廠和馬太鞍濕地生態館在一次出遊中同時被安排進旅程的機率比隨機發生高。而增益值大於 1 者稱為正相關，反之則稱為負相關，因此增益值通常以 1 做為門檻值。以表 7 的第 1 筆資料為例，可看出該樣態的信賴度為 1，而 D 發生的機率為 0.75，因此該樣態的增益值為 $1/0.75=1.333$ 。而第 2 筆資料則可看出該樣態的信賴度為 0.667，而 C 發生的機率為 0.5，因此第 2 筆樣態的增益值為 $0.667/0.5=1.334$ ，以此類推。此時要注意的是若有任一樣態的增益值小於 1，則須考慮將該樣態剔除。

表 7 高頻樣態結果表

樣態結果	支持度	信賴度	增益值
若 C 則 D	0.5	1	1.333
若 D 則 C	0.5	0.667	1.334
若 C 則 E	0.5	1	1.333
若 E 則 C	0.5	0.667	1.334
若 D 則 E	0.75	1	1.333
若 E 則 D	0.75	1	1.333
若 C、D 則 E	0.5	1	1.333
若 C、E 則 D	0.5	1	1.333
若 D、E 則 C	0.5	0.667	1.334
若 C 則 D、E	0.5	1	1.333
若 D 則 C、E	0.5	0.667	1.334
若 E 則 C、D	0.5	0.667	1.334

4.5 循序樣式探勘

循序樣式探勘為關聯規則的一種延伸，用以探勘具有時間性的資料之間是否存有關聯性。和關聯規則不同之處在於，循序樣式探勘少了信賴度及增益兩個門檻值，在資料探勘時僅須考慮樣態的支持度是否有高於最小支持度。另外，循序樣式探勘所得到的結果具有時間性，也就是說若得到的高頻樣態為「若 A 則 B」，則表示先發生 A 事件的話，下一個發生的事件有高機率為 B。

圖 5 為循序樣式探勘的範例圖，左邊表格為 4 位用戶以及其被找到的停留點和時間戳

記表，而右邊表格則是在最小支持度 0.5 的設定下能找到的 3 個高頻樣態。根據第 4.4 小節對支持度的說明，可看出樣態「七星潭風景區」在 4 位用戶中有 3 位拜訪過，因此其支持度為 $3/4=0.75$ ，大於設定的最小支持度，也就是說七星潭風景區在此範例中為一高頻樣態；另一方面，樣態「花蓮糖廠」在 4 名用戶中僅有 1 名用戶曾拜訪過，其支持度為 $1/4=0.25$ ，並沒有超過設定的最小支持度，因此花蓮糖廠並不算是一高頻樣態。本範例能找 3 個高頻樣態，如圖 5 右邊的表格所示，其中第 3 個高頻樣態即說明先拜訪七星潭風景區的人，有極大的機率會再去拜訪煙波大飯店，而由於本研究之目標為找尋頻繁被拜訪的興趣點組合，因此僅含有一個興趣點的高頻樣態並不會進入到後續比對臺灣好行路線的環節。

用戶	時間戳記	興趣點名稱
1	1	七星潭風景區
1	2	太魯閣國家公園
1	3	煙波大飯店
2	1	七星潭風景區
2	2	鯉魚潭
3	1	花蓮糖廠
3	2	花蓮火車站
3	3	煙波大飯店
4	1	七星潭風景區
4	2	煙波大飯店

高頻樣態	支持度
七星潭風景區	0.75
煙波大飯店	0.75
{七星潭風景區、煙波大飯店}	0.5

圖 5 循序樣式探勘範例圖

4.6 比對臺灣好行之服務路線

透過第 4.4 以及第 4.5 小節對資料做探勘所得到的結果，可將其與臺灣好行於花蓮地區之兩條路線所行經的站點進行比對，即可得知是否有地方存有服務缺口。以關聯規則的結果為例，假設某一樣態的先導規則與後繼規則都能被臺灣好行所服務到，則該規則將被視為是可被公共運輸服務到之樣態；反之，若先導規則或後繼規則中有任一規則無法完全被臺灣好行的服務路線所滿足，則該規則將被視為可能存有服務缺口之樣態。

圖 6 為循序樣式探勘結果應用於臺灣好行服務建議之範例。須注意由於循序樣式探勘具有時間性，因此並不能直觀的像關聯規則的結果直接比對是否有沒有被服務到之規則，還得注意樣態的順序與路線行經站點的順序是否一樣。圖 6 上方表格為假定臺灣好行行經的 4 個站點，分別為站點 A、站點 B、站點 C 以及站點 D，且該路線為單方向運行，而下方表格則為循序樣式探勘之結果樣態。可看出第 2 及第 3 個結果樣態皆能被服務到，而第

1 個樣態則可建議交通當局於站點 A 和站點 B 之間新設一個站點 E，最後，即便第 4 個樣態的兩個元素皆能被臺灣好行路線服務到，但由於該樣態的順序與路線行經的順序不同，因此第 4 個樣態也得算是比對失敗，即為服務缺口。

現有公共運輸路線	A->B->C->D
----------	------------

高頻樣態	支持度
A->E->B	0.85
A->C	0.8
B->C	0.6
C->B	0.6

圖 6 循序樣式探勘應用於臺灣好行路線建議範例圖

五、結果分析與討論

本節說明套用第四章所述之研究方法於本研究所取得之 3 個周末的信令資料之探勘結果，分為兩個部分：潛在公共運輸觀光需求用戶分析結果，以及該結果與臺灣好行路線比對後之服務改善建議。由於本研究之目的為探索觀光用戶於觀光景點之間移動的行為，因此在高頻樣態結果的呈現上將不會顯示出含有非觀光景點之興趣點的樣態，也就是不包含飯店以及臺鐵火車站的興趣點。

5.1 潛在公共運輸需求觀光用戶分析結果

表 8-1 ~ 3 為信令資料探勘所得到的潛在公共運輸需求觀光用戶數，以及關聯規則和循序樣式探勘結果之樣態數，容忍距離分別設定為 100 公尺、200 公尺及 500 公尺。本研究分別探討 3 個不同興趣點停留時間門檻值：1 分鐘、2 分鐘以及 5 分鐘的資料探勘結果，停留時間乃依旅客的觀光行為而設定。本研究起初設定的停留時間為 20 分鐘，而結果不盡理想，幾乎比對不到用戶在某個景點有成功停留的情況，本研究認為此情況可能是和觀光客的旅遊行為有關。現階段在各風景區大部分皆有提供 WiFi 服務，也因此遊客已習慣於抵達景點時便開始尋找是否有 WiFi，一旦旅客連上 WiFi，便會暫停使用本身的 4G 網路連線，故不會有信令資料記錄產生。另遊客在景點遊玩時一開始不會頻繁使用網路服務，多為景點觀賞、拍照或參與該景點之活動，此時手機並不會和基地台產生資料傳遞，只有到離開前的空檔才會開始上網（例如打卡），故在景點停留期間遊客手機和 4G 基地台的資料交換頻率並不會太高，也因此難以獲得大量連續的信令資料。考慮到上述原因，本研究

便從 1 分鐘開始設定停留時間門檻，而當停留時間門檻達 5 分鐘時，能判定的停留景點數已大幅下降，因此便不再將停留時間門檻值上調。

表 8-1 潛在公共運輸需求觀光用戶數與資料探勘結果樣態數 (100 公尺)

容忍距離	100					
興趣點停留時間 (min)	1		2		5	
車站停留時間 (min)	1					
潛在公共運輸需求用戶數	26,572 (42%)		27,468 (43.44%)		27,969 (44.23%)	
支持度	0.005	0.01	0.005	0.01	0.005	0.01
信賴度	0.2					
高頻樣態數 (關聯規則)	61	28	42	21	29	12
高頻樣態數 (循序樣式探勘)	47	13	27	8	18	6

表 8-2 潛在公共運輸需求觀光用戶數與資料探勘結果樣態數 (200 公尺)

容忍距離	500					
興趣點停留時間 (min)	1		2		5	
車站停留時間 (min)	1					
潛在公共運輸需求用戶數	32,488 (52.66%)		35,065 (56.84%)		36,217 (58.70%)	
支持度	0.005	0.01	0.005	0.01	0.005	0.01
信賴度	0.2					
高頻樣態數 (關聯規則)	162	67	109	51	74	36
高頻樣態數 (循序樣式探勘)	92	33	69	26	47	15

表 8-3 潛在公共運輸需求觀光用戶數與資料探勘結果樣態數 (500 公尺)

容忍距離	500					
興趣點停留時間 (min)	1		2		5	
車站停留時間 (min)	1					
潛在公共運輸需求用戶數	43,669 (69.06%)		48,362 (76.48%)		49,799 (78.76%)	
支持度	0.005	0.01	0.005	0.01	0.005	0.01
信賴度	0.2					
高頻樣態數 (關聯規則)	4594	1380	1367	460	441	186
高頻樣態數 (循序樣式探勘)	749	267	376	133	208	76

此外，考慮到一般民眾在臺鐵車站以及觀光景點的停留時間會有落差，因此額外替臺鐵車站設定其停留時間門檻值為 1 分鐘，也就是說只要用戶於臺鐵車站有停留 1 分鐘的情況發生，即將該用戶判定為在該車站有停留行為。參考文獻中常用之設定值，最小支持度的設定上有 0.005 以及 0.01 兩種，而最小信賴度則設定為 0.2。在 100 公尺的情況下，可發現得到的比對結果與嚴格比對的結果相差不遠，即在嚴格比對結果中顯示「NA」的結果並無明顯改善的情況。而在 500 公尺的容忍距離下，會產生信令資料在鄰近的兩個景點交互出現的情況，本研究認為是由於容忍距離放鬆過多，導致部分「漂移」的資料點反而能成功比對到附近景點。而在 200 公尺的容忍距離下，無論是景點成功比對的情況，或是景點連續出現的順序等皆較符合正常的旅次行為，因此選定 200 公尺作為容忍距離的放鬆長度。

電信業者所提供的信令資料經過判定共有 61,696 名觀光用戶（居住地不在花蓮地區），在 3 種興趣點停留時間門檻值之下，分別能找到 32,488 名用戶、35,065 名用戶以及 36,217 名用戶為潛在公共運輸需求觀光用戶，分別占全部觀光用戶的 52.66%、56.84% 以及 58.7%。

5.1.1 關聯規則結果

從表 8-2 可看出在興趣點停留時間門檻值 1 分鐘、最小支持度 0.005 的情況下，有 162 個高頻樣態被找出，而興趣點停留時間門檻值 1 分鐘、最小支持度 0.01 的情況下能找到 67 個高頻樣態。在興趣點停留時間門檻值為 2 分鐘的情況下，最小支持度 0.005 以及 0.01 則分別能找到 109 個以及 51 個高頻樣態。在興趣點停留時間門檻值為 5 分鐘、最小支持度 0.005 和 0.01 的情況下，分別能找到 74 個以及 36 個高頻樣態。

本研究利用 R 語言完成關聯規則探勘，R 語言除可進行關聯規則探勘分析外，更提供將探勘結果視覺化的功能，如：矩陣圖、網路分析圖等。關聯規則結果的矩陣圖以圖 7 為範例，其中圖 7 為興趣點停留時間門檻值 2 分鐘，且最小支持度為 0.01 的關聯規則 LHS/RHS 矩陣圖。矩陣圖的橫軸（X 軸）為先導規則而縱軸（Y 軸）則為後繼規則，圖中的各個圓圈表示各個結果樣態，而圓圈的大小隨該樣態的支持度越高而越大，圓圈的顏色則隨則樣態的增益值越高而顏色越深，因此可看出圖 7 是依據增益值排序，且增益值大小由左上到右下遞減。以砂卡噹步道為例（RHS），在前往砂卡噹步道之前，旅客會先到花蓮文化創意園區與七星潭（LHS 1），及先到七星潭風景區與太魯閣國家公園（LHS 4），這兩個規則具有較高的增益值。

關聯規則網路分析圖如圖 8 範例所示，其中圖 8 同樣為興趣點停留時間門檻值 2 分鐘、最小支持度 0.01 的網路分析圖。圖中一個圓圈代表著一個樣態結果，而每個圓圈會被兩個景點所連結著，指向圓圈的景點為先導規則，而從圓圈指向的景點則為後繼規則。與矩陣圖同理，圓圈的大小代表該樣態的支持度大小，圓圈顏色越深者表示樣態增益值越高。以東大門國際觀光夜市為例，由於沒有景點指向東大門國際觀光夜市，即說明東大門國際觀光夜市是先導規則，並無法探勘出旅客在前往東大門夜市前會先探訪的景點。同時，由於

並沒有從文化創意產業園區指向東大門夜市的箭頭，因此即便有用戶的旅運行為是先到文化創意產業園區再到東大門夜市，但「若拜訪文化創意產業園區則會拜訪東大門夜市」規則並不存在；而圖中存在東大門夜市往文化創意產業園區的箭頭，代表「若拜訪東大門夜市則會拜訪文化創意產業園區」的規則被探勘尋得。

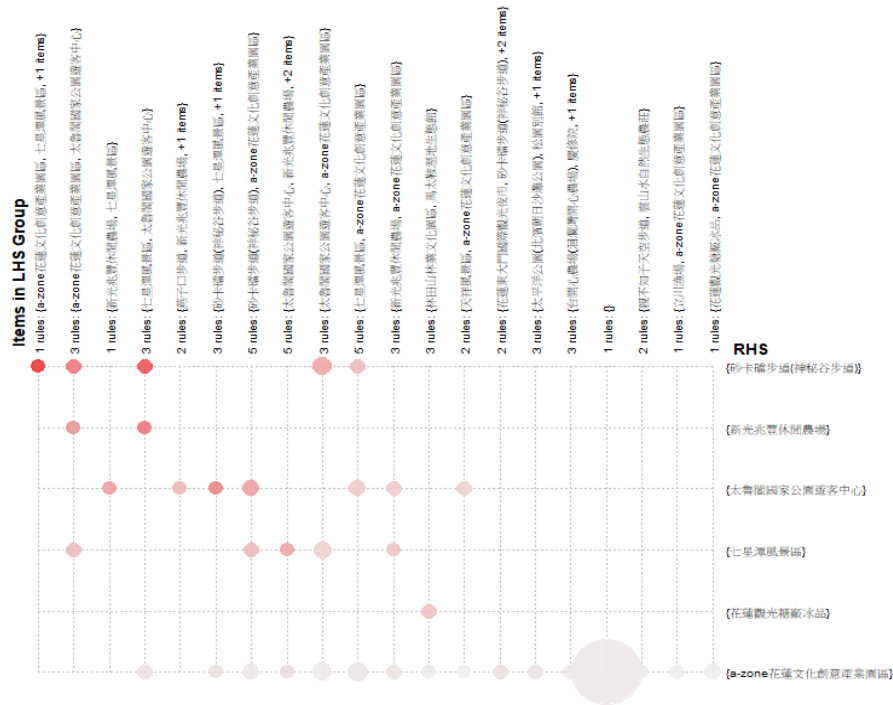


圖 7 關聯規則 LHS/RHS 矩陣圖

5.1.2 循序樣式探勘結果

根據表 8-2 可看出在興趣點停留時間門檻值 1 分鐘、最小支持度 0.005 的情況下，有 92 個高頻樣態被找出，而興趣點停留時間門檻值 1 分鐘、最小支持度 0.01 的情況下能找到 33 個高頻樣態。在興趣點停留時間門檻值為 2 分鐘的情況下，最小支持度 0.005 以及 0.01 則分別能找到 69 個以及 26 個高頻樣態。在興趣點停留時間門檻值為 5 分鐘、最小支持度 0.005 和 0.01 的情況下，分別能找到 47 個以及 15 個高頻樣態。

而循序樣式探勘不似關聯規則的結果有多種視覺化的圖表能呈現，僅能以表格的方式呈現最後的結果。表 9 為興趣點停留時間門檻值 2 分鐘、最小支持度 0.01 的循序樣式探勘結果表之部分結果，每一列包含兩個欄位，而每一列即為一個高頻樣態結果，其中第 1 欄為該樣態之先發生的元素（先拜訪之景點），而第 2 欄則為同樣樣態中後發生的元素。以第 1 列為例，表示遊客拜訪七星潭風景區後會接著拜訪新光兆豐休閒農場。

5.1.3 與臺灣好行路線比對結果

臺灣好行為交通部觀光局針對臺灣地區所設計之觀光旅遊公車服務，目前在花蓮地區有兩條臺灣好行的路線在運行，分別為縱谷花蓮線以及太魯閣線。在縱谷花蓮線部分，去程自臺鐵花蓮站發車，行經慶修院、鯉魚潭潭北遊客中心、潭南遊憩區、立川魚場、豐華再現館、新光兆豐休閒農場、臺鐵鳳林站、林田山林業文化園區、臺鐵光復站、花蓮觀光

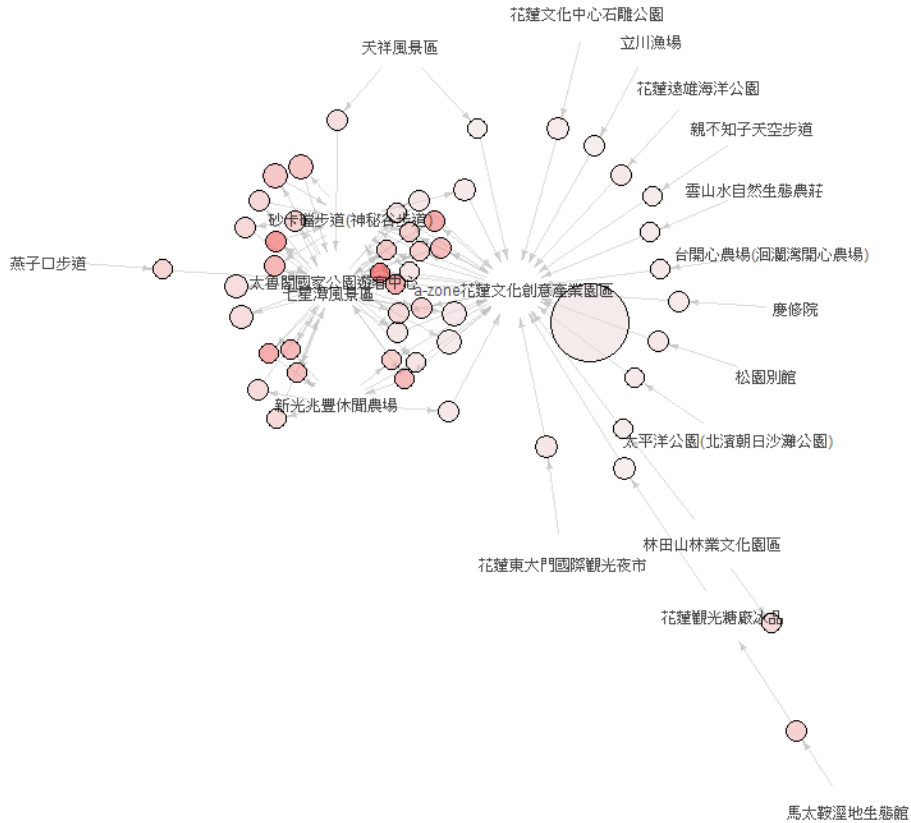


圖 8 關聯規則網路分析圖

表 9 循序樣式探勘結果表

景點 1	景點 2
七星潭風景區	新光兆豐休閒農場
太魯閣國家公園遊客中心	新光兆豐休閒農場
七星潭風景區	砂卡礑步道(神秘谷步道)
太魯閣國家公園遊客中心	砂卡礑步道(神秘谷步道)

糖廠以及大農大富平地森林園區，且此路線的去回程所服務之景點相同，僅行駛方向相反。另一方面，在太魯閣線的部分，去程亦自臺鐵花蓮站發車，行經七星潭、臺鐵新城站、太魯閣、太魯閣遊客中心、砂卡礑、布洛灣、燕子口、綠水以及天祥，需注意除砂卡礑之外，其餘景點皆在回程時有經過。

本研究將 5.1.1 與 5.1.2 所得到的關聯規則結果與循序樣式探勘結果分別與兩條臺灣好行路線所經過的站點做比對，並檢視遊客的旅運需求是否都能被滿足。在關聯規則結果的部分，LHS 與 RHS 中所包含的所有景點皆必須被臺灣好行的這兩條路線服務到才算比對成功，假使某樣態的 LHS 有 3 個景點，當中兩個景點能被臺灣好行路線服務，一個景點在服務範圍外，則該樣態依然屬於比對失敗。在循序樣式探勘結果的部分，與關聯規則結果比對方式同理，須樣態中包含的全部景點都能被路線服務到，同時由於此結果具有時間性，因此結果樣態中的元素順序須與路線行經站點的順序一致，該樣態才算是比對成功。

表 10 為關聯規則結果樣態與臺灣好行路線比對之結果表。最上方一列以區分每個停留時間門檻值與最小支持度的組合，第 2 列為各個狀況下能被服務到的樣態數，而第 3 列為各個狀況下總共得到的樣態結果數量，最後一列則為能被服務到的樣態數佔全部樣態數中的比例。

表 11 為循序樣式探勘結果樣態與臺灣好行路線比對之結果。和表 9 的呈現方式相同，以第 1 列區分各個停留時間門檻值及最小支持度的情況，再以第 2、3 以及第 4 列說明各個情況下能被服務到的樣態數、總樣態數以及能被服務到的樣態數佔全部樣態的比例。

表 10 關聯規則結果與臺灣好行路線比對結果

停留時間 (分鐘)	1	2	5	1	2	5
支持度	0.005			0.01		
有服務到的規則筆數	56	24	41	21	25	15
規則總筆數	162	67	109	51	74	36
有服務到的規則比例	34.6%	35.8%	37.6%	41.2%	33.7%	41.7%

表 11 循序樣式探勘結果與臺灣好行路線比對結果

停留時間 (分鐘)	1	2	5	1	2	5
支持度	0.005			0.01		
有服務到的規則筆數	30	27	20	10	9	6
規則總筆數	92	69	47	33	26	15
有服務到的規則比例	32.6%	39.1%	42.5%	30.3%	34.6%	40.0%

5.2 服務改善建議

根據 5.1 節的各個潛在公共運輸需求觀光用戶的關聯規則與循序樣式探勘結果，與臺灣好行路線進行比較之結果如表 12。表 12 為在 3 種興趣點停留時間門檻值之下，潛在公共運輸需求觀光用戶之規則結果所比對出無法被臺灣好行服務之景點，可看出隨著興趣點停留時間門檻值的增加，能找到未被臺灣好行服務到的景點則越少。

表 12 潛在公共運輸需求觀光用戶關聯規則結果中未被臺灣好行服務之興趣點

1 分鐘	2 分鐘	5 分鐘
花蓮文化創意產業園區	花蓮文化創意產業園區	花蓮文化創意產業園區
花蓮文化中心石雕公園	花蓮文化中心石雕公園	花蓮文化中心石雕公園
東大門國際觀光夜市	東大門國際觀光夜市	東大門國際觀光夜市
花蓮遠雄海洋公園	花蓮遠雄海洋公園	花蓮遠雄海洋公園
松園別館	松園別館	松園別館
親不知子天空步道	親不知子天空步道	
馬太鞍溼地生態館	馬太鞍溼地生態館	
洄瀾灣開心農場	洄瀾灣開心農場	
新社梯田	新社梯田	
太平洋公園		
九曲洞隧道		

在眾多無法被臺灣好行路線服務的景點中，花蓮文化創意產業園區是目前看來最需要被改善的一個景點。從圖 8 亦可看出，花蓮文化創意產業園區位在圖中較中間的位置，且該景點同時被多個景點指向，表示花蓮文化創意產業園區與其餘景點的關聯性較高，此外，透過計算改善後能服務到的樣態在全部樣態中的站比，也可看出花蓮文化創意產業園區是當中較需要被改善的一個景點。在探勘出文化創意產業園區為當前最需要被納入路線之景點後，本研究進一步根據其地理位置以及現有的兩條路線進行探討。文化創意產業園區位於花蓮火車站東南方約 2.5 公里處，行車時間約需 10 分鐘；而花東縱谷線駛離花蓮火車站後的第 1 站為鯉魚潭，位於花蓮火車站西南方約 18 公里處，行車時間約 30 分鐘，且考慮到從文化創意產業園區行駛至鯉魚潭約 25 分鐘來看，將花蓮文化創意產業園區加入是花東縱谷線是可行且不會與現況相差太大的方案。而太魯閣線駛離花蓮火車站後是往北行駛至七星潭，因此若將文化創意產業園區納入至太魯閣線會產生繞路的情況，但考量到文化創意產業園區與花蓮火車站距離不遠，若交通路線規畫得宜，仍可考慮將該景點加入至太魯閣線路線之中。

表 13 為將花蓮文化創意產業園區假設為能被臺灣好行服務到的景點後之潛在公共運輸需求用戶關聯規則結果與臺灣好行服務路線比對結果，而表 14 則是同樣加入花蓮文化

創意產業園區後之潛在公共運輸需求用戶循序樣式探勘結果與臺灣好行服務路線比對結果表，可看出在關聯規則結果的部分，幾乎八成以上的結果樣態都能被臺灣好行服務到，甚至與表 10 及表 11 相比較，更可以發現能被服務到的樣態數之占比有翻倍的現象。同時，花蓮文化創意產業園區距離花蓮火車站僅約 3 公里的路程，因此本研究建議可將花蓮文化創意產業園區納入臺灣好行的服務路線中。

表 13 潛在公共運輸需求用戶之關聯分析結果與臺灣好行服務路線比對結果表

停留時間 (分鐘)	1	2	5	1	2	5
支持度	0.005			0.01		
有服務到的規則筆數	124	58	91	42	64	30
規則總筆數	162	67	109	51	74	36
有服務到的規則比例	76.5%	86.6%	83.5%	82.3%	86.4%	83.3%

表 14 潛在公共運輸需求用戶之循序樣式探勘結果與臺灣好行服務路線比對結果表 (加入花蓮文化創意產業園區)

停留時間 (分鐘)	1	2	5	1	2	5
支持度	0.005			0.01		
有服務到的規則筆數	56	50	37	23	19	13
規則總筆數	92	69	47	33	26	15
有服務到的規則比例	60.9%	72.5%	78.7%	69.7%	73.1%	86.7%

六、結論與建議

本研究採用關聯規則與循序樣式探勘兩種方法，針對電信業者所提供之信令資料進行資料探勘，找出潛在公共運輸需求之觀光客在花蓮地區各主要景點間之移動需求與樣態。經由關聯規則，在興趣點停留時間門檻值 1 分鐘、最小支持度 0.005 的情況下，有 162 個高頻樣態被找出，而興趣點停留時間門檻值 1 分鐘、最小支持度 0.01 的情況下能找到 67 個高頻樣態；在興趣點停留時間門檻值為 2 分鐘的情況下，最小支持度 0.005 以及 0.01 則分別能找到 109 個以及 51 個高頻樣態；在興趣點停留時間門檻值為 5 分鐘、最小支持度 0.005 和 0.01 的情況下，分別能找到 74 個以及 36 個高頻樣態。經由循序樣式探勘，在興趣點停留時間門檻值 1 分鐘、最小支持度 0.005 的情況下，有 92 個高頻樣態被找出，而興趣點停留時間門檻值 1 分鐘、最小支持度 0.01 的情況下能找到 33 個高頻樣態；在興趣點停留時

間門檻值為 2 分鐘的情況下，最小支持度 0.005 以及 0.01 則分別能找到 69 個以及 26 個高頻樣態；在興趣點停留時間門檻值為 5 分鐘、最小支持度 0.005 和 0.01 的情況下，分別能找到 47 個以及 15 個高頻樣態。

接著與臺灣好行的路線進行比對，提出相關之服務改善建議。本研究分析多數樣態，提出於臺灣好行路線中先行增設「花蓮文化創意產業園區」站點的建議；但若以關聯規則或是循序樣式探勘的分析觀點，凡是最終被找出的高頻樣態，都應該被視為是許多遊客可能會前往的景點組合，因此在法令與預算等相關條件許可下，主管機關或許可考慮增設除了花蓮文化創意產業園區之外的景點，如表 12 所列出的各個景點皆可納入服務改善之考量。

此外，本研究由於研究經費與時程限制仍有未盡之處，茲提出以下建議供後續相關研究參考。首先在資料方面，本研究僅蒐集 107 年 5 月 19 到 20 號、5 月 26 到 27 號以及 6 月 2 號到 3 號這 3 個周末的手機信令資料，資料內容相對侷限，未能夠充分蒐集花蓮縣旅遊的旺季—夏季及冬季的完整資料；且本研究僅有單一電信業者來源，未能同時取得其他家電信業者之手機信令資料，都讓本研究在資料分析的全面性有所侷限，建議後續相關研究可和更多的電信業者保持緊密合作關係，以利取得更大量數據供後續分析。

接著，本研究建議可納入更多元的旅次目的分類，本研究僅將用戶區分為觀光用戶及當地用戶，然而，非居住於花蓮地區的用戶拜訪花蓮的目的可能不只有觀光，如：商務需求、返鄉探親等，因此若能將用戶的分類更加多元且準確，則可能可以獲得更精確的分析結果。

再者，本研究建議判別潛在公共運輸需求觀光用戶的方法能更精進，由於資料與時程限制，因此研擬出以用戶的第一個或最後一個停留點是否為火車站的方法，來判別該用戶是否為潛在公共運輸需求用戶；然而，若能取得花蓮地區以及聯外的火車軌道網格，則能更進一步比對用戶軌跡以及軌道位置，來確認用戶是否是有搭乘火車的行為，使得在判別用戶是否為潛在公共運輸需求用戶時能更精確。

最後，藉由本研究所提出之關聯規則與循序樣式勘方法，後續的研究者可應用此些方法分析不同供需特性觀光區域的公共運輸系統，進一步驗證此資料探勘方法於不同旅遊環境下之適宜性。

參考文獻

1. 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.
2. 國家發展委員會，106 年持有手機民眾數位機會調查報告，民國 106 年。
3. 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 106 年。
4. Hollecsek, T., Anh, D. T., Yin, S., Jin, Y., Antonatos, S., Goh, H. L., Low, S., and ShiNash,

- A., “Traffic Measurement and Route Recommendation System for Mass Rapid Transit (MRT)”, KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, 2015, pp. 1859-1868.
5. Demissie, M. G., Phithakkitnukoon, S., Sukhvibul, T., Francisco, A., Rui, G., and Carlos, B., “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.
 6. Ahas, R., Aasa, A., Silm, S., and Tiru, M., “Mobile Positioning Data in Tourism Studies and Monitoring: Case Study in Tartu, Estonia”, *Information and Communication Technologies in Tourism 2007*, Springer, Vienna, 2007, pp 119-128.
 7. Trasarti, R., Olteanu-Raimond, A-M., Nanni, M., Couronné, T., Furletti, B., Giannotti, F., Smoreda, Z., and Ziemlicki, C., “Discovering Urban and Country Dynamics from Mobile Phone Data with Spatial Correlation Patterns”, *Telecommunications Policy*, Vol. 39, 2015, pp. 347-362.
 8. Mamei, M. and Colonna M., “Analysis of Tourist Classification from Cellular Network Data”, *Journal of Location Based Services*, Vol. 12, No. 1, 2018, pp. 19-39.
 9. Holleczeck, T., Yu, L., Lee, J. K., Senn, O., Ratti, C., and Jaillet, P., “Detecting Weak Public Transport Connections from Cellphone and Public Transport Data”, in Proceedings of the Third ASE International Conference on Big Data Science and Computing (BigDataScience 2014), Beijing, China, 2014, pp. 1-8.
 10. Chen, Q., Hu, Z., Su, H., Tang, X., and Yu, K., “Understanding Travel Patterns of Tourists from Mobile Phone Data: A Case Study in Hainan”, 2018 IEEE International Conference on Big Data and Smart Computing, 2018, pp. 45-51.
 11. Bekhor, S., Cohen, Y., and Solomon, C., “Evaluating Long-Distance Travel Patterns in Israel by Tracking Cellular Phone Positions”, *Journal of Advanced Transportation*, Vol. 47, 2013, pp. 435-446.
 12. 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.