

國立交通大學
運輸與物流管理學系

碩士論文

手機信令資料於公共運輸潛在觀光需求之分析

The analysis of latent tourism demand of public transportation
by using mobile phone data

研究生：李晟豪

指導教授：王晉元

中華民國一〇八年七月

手機信令資料於公共運輸潛在觀光需求之分析
The Analysis of Latent Tourism Demand of Public Transportation by using
mobile phone data

研 究 生：李 晟 豪

Student：Cheng-Hao, Lee

指導教授：王 晉 元

Advisor：Jin-Yuan, Wang

國立交通大學
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 Chiao Tung University

in Partial Fulfillment of the Requirements

For the Degree of

Master

In

Traffic and Transportation

July 2019

Hsinchu, Taiwan, Republic of China

中 華 民 國 一 ○ 八 年 七 月

手機信令資料於公共運輸潛在觀光需求之分析

學生：李晟豪

指導教授：王晉元

國立交通大學運輸與物流管理學系碩士班

摘要

觀光公共運輸系統的路線規畫是否符合觀光客實際在遊玩時的旅運行為是非常重要的課題，而本研究即是在探討觀光客出遊時的旅運行為與觀光公共運輸路線是否存在差異，找出觀光公共運輸的服務缺口並提出相對應之公共運輸路線服務改善建議。本研究以觀光客出遊時的信令資料為基礎，分析出其旅運需求後再與觀光公共運輸系統路線進行供需比對以找出是否存在服務缺口，最後提出相對應之路線服務改善建議。研究結果顯示特定地區的哪些景點為觀光客常拜訪但當地觀光公共運輸系統無法服務之景點，相關單位可將分析結果做為未來在調整觀光公共運輸路線或新闢服務路線時的參考依據。

關鍵詞:信令資料、觀光旅運需求、關聯規則分析

The analysis of latent tourism demand of public transportation by using mobile phone data

Student: Cheng-Hao Lee

Advisor: Jin-Yuan Wang

Department of Transportation and Logistics Management
National Chiao Tung University

ABSTRACT

Whether the route of the sightseeing public transportation system is in line with the actual travel of the tourists during the tour is a very important issue. This study is to explore whether the travel operation of the tourist travels is different from the public transportation route for sightseeing, finding out the service gap for public transport and propose corresponding improvements to the public transport route service. Based on the phone data of tourists during their travels, this study analyzes their travel demand and then compares the supply and demand with the tourism public transport system route to find out whether there is a service gap. Finally, the corresponding route service improvement suggestions are proposed. The results of the study show that the scenic spots in a particular area are attractions that tourists often visit but cannot be served by the local tourism public transportation system. Concerned departments can use the analysis results as a reference for future adjustments to the public transportation routes or new service routes.

Keywords: phone data, tourism travel demand, Association rule analysis

致謝

兩年的研究生生活過得很快，剛升上研究所時的不安全感、研究進度落後時的焦慮感等，彷彿都還只是昨天的體驗。而如今能突破困境、順利完成本研究，首先要感謝我的指導教授王晉元博士。老師的指導風格一向都是放手讓學生去發揮，並不會對學生的想法有過多的限制，也因此每當研究有想嘗試的新方向時，老師總是給予全力的支持並且認真地和我一起對該方法進行討論。此外，也很謝謝老師對我的耐心指導，有時我對於一個問題反應不過來時，常常會花到兩到三周的時間與老師討論以找尋解決方法，而過程中老師不但沒有對我的驚鈍感到不耐煩，反而是非常有耐心且細心地將我的思緒引導至對的方向上。兩年下來，受老師的指導與照顧真的太多太多，不只是學課業上，甚至在待人處事上老師也與我談過不少，真的非常感謝老師。

謝謝口試委員盧宗成老師以及蘇昭銘老師。兩位老師在口試時提供了我許多還能再精進的部分，也因為有兩位老師的幫助，我才能更加全面地審視這篇研究。同時，兩位老師提供的來自不同角度所看到的問題，也讓這篇研究在最後能更加完整，非常感謝兩位老師的幫助。

謝謝研究室的夥伴依庭、玟姍、傳軒、曲平、崑育、筠筑、煜民、宛樺、筱彤，在我研究碰到困難時，能與我一起討論、解決難題，若沒有大家的幫助，我想我不會有如此的成果，也祝大家未來一切順利並朝著自己的目標邁進。

最後我要感謝我的家人，兩年的時間雖然不長，但過程中的繁忙導致我鮮少回到家中陪伴家人，兩年的時間便也顯得意外長久。這兩年來，家人也全力支持我完成我的研究生生涯，甚至在我沒時間回家時親自前來新竹給我鼓勵、為我加油。若沒有家人如此的陪伴及支持，我想我很有可能會支持不下去，中途放棄。如今我已完成研究，一直以來辛苦照顧我的家人們都辛苦了，謝謝你們。

在交大攻讀碩士學位的這兩年來受到非常多人的照顧及幫助，這份研究結果絕非是我一人完成的，僅感謝所有幫助過我的老師、同學、家人。

李晟豪 謹誌

2019 年7 月

目錄

摘要.....	i
ABSTRACT.....	ii
致謝.....	iii
目錄.....	iv
表目錄.....	v
圖目錄.....	vi
第一章 緒論.....	1
1.1 研究動機.....	1
1.2 研究目的.....	1
1.3 研究範圍.....	2
1.4 研究流程.....	2
第二章 文獻回顧.....	5
2.1 手機信令資料應用與用戶特性.....	5
2.2 手機信令資料與公共運輸服務.....	10
2.3 小結.....	13
第三章 分析方法建立.....	14
3.1 資料蒐集.....	15
3.1.1 手機信令資料與網格資料.....	15
3.1.2 興趣點地理資訊資料.....	17
3.2 建立觀光用戶旅運資訊.....	17
3.2.1 容忍距離比對.....	18
3.2.2 停留時間計算.....	20
3.3 篩選第一類用戶.....	21
3.4 關聯規則分析.....	23
3.4.1 關聯規則分析.....	23
3.4.2 雜訊問題.....	27
3.4.3 單點式關聯規則分析.....	28
3.5 比對台灣好行服務路線.....	30
3.6 小結.....	35
第四章 實際範例測試.....	36
4.1 興趣點停留時間門檻值測試結果.....	36
4.2 兩類別用戶篩選結果.....	37
4.3 改善樣態數之測試結果.....	39
4.4 樣態支持度測試結果(一般關聯規則).....	44
4.5 樣態支持度測試結果(單點式關聯規則).....	46
4.6 台灣好行服務建議.....	50
第五章 結論與建議.....	56
5.1 結論.....	56
5.2 建議.....	57
參考文獻	59

表目錄

表 2.1 旅次目的分類比例.....	9
表 3.1 信令資料格式.....	16
表 3.2 信令資料範例表.....	16
表 3.3 網格中心經緯度資料範例表.....	16
表 3.4 興趣點地理資訊資料格式.....	17
表 3.5 興趣點地理資訊資料範例表.....	17
表 3.6 用戶旅運資訊範例表.....	23
表 3.7 關聯規則示範資料庫.....	24
表 3.8 高頻樣態結果表.....	24
表 3.9 十二條候選規則之信賴度結果表.....	25
表 3.10 關聯規則示範結果表.....	26
表 3.11 關聯規則產生雜訊之示範資料庫.....	27
表 3.12 興趣點拜訪人數統計表.....	28
表 3.13 太魯閣風景區之子資料庫範例.....	30
表 3.14 台灣好行路線範例路線表.....	31
表 3.15 關聯規則分析結果於台灣好型服務比對之範例表.....	31
表 4.1 取得之觀光用戶信令資料日資料量表.....	36
表 4.2 興趣點停留時間門檻值與用戶平均停留興趣點數量對應表.....	37
表 4.3 兩類別用戶人數表.....	38
表 4.4 興趣點停留時間門檻值5分鐘，全體觀光用戶關聯規則結果表 (最小支持度0.01).....	40
表 4.5 全體觀光用戶關聯規則結果樣態與台灣好行路線比對結果表.....	43
表 4.6 第一類用戶關聯規則結果樣態與台灣好行路線比對結果表.....	43
表 4.7 第二類用戶關聯規則結果樣態與台灣好行路線比對結果表.....	44
表 4.8 全體觀光用戶興趣點支持度結果表.....	45
表 4.9 第一及第二類用戶興趣點支持度結果表.....	46
表 4.10 興趣點停留時間門檻值1分鐘觀光景點拜訪人數表(全體觀光用戶).....	47
表 4.11 單點式關聯規則興趣點支持度結果表(興趣點拜訪人數前50%).....	48
表 4.12 單點式關聯規則興趣點支持度結果表(興趣點拜訪人數前25%).....	49

圖目錄

圖 1-1 研究流程圖.....	4
圖 2-1 觀光客拜訪完塔圖後拜訪的前10大地點分布及人口流量圖.....	5
圖 2-2 2004年及2005年6、7月份愛沙尼亞觀光人數分布圖.....	5
圖 2-3 各類別工作地用戶間之旅運行為的漢明距離長條圖.....	6
圖 2-4 找尋停留點範例圖.....	7
圖 2-5 手機信令資料推估道路流量分布與人口普查資料推估道路流量比較圖...8	
圖 2-6 塞納-聖但尼與周圍關聯性省分示意圖.....	8
圖 2-7 ALLAboard 系統結構圖.....	10
圖 2-8 捷運路線資訊系統開發示意圖.....	11
圖 3-1 分析方法流程圖.....	14
圖 3-2 用戶旅運資料建立流程圖.....	18
圖 3-3 容忍距離比對示意圖.....	19
圖 3-4 容忍距離比對範例圖.....	20
圖 3-5 停留時間計算示意圖.....	21
圖 3-6 單點式關聯規則分析流程圖.....	29
圖 3-7 樣態支持度統計範例圖.....	33
圖 4-1 興趣點停留時間門檻值提升與第一類用戶數提升示意圖.....	39
圖 4-2 台灣好行路線圖.....	42
圖 4-3 花蓮火車站、慶修院、花蓮文化創意產業園區以及七星潭風景區位 置分佈圖.....	50
圖 4-4 花蓮火車站、慶修院、花蓮文化創意產業園區旅行時間示意圖.....	52
圖 4-5 花蓮觀光糖廠、馬太鞍濕地生態館、大農大富平地森林園區以及瑞 穗牧場位置分佈圖.....	53
圖 4-6 花蓮觀光糖廠、馬太鞍濕地生態館以及大農大富平地森林園區旅行 時間示意圖.....	54
圖 4-7 大農大富平地森林園區至瑞穗牧場之旅行時間示意圖.....	55

第一章 緒論

1.1 研究動機

觀光公共運輸系統的路線規劃是否符合民眾的實際需求常是交通當局所關心的一個議題。若路線設計與民眾的出遊時的實際需求有明顯落差，將會使民眾使用觀光公共運輸系統的意願降低，可能造成資源的浪費，更可能由於私人運具的大量使用而衍生出觀光勝地周邊道路堵塞、停車位不足等問題。反之，若路線設計能貼合民眾需求，則可以吸引更多的民眾於遊玩時使用公共運輸系統來移動，進而使私人運具使用減少而讓道路更加通暢，甚至能吸引更多無駕駛私人運具能力的遊客前往觀光。因此，規劃完善且路線符合民眾實際需求的觀光公共運輸系統是交通當局一個重要的課題。

為了得知民眾於景點之間的移動情況，可以透過調查民眾於出遊時的旅運行為來進行分析。傳統的旅運行為調查有日記記錄(paper travel diary)、問卷調查或電話訪問等方法。然而，傳統的旅運調查行為不僅在人力、時間或金錢上的成本消耗龐大，甚至會有受訪者隱瞞行蹤、忘記記錄導致資料有缺失的情況發生。Zahedi and Shafahi(2017)也提到傳統的交通資料調查方式存有費時、不便利、問卷回收率低、使受訪者感到不舒服、調查結果受到受訪者教育程度影響大…等問題。

隨著科技進步，手機的普及率大幅提升，代表著行動通訊的資料量也有所增加。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，截至 2017 年 7 月，台灣平均每人約擁有 1.2 個手機門號，當中 4G 用戶占了 73%，而使用行動上網的用戶也從 2011 年的 587 萬戶攀升至 2017 年 7 月底的 2270 萬戶，行動上網率達 79%。此成長現象代表手機上網的服務在台灣民眾的生活佔了不可或缺的一個地位，也表示將帶來極為可觀的行動通訊資料。

手機信令資料為民眾使用手機時與手機基地站互動時所留下之時空紀錄，並可作為一種分析旅運行為的資料來源。每當手機使用者從事 4G 網路活動，或是移動時導致手機與電信基地站的連結有所改變等行為時，電信基地站便會記錄使用者發生這些行為的時間及地點。此時，若能蒐集使用者一天完整的信令資料，則可以知道該使用者在當天的移動軌跡，進而得知其一天的旅次鏈，並達到蒐集旅運行為資料的目的。

因此，本研究主要以觀光客之手機信令資料為基礎，先建立各別用戶的觀光旅運行為後，透過資料科學的技術找出景點之間的關聯性。接著，比對現有的公共運輸供給來找出是否存有未被滿足的觀光旅運需求存在。最後，根據比對結果及產生各個服務缺口的景點，分析其所在位置以及可及性等因素，提出相對應的觀光公共運輸路線改善建議。

1.2 研究目的

本研究主要目的為對取得的觀光用戶信令資料進行旅運行為分析，比對觀光客的

實際旅運需求與觀光公共運輸系統路線是否有差異，再根據比對結果檢視觀光公共運輸系統是否有服務缺口，並提出相對應之改善建議。為了完成上述目的，本研究提出「單點式關聯規則」以及兩種解讀資料探勘結果之方法，關於單點式關聯規則以及解讀資料探勘結果之方法會在後續章節進行詳細說明。

此外，為了使分析結果有更全面的解讀，本研究進一步將全部的觀光用戶再細分為旅遊過程中是以公共運具抵達或離開觀光地的觀光客(以下簡稱為第一類用戶)以及使用私人運具往返觀光地的觀光客(以下簡稱為第二類用戶)，並分別進行旅運行為分析以及與觀光公共運輸系統比對。第一類用戶的研究結果可用以檢視有使用公共運輸的觀光族群其旅運行為與觀光公共運輸路線是否有差異；第二類用戶的比對結果則是檢視較無意願使用公共運輸的觀光族群其旅運行為是否能被現有觀光公共運輸系統滿足。透過將全部的觀光用戶再細分成兩個類別，即可觀察使用不同運具的用戶之間其旅運行為是否有差異，更可將兩類別用戶與觀光公共運輸的比對結果作為本研究最後提出服務改善建議之依據。

1.3 研究範圍

本研究所使用之信令資料為電信公司提供的手機信令資料。電信公司所提供之信令資料包含使用者識別碼、時間與日期以及使用者所在位置，三個欄位的資料。此外，蒐集資料的用戶會選擇其旅途至多為三天，且期間包含周末二日之用戶。關於信令資料的資料格式等資訊會在後續的章節進行詳細說明。

本研究的公共運輸對象為花蓮縣境內營運的兩條台灣好行路線，分別為縱谷花蓮線以及太魯閣線。台灣好行為交通部觀光局於2010年推出的專為旅遊規劃設計之接駁公車服務。服務內容主要是以台灣各大景點附近的各大台鐵、高鐵站為出發點，接送旅客至附近的觀光景點旅遊、觀光。

本研究的研究場域為單一縣市，不考慮跨縣市的觀光旅運行為。由於台灣好行是交通部輔導各地方政府來推動的於各地台、高鐵車站與各大景點間的接駁服務，因此各地的台灣好行服務皆沒有跨縣市的路線，即各縣市的台灣好行路線僅會服務該縣市內的景點與車站間的接駁，故本研究亦將研究場域設定為單一縣市。

1.4 研究流程

本研究包含以下4個流程，如圖1-1所示，各流程詳細說明如下：

(1)文獻回顧

此步驟為收集及回顧過往之相關研究，作為本研究之參考依據。本研究所回顧之文獻分為以下兩個方向，首先是國內外利用手機信令資料應用於探討用戶特性或用戶移動行為的相關文獻，用以檢視本研究之研究方向。另一方面則是回顧手機信令資料應用於公共運輸服務之相關文獻，以探討本研究最後用於改善公共運輸服務路線之可

行性。

(2) 分析方法建立

根據研究目的，建立一套分析資料的方法。其中包括資料處理、篩選出可能會使用公共運輸的潛在用戶、探索頻繁被拜訪的觀光景點組合以及與現有公共運輸系統進行供需比對等步驟。

(3) 實際範例測試

本研究選擇花蓮縣作為實際範例測試的場域。花蓮縣位於台灣東部，具豐富的生態資源、壯麗的地形景觀以及多元的人文特色，每年皆吸引大量的觀光客前往。同時，根據行政院交通部觀光局所發布的 2018 年國內主要觀光遊憩據點遊客人數月別統計資料顯示，花蓮縣的景點其拜訪人口皆能在各個類別名列前茅，舉例來說，台八線沿線景觀區(如太魯閣、燕子口等)在 2018 年時共有 3104277 人前往拜訪，在「國家公園」的分類裡共 36 個景點中排名第一；而鯉魚潭風景區也在「國家級風景特定區」的分類裡共 76 個景點中排名第 6，於 2018 年共有 1103360 人前往拜訪。故本研究認為選擇以花蓮作為研究的場域是非常合適的。

在此步驟本研究使用電信公司所提供之花蓮縣的觀光用戶信令資料，以及所蒐集的花蓮興趣點資訊，套用進本研究發展的資料分析方法。在比對台灣好行路線時也將以運行在花蓮地區的兩條台灣好行路線為主要分析對象，分別為縱谷花蓮線以及太魯閣線。縱谷花蓮線從花蓮台鐵站出發，行經慶修院、鯉魚潭、立川漁場、雲山水、新光兆豐休閒農場、台鐵鳳林站、林田山林業文化園區、台鐵光復站、花蓮觀光糖廠以及大農大富平地森林園區，回程時行經的站點相同；太魯閣線從花蓮台鐵站出發，行經七星潭、台鐵新城站、太魯閣、砂卡礑、布洛灣、燕子口、綠水以及天祥，回程時不行經布洛灣，取而代之的是長春祠步道，其餘站點皆與去程相同。最後，再根據比對的結果探討此兩條路線是否有可以改善的地方。

(4) 結論與建議

根據實際範例測試的結果以及研究過程中所發現之問題進行探討。綜合以上各個步驟在執行時所遇到的困難提出解決辦法供後續相關研究參考，如：資料蒐集不易、資料缺失、研究方法有可精進之處…等。除此之外，也可以對後續之相關研究提供建議。

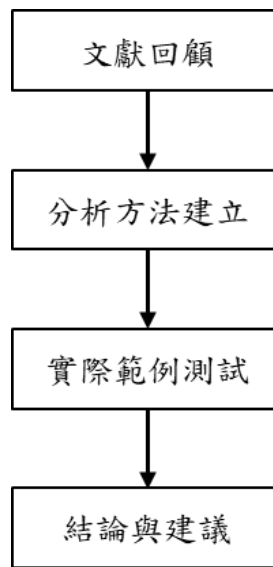


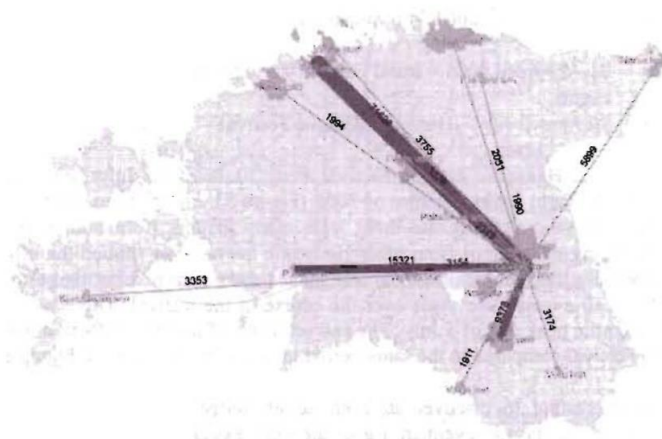
圖 1-1 研究流程圖

第二章 文獻回顧

本研究之目的為透過手機信令資料及興趣點地理資訊探討觀光客的旅運行為，並檢視公共運輸於觀光景點的服務是否有可以改善的地方。因此，本章節將對於應用手機信令資料於探討用戶特性或用戶移動行為的文獻進行回顧，也會回顧應用手機信令資料於公共運輸服務相關之文獻，作為本研究的參考依據。

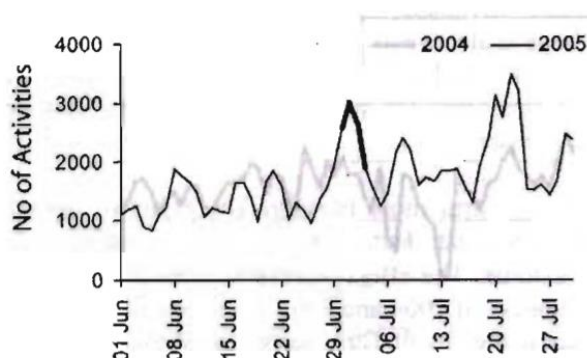
2.1 手機信令資料應用與用戶特性

Ahas 等人(2007)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。該研究先利用取得的信令資料中「國籍」的欄位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、地理位置分布以及境內舉辦大型活動時，參與的觀光客數量。圖 2-1 顯示愛沙尼亞境內的觀光客在拜訪塔圖後會再拜訪的地點以及前往該地點的人數，並取前 10 大熱門的地點顯示。而圖 2-2 則是顯示 2004 年及 2005 年 6、7 月份拜訪愛沙尼亞的觀光客人數分布圖，可看出於 6 月 30 號到 7 月 1 號時 2005 年的拜訪人數明顯高於 2004 年的，而會有此現象原因為 2005 年的 6 月 30 號到 7 月 1 號當地舉辦了漢薩同盟節的活動 (Hanseatic League Festival)。



資料來源: Ahas et al. (2007)

圖 2-1 觀光客拜訪完塔圖後拜訪的前 10 大地點分布及人口流量圖

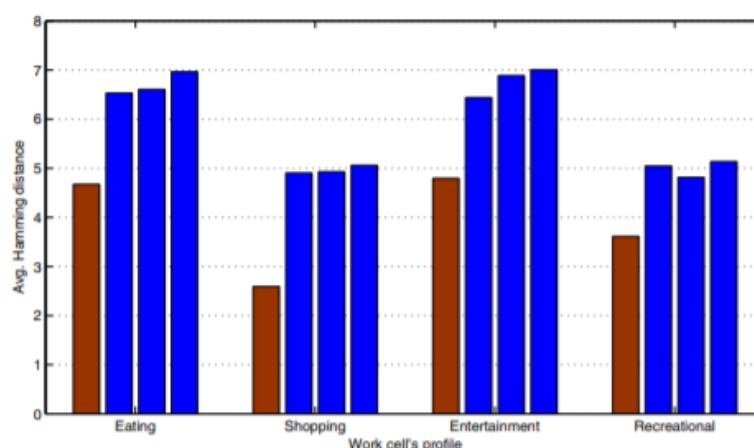


資料來源: Ahas et al. (2007)

圖 2-2 2004 年及 2005 年 6、7 月份愛沙尼亞觀光人數分布圖

Gur 等人(2009)利用手機信令資料推估以色列民眾的旅次起迄點以及各地區之間的旅運量，並利用該結果進行交通量預測，而最終的預測結果可供交通當局作為公共運輸路線規劃的參考依據。此研究以 20 分鐘為門檻，判斷用戶是否有停留行為的發生並進行旅次的切割，接著判斷各個用戶一周內停留時間最長的交通分析區作為其居住地理位置，並以每個用戶的每趟旅次是否從居住地出發或返回進行旅次的分類。隨後，此研究將各種類的旅次統整，配合大眾運輸系統資料、路網資料等，並利用回歸模型進行旅次量的預測及其分佈。此研究之模型可在地區之間的人口數、政策或是交通路網發生改變時，預測改變後的情況其交通需求會有何改變。

Phithakkitnukoon 等人(2010)透過手機信令資料探討民眾的旅運行為，並在將其分群後，分析各群的用戶之間的旅運行為是否有差異，而此研究的研究場域為波士頓都會區。此研究首先設定距離門檻值與時間門檻值，並透過此兩門檻值來判斷個別用戶一天的旅次鍊及停留地點，隨後即可根據各個用戶於白天時最常出現的地點來判斷其工作地。接著將研究場域切割成數個 500 乘 500 公尺大小的網格，並透過 Python 套件以及地理資訊系統將每個網格根據其範圍內各個種類的興趣點所佔有的比例予以分類。舉例來說，某一網格內大部分店家為餐廳、咖啡店或麵包店，則該網格即分類為飲食的類別，而此研究一共有四個類別，分別為飲食、購物、娛樂以及休閒活動。最後，透過比對用戶一天內的各個停留地點以及該地點是在哪個分類的網格裡，即可得知用戶一天的旅次鍊是由哪些活動所組成的。此外，此研究亦根據用戶的工作地所在網格，分析各類別工作地的用戶其一天的旅運行為是否存有差異，此研究透過漢明距離(Hamming distance)來計算各類別間的差異性大小，而計算結果如圖 2-3 所示。圖 2-3 的橫軸為四個分類的工作地，縱軸為漢明距離，長條圖顏色為咖啡色的表示是組內距離，而藍色的為組間距離，可看出四個類別的組內距離皆小於組間距離，說明工作地在相同類別網格的用戶其旅運行為越相近。



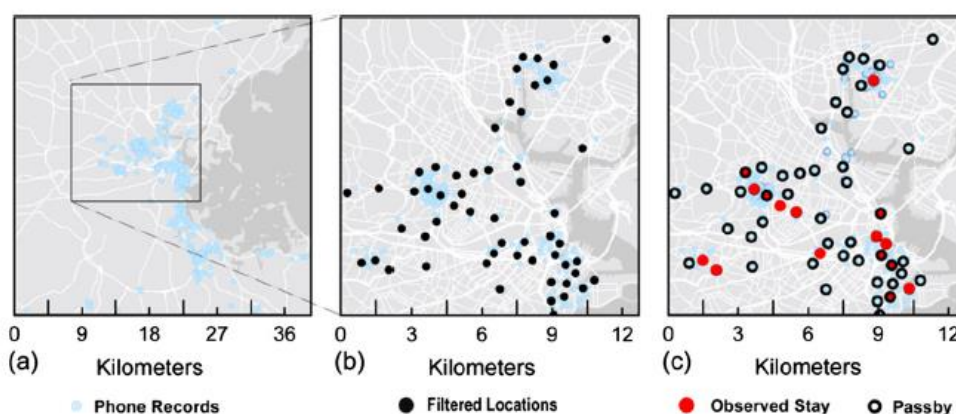
資料來源: Phithakkitnukoon et al. (2010)

圖 2-3 各類別工作地用戶間之旅運行為的漢明距離長條圖

Nanni 等人(2013)透過手機信令資料找出用戶的重現性移動行為，即通勤需求，並以此建立所有用戶的旅運起迄矩陣，最後再透過模擬軟體進行交通量指派，藉以了解象牙海岸和阿必尚(象牙海岸的最大都市)的交通需求狀況。此研究首先欲找出用戶的居住地，方法為檢視用戶一天之中最常被記錄到的位置，但考量到信令資料會有飄

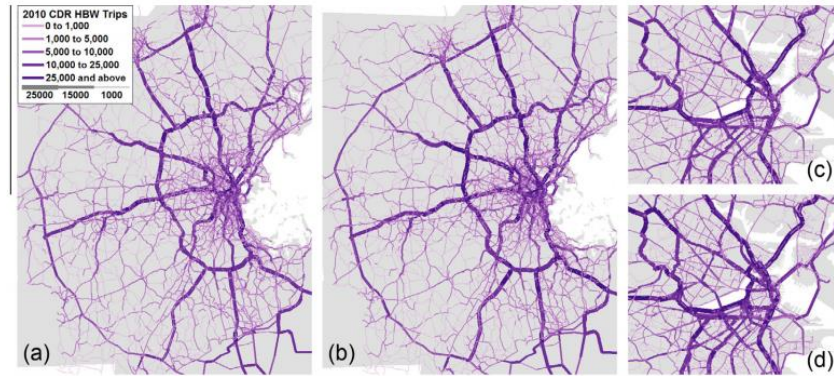
移的現象發生，因此用戶的居住地為一天之中最常被記錄到的位置及其附近的區域。接著，找尋居住區域以外，用戶最常出現的地點，並將其視為用戶的工作地或是主要活動區域。隨後，根據用戶每天於居住區域及工作地之間的移動行為進行統計(即通勤旅次)，並根據所有用戶數替每個起迄對進行標準化的動作，使每個起迄對的旅運量得以推估得更加準確。最後，根據上述得到的統計結果以建立象牙海岸和阿必尚的旅運起迄矩陣，並配合模擬軟體進行道路的交通量指派，即可了解此兩個地區的交通需求狀況。

Alexander 等人(2015)利用手機信令資料道路資訊，發展出一套相較於傳統旅運行為調查更為彈性且有效率的調查旅次起迄點之方法，而該研究的研究場域為波士頓都會區。該研究首先透過集群分析將信令資料根據空間的相似性進行合併，再根據停留時間是否超過研究人員設計的停留時間門檻來判斷各資料點為停留點(Stay area)或是經過點(Passby area)，範例如圖 2-4 所示。圖 2-4(a)表示此研究所界定之研究場域大小，圖 2-4(b)則是將該範圍內收集到的信令資料透過集群分析後之資料分佈結果，而圖 2-4(c)則是將圖 2-4(b)的各點根據用戶的停留時間來判斷是屬於停留點或是經過點，其中紅色的點為停留點，而空心的點則是經過點。接著再根據各停留點被記錄的時間及位置判斷其為居住地、工作地或其他旅次目的地。最後再與美國各機關之傳統旅運調查資料進行比對，得知研究結果與傳統調查結果相似，且研究結果與傳統調查結果之相關係數也趨近於 1，比對結果如圖 2-5 所示，其中圖 2-5(a)和圖 2-5(c)為手機信令資料分別在波士頓大都會區及市中心之道路流量分布圖，而圖 2-5(b)和圖 2-5(d)則為美國聯邦公路局於 2013 年的運輸規劃人口普查調查結果，可看出圖 2-5(a)與圖 2-5(b)在圖像的呈現上極為相似，而圖 2-5(c)與圖 2-5(d)亦同，因此可說明透過分析手機信令資料已可達到傳統旅運調查之效果。



資料來源: Alexander et al. (2015)

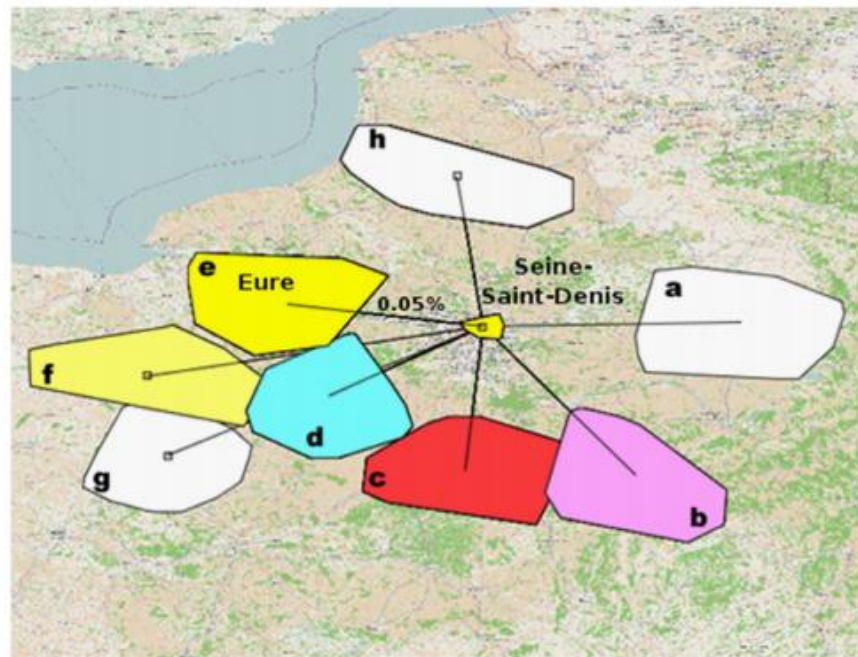
圖 2-4 找尋停留點範例圖



資料來源: Alexander et al. (2015)

圖 2-5 手機信令資料推估道路流量分布與人口普查資料推估道路流量比較圖

Trasarti 等人 (2015) 以用戶信令資料為基礎，透過循序樣式探勘的方法，找出一地區中具有高度關聯性的地點。此研究根據每次實驗區域規模的不同，會設計不同的門檻值來找出彼此之間具關連性的地點。接著再根據研究區域範圍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。研究場域為法國，而每次實驗的區域範圍有大規模的省、城市以及小規模的地區(如法國戴高樂機場與鄰近地區之關聯)。圖 2-6 是以法國的塞納-聖但尼省為例，透過此研究方法找出之與其具有相關性的地區。



資料來源: Trasarti et al. (2015)

圖 2-6 塞納-聖但尼與周圍關聯性省分示意圖

Zahedi and Shafahi (2017)以手機時空資訊資料為基礎，發展一套推估使用者行為的模型。有鑑於保護使用者的隱私，該研究所使用的位置資訊是與使用者手機連接的基地站的位置，而非使用者實際的經緯度位置。接著以 20 分鐘為門檻值找出各個用戶在一天之內的所有停留地點，再判斷各個停留點何者為居住地以及工作地，最後再

判斷其餘停留點的旅次目的為何。表 2.1 為西拉當地政府於 2000 年做的旅次目的調查資料和手機信令資料結果比較表，其中 W 為工作旅次、E 為就學旅次、P 為接送旅次、C 為商務旅次、R 為聚會或休閒活動旅次、A 為行政作業旅次、H 則為家旅次。可從表中看出各個旅次目的之比例在兩種調查方法下大致相同，因此可說明手機時空資訊(手機信令資料)可用於判斷旅次目的。

表 2.1 旅次目的分類比例

旅次目的	W	E	P	C	R	S	A	H
手機信令資料	13.35	13.5	4.45	9.5	6.45	3.98	3.85	42.92
傳統調查資料	19.76	13.5	2.59	7.3	3.12	2.8	2.29	48.03

單位：百分比(%)

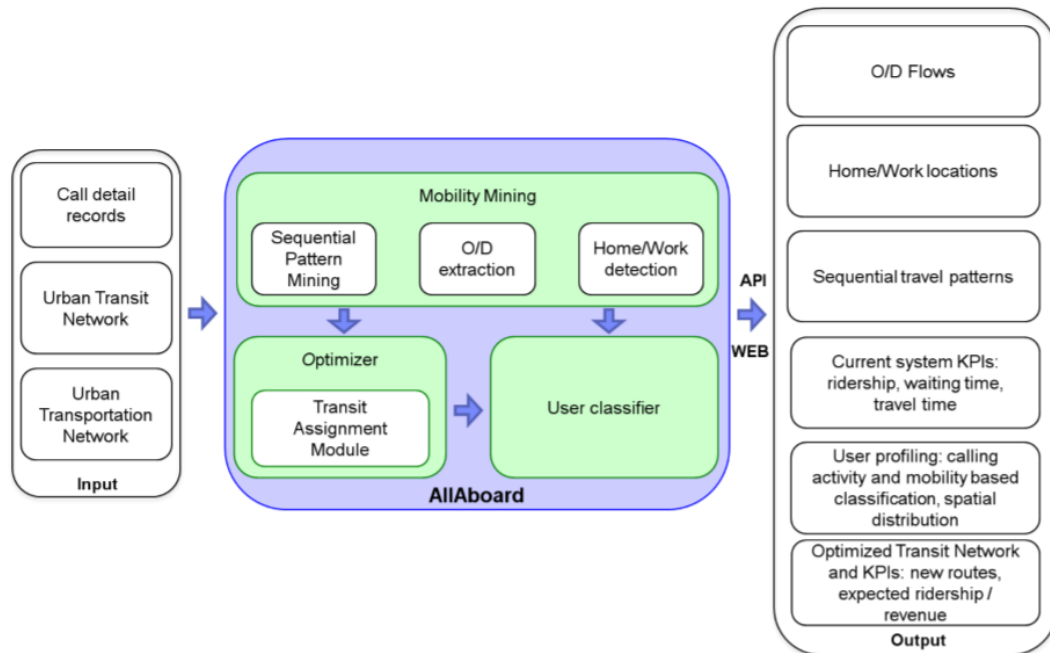
資料來源：Zahedi and Shafahi (2017)

謝雅晴(2018)利用手機信令資料以及汽車客運即時動態的時空軌跡以得到汽車客運、非汽車客運以及總旅運的起迄矩陣，研究場域為花蓮，使用之信令資料其位置解析度為二級發布區。此研究首先對信令資料進行插捕的動作，讓每個用戶在每分鐘皆能比對到一個二級發布區。接著以 20 分鐘作為判斷用戶是否有停留行為的發生的依據，並同時進行旅次的切割。隨後將每個旅次中每分鐘的位置資訊與汽車客運的時空軌跡資訊進行比對，並將滿足「成功比對的資料數大於兩筆(即至少兩分鐘的信令資料所在位置與汽車客運軌跡是相同的)」以及「第一筆與最後一筆比對成功之信令資料其時間間隔為 5 分鐘以上」兩個條件的旅次再計算其符合率，也就是整段旅次與汽車客運軌跡比對成功的資料數比例。最後，將符合上述兩個條件且符合率大於所設定之門檻的旅次視為搭乘汽車客運的旅次，其於旅次則視為非搭乘汽車客運之旅次，並統計兩種旅次於各起迄對之間的數量，即可得到汽車客運起迄矩陣、非汽車客運起迄矩陣以及總旅運起迄矩陣。

李泓儒(2018)利用手機信令資料配合軌道運具的班表資料，研擬出一個可得到軌道運具之起迄矩陣的方法。此研究之軌道運具研究對象為台灣鐵路以及高雄捷運兩者，而使用之信令資料的位置解析度為二級發布區。此研究首先對每個用戶之信令資料進行插補，並以 20 分鐘作為判斷用戶是否有停留行為發生及作為旅次切割的依據。接著將台鐵以及高捷的車站位置轉換成其所在的二級發布區以便與信令資料進行空間上的比對，並配合各自的班表資料來與信令資料進行時間上的比對。在台鐵的比對方面，若某個旅次其原始的信令資料(即未插捕前的信令資料)，能與台鐵的時空軌跡資料成功比對到至少兩個不同的站點，則將該旅次判定為搭乘台鐵的旅次；比對高捷的部分，則是在判斷某旅次是否能與高捷的時空軌跡資料成功比對到至少兩個站點後，還需再檢查比對成功的兩筆信令資料其時間間隔是否大於 5 分鐘以及該旅次與高捷軌跡比對成功的資料數比例是否大於門檻值。最後，整理比對成功的旅次並進行統計，即可得到台鐵以及高捷的起迄矩陣。

2.2 手機信令資料與公共運輸服務

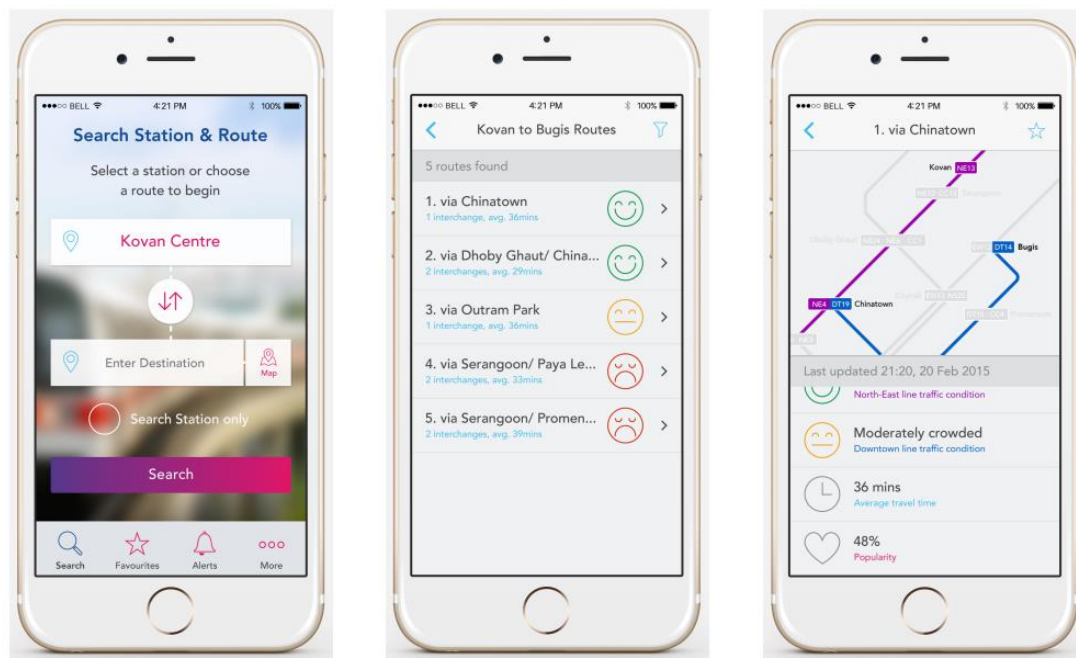
Berlingerio 等人(2013)與 IBM 共同開發 ALLAboard 系統，此系統架構如圖 2-7 所示。圖 2-7 左方區域為資料輸入端，可在此步驟輸入手機信令資料、都市的交通運輸路網等資料。接著，圖 2-7 中間區域為分析階段，此系統主要有兩個核心模組，分別為移動性的探勘(Mobikity Mining)以及最佳化(Optimizer)。最後，圖 2-7 右方區域為輸出之結果，可得到旅運起迄矩陣、居住地及工作地、旅運行為等資訊。此研究在象牙海岸的阿必尚進行實測，結果顯示若新設置四條道路，則可替整個城市省下 10% 的總旅行時間。



資料來源: Berlingerio et al. (2015)

圖 2-7 ALLAboard 系統結構圖

Holleczech 等人(2015)利用手機信令資料搭配捷運的班表，發展出一個能即時回報捷運系統當時的路網情況的系統，回報的路網資訊包含:月台等候人數、轉運站當時欲轉換路線之人潮以及目標起迄對的預期旅行時間，讓民眾可以根據當時各路線的服務狀況選擇其欲搭乘的路線，避免尖峰時間由於民眾皆欲搭乘同一路線導致捷運系統癱瘓或造成民眾浪費太多時間在等待。圖 2-8 為程式開發示意圖，可看到在選定起迄站之後，系統會找出幾個可搭乘的路線並顯示其旅行時間以及擁擠程度。而在選擇路線後，系統會顯示該路線之路線圖以及路線的其他資訊。



資料來源: Holleczech et al. (2015)

圖 2-8 捷運路線資訊系統開發示意圖

Demissie 等人(2016)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，第一步將蒐集到的信令資料根據用戶特性進行篩選，挑選出一天有超過 75%的時間在與基地站互動之用戶，以及每周與基地站互動次數少於 1000 次之用戶。接著針對各個用戶探勘出各自除了居住地最常前往的 5 個地點，並以用戶的居住地以及這些地點所連接之路線作為公共運輸的潛在需求，最後再比對公車路線於這些起迄對的服務量來檢視是否有供需失衡或甚至沒有公共運輸服務的地方。根據研究結果指出，該地區有 47 條路線須改善或新設。

交通部運輸研究所(2017)執行旅運時空資料分析與公共運輸服務應用發展計畫，此計畫以信令資料為基礎，並將其應用於多種交通運輸相關的研究，而此研究計畫所使用的信令資料位置解析度為二級發布區。此計畫首先利用手機信令資料配合公共運輸的動態班表資訊以求得公路客運、高雄捷運、台灣鐵路以及私人運具的起迄矩陣，並以產製出的起迄矩陣作為各地區的需求量並找出公共運輸的服務缺口，藉以提出相關的改善策略。除此之外，此研究計畫同時利用手機信令資料推估各個地區的常住人口數，以作為交通建設或公共設施建設時的選址依據。最後，透過信令資料以及資料科學方法找出景點之間的關聯性，並以具有關聯性的地區作為觀光運輸的需求，以提供交通當局在規劃或調整觀光公共運輸路線時的參考依據。

交通部運輸研究所(2018)執行應用旅次特性大數據精進公共運輸服務計畫，此研究計畫利用信令資料來推估每個用戶之居住地、進行旅次鏈分析、判斷用戶所使用之運具以及用以推估觀光公共運輸系統之服務缺口，而此研究計畫使用的信令資料位置解析度為 50*50 公尺的範圍(以下將此 2500 平方公尺的範圍簡稱為網格)。此計畫首先精確地找出用戶的居住地位於哪個網格之中，隨後將用戶整天的信令資料列出，並判斷用戶離開居住地的時間以及抵達目的地(即工作地或學校)的時間，藉此來判斷每個

用戶的就學或是工作旅次，而將全部用戶的就學或工作旅次進行彙整，即可得到全體用戶之就學或工作旅次起迄矩陣。此研究計畫與運研所(2017)同樣有運具判斷的研究，但由於資料位置解析度的精進，此研究計畫除了可直接將用戶移動軌跡與運具移動軌跡直接進行比對以增加判斷的準確率，亦可藉由判斷用戶是否在移動的過程中停留行為來進行運具轉乘的分析。此外，此研究計畫亦透過信令資料對用戶特性進行篩選，判斷所取得的用戶中哪些為觀光用戶，並針對篩選出來的觀光用戶進行觀光旅運分析以及資料探勘分析，藉以找出觀光公共運輸的服務缺口並提出相關的服務建議。上述關於居住地分析、就學或工作旅次分析以及轉乘分析，皆會在最後以圖象化的方式呈現，並與公共運輸路線進行疊圖後，即可透過圖像得知哪些地方存有旅運需求卻沒有公共運輸路線服務，或是有特定地方發生供需失衡的情況，藉以供交通當局作為改善公共運輸環境的參考依據

施冠毅(2018)利用手機信令資料找出具有重現性需求的用戶，再透過找到的用戶來判斷哪些地點之間具有較高的旅運需求。研究場域為高雄地區，資料取得時間為2017年4月24號至4月28號共計5天的資料，空間位置解析度為二級發布區。此研究首先透過信令資料判斷用戶在研究期間的這5天內於高雄地區的移動模式，並根據用戶在5天內移動模式的差異程度分成4群，再挑選移動模式差異較小的兩群用戶作為「可能具有重現性需求」的用戶來進行下一步分析。最後，透過循序樣式探勘的方法，分析具有重現性需求的用戶較常在高雄地區的哪些二級發布區之間進行移動，即可得知高雄市的哪些地區之間有較高且重現性的旅運需求，而交通當局即可根據此結果來規劃市內公共運輸的路線及班次。

2.3 小結

根據第 2.1 及第 2.2 小節，發現手機信令資料可實際應用於交通運輸的領域。無論是在推估道路流量、推估使用者旅次目的或使用的運具…等應用，而研究結果與傳統旅運行為調查資料相比對後，能得到大致相符的結果，也說明手機信令資料的可行性。

不僅如此，手機信令資料在蒐集的過程比起傳統調查資料有著更容易取得、資料更加全面的優點。Ahas (2007)提到手機信令資料將顯示一使用者的軌跡並可推估其從事的活動與停留的地點，且在資料取得方面也相對省成本跟簡易。Zahedi and Shafahi (2017)也指出一些在深夜或是使用者認為不重要的旅次，如：到雜貨店買東西，在傳統調查資料中可能會因為使用者的人為因素而蒐集不到，但對於手機信令資料而言並不存在此問題。

根據回顧的文獻，可發現運研所(2017、2018)雖然已透過信令資料並配合資料科學的方式得到景點間的關聯性，並以此作為觀光需求以提出觀光公共運輸的服務改善建議。然而，過程中該研究僅利用關聯規則進行資料探勘，並透過找尋「哪個景點與最多景點有著關聯性」來做為最後提出的新增站點之依據。本研究認為，該研究方法雖能得到觀光客的移動行為並提出服務建議，但也有可能產生部分問題或忽略部分結果，例如：觀光客的移動行為通常不具有重現性，因此若直接對取得的全體觀光客旅運行為進行關聯規則的資料探勘，則結果可能僅會顯示最多人拜訪的景點以及與之具關連性的景點，部分熱門景點可能因此被忽略。不僅如此，該研究最終提出的服務建議僅依據「哪個景點與最多景點有著關聯性」來做判斷，然而依照關聯規則的原理，兩者之間關連性的強烈與否也應作為評選結果是否為服務缺口並將其改善的依據之一。舉例來說，若「花蓮觀光糖廠」與「馬太鞍濕地生態館」具有強烈的關聯性(即多數觀光客拜訪花蓮時會將此兩景點一併規劃進旅程中)，但兩者卻僅與彼此有關連而與其他景點較無關聯性；另一方面，「花蓮文化創意產業園區」雖然與各個景點的關聯性並不強烈，但它與眾多景點皆存有些微的關聯性，則運研所(2018)的研究方法僅會考量是否將花蓮文化創意產業園區規劃進台灣好行的路線內，而忽略了糖廠與馬太鞍濕地生態館的部分。因此，本研究將以運研所(2018)對景點找出關聯性之研究為基礎，將其過程改善(即提出單點式關聯規則)，並在最終提出服務建議時也額外提出另一個評量標準，使結果更具可信度。

第三章 分析方法建立

本研究之目的為從觀光用戶的手機信令資料中先分析其旅次行為，找出頻繁被拜訪的景點組合後，再與現有的台灣好行服務路線進行比對，以檢視該觀光接送服務的完整性。由於本研究是針對觀光客在景點與景點之間的移動做為比對台灣好行路線服務的依據，因此在進行資料探勘時，會以頻繁被拜訪的多個景點組成的組合為主。為完成上述目的，本研究發展一套研究方法，流程如圖 3-1 所示。

本研究之分析方法的第一個步驟為資料蒐集，包含手機信令資料、網格中心經緯度資料和興趣點地理位置資訊。第二步驟為將觀光用戶之信令資料與興趣點的地理資訊資料進行比對以得知用戶是否有出現在特定興趣點的紀錄。第三個步驟則是計算用戶在各興趣點被記錄到的時間長度，並判斷用戶於該興趣點是否有停留的行為發生。第四及第五步驟則是根據本研究研擬之判斷邏輯，進一步篩選觀光用戶中符合本研究所定義之第一類用戶，隨後將各類別用戶的旅運資訊進行一般關聯規則分析以及單點式關聯規則分析以找出高頻的興趣點旅運需求樣態。最後，將得到的結果與現有之台灣好行觀光巴士路線進行比對以檢視是否存有未被滿足的觀光旅運需求，並根據比對結果提出相對應之服務改善建議。

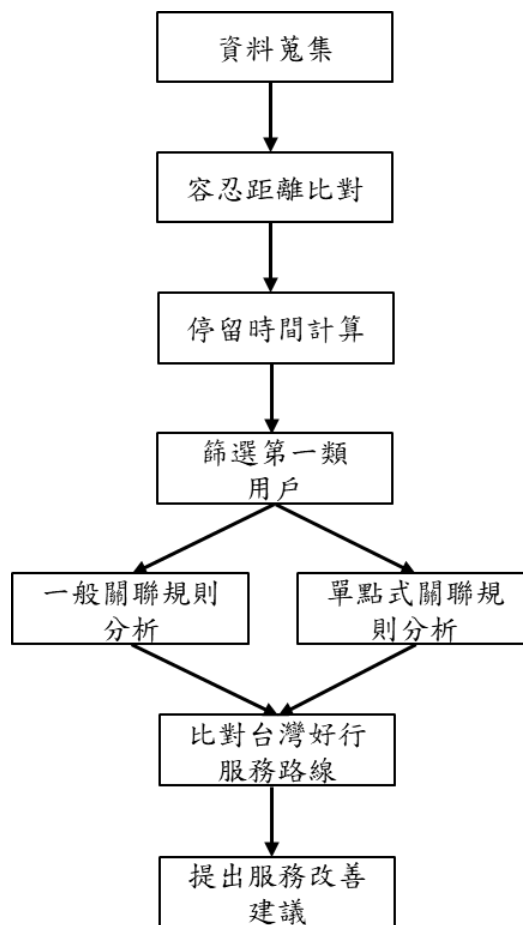


圖 3-1 分析方法流程圖

3.1 資料蒐集

本研究所蒐集的資料共有三種，分別是手機信令資料、網格中心經緯度資料，與花蓮興趣點地理資訊資料。手機信令資料為本研究透過電信公司所取得之花蓮縣三個周末的觀光用戶信令資料。網格中心經緯度資料為紀錄每個網格其中心位置經緯度的資料。最後，花蓮興趣點資料為本研究用以判斷觀光用戶是否有在特定興趣點停留之資料，當中紀錄花蓮各大名勝景點之位置資訊。以下將對於上述各種資料進行詳細說明。

3.1.1 手機信令資料與網格資料

手機信令資料為手機使用者使用手機透過行動網路上網，或是在移動時使得手機與連結的電信基地站有改變…等手機與電信基地站有互動的行為時，其當下的時間及位置會被電信公司記錄下來之資料。除此之外，若手機使用者長時間沒有發生與基地站有互動之行為，例如：使用 Wi-Fi 上網、待機等非關機的情況，電信公司仍會以每小時為間隔主動地蒐集用戶之信令資料。

除此之外，隨著行動通訊技術的發展，手機電信基地站已能夠結合高精度的地理資訊系統來定位手機使用者使用網路行為時的位置。早期的手機信令資料其位置資訊解析度大多以電信基地站服務的範圍來表示，而現今則可以把所在地區再劃分為數個小區域來做為位置資訊的解析度。本研究取得的信令資料其位置資訊解析度即為 50 乘 50 公尺的區域，而本研究稱此方格稱為「網格」，並將手機使用者的位置資訊以所在的網格表示。

表 3.1 為本研究取得之手機信令資料的資料格式，包含加密後的使用者識別碼、日期與時間以及所在網格共三個欄位。其中第一欄為加密後的使用者識別碼，電信公司為避免用戶的個資洩漏，在提供信令資料時會先將用戶的門號加密處理為 64 位英數字混和的亂碼。第二欄為該筆信令資料被基地站紀錄的日期與時間，即代表手機使用者在該日期時間有使用手機從事與 4G 網路有關連的活動。第三欄則為手機使用者與 4G 網路有所互動時所處的網格，而網格以 9 位的數字代碼表示。

表 3.2 為電信公司提供的手機信令資料範例，其中已被加密之使用者識別碼以後 6 碼表示，並在本研究之後的範例圖表中皆以此方式表示使用者識別碼。以表 3.2 第一列為例，表示 lae2v3 這名用戶在西元 2018 年 5 月 19 日的上午 8 點 22 分時，所處的位置為編碼 214707015 的網格。

表 3.1 信令資料格式

欄位	格式
使用者識別碼	64位英數字
日期與時間	西元年/月/日 小時:分鐘:秒
網格	9位純數字

表 3.2 信令資料範例表

加密後之使用者識別碼	日期與時間	網格
1ae2v3	2018/05/19 08:22:00	214707015
0003ab	2018/05/19 08:24:00	213627013
e654ba	2018/05/19 08:25:00	254715036
0003ab	2018/05/19 08:32:00	242968003
019cf5	2018/05/19 08:51:00	265717038
...	...	...

除此之外，電信公司同時提供網格度資料，表 3.3 為網格資料範例表，其中包含網格編號、一級和二級行政區(省、直轄市、縣、市)、三級行政區(鄉、鎮、縣轄市)、四級行政區(村、里)以及該網格中心點之經緯度。此份資料顯示各個網格其中心點之經緯度為何，可用以計算網格之間的距離。

表 3.3 網格中心經緯度資料範例表

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
265717038	花蓮縣	富里鄉	豐南村	121.2609952	23.12547216
...	...	...	...	...	...

3.1.2 興趣點地理資訊資料

表 3.4 為興趣點之空間資料格式，包含花蓮縣境內興趣點的中文名稱、興趣點代碼以及興趣點所包含之網格。第一欄為花蓮縣境內 55 個觀光景點以及 25 個台鐵火車站，共計 80 個興趣點之中文名稱，其中 55 個觀光景點的選取方式為交通部運輸研究所根據 Google Map 和 TripAdvisor 兩大網站上的遊客評論所產生。第二欄為各個興趣點所對應之代碼，以 2 位純數字表示，編號方式依照 80 個興趣點的位置由北到南依序給予 01 到 80 的代碼。第三欄為興趣點所包含的網格，由於一個興趣點的面積通常會超過 2500 平方公尺，意即一個興趣點的面積會比一個網格的面積大，因此同一個興趣點會對應到多個網格代碼。以表 3.5 興趣點地理資訊資料為範，從上述說明可得知台鐵和平站之代碼為 01，共包含了 22 個網格，表 3.5 僅顯示其中 5 個網格。

表 3.4 興趣點地理資訊資料格式

欄位	格式
興趣點中文名稱	中文
興趣點代碼	2位純數字
網格	9位純數字

表 3.5 興趣點地理資訊資料範例表

興趣點中文名稱	興趣點代碼	網格
台鐵和平火車站	01	300197070
台鐵和平火車站	01	300197079
台鐵和平火車站	01	300197080
台鐵和平火車站	01	300263044
台鐵和平火車站	01	300263053
...	...	...

3.2 建立觀光用戶旅運資訊

為了後續分析之所需，必須先替各觀光用戶建立其旅運資訊，也就是判斷各用戶於花蓮的造訪了哪些興趣點。此步驟利用到 3.1.1 小節所取得的觀光用戶信令資料，配合興趣點地理資訊資料以及網格中心點資料來完成。此外，此步驟分成兩個程序，分別為容忍距離比對以及停留時間計算，也就是圖 3-1 的第二及第三步驟，流程及所需之資料如圖 3-2 所示，以下將詳細說明這兩個程序之內容。

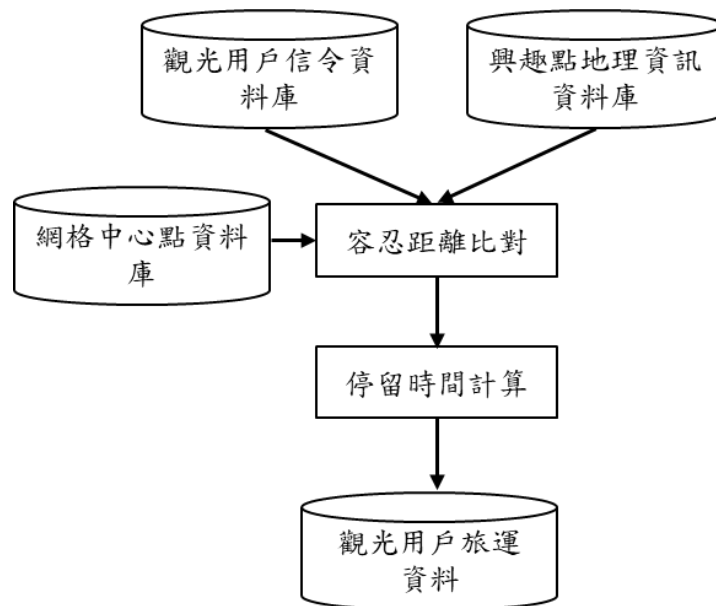


圖 3-2 用戶旅運資料建立流程圖

3.2.1 容忍距離比對

為了得知觀光用戶是否有在興趣點出現或是有停留的行為，本研究透過比對觀光用戶的信令資料和興趣點地理資訊資料來判斷。根據兩個資料庫皆含有的「網格」的欄位，即可比對用戶所在網格是否與興趣點所包含的網格一致，若有，則表示該用戶在該時間點出現在比對到的興趣點。

然而，手機信令資料常因手機訊號受到地形影響或是受天候影響等因素而與基地站之間的連結強弱不定，進而產生飄移的現象，導致被偵測到的位置資訊並不是使用者確切的所在位置。為解決此問題，本研究在進行信令資料與興趣點資料比對時使用「容忍距離」這一個放鬆條件，以解決飄移之影響。此程序的比對方法即是檢查每一筆信令資料，其周圍的特定範圍內是否有涵蓋到某一興趣點的任一網格，若有，則判定為比對成功，而該特定範圍即為容忍距離。圖 3-3 為容忍距離比對示意圖，假設圖中的九宮格為七星潭風景區的範圍，且有三筆信令資料發生於七星潭風景區的附近，其位置分別為圖中的 A、B 以及 C 三個點，在進行比對時，可看出由於 A 點其位置坐落於七星潭風景區內，因此視為比對成功；B 點雖然其位置不在七星潭風景區內，但該點與七星潭風景區的距離小於所設定的容忍距離，因此亦視為比對成功；而 C 點則由於與七星潭風景區的距離過遠，兩者的距離已大於所設定的容忍距離，因此視為比對失敗。網格之間距離的計算可透過第 3.1.1 節提到的網格資料，利用當中所記錄各網格的中心點經緯度資訊，即可得出網格之間的距離。而關於容忍距離的設定，本研究參考運研所(2018)之研究內容，將其設定為 200 公尺。

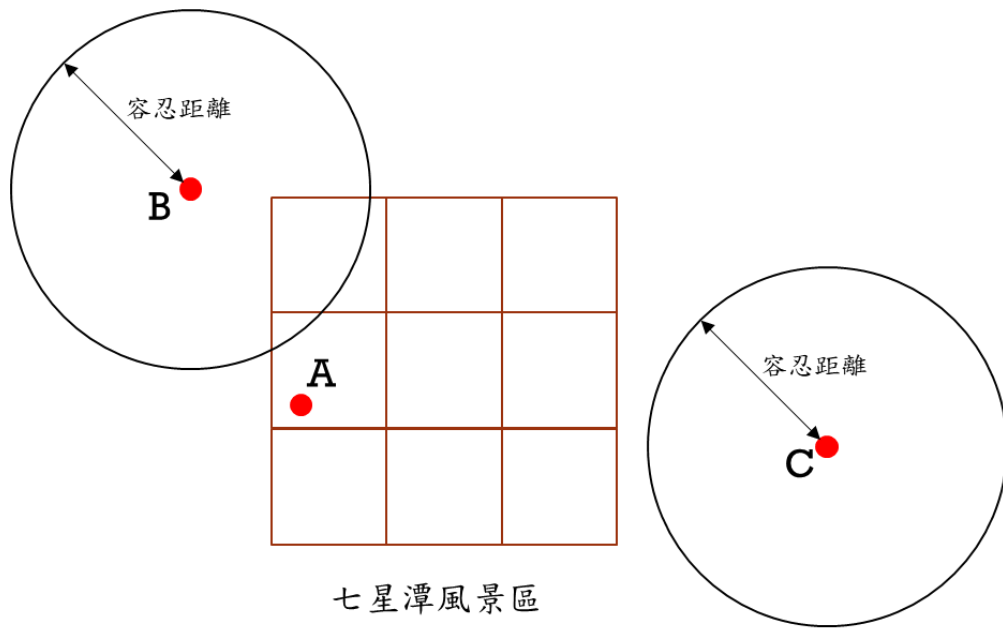


圖 3-3 容忍距離比對示意圖

比對流程為利用網格資料，計算每筆信令資料其網格的中心經緯度與各個興趣點所包含的網格是否有任一在 200 公尺以內者，若有，則判定該筆信令資料與該興趣點為比對成功。以圖 3-4 為例，左上方的表格為觀光用戶信令資料，而圖 3-4 右上方的表格則為興趣點地理資訊資料留下興趣點中文名稱和所包含網格的表格，此範例表格只取 5 個興趣點並各只取一個所包含的網格作為範例展示。將此兩表格比對之後，可發現觀光用戶信令資料表格中第 1、3 及 5 列的用戶所在網格，和興趣點地理資訊資料表格中的花蓮觀光糖廠以及花蓮火車站有網格代碼相同的情況，即表示用戶 1ae2v3 曾在 2018 年 5 月 19 號的早上 8 點 25 分和 2018 年 5 月 19 號的早上 8 點 26 分出現在花蓮觀光糖廠，而用戶 8qs12r 則在 2018 年 5 月 19 號的早上 8 點 30 分出現在花蓮火車站。除此之外，假設用戶 019cf5 的所在網格在經過計算後，與花蓮火車站的網格距離小於 200 公尺，因此即便用戶 019cf5 的所在網格與花蓮火車站的網格代碼不同，此筆信令資料仍判定為比對成功，表示用戶 019cf5 在 2018 年 5 月 19 號的早上 8 點 28 分出現在花蓮火車站。而用戶 8qs12r 的所在網格經過計算後，其 200 公尺範圍內則無法比對到任何興趣點的網格，因次判定為該名用戶在該時間沒有在此次研究的 80 個興趣點出現，並以「NA」表示此情況，如圖 3-4 下方表格的用戶 8qs12r 之興趣點中文名稱欄位處所示。在此須注意若信令資料中心點方圓 200 公尺以內比對到多個不同的興趣點所包含的網格，則以距離最近的興趣點做為該筆信令資料之比對結果。

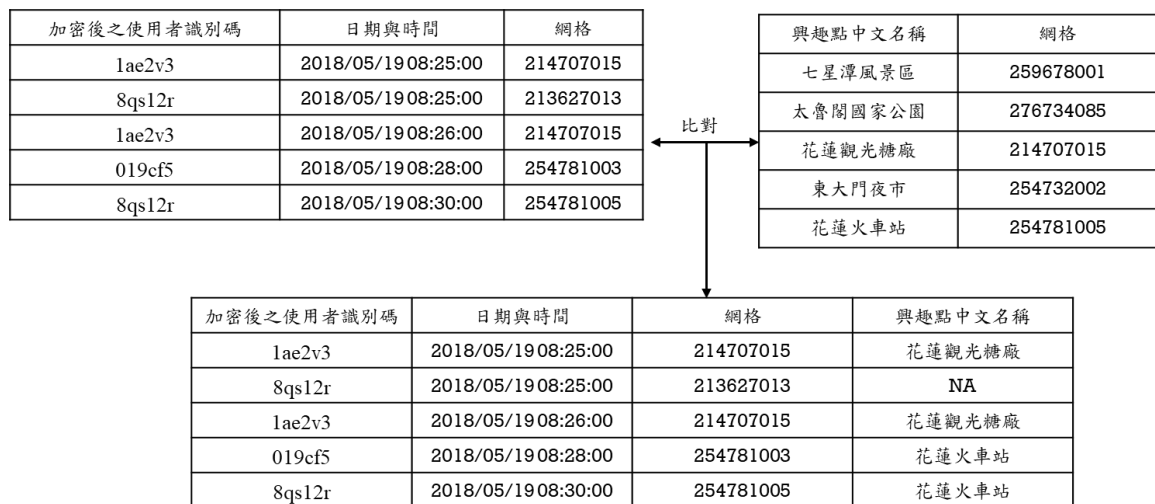


圖 3-4 容忍距離比對範例圖

3.2.2 停留時間計算

為了區分比對成功的信令資料是手機使用者剛好經過該景點，亦或是有停留的行為發生，本研究接著計算於同一興趣點之停留時間，用以判斷用戶是否有在特定地點停留。首先即是設定停留的時間門檻值(以下稱此門檻值為興趣點停留時間門檻)，接著取出單一用戶之所有信令資料，並計算該用戶是否有出現於某一興趣點停留超過興趣點停留時間門檻的情況發生；若有，則判定該用戶於該興趣點有停留的行為，並將每個用戶依序完成上述動作，即可得到所有觀光用戶的觀光旅運資訊。為了避免信令資料的飄移問題導致同一個興趣點的連續多筆信令資料中間被漂移訊號切開，使得最終用戶被判定在同一興趣點有連續兩次或以上的停留行為，本研究將發生飄移狀況的信令資料剔除後，再計算用戶於該興趣點的停留時間。

以圖 3-5 為例，左邊的表格為某一用戶在 2018 年 5 月 19 日早上 8 點 22 分到 9 點 31 分之間的信令資料。假設此時興趣點停留時間門檻設定為 10 分鐘，則可看出該用戶自當天早上 8 點 24 分到 8 點 35 分，以及 9 點 01 分至 9 點 11 分皆有被比對到出現在七星潭風景區，雖中間有出現一筆被記錄為「NA」者，即上午 8 點 51 分的信令資料，但本研究將出現於相同興趣點中間的「NA」資料視為是飄移資料並將其剔除，即可發現用戶於早上 8 點 24 分到 9 點 11 分只出現在七星潭風景區，並未出現在其他的興趣點，因此可計算出該用戶於七星潭風景區停留 47 分鐘，且由於 47 分鐘有超過起初設定的 10 分鐘門檻，故判定為停留，並將此停留行為記錄於圖 3-5 右方的表格。同理，可得知該用戶於太魯閣國家公園同樣有停留行為，因此亦將太魯閣國家公園紀錄於圖 3-5 的右方表格。將每一用戶的信令資料完成此動作後，即可得到全部觀光用戶的旅運資訊並進行下一步的作業。

日期與時間	興趣點中文名稱		興趣點中文名稱	停留時間
2018/05/19 08:22:00	NA			
2018/05/19 08:24:00	七星潭風景區			
2018/05/19 08:25:00	七星潭風景區			
2018/05/19 08:35:00	七星潭風景區	→	七星潭風景區	47
2018/05/19 08:51:00	NA		太魯閣國家公園	10
2018/05/19 09:01:00	七星潭風景區			
2018/05/19 09:11:00	七星潭風景區			
2018/05/19 09:15:00	NA			
2018/05/19 09:21:00	太魯閣國家公園			
2018/05/19 09:31:00	太魯閣國家公園			

圖 3-5 停留時間計算示意圖

除此之外，若用戶的旅運行為顯示其曾拜訪某一興趣點兩次或以上，本研究將只記錄其中停留時間最長者。由於替用戶建立其旅運行為之目的為後續的關聯規則分析所需，而在關聯規則的探勘中，僅會探勘不同元素之間的關聯性，也就是說不會出現「若 A 則 A」的結果樣態，也因此替用戶建立其旅運行為表時，旨在找尋用戶曾於那些興趣點有停留行為，而用戶於某興趣點停留的次數便不在考慮範圍內。針對用戶多次拜訪的同一興趣點的情況，本研究以當中停留時間最長者為該用戶於該興趣點的停留紀錄，並將其餘的停留行為刪除，舉例來說，假設透過某用戶兩天的信令資料共可判定其有 3 個停留行為，分別為在花蓮糖廠停留 60 分鐘、在太魯閣風景區停留 70 分鐘以及在花蓮糖廠停留 20 分鐘，根據上述，本研究會將用戶於花蓮糖廠停留較短的紀錄剔除，也就是刪除用戶於花蓮糖廠停留 20 分鐘的紀錄，只留下此用戶於太魯閣風景區停留 70 分鐘以及在花蓮糖廠停留 60 分鐘兩個停留紀錄。

3.3 篩選第一類用戶

本研究除了針對全體觀光用戶的旅運行為進行資料探勘，亦會將觀光用戶再細分兩個族群(第一類用戶以及第二類用戶)，再依序進行關聯規則分析，故本章節將介紹本研究是如何篩選出第一類用戶，而其餘的觀光用戶則為第二類用戶。本研究定義第一類用戶即為使用公共運具抵達(或離開)觀光縣市之用戶，有此定義為考慮到使用私人運具前往觀光地區的民眾較不太可能在當地使用公共運輸系統；然而，對於使用公共運輸系統前往或離開觀光地區的遊客而言，有較大的比例會在觀光地區使用公共運輸系統來移動。

本研究篩選第一類用戶的方法即為判斷用戶於周末兩天的信令資料中，第一個停留或最後一個停留是否為觀光縣市的公共運輸系統站點，且停留時間大於設定之門檻值(以下稱此門檻值為運輸場站停留時間門檻)。本研究假設一般民眾利用公共運輸系

統出遊時，在運具上或公共運輸系統的站點會有使用手機的行為，如：查詢欲轉乘的運具、查詢當地的觀光景點或是等車的空檔上網看影片等，因此，本研究透過民眾在使用公共運具出遊時會在公共運輸系統的站點產生該旅程第一個或最後一個停留點的特性，來篩選全體觀光用戶中屬於第一類用戶者。然而，考量到一般民眾於興趣點遊玩時和在運輸場站停留的時間有一定的差異，且通常在興趣點遊玩的時間會遠大於在運輸場站停留的時間，因此上述提到的運輸場站停留時間門檻在本研究有著「至多等於興趣點停留時間門檻」的限制。

舉例來說，若用戶的第一個或最後一個停留點為火車站時，即判定此用戶為利用火車往返觀光地之用戶，符合第一類用戶之定義。同理，若用戶的第一個或最後一個停留點為觀光當地的飛機場，則判定該用戶為利用飛機往返的用戶，亦為第一類用戶。但在公路客運的部分，由於公路客運行駛的路線通常會先經過市區內的部分景點後才抵達當地的客運總站，並不像台鐵或飛機在進入各縣市後會直接抵達其運輸場站。因此在判定用戶是否為利用公路客運往返觀光地時，除了辨別用戶是否有在，還需再檢視用戶的移動軌跡與該縣市聯外的公路客運系統，在該縣市內的移動軌跡是否一致，辨別用戶是否為利用公路客運往返觀光地之用戶。

本研究透過表 3.6 為範例說明如何檢查用戶的第一個或最後一個停留點是否為公共運輸系統的運輸場站，並以台鐵為例。根據第 3.2 小節得到的觀光用戶旅運資訊結果，檢查每個用戶的第一個或最後一個停留點是否為花蓮境內的 24 個火車站之一，若用戶的旅運行為顯示其第一個或最後一個停留點為台鐵車站，則該用戶判定為第一類用戶，反之則為第二類用戶。表 3.6 為用戶旅運資訊範例表，從表中可看出用戶 1ae2v3 於周末兩天的第一個停留點發生在花蓮火車站，表示用戶 1ae2v3 為本研究所定義之第一類用戶。同理，用戶 8qsl2r 於周末兩天的最後一個停留點發生在富里火車站，因此用戶 8qsl2r 亦為第一類用戶。然而，用戶 019cf5 在兩天的停留點中皆無火車站，表示該用戶並沒有在火車站有停留的行為發生，並非本研究所定義之第一類用戶，因此歸類為第二類用戶。而用戶 abb32s 雖然旅途中曾在花蓮火車站停留，但此停留行為並非是該旅途的第一個或最後一個停留，因此用戶 abb32s 亦不屬於第一類用戶，而是第二類用戶。

表 3.6 用戶旅運資訊範例表

用戶	興趣點中文名稱
1ae2v3	花蓮火車站
1ae2v3	太平洋公園
1ae2v3	東大門夜市
8qs12r	花蓮觀光糖廠
8qs12r	富里火車站
019cf5	太魯閣國家公園
019cf5	太平洋公園
abb32s	慶修院
abb32s	花蓮火車站
abb32s	花蓮觀光糖廠

利用此方法尋找全體用戶中何者為第一類用戶與第二類用戶之前提假設為假設用戶在運輸場站會有使用手機上網或是其他與基地站互動的行為以及用戶旅途的第一天或最後一天需至少一天被包含於資料蒐集期間。第一個假設係為了避免用戶在運輸場站沒有使用手機的行為，導致沒有信令資料的產生進而造成有遺漏的情況發生。而第二個假設則是考量到資料蒐集期間若不包含民眾旅遊時的第一天或最後一天，便無法得知用戶旅途的第一個或最後一個停留點為何，導致無法透過此邏輯判斷用戶是否為第一類用戶，如本研究於研究範圍即提及本研究會針對旅途至多為三天且包含周末二日之用戶進行信令資料的蒐集，此時若資料蒐集期間選擇在周末二日，則無論用戶的旅途期間為周末二日或星期五、六及日三天，皆可透過信令資料得知用戶整趟旅途的第一個或最後一個停留點為何。

3.4 關聯規則分析

關聯規則為資料探勘中常被使用以找出資料間相關性的方法，固本研究透過關聯規則分析找出在觀光客之中較具有關聯性的興趣點組合，並將其視為觀光客之旅運需求。本章節首先會對關聯規則進行介紹，接著說明本研究在應用關聯規則分析時所面臨的問題。最後，針對上述提及之問題本研究提出一套解決的方法，亦會在此節一併做說明。

3.4.1 關聯規則分析

關聯規則為資料探勘中常被使用以找出資料間相關性的方法，而關聯規則透過設定最小支持度、最小信賴度以及增益值三個門檻值來完成搜索資料庫中具有關聯性的資料集合。以下透過表 3.7 的示範資料庫進行三種門檻值以及關聯規則分析的說明。

表 3.7 為四名用戶之興趣點表格。以表 3.7 第一列為例，表示用戶 1 時拜訪過砂卡礑步道以及 a-zone 花蓮文化創意產業園區兩個興趣點。此外，為了便於後續的說明，表 3.7 將各個興趣點給予相對應的英文代碼，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，並在後續的說明以代碼表示各個興趣點。

表 3.7 關聯規則示範資料庫

使用者	興趣點集合
1	砂卡礑步道(A)、a-zone 花蓮文化創意產業園區(B)
2	七星潭風景區(C)、煙波大飯店(D)、花蓮火車站(E)
3	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)
4	煙波大飯店(D)、花蓮火車站(E)

支持度(support)為某一種資料樣態(pattern)在資料庫中佔的比例，即 $\text{support}(A)=P(A)$ ，而資料樣態可以以單一元素或是多個元素所組成。以表 3.7 為例，A 與 B 皆可各自成為單一元素的資料樣態，且由於 A 與 B 曾同時被用戶 1 拜訪過，因此 A 與 B 可組成兩個元素的資料樣態。同理，C、D、E 三個興趣點亦可各自成為單一元素的資料樣態，亦可兩兩組合或是三者一起成為多元素構成的資料樣態。最小支持度則是設定一個門檻值來檢視哪些資料樣態的支持度是足夠大的，即表示該資料樣態在資料庫中佔有一定的比例，而此類支持度大於等於最小支持度的資料樣態稱為高頻樣態，也是作為後續找出關聯規則結果之基礎。由於本研究利用關聯規則分析的目的為找出彼此具有關聯性的興趣點組合，故不考慮僅含一個元素的高頻樣態結果，因此在最小支持度設定為 0.5 時可找出表 3.8 所列出的 4 個由至少 2 個元素所組成的資料樣態。以表 2.3 第一列為例，資料樣態 {C、D} 其支持度，也就是 $P(C \cap D)$ ，可透過表 3.7 看出同時拜訪過興趣點 C 及 D 的用戶在全部 4 個用戶中占有 2 個，因此資料樣態 {C、D} 的支持度為 $2/4=0.5$ ，符合高頻樣態之條件。

表 3.8 高頻樣態結果表

高頻樣態結果	支持度
{C、D}	0.5
{C、E}	0.5
{D、E}	0.75
{C、D、E}	0.5

對於關聯規則之結果，一般會解讀為「若甲則乙」，其中甲稱為先導規則 (antecedent, left-hand-side, LHS)，乙則稱為後繼規則 (consequent, right-hand-side, RHS)。而關聯規則分析的其中一步驟即是將上一段落所找尋到之高頻樣態中的元素分別作為先導規則與後繼規則，並將其視為最終結果規則的候選規則，隨後再透過最小信賴度以及增益值篩選哪些候選規則為最後可用以分析和預測的結果規則。在將高頻樣態中的元素歸類為先導規則以及後繼規則時，若高頻樣態中含有的資料元素大於等於兩個，則會因為決定何者為先導規則又何者為後繼規則而產生多個候選規則。舉例來說，高頻樣態結果{C、D}可產生「若 C 則 D」以及「若 D 則 C」兩個候選規則；同理，高頻樣態結果{C、D、E}則可產生出「若 C 則 D、E」、「若 D 則 C、E」、「若 E 則 C、D」、「若 C、D 則 E」、「若 C、E 則 D」以及「若 D、E 則 C」共六個候選規則

此時，為了得知各個候選規則何者為最終可用以預測及分析之結果規則，即需要最小信賴度一門檻進行第一步的篩選。信賴度為衡量某規則在先導規則發生時，能正確推得其後繼規則正確性之信心程度的依據，也可將其視為是衡量規則可信度的指標，而最小信賴度即是設定一個規則可信度的門檻值，以避免可信度不夠高之候選規則被判定為最終的結果規則。信賴度的計算方法為「在先導規則發生的前提下，後繼規則發生的條件機率」，也就是 $\text{Confidence}(\text{LHS} \Rightarrow \text{RHS}) = P(\text{RHS} | \text{LHS})$ 。以高頻資料樣態{C、D}為例，此樣態可產生「若 C 則 D」以及「若 D 則 C」兩個候選規則，而此兩個候選規則之信賴度又分別為 $P(D | C)$ 以及 $P(C | D)$ ，因此，可透過表 3.7 得知此兩候選規則的信賴度分別為 1 以及 0.667。對表 3.8 之四個高頻樣態所產生的十二條候選規則分別計算其信賴度，即會得到如表 3.9 之計算結果。倘若此時最小信賴度設定為 0.5，則表 3.9 所列出之十二條候選規則皆能視為具有高可信度之候選規則，並再檢視其增益值以判斷何者為最終可用以預測以及分析之結果規則。

表 3.9 十二條候選規則之信賴度結果表

候選規則	支持度	信賴度
若 C 則 D	0.5	1
若 D 則 C	0.5	0.667
若 C 則 E	0.5	1
若 E 則 C	0.5	0.667
若 D 則 E	0.75	1
若 E 則 D	0.75	1
若 C、D 則 E	0.5	1
若 C、E 則 D	0.5	1
若 D、E 則 C	0.5	0.667
若 C 則 D、E	0.5	1

若 D 則 C、E	0.5	0.667
若 E 則 C、D	0.5	0.667

增益值為最後確認高可信度之候選規則用以預測結果時會比隨機發生的機率高多少的值，計算方式為 $P(RHS|LHS)/P(RHS)$ ，且增益值的門檻值設定通常為 1。舉例來說，假設一候選規則為「拜訪太魯閣則拜訪七星潭」，又該候選規則的信賴度為 0.5，且拜訪七星潭的觀光客在全體觀光客中佔了四分之一，也就是 0.25，此時該候選規則的增益值即為 $0.5/0.25=2$ ，表示太魯閣跟七星潭被前往花蓮的觀光客同時安排進旅程的機率比隨機發生來的大。以表 3.10 第一筆資料為例，該規則的信賴度為 1，同時根據表 3.7 可得知 D 發生的機率為 $3/4=0.75$ ，因此該結果規則的增益值為 $1/0.75=1.333$ ，符合增益值需大於 1 的限制。

表 3.10 關聯規則示範結果表

結果規則	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

透過上述的三個關聯規則門檻值，也就是最小支持度、最小信賴度以及增益值，即可藉由關聯規則分析得到資料庫間具有關聯性的資料集。本研究即利用關聯規則分析找出興趣點之間的關聯性，並將得到的結果作為觀光用戶的旅運需求，接著進行需求與供給(即台灣好行路線)的相互比對，得到是否有地方存有觀光公共運輸服務缺口，最後再提出相對應的服務改善建議。

3.4.2 雜訊問題

傳統關聯規則分析的操作方法通常為針對整個資料庫做資料探勘，找出具關連性的資料集，然而由於觀光客之間的旅運行為差異大，因此若直接對整個資料庫進行關聯規則分析，會面臨到要將最小支持度設得較低才能找到足夠數量的規則的問題。但若將最小支持度設定得越低，探勘出的結果則會產生非頻繁被拜訪的興趣點組合，而此種狀況本研究稱為「雜訊」。雜訊的產生會直接影響到後續分析之結果，舉例來說，在最後利用分析結果以改善台灣好行路線的步驟時，若加入至路線的興趣點為雜訊樣態之結果，則可能導致路線調整後依然無法滿足多數觀光客的實際旅運需求，造成資源的浪費。根據上述，可得知雜訊的產生在使用關聯規則進行資料分析時是一個重要的問題。

表 3.11 為根據第 3.2 節得到的觀光用戶旅運資訊，並將其以使用者為單位的呈現結果。以表 3.11 第一列為例，表示用戶 2c03a1 拜訪過太魯閣風景區以及花蓮文化創意產業園區兩個興趣點。由於觀光客的旅運行為差異大，為了找出具關連性之興趣點組合，需將最小支持度設定至 0.2(十名用戶中至少有兩名有該旅運行為)才能探勘到關聯規則之結果，即「若拜訪太魯閣風景區則拜訪花蓮文化創意產業園區」以及「若拜訪吉蒸牧場則拜訪七星潭風景區」。

此時觀察各興趣點的拜訪人數，如表 3.12 所示，可看出吉蒸牧場與七星潭風景區並不如太魯閣風景區或花蓮文化創意產業園區是較多人拜訪的興趣點。關聯規則之探勘結果會成為最終提出觀光公共運輸路線服務改善建議之依據，而如同「若拜訪吉蒸牧場則拜訪七星潭風景區」此類被拜訪人數較少的規則，在改善後所能帶來的效益會遠不及如「若拜訪太魯閣風景區則拜訪花蓮文化創意產業園區」此類被拜訪人數較高的規則，甚至造成資源的浪費。而類似此種拜訪人數較低但具關連性的規則，即本研究所稱的「雜訊」。

表 3.11 關聯規則產生雜訊之示範資料庫

使用者	興趣點集合
2c03a1	太魯閣風景區、花蓮文化創意產業園區
4cdea3	七星潭風景區、吉蒸牧場、慶修院、馬太鞍濕地生態館
42e49e	砂卡礑步道、太魯閣風景區、慶修院
1fa735	太魯閣風景區、花蓮文化創意產業園區、花蓮火車站
32sd7f	太魯閣風景區、吉蒸牧場、七星潭風景區
66cr8d	慶修院、花蓮文化創意產業園區
aer4tq	花蓮海洋公園、太魯閣風景區、太平洋公園

acv715	馬太鞍濕地生態館、太魯閣風景區
2cfg27	太平洋公園、花蓮文化創意產業園區
eesucl	馬太鞍濕地生態館、花蓮文化創意產業園區

表 3.12 興趣點拜訪人數統計表

興趣點中文名稱	拜訪人數
太魯閣風景區	6
花蓮文化創意產業園區	5
慶修院	3
馬太鞍濕地生態館	3
太平洋公園	2
花蓮火車站	2
吉蒸牧場	2
七星潭風景區	2
砂卡礑步道	1
花蓮海洋公園	1

3.4.3 單點式關聯規則分析

為解決支持度設定過低導致雜訊增生的問題，本研究提出「單點式關聯規則分析」之方法。單點式關聯規則分析透過先找出資料庫中同樣有拜訪特定興趣點的用戶以成立子資料庫，藉以讓子資料庫裡用戶間的旅運行為更加相近，進而提升資料探勘時所設定的最小支持度，避免因最小支持度設定過低造成的雜訊問題。圖 3-6 為單點式關聯規則分析流程圖，先對資料庫中的每個資料進行個數統計，接著再替資料量較多者進行子資料庫的建立，最後再各別對每個子資料庫進行關聯規則分析，即為本研究提出之單點式關聯規則分析之步驟。

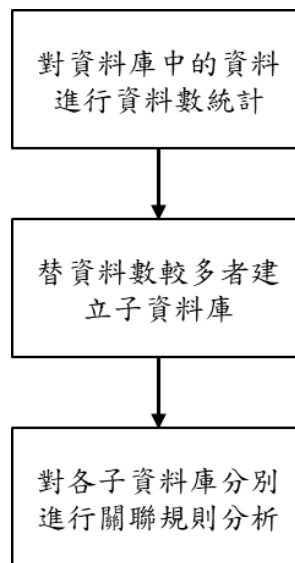


圖 3-6 單點式關聯規則分析流程圖

以表 3.11、表 3.12 以及表 3.13 為例，進行單點式關聯規則分析的示範。在進行單點式關聯規則時，首先對各個景點進行拜訪人數統計，表 3.12 即為表 3.11 的人數統計結果表。隨後挑選拜訪人數排名前 50%以及前 25%的興趣點分別進行子資料庫的篩選，以挑選拜訪人數排名前 50%的興趣點為例，也就是挑選表 3.12 中十個興趣點裡拜訪人數排名為前五的興趣點，即太魯閣風景區、花蓮文化創意產業園區、慶修院、馬太鞍濕地生態館以及太平洋公園，並依序對上述的五個興趣點建立其各自的子資料庫，而表 3.13 即是以太魯閣風景區為例的子資料庫。透過表 3.12 可得知原本表 3.11 的十名觀光用戶中，有六名拜訪過太魯閣風景區，因此表 3.13 會有該六名用戶及其興趣點集合之資料。替上述五個興趣點建立完如表 3.13 的子資料庫後，再依序對每個子資料庫進行傳統的關聯規則分析，即可完成本研究提出之單點式關聯規則分析。

單點式關聯規則即是對各個子資料庫分別進行關聯規則分析之方法，且由於欲進行資料探勘的資料庫是已經做過篩選動作的子資料庫，因此在設定最小支持度時可設定得較起初對整個資料庫進行關聯規則分析時來得高。以表 3.11 和表 3.13 為例，可看出表 3.11 的資料庫須將最小支持度設定為 0.2，也就是十名用戶中至少有兩名用戶有該旅運行為，才能找到「若拜訪太魯閣風景區則拜訪花蓮文化創意產業園區」結果樣態，且會有「若拜訪七星潭風景區則拜訪吉蒸牧場」的雜訊產生。然而透過單點式關聯規則的方法，對表 3.13 進行關聯規則分析，可發現欲找出「若拜訪太魯閣風景區則拜訪花蓮文化創意產業園區」結果樣態時，其最小支持度可提高設定至 0.33(六名用戶中有兩名用戶同時拜訪此兩興趣點)，且由於「若拜訪七星潭風景區則拜訪吉蒸牧場」之樣態在此資料庫的支持度為 0.167，也就是六名用戶中僅有一名用戶有此旅運行為，因此此樣態便不會被探勘出來，說明單點式關聯規則對於雜訊的排除是可行的。

表 3.13 太魯閣風景區之子資料庫範例

使用者	興趣點集合
2c03a1	太魯閣風景區、花蓮文化創意產業園區
42e49e	砂卡礑步道、太魯閣風景區、慶修院
1fa735	太魯閣風景區、花蓮火車站、花蓮文化創意產業園區
32sd7f	太魯閣風景區、吉蒸牧場、七星潭風景區
aer4tq	花蓮海洋公園、太魯閣風景區、太平洋公園
acv715	馬太鞍濕地生態館、太魯閣風景區

另一方面，從表 3.11 以及表 3.12 可得知拜訪過砂卡礑步道以及花蓮海洋公園的觀光用戶都各僅有一人，此時，若沒有先替各個興趣點進行人數統計，即直接進行單點式關聯規則分析，則可發現與砂卡礑步道或是花蓮海洋公園有關的樣態其支持度皆非常高，但實際拜訪人數卻只有一人。以花蓮海洋公園為例，若替花蓮海洋公園建立子資料庫，則該子資料庫中只會含有用戶 aer4tq 及其拜訪的興趣點集合資料，也因此花蓮海洋公園、太魯閣風景區以及太平洋公園三者任兩者構成的組合，亦或是三者一起的結果樣態，其樣態支持度皆為 1.0，在最終分析流程上此結果會被視為具有非常強烈的關聯性且應被改善，但實際拜訪的人數僅有一人，也就是說若加入至台灣好行路線的興趣點為上述三個興趣點，則可預期路線改善後能得到的效益並不大。因此在進行單點式關聯規則時，需先篩選原始資料庫中可建立子資料庫之元素，不可直接替全部的元素都建立其子資料庫。

根據上述範例，可看出單點式關聯規則確實能透過篩選子資料庫的方式讓雜訊的問題得以解決。而「雜訊」出現的機率降低，即可以使在最後的分析步驟時能避免雜訊所造成的影響，如非頻繁被拜訪的興趣點在最後被視為是需要改善的服務缺口，進而導致在調整路線後依舊無法提升觀光公共運輸的使用率，造成資源的浪費。不僅如此，根據關聯規則的原理，擁有越高支持度的樣態表示該樣態在整個資料庫出現的比例越高，也代表樣態中的資料其關聯性越強，而單點式關聯規則亦可有效地提升最小支持度的設定，藉以發現資料庫中擁有較高關聯性的資料集。

3.5 比對台灣好行服務路線與提出服務改善建議

透過第 3.4 小節的資料探勘步驟，即可得知目標用戶之間較常拜訪的興趣點組合為何，將其視為觀光運輸服務潛在需求，再與台灣好行現有的路線進行比對，以找出是否存有觀光公共運輸服務缺口。比對方式為若關聯規則結果的先導規則以及後繼規則皆能被台灣好行路線服務，則該結果樣態即算是可被滿足；反之，若先導規則或後繼規則中有任一無法被台灣好行路線服務，即算是無法被滿足。部分樣態的先導規則會是由多個興趣點所組成得集合而非僅含單一興趣點，因此在比對時須注意該先導規則所包含的所有興趣點皆能被台灣好行路線服務才算是可被滿足。此外，由於本研究比對時是將觀光用戶兩天的信令資料與台灣好行路線進行比對，表示該旅運行為是

可以分成多天完成的，並不侷限於單日甚至同一條觀光路線上，也因此在此比對時不僅限皆要同時被同一條台灣好行路線服務才算是可被滿足，只要先導規則與後繼規則皆能被台灣好行路線所服務到，該樣態即算是可被滿足。

以表 3.14 和表 3.15 為例，說明本研究將關聯規則結果於比對台灣好行路線之方法，其中表 3.14 為台灣好行範例路線表，表 3.15 為關聯規則分析結果於台灣好行服務比對之範例表。根據表 3.14 假設觀光當地有兩條台灣好行路線服務，分別為台灣好行路線 1 以及台灣好行路線 2，且路線 1 依序經過興趣點 A、B、C、D，再從興趣點 D 折返駛回興趣點 A，而路線 2 則依序經過興趣點 1、2、3、4，再從興趣點 4 折返駛回興趣點 1。關聯規則分析結果如表 3.15 所示共有九個結果樣態，根據上一段落所提及之比對方式，可看出樣態編號第 1 及第 2 的結果樣態其先導規則與後繼規則可被路線 1 所服務，因此在可否被滿足欄位處以「V」表示。編號第 3 的結果樣態其先導規則與後繼規則可被路線 2 所服務，而編號第 4、第 5 以及第 6 的結果樣態雖然無法被單一條路線所服務，但若同時將兩條路線納入考量，則此三個結果樣態的先導規則與後繼規則皆能被當地的觀光公共運輸系統所服務，故上述的六個樣態皆視為可被服務，並在可否被服務欄位處標記「V」。而編號第 7、第 8 以及第 9 的結果樣態部分，可看出即便同時比對兩條台灣好行的路線，此三個樣態皆有至少一個元素無法被服務到，如編號 7 樣態的站點 F、編號 8 以及編號 9 樣態的站點 6，因此此三個樣態為無法被滿足之樣態，在可否被滿足欄位便不予以標記。

表 3.14 台灣好行路線範例路線表

台灣好行路線 1	A→B→C→D→C→B→A
台灣好行路線 2	1→2→3→4→3→2→1

表 3.15 關聯規則分析結果於台灣好行服務比對之範例表

樣態編號	關聯規則分析結果		可否被滿足
	LHS	RHS	
1	A	C	V
2	{C、D}	A	V
3	{2、4}	1	V
4	1	C	V
5	{A、D}	2	V
6	{A、3}	C	V
7	B	F	
8	{1、6}	3	

9	{B、4}	6	
---	-------	---	--

最後，蒐集所有無法被滿足的樣態，並透過本研究所提出的兩個評判方式進行分析後，得出哪些興趣點於台灣好行的路線上增加對整體觀光公共運輸系統是有益的，達到將研究結果回饋於現有觀光公共運輸服務環境的目的。本研究根據「改善樣態數」以及「樣態支持度」兩種評判方法分別檢視是否有需要納入台灣好行路線的興趣點。改善樣態數即表示若增加某興趣點在台灣好行的路線中，所增加的可被滿足樣態數量。以表 3.15 為例，若新增興趣點 F 進台灣好行的路線中，能改善的樣態僅有編號 7 一筆，但若新增的興趣點為站點 6，則三筆原先無法被滿足的樣態中能有編號 8 以及編號 9 兩筆得到改善。以改善樣態數的觀點而言，改善興趣點 6 的效益來得比改善興趣點 F 高，故在最終選擇興趣點 6 作為改善的建議。

樣態支持度為本研究利用關聯規則之結果有著越高支持度表示該樣態中的資料其關聯性越強的特性，作為衡量興趣點改善優先順序的依據之一。本研究透過重複操作一般關聯規則分析，並隨著操作次數逐漸提升設定的最小支持度，且在每完成一次關聯規則分析時即對結果進行與路線的比對，藉以得知各個無法被滿足的興趣點其支持度為何。舉例來說，當最小支持度設定為 0.01 時，可找到 X 以及 Y 兩個無法被滿足的興趣點，但當最小支持度提升至 0.02 時，則只能找到 Y 一個無法被滿足的興趣點，此狀況說明含有興趣點 X 的樣態其支持度為一個大於 0.01 但小於 0.02 的值，也說明含有興趣點 Y 的樣態其支持度會相較於 X 大，與樣態中其他興趣點關連性亦較 X 強，因此在資源有限的情況下應優先改善興趣點 Y。

測試方式為設定一個最小支持度的起始值及其提升間隔，隨後將用戶的旅運資訊自最小支持度的起始值開始執行一般關聯規則分析，完成後隨即將最小支持度的設定提升一個間隔，並依更新後的最小支持度再度對用戶的旅運資訊執行關聯規則分析，直至被探勘出之無法被滿足的興趣點僅剩一個。每當完成一次關聯規則分析時便立即與台灣好行路線進行比對，找出哪些興趣點為無法被台灣好行所服務之興趣點並記錄其當時執行一般關聯規則的最小支持度。當完成多次的關聯規則分析後，即可藉由觀察統計結果得知各個造成樣態無法被滿足的興趣點會在支持度提升至哪個特定值時消失於統計表上，也說明該興趣點所具有的支持度至高為何。

最小支持度起始值設定為欲探勘出結果規則時的最下限值(以下稱為單位支持度)，也就是說資料庫中至少一名用戶有該旅運行為的比例。舉例來說，若資料庫中共有 50 名用戶，最小支持度的起始值則設定為 $1/50=0.02$ 。在此設定下，由於最小支持度起始值的設定為資料庫探勘到具關連性樣態的最下限值，因此可確實地觀察到資料庫中所有樣態的支持度為何。而提升區間的部分，亦設定為單位支持度或是其倍數，設定為單位支持度係以各資料庫在進行關聯規則分析時的最小單位為考量來設計，舉例來說，同樣假設資料庫中有 50 名用戶，但探勘具關連性樣態的條件從至少一名用戶有該旅運行為變為至少兩名用戶有該旅運行為，表示最小支持度從 0.02 提升至 0.04，也就是提升一個單位支持度。然而，考量到需進行資料探勘的資料庫其資料量通常較大，加上最小支持度增減一兩個最小單位對於探勘結果的影響通常也較低，因此，在提升區間的設定亦可適度地設定為單位支持度的倍數值來減少資料探勘時所需花費的時間。

以圖 3-7 為範例說明，且假設資料庫中有 50 名用戶，因此最小支持度的起始值為 0.02，而提升區間也設定為 0.02。圖 3-7 左方的四個表格為在各個不同的最小支持度時能找出使樣態無法被滿足之興趣點，如：當最小支持度為 0.02 時，探勘出來的樣態中無法被台灣好行所滿足的興趣點有興趣點 E、F、G、H、5、6 以及興趣點 7。而從圖 3-7 可得知，當最小支持度提升至 0.04 時，興趣點 E 及興趣點 5 已無法再被探勘到，也就是說當最小支持度提升至 0.04 時，結果樣態中已不包含含有興趣點 E 及興趣點 5 的樣態，此時便在圖 3-7 右方表格標記興趣點 E 以及興趣點 5 於最小支持度 0.04 時無法再被探勘出來，以此類推，即可得知興趣點 6、7 以及興趣點 F 於最小支持度為 0.10 時無法再被探勘出來。而當最小支持度提升至僅能探勘出一個興趣點無法被滿足時，即停止繼續提升最小支持度並同時停止關聯規則分析的執行。有此設定的原因為若最小支持度已提升至僅能探勘出一個無法被滿足的興趣點，說明含有該興趣點之樣態為資料庫中擁有最高的支持度者，亦表示在改善優先順序上已找到最需要被改善者，故無需再繼續提升最小支持度及執行關聯規則分析。最終可透過支持度統計表，即圖 3-7 右方表格，得知各個興趣點所擁有的支持度至多為何，並以此作為判斷改善興趣點的優先順序依據，來提出相對應之服務建議。

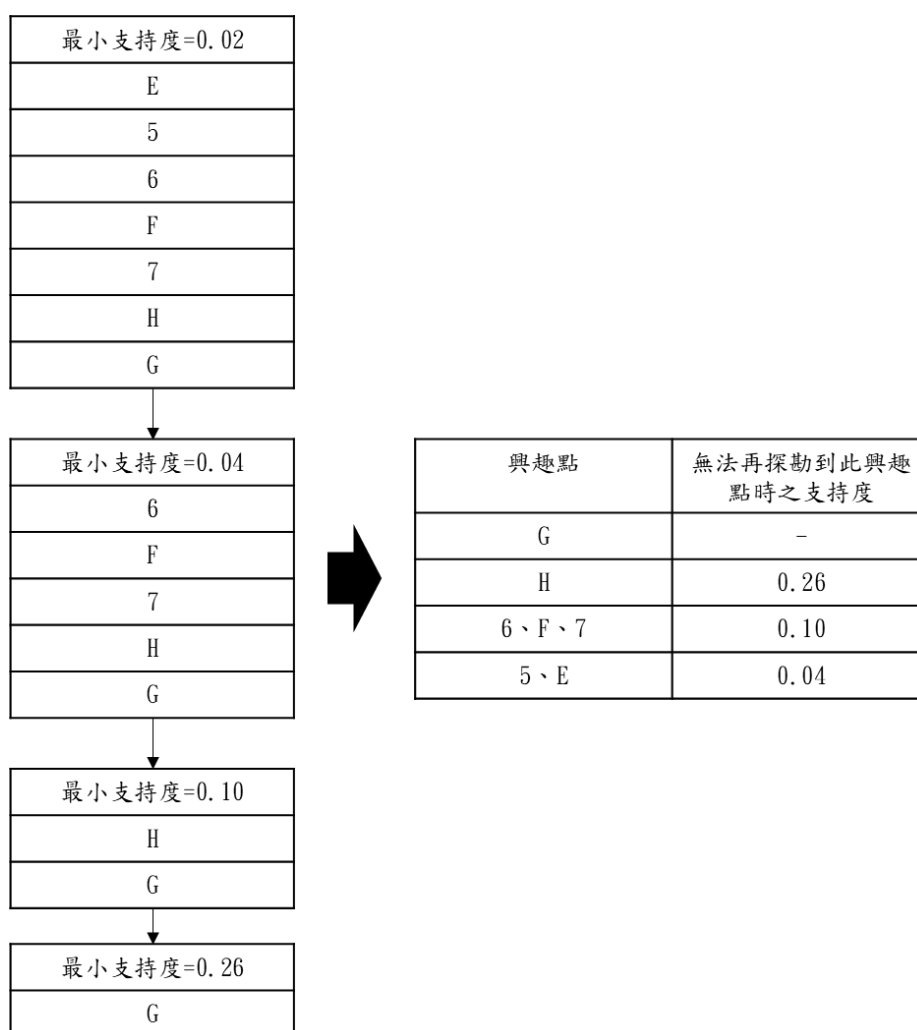


圖 3-7 樣態支持度統計範例圖

由於單點式關聯規則是藉由先篩選子資料庫後並對每一個子資料庫執行關聯規則的一種衍生方法，為了避免在統計時有重複計算、忽略計算等情況發生，「改善樣態數」的評判方法不適用於單點式關聯規則結果的分析。舉例來說，若興趣點 Z 為一個無法被台灣好行路線服務的興趣點、興趣點 A 以及 B 則皆能被台灣好行路線所服務，且「若 A 、 B 則 Z 」為一個結果樣態，則在計算改善樣態數時，興趣點 Z 會被計算為可改善一個樣態。但由於單點式關聯規則會替各個興趣點建立子資料庫，也就是說「若 A 、 B 則 Z 」樣態可能會重複出現在興趣點 A 的子資料庫分析結果、興趣點 B 的子資料庫分析結果以及興趣點 Z 的子資料庫分析結果，此時興趣點 Z 的改善樣態數會從一個樣態提升至三個樣態。另一方面，由於單點式關聯規則會先進行興趣點的人數統計，此時倘若興趣點 A 、 B 及 Z 皆非為拜訪人數排序在前 50% 之興趣點，則「若 A 、 B 則 Z 」樣態會因此無法被探勘到，而興趣點 Z 的改善樣態數也從一個樣態變為零個樣態。根據上述原因，本研究在進行「改善樣態數」之統計時，將不會包含單點式關聯規則的結果。

部分結果樣態為 LHS 與 RHS 彼此互換之樣態，如「若 A 則 B 」與「若 B 則 A 」，本研究將此類由相同興趣點組成的樣態視為彼此獨立的樣態，並各別分析其改善樣態數以及樣態支持度。此類由相同興趣點組成之樣態，除了其支持度為相同之外，信賴度以及增益值皆有可能不同，如前者之信賴度為 $P(B|A)$ 而後者則為 $P(A|B)$ ，也就是說此類型的樣態在最終的探勘結果中可能皆被探勘出或僅探勘出其中一個。因此，即便探勘結果中存在兩個樣態惟 LHS 與 RHS 彼此互換，也不可將其視為相同的樣態，需視為是彼此不同且獨立的樣態並進行後續的分析。

將關聯規則分析結果與台灣好行路線進行比對，再分析各結果的改善樣態數以及樣態支持度後，本研究將較需要被改善的興趣點進行空間與時間的分析，以提出服務改善建議。空間分布的部分為觀察需要被改善的興趣點其地理位置，並決定將建議其加入至該縣市中的哪一條台灣好行路線中。時間部分則是觀察將特定興趣點加入至台灣好行路線後，增加的旅行時間是否有比原先各站之間的平均旅行時間高，舉例來說，若要在原路線的特定路段 $A \rightarrow B$ 中新增一興趣點 C ，使該路段變為 $A \rightarrow C \rightarrow B$ ，則本研究將觀察 $A \rightarrow C$ 的旅行時間與 $C \rightarrow B$ 的旅行時間是否有任一段比原先的站間平均旅行時間長，若有，則建議該興趣點的改善方式採用部分班次延駛或繞駛的方式進行；若沒有，則建議將該興趣點加入至常規的路線中。

3.6 小結

根據本章節之敘述，可得知本研究之主要概念為透過信令資料得出觀光客的旅運行為，再藉由資料探勘的方法得到觀光客的旅運需求，透過與台灣好行路線的比對以得知是否有服務缺口的存在，並提出相對應之改善建議。而本研究為將運研所(2018)之研究為基礎，將其資料探勘以及結果分析的部分作精進之研究。

在資料探勘的部分，運研所(2018)透過關聯規則分析來得知用戶的旅運需求。然而如第 3.4.2 節所述，傳統的關聯規則分析可能會因資料庫中的資料較複雜導致雜訊的產生，此時若沒有適時的判斷出部分探勘結果實屬於雜訊並將其作為最後提出服務改善的依據，則可能導致改善後成效不彰等情況發生。本研究為改善此問題，提出單點式關聯規則分析，透過先一步篩選的動作藉以讓子資料庫中的資料較為單純，使得最小支持度能適度地提升，並有效地紓解雜訊的產生。

另一方面，運研所(2018)針對探勘節的解讀僅利用改善樣態數一指標進行。而改善樣態數的意義即為觀察哪個興趣點與最多台灣好行路線的興趣點有關聯性，此指標固然是在分析時一個優秀的指標，然而本研究認為只使用改善樣態數會有著評判方式過於單一，導致部分重要的結果被忽略的情況發生，因此新增了樣態支持度作為第二個結果分析的指標。樣態支持度在探討興趣點與台灣好行路線的關聯性強烈與否，並不考慮是否與路線上的多個興趣點有關聯性存在，也因此可以找出部分只用單一指標無法分析出的結果。舉例來說，部分興趣點與台灣好行路線存有很強烈的關聯性，但由於該興趣點與台灣好行路線上具有關聯性的站點只有少數幾個甚至只有一個，導致此興趣點在只使用改善樣態數一指標的情況下會被視為是改善優先順序較低的興趣點。但若同時使用改善樣態數以及樣態支持度兩個指標，則可以避免上述得情況發生，使分析結果更全面。

第四章 實際範例測試

本節將說明取得的資料實際套用進第三章之研究方法的結果。本研究取得之花蓮地區的觀光用戶信令資料為 2018 年 5 月 19 日至 5 月 20 日、5 月 26 日至 5 月 27 日以及 6 月 2 日至 6 月 3 日共計三個周末的信令資料，共計有 61696 名觀光用戶，而六天的資料量如表 4.1 所示，表 4.1 第一列為此三個周末之日期，而第二列則為該日期所含有的信令資料筆數，舉例來說，本研究所蒐集之 5 月 19 日的觀光用戶信令資料，共有 7691579 筆，而 5 月 20 日則有 3913119 筆資料。此外，由於本研究之目的為探討用戶於觀光景點間的移動行為，因此分析結果將不會呈現出包含非觀光景點之樣態，意即結果樣態無論是先導規則或後繼規則皆不會含有台鐵車站的興趣點，台鐵車站的興趣點資料僅用於判斷用戶是否為第一類用戶之用途。

表 4.1 取得之觀光用戶信令資料日資料量表

日期	5/19	5/20	5/26	5/27	6/02	6/03
資料數	7691579	3913119	10919705	10987335	9286832	9447877

本章節將分為以下六個部分呈現，分別為興趣點停留時間門檻值測試結果、兩個類別用戶之篩選結果、改善樣態數測試結果、一般關聯規則分析的樣態支持度測試結果、單點式關聯規則分析的樣態支持度測試結果以及台灣好行路線改善建議。將觀光用戶再分為兩個類別的主要目的為透過分析不同類別用戶的旅運行為以及資料探勘的結果，藉以比較使用不同交通工具的旅客是否會有不同的旅運行為；同時，不同群用戶的分析結果也可以作為最後提出台灣好行路線改善建議的依據。此外，三種不同評量方式的測試結果，也就是改善樣態數以及兩種關聯規則的樣態支持度，皆會包含三種類別用戶之測試結果，分別為本研究所取得之全體觀光用戶、使用火車往返觀光地之用戶（第一類用戶）以及使用私人運具往返觀光地之用戶（第二類用戶）。

4.1 興趣點停留時間門檻值測試結果

本研究參考運研所(2018)於探討觀光用戶旅運行為研究所設定的興趣點停留時間門檻值，分別對興趣點停留時間門檻值 1、2 以及 5 分鐘三種不同情況的門檻值，且額外對興趣點停留時間門檻值 10 分鐘進行用戶平均興趣點停留數量的測試。表 4.2 為本研究在使用不同停留時間門檻值的情況下，平均每位用戶在多少個興趣點停留之對應表，而用戶平均停留的興趣點數量在計算時，已將台鐵車站排除。從表 4.2 可看出當停留時間門檻值為 1 分鐘時，平均每位用戶在扣除台鐵車站的停留後，約會於 1.35 個興趣點進行停留；當興趣點停留時間門檻值提升至 5 分鐘時，平均每位用戶會在約 1.18 個興趣點進行停留；但當興趣點停留時間門檻值提升至 10 分鐘時，平均每位用戶僅會在 1.04 個興趣點停留。

從表 4.2 可得知在興趣點停留時間門檻值從 2 分鐘提升至 5 分鐘時，用戶的平均停留點數量便已有一定程度的減少，而當興趣點停留時間門檻至再提升至 10 分鐘時，用戶平均停留興趣點數量已約等於 1。此現象說明當停留時間門檻值設定為 10 分鐘

時，用戶除了台鐵車站外，在其兩天的旅遊途中僅會拜訪一個興趣點，本研究認為此狀況較不符合實際情況。此外，若用戶在旅遊途中僅會拜訪一個興趣點，便失去了利用關聯規則分析找出興趣點間關聯性的意義。因此，本研究認為 10 分鐘的門檻值設定為過高，並參考運研所(2018)之相關研究，將興趣點停留時間門檻值設定為 1、2 以及 5 分鐘。

表 4.2 興趣點停留時間門檻值與用戶平均停留興趣點數量對應表

興趣點停留時間門檻值	用戶平均停留的興趣點數量
1 分鐘	1.35
2 分鐘	1.34
5 分鐘	1.18
10 分鐘	1.04

4.2 兩類別用戶篩選結果

由於本研究所選取之研究場域為花蓮，而能抵達花蓮的公共運輸系統幾乎只能仰賴台鐵，因此本研究在進行實測時，更假設第一類用戶即為使用台鐵往返花蓮的用戶，而其餘不屬於第一類的用戶皆判斷為第二類用戶。能抵達花蓮的公共運輸系統除了台鐵還有公路客運以及飛機，但由於聯外的公路客運僅能抵達台中梨山以及台東地區，以台灣大多數人口分布在西半部都會區的情況來說，觀光客較無可能選擇公路客運前往花蓮；同理，飛機的部分不但僅能與高雄進行往返，且每日去回程的班機皆只有一班，也因此觀光客較無機會選擇利用飛機往返花蓮。根據上述，本研究便可推論非利用台鐵往返花蓮地區之民眾有很大的可能為利用私人運具往返花蓮之族群，也就是第二類用戶。

使用台鐵往返花蓮之用戶，即為本研究所稱的第一類用戶，為本研究所判斷其使用台鐵前往或離開花蓮之觀光用戶。第一類用戶的研究結果可用以檢視有使用公共運輸的觀光族群其旅運行為與觀光公共運輸路線是否有差異，而第二類用戶，也就是使用私人運具往返觀光的之用戶，其比對結果則可檢視較無意願使用公共運輸的觀光族群其旅運行為是否被現有觀光公共運輸系統滿足，也因此本研究將取得之觀光用戶再細分成兩個類別，藉以觀察使用不同運具的用戶之間其旅運行為是否有差異，更可將兩類別用戶與觀光公共運輸的比對結果作為本研究最後提出服務改善建議之依據。而本節即呈現本研究對於第一類用戶以及第二類用戶的篩選結果。

表 4.3 為將三周共計 61696 名觀光用戶根據不同的時間門檻值所得到的兩類別用戶人數表，其中第一列是容忍距離的長度；第二列為三種興趣點停留時間門檻值，分別有 1 分鐘、2 分鐘以及 5 分鐘；第三列為車站停留時間門檻值，本研究根據第三章提及之「車站停留時間門檻值不得大於興趣點停留時間門檻值之限制」，以及運研所(2018)之研究內容，將其設定為 1 分鐘；第四列為最終能篩選出的使用台鐵往返觀光地之用戶，舉例來說，若興趣點停留時間門檻值為 1 分鐘，則能篩選出 35065 名具公

共運輸服務需求之用戶用戶，且佔全體觀光用戶的百分之 56.86。從表 4.3 中可看出，由於興趣點停留時間門檻值的不同，最終得到的第一類用戶數也會跟著改變，會有此現象的原因為隨著興趣點停留時間門檻值的增加，用戶便會被判斷到在越少的興趣點停留，另一方面，由於車站停留時間門檻值並沒有變動，用戶被判斷到在車站的停留行為並不會發生改變，也就是說，隨著停留的興趣點減少但停留的車站數不變，用戶的旅運行為便會越趨向於第一個或最後一個停留點為車站，即用戶的旅運行為會越趨向於本研究用以判斷第一類用戶的旅運行為。

以圖 4-1 為例，圖 4-1 左方表格為興趣點停留時間門檻值為 1 分鐘時，某用戶於拜訪花蓮時的旅運行為表。可發現此用戶由於其第一個及最後一個停留點皆非台鐵車站，因此被定義為第二類用戶；然而，隨著興趣點停留時間門檻值提升至 5 分鐘，此用戶的旅運行為變為圖 4-1 右方表格，可發現因用戶於太魯閣風景區的停留時間不足 5 分鐘，導致在興趣點停留時間門檻值為 5 分鐘的情況下此用戶的第一個停留點為花蓮火車站，也使得此用戶從原本的第二類用戶變為第一類用戶，故隨著興趣點停留時間門檻值的提升，第一類用戶的數量亦會跟著提升。此外，在每種情境下剩餘沒被篩選到之觀光用戶即為使用私人運具往返觀光地之用戶，也就是第五列所展示的人數。以表 4.3 為例，在興趣點停留時間門檻值 1 分鐘的情況下能篩選出 35065 名第一類用戶，因此在興趣點停留時間門檻值 1 分鐘的情況下有 26631 名使用私人運具往返觀光地之用戶，即有 26631 名第二類用戶，以此類推。

表 4.3 兩類別用戶人數表

容忍距離	200		
興趣點停留時間(min)	1	2	5
車站停留時間(min)	1		
使用台鐵往返觀光地之用戶 (第一類用戶)	35065(56.86%)	35210(57.07%)	36439(59.06%)
使用私人運具往返觀光地之用戶 (第二類用戶)	26631(43.14%)	26486(42.93%)	25257(40.97%)

興趣點停留時間門檻值:1			興趣點停留時間門檻值:5	
景點中文名稱	停留時間		景點中文名稱	停留時間
太魯閣風景區	1		花蓮火車站	3
花蓮火車站	3		花蓮糖廠	35
花蓮糖廠	35		七星潭風景區	23
七星潭風景區	23			

圖 4-1 興趣點停留時間門檻值提升與第一類用戶數提升示意圖

4.3 改善樣態數之測試結果

本節將說明將用戶旅運行為的關聯規則分析結果與台灣好行路線比對後，所探勘出的各個能藉由將其新增至台灣好行路線使部份結果樣態能得到改善之興趣點及其能改善的樣態數量。且如同第 3.4 節所述，由於單點式關聯規則分析之結果不適合以樣態改善數作為評量指標，因此在樣態改善數的部分僅會以一般的關聯規則分析之結果來做呈現。

此外，本研究只針對單一興趣點的改善進行探討，並不考慮同時改善兩個興趣點的情況。有此限制的原因為根據興趣點停留時間門檻值 5 分鐘且最小支持度為 0.01 的情況下，全體觀光用戶的關聯規則結果，如表 4.4 所示，可發現關聯規則之結果大多數的樣態為 1 對 1 樣態，即先導樣態和後繼樣態中各只含有一個元素，此時若兩個興趣點皆無法被台灣好行路線所服務，說明此樣態所包含的興趣點與原有的台灣好行路線景點的關聯性不高，若恣意將此兩興趣點加入至台灣好行路線中，不但無法確保其改善效益，甚至因新增站點而增加的額外旅行時間可能會使原先有意使用台灣好行的民眾變為使用其他交通方式。另一方面，若樣態中的元素皆為無法被台灣好行所服務的興趣點，則可透過「新增路線」的方式使其得到改善，然而本研究是以改善現有的台灣好行路線為研究目的，增設新路線之方法並不在本研究之範圍內。同時，由於本研究最後會利用此分析結果作為路線改善的依據，考量到實務面通常會有資源的限制，以及在原路線加入兩個興趣點所帶來的時間成本影響會遠大於只加入一個興趣點所帶來的影響，而關於時間等成本的詳細計算亦不在本研究之研究範圍內，故本研究只針對單一興趣點的改善進行探討，並不考慮同時改善兩個興趣點的情況。

表 4.4 為興趣點停留時間門檻值 5 分鐘且最小支持度為 0.01 的情況下，全體觀光用戶的關聯規則結果表，以此表為範例說明本研究進行改善樣態數的測試過程。圖 4-2 為花蓮地區兩條運行的台灣好行路線圖，圖片上方黃色的站點為縱谷花蓮線行經的站點，下方綠色的站點則為太魯閣線所行經的站點，而圖中的紅色箭頭為去程時所行經的站點順序，綠色箭頭則標示出回程時所行經的站點順序。縱谷花蓮線去回程所行經的站點一致，太魯閣線回程時除了布洛灣改為長春祠步道外，其餘站點也是與去程相同。根據本研究使用的比對方法，可得知表 4.4 的 38 個結果樣態中，可被台灣好行路

線服務者，也就是先導樣態與後繼樣態皆為圖 4-2 所提及之站點，共有 14 個樣態，分別為樣態編號第 1、3、4、5、8、16、17、18、20、23、24、28、29 以及第 30 號者，而其餘編號的樣態則因為至少有一個元素無法被台灣好行路線所服務而被視為無法被滿足，舉例來說，樣態編號第 36 號樣態，由於其先導樣態(LHS)含有一個無法被服務的興趣點(花蓮文化創意產業園區)，因此即使編號 36 的樣態其後繼樣態以及其先導樣態中的其中一個元素(太魯閣國家公園遊客中心)能被台灣好行路線服務，此樣態仍然判定為無法被滿足。

根據表 4.4 以及上一段落的敘述，可得知當中共有 24 個樣態無法被台灣好行滿足。此時，若將花蓮文化創意產業園區加入至台灣好行的路線中，可發現有 15 個樣態因而從原先無法被滿足的狀態轉變為可被滿足，而此 15 個樣態分別為樣態編號第 6、9、13、19、21、22、25、26、27、31、32、33、34、35 以及第 36 號。同理，若將馬太鞍濕地生態館加入至台灣好行的路線中，則有 1 個樣態因此從無法被滿足的狀態轉變為可被滿足，該樣態為樣態編號第 7 號。在此需注意此步驟暫不討論將興趣點加入至哪一條路線中，僅討論將興趣點加入至台灣好行路線後對於關聯規則分析結果的影響，而將興趣點加入至哪條路線的分析會在後續的步驟提及。

表 4.4 興趣點停留時間門檻值 5 分鐘，全體觀光用戶關聯規則結果表
(最小支持度 0.01)

樣態編號	LHS	RHS
1	岳王亭	天祥風景區
2	太平洋公園	a-zone 花蓮文化創意產業園區
3	燕子口步道	太魯閣國家公園遊客中心
4	新光兆豐休閒農場	七星潭風景區
5	新光兆豐休閒農場	太魯閣國家公園遊客中心
6	新光兆豐休閒農場	a-zone 花蓮文化創意產業園區
7	馬太鞍溼地生態館	花蓮觀光糖廠
8	林田山林業文化園區	花蓮觀光糖廠
9	林田山林業文化園區	a-zone 花蓮文化創意產業園區
10	花蓮遠雄海洋公園	a-zone 花蓮文化創意產業園區
11	花蓮東大門國際觀光夜市	a-zone 花蓮文化創意產業園區
12	松園別館	a-zone 花蓮文化創意產業園區
13	慶修院	a-zone 花蓮文化創意產業園區
14	親不知子天空步道	a-zone 花蓮文化創意產業園區

15	花蓮文化中心石雕公園	a-zone 花蓮文化創意產業園區
16	砂卡礑步道	七星潭風景區
17	砂卡礑步道	太魯閣國家公園遊客中心
18	太魯閣國家公園遊客中心	砂卡礑步道
19	砂卡礑步道	a-zone 花蓮文化創意產業園區
20	天祥風景區	太魯閣國家公園遊客中心
21	天祥風景區	a-zone 花蓮文化創意產業園區
22	立川漁場	a-zone 花蓮文化創意產業園區
23	七星潭風景區	太魯閣國家公園遊客中心
24	太魯閣國家公園遊客中心	七星潭風景區
25	七星潭風景區	a-zone 花蓮文化創意產業園區
26	花蓮觀光糖廠	a-zone 花蓮文化創意產業園區
27	太魯閣國家公園遊客中心	a-zone 花蓮文化創意產業園區
28	七星潭風景區, 砂卡礑步道	太魯閣國家公園遊客中心
29	太魯閣國家公園遊客中心, 砂卡礑步道	七星潭風景區
30	七星潭風景區, 太魯閣國家公園遊客中心	砂卡礑步道
31	太魯閣國家公園遊客中心, 砂卡礑步道	a-zone 花蓮文化創意產業園區
32	a-zone 花蓮文化創意產業園區, 砂卡礑步道	太魯閣國家公園遊客中心
33	a-zone 花蓮文化創意產業園區, 太魯閣國家公園遊客中心	砂卡礑步道
34	七星潭風景區, 太魯閣國家公園遊客中心	a-zone 花蓮文化創意產業園區
35	a-zone 花蓮文化創意產業園區, 七星潭風景區	太魯閣國家公園遊客中心
36	a-zone 花蓮文化創意產業園區, 太魯閣國家公園遊客中心	七星潭風景區

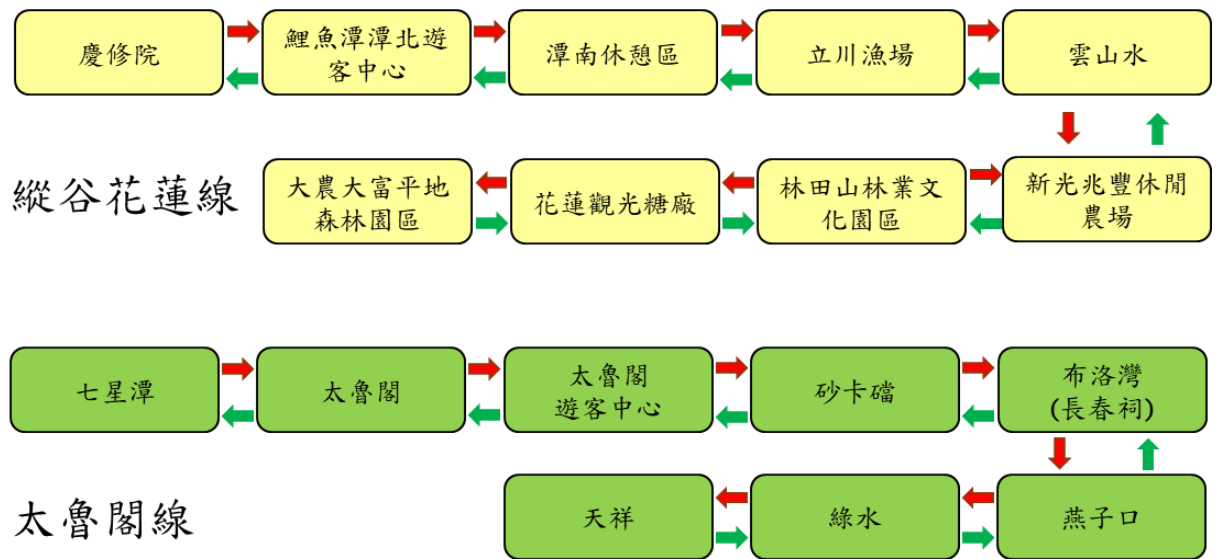


圖 4-2 台灣好行路線圖

本研究所設定用以進行關聯規則分析的最小支持度為 0.005 以及 0.01 兩種，搭配上 3 種停留時間的門檻值，共會產生 6 種的情境，而表 4.5 即是此 6 種情境的樣態改善數結果表。從表 4.5 的第一列為 6 種情境的組合，以停留時間-最小支持度的方式呈現，第二列為特定情境下所能探勘到的結果規則數量，第三列為特定情境下有多少筆的樣態能被台灣好行的路線所服務，第四列即為能被服務的樣態數量佔該情境下全部樣態數量的比例，而從第五列開始為若加入特定的興趣點至台灣好行的服務路線中，各個情境下能有多少筆原先無法被台灣好行路線服務的樣態轉變為能被服務。以表 4.5 最後一欄為例，表示在停留時間門檻值 5 分鐘且最小支持度為 0.01 的情況下，有 36 個結果樣態被探勘出來，且當中有 14 筆樣態能被台灣好行的路線所服務，佔全部樣態數的 38.9%，其中若將花蓮文化創意產業園區加入至台灣好行的路線中，則會有 15 筆原先無法被台灣好行路線所服務的樣態因而轉變為可以被服務，倘若加入台灣好行路線的興趣點是馬太鞍濕地生態館而非花蓮文化創意產業園區，則能從原先無法被服務的狀態轉為能被服務的樣態的樣態數便只剩一筆。

表 4.5 第 5 到第 9 列所列出的 5 個興趣點為興趣點停留時間門檻值為 1 分鐘且最小支持度為 0.005 時，只要將其加入至台灣好行的路線中，便能有樣態因此得到改善之興趣點，且此情境下能找到符合上述條件之興趣點亦只有此 5 個。舉例來說，若只九曲洞隧道加入台灣好行的服務路線中，在興趣點停留時間門檻值為 1 分鐘且最小支持度為 0.005 的 168 個結果樣態中，便能有 3 個樣態因此從原先的無法被台灣好行路線所服務轉變為可被服務。且表 4.5 從左往右，對於找到結果樣態的條件即越嚴苛，如：停留時間門檻值的增加或是最小支持度的提升，也因此某些興趣點從原先能被探勘到變為無法被探勘出來，如：包含九曲洞隧道的樣態在興趣點停留時間門檻值為 1 分鐘且最小支持度為 0.005 時能被探勘出來，但當最小支持度上升為 0.01，或是停留時間門檻值上升為 2 分鐘時，便不再能探勘出包含九曲洞隧道的樣態，而對於這些在特定情境中無法被探勘出的興趣點，本研究以「-」表示。某些興趣點雖然在特定情境下能被探勘出來且視為服務缺口，但由於該興趣點所屬樣態的改善牽涉到的興趣點有 2 個或以上，故表 4.5 將不會列出此類型的興趣點，以表 4.4 為例，編號第 11 號樣態為

「若拜訪東大門夜市，則拜訪花蓮文化創意產業園區」，此時，由於僅改善東大門夜市或是花蓮文化創意產業園區皆無法使此樣態因此轉變為能被台灣好行所服務之樣態，因此如同東大門夜市此類型的興趣點將不會於表 4.5 中列出。

表 4.6 以及表 4.7 分別為第一類用戶關聯規則結果樣態與台灣好行路線比對結果表以及第二類用戶關聯規則結果樣態與台灣好行路線比對結果表。從表 4.5、表 4.6 以及表 4.7 可發現，由於興趣點停留時間門檻值 1 分鐘且最小支持度 0.005 為 6 種情境中對於找出結果規則條件最寬鬆者，也因此無論是哪種分類的用戶，皆可在在此情境找出最多可用來改善台灣好行路線之興趣點。隨著興趣點停留時間門檻值的提升，或是最小支持度的提升，各分類的用戶在最終皆會只剩下花蓮文化創意產業園區以及馬太鞍濕地生態館兩個興趣點可用以將樣態進行改善，且當中又以花蓮文化創意產業園區能改善的樣態數最多。因此，根據本節欲探討的目的，也就是欲找出能改善最多樣態數的興趣點作為首要改善目標，本研究建議若欲改善台灣好行的路線使其能更貼近觀光民眾的需求，則首先應將花蓮文化創意產業園區加入至台灣好行的路線中，而資源許可的狀況下，則可另外考慮將馬太鞍濕地生態館也一併加入至路線中。

表 4.5 全體觀光用戶關聯規則結果樣態與台灣好行路線比對結果表

停留時間(分鐘)-支持度	1-0.005	1-0.01	2-0.005	2-0.01	5-0.005	5-0.01
規則總筆數	168	57	117	49	89	36
能被服務之規則數	64	22	51	20	35	14
能被服務之規則數佔比(%)	38.1	38.6	43.6	40.8	39.3	38.9
花蓮文化創意產業園區	57	22	44	18	38	15
九曲洞隧道	3	-	-	-	-	-
馬太鞍濕地生態館	3	1	2	1	1	1
花蓮文化中心石雕公園	2	-	-	-	-	-
瑞穗牧場	1	-	1	-	1	-

表 4.6 第一類用戶關聯規則結果樣態與台灣好行路線比對結果表

停留時間(分鐘)-支持度	1-0.005	1-0.01	2-0.005	2-0.01	5-0.005	5-0.01
規則總筆數	223	86	134	65	111	51
能被服務之規則數	82	40	53	26	38	18

能被服務之規則數佔比(%)	36.8	46.5	39.6	40.0	34.2	35.3
花蓮文化創意產業園區	73	31	59	28	55	25
花蓮文化中心石雕公園	11	-	-	-	-	-
九曲洞隧道	5	-	-	-	-	-
太平洋公園	3	-	-	-	-	-
馬太鞍濕地生態館	3	1	3	1	2	1
親不之子天空步道	1	-	1	-	1	-
瑞穗牧場	1	-	1	-	1	-

表 4.7 第二類用戶關聯規則結果樣態與台灣好行路線比對結果表

停留時間(分鐘)-支持度	1-0.005	1-0.01	2-0.005	2-0.01	5-0.005	5-0.01
規則總筆數	92	38	91	38	53	24
能被服務之規則數	36	15	36	14	19	9
能被服務之規則數佔比(%)	39.1	39.5	39.6	36.8	35.8	37.5
花蓮文化創意產業園區	27	13	26	13	16	6
馬太鞍濕地生態館	2	1	1	1	2	1
九曲洞隧道	1	-	1	-	-	-
瑞穗牧場	1	-	1	-	1	-

4.4 樣態支持度之測試結果(一般關聯規則)

樣態支持度為經過關聯規則分析後所探勘出來無法被服務的興趣點，其需要被改善的優先順序為何的另一個指標。如第 3.4.3 節所述，若樣態擁有越高的支持度，表示觀光客有越高的機率將該樣態中的所有元素，也就是該樣態所包含的所有興趣點一併安排進旅程裡，故本研究認為在資源有限的情況下，應選擇擁有較高支持度的興趣點作為改善的目標。本研究透過重複執行關聯規則分析並隨著執行的次數逐步提升所設定的最小支持度來找出各個結果樣態的支持度，而本節即展示一般關聯規則結果分析在三個用戶分類以及三種興趣點停留時間門檻值情況下的樣態支持度結果。

本研究於一般關聯規則分析的樣態支持度部分設定的最小支持度起始值為 0.005，而提升間距為每次提升 0.002。最小支持度區間的下限設計不採用單位支持度(也就是 $1/61696=0.00002$)而是採用 0.005 為考量到在第 4.3 節樣態改善數的關聯規則分析時

本研究採用 0.005 以及 0.01 作為最小支持度的設計，故取當中較小的值作為此節研究的最小支持度起始值。另一方面，在最小支持度提升區間的方面，本研究起初以單位支持度 0.00002 作為遞增區間，但從結果來看可發現各無法被滿足之興趣點其支持度間距皆會大於 0.001，考量到時間成本的因素，本研究最終以 0.001 作為最小支持度的遞增區間。

表 4.8 為全體觀光用戶進行多次關聯規則分析後的樣態支持度統計表，表中顯示全體觀光用戶於 3 中不同的停留時間門檻值狀況下，各個探勘出來的興趣點所擁有的最大支持度為何。與第 4.3 節統計樣態改善數時的邏輯相同，當樣態無法被台灣好行路線滿足係由於樣態中的單一興趣點導致，才會計算該興趣點之支持度並記錄於表 4.8。從表 4.8 可得知，當興趣點停留時間門檻值為 1 分鐘且最小支持度為 0.005 時，可藉由將花蓮文化創意產業園區、馬太鞍濕地生態館、九曲洞隧道、瑞穗牧場以及花蓮文化中心石雕公園加入至台灣好行路線中來使得特定的樣態自無法被滿足轉變為可被滿足，但當最小支持度提升至 0.006 時，將花蓮文化中心石雕公園加入至台灣好行路線已無法讓特定的樣態從無法被滿足轉變為可被滿足，說明此興趣點的支持度介於 0.005 至 0.0059 之間。同理，當最小支持度提升至 0.019 時，已無法再藉由加入馬太鞍濕地生態館至台灣好行的路線中來達到改善樣態的目的。

而當最小支持度提升至特定值時，探勘結果顯示僅能透過加入花蓮文化創意產業園區至台灣好行路線以達到改善樣態的目的，也就是達到停止繼續操作關聯規則分析的條件，因此其標記為「-」。另一方面，由於在興趣點停留時間門檻值為 2 分鐘以及 5 分鐘的情況下，即便最小支持度為 0.005，皆無法藉由新增花蓮文化中心石雕公園以及九曲洞隧道來達到改善樣態的目的，因此此二個興趣點在 2 分鐘以及 5 分鐘的情境下其支持度標註為「NA」。

表 4.8 全體觀光用戶興趣點支持度結果表

興趣點 \ 情境(分鐘)	1	2	5
花蓮文化創意產業園區	-	-	-
馬太鞍濕地生態館	0.019	0.019	0.017
瑞穗牧場	0.008	0.009	0.009
九曲洞隧道	0.007	NA	NA
花蓮文化中心石雕公園	0.006	NA	NA

表 4.9 第一及第二類用戶興趣點支持度結果表

情境(類別-分鐘) 興趣點	第一類-1	第一類-2	第一類-5	第二類-1	第二類-2	第二類-5
花蓮文化創意產業園區	-	-	-	-	-	-
馬太鞍濕地生態館	0.019	0.019	0.019	0.019	0.019	0.015
花蓮文化中心石雕公園	0.011	NA	NA	NA	NA	NA
親不知子天空步道	0.009	0.009	0.011	NA	NA	NA
九曲洞隧道	0.009	NA	NA	0.007	0.007	NA
瑞穗牧場	0.007	0.007	0.007	0.011	0.011	0.009
太平洋公園	0.006	NA	NA	NA	NA	NA

表 4.9 為第一及第二類用戶興趣點支持度結果表，呈現方式與表 4.8 相同，惟第一列的屬性欄部分從僅顯示興趣點停留時間門檻值轉變為用戶類別與興趣點停留時間門檻值一併顯示。綜合表 4.8 及表 4.9，即可得知三個分類的用戶分別在特定的興趣點停留時間門檻值之下的興趣點支持度為何，且可觀察到無論在何種情境下，花蓮文化創意產業園區皆是擁有最高支持度者，因此若在資源有限的情況下，就樣態支持度的角度來看，應優先選擇花蓮文化創意產業園區進行改善。此外，可看出馬太鞍濕地生態館所擁有的樣態支持度約為 0.2，可將其視為擁有高支持度的興趣點，故若交通當局的資源允許的話，本研究認為馬太鞍濕地生態館也應受到重視並將其加入至台灣好行的路線中。

4.5 樣態支持度之測試結果(單點式關聯規則)

此章節與第 4.4 節一樣為透過找尋樣態的支持度，藉以得知各個被視為服務缺口且需改善的興趣點其擁有的支持度大小為何，並以此做為衡量各個興趣點被改善的優先順序，使本研究提出的服務建議更有可信度且可行性。然而，此章節與第 4.4 節不同處在於此節之結果為透過本研究提出之單點式關聯規則的測試結果，而單點式關聯規則可透過篩選子資料庫的方式，藉以過濾掉雜訊，甚至也因此可讓原本已具有強烈關聯性的樣態得以有更高的支持度，更能凸顯該規則的強烈關聯性。

在進行單點式關聯規則分析時，為避免某些興趣點的拜訪人數過少而導致與該興趣點有關的樣態會有較高的支持度使得最終結果失真，本研究首先替 3 種類別用戶以及 3 種停留時間門檻值共 9 種的組合各自建立其 55 個觀光景點進行拜訪人數統計表，並僅針對拜訪人數前一半以及前四分之一的觀光景點分別進行單點式關聯規則分析。表 4.10 為拜訪人數統計表範例，以全體觀光用戶、興趣點停留時間門檻值 1 分鐘的統計結果為範例展示，第一、三欄為觀光景點的名稱，第二、四則為相對應景點的拜訪人數。舉例而言，花蓮文化創意產業園區在全體觀光用戶中有 13367 名曾在研究期間內拜訪過，而太魯閣國家公園則有 5321 名用戶拜訪過，以此類推。表 4.10 僅拜訪人數排名前百分之 50 的景點，因此僅含有 28 個景點。

表 4.10 興趣點停留時間門檻值 1 分鐘觀光景點拜訪人數表(全體觀光用戶)

觀光景點	拜訪人數	觀光景點	拜訪人數
花蓮文化創意產業園區	13367	燕子口步道	1669
太魯閣國家公園	5321	新光兆豐休閒農場	1467
花蓮觀光糖廠	4762	新社梯田	1462
七星潭風景區	4166	鯤魚潭風景遊憩區	1427
花蓮文化中心石雕公園	3554	林田山林業文化園區	1344
天祥風景區	3289	雲山水	1312
立川漁場	3105	靜浦北迴歸線界標	1304
砂卡礑步道	2640	馬太鞍濕地生態館	1141
東大門國際觀光夜市	2620	岳王亭	1124
松園別館	2603	石梯坪	1077
慶修院	2114	錐麓古道	1002
親不知子天空步道	2041	台開心農場	957
花蓮遠雄海洋公園	2027	瑞穗牧場	881
太平洋公園	2005	舊長虹橋	790

隨後，替欲進行單點式關聯規則的興趣點建立其各自的子資料庫，並如同第 4.4 節的方式，對每個子資料庫執行多次的關聯規則分析，以得到興趣點的樣態支持度，藉以作為衡量興趣點改善優先順序的依據。舉例來說，若欲進行興趣點拜訪人數前 50% 之單點式關聯規則分析，則需替表 4.10 中顯示的所有 28 個興趣點各自建立其子資料庫後，再依序對各個子資料庫進行多次關聯規則分析；若欲進行興趣點拜訪人數前 25% 之單點式關聯規則分析，則需建立子資料庫的興趣點僅有表 4.10 左半邊所列出的興趣點，並同樣對這些興趣點進行多次關聯規則分析。

最小支持度起始值設定為單位支持度，並採用拜訪人數最多者的單位支持度為選擇。舉例來說，若欲執行興趣點停留時間門檻值 1 分鐘的全體觀光用戶單點式關聯規則分析，則此情境由於拜訪人數最多者為花蓮文化創意產業園區，又其子資料庫的單位支持度為 $1/13367$ ，約等於 0.0001，因此在操作此情境之單點式關聯規則分析時，最小支持度起始值以及提升間隔皆設定為 0.0001。同理，本研究再計算其他情境時，拜訪人數最多的興趣點皆為花蓮文化創意產業園區，且單位支持度皆約為 0.0001，因此下方表格呈現結果時皆會顯示到小數點後第四位數。

表 4.11 為 3 個類別觀光用戶之興趣點拜訪人數前 50% 之單點式關聯規則興趣點支持度結果表，並顯示支持度大於 0.2 者，而顯示「NA」者表示包含該興趣點的樣態在特定情境下皆沒有支持度大於 0.2 者。可透過表 4.11 得知，當進行單點式關聯規則，且興趣點拜訪人數取前 50% 時，馬太鞍濕地生態館的支持度無論在何種情境皆會是最高者。不僅如此，亦可從表 4.11 得知，當用戶類別為第二類用戶時，瑞穗牧場所擁有的支持度會接近花蓮文化創意產業園區。根據第 4.4 節的表 4.8 可得知，在沒有進行子資料庫篩選的情況下，雖然可看出馬太鞍濕地生態館與台灣好行路線站點具有高關聯性，但由於雜訊的問題存在，導致此強烈的關聯性在重要性的裁定上終究不及於花蓮文化創意產業園區與台灣好行的站點，然而，藉由單點式關聯規則篩選子資料庫的步驟以讓雜訊問題減少並放大具有強烈關聯性者，可因此使馬太鞍濕地生態館的支持度超越花蓮文化創意產業園區的支持度，進而變成改善優先順序第一之興趣點。同時，也因為馬太鞍濕地生態館所擁有的支持度提升以及瑞穗牧場所擁有的支持度提升一事，可證實本研究提出之單點式關聯規則於減少雜訊問題以及放大強烈關聯性者是可行的。

另一方面，表 4.12 為 3 個類別觀光用戶之興趣點拜訪人數前 25% 之單點式關聯規則興趣點支持度結果表，並顯示支持度大於 0.1 者。將此表與表 4.11 相比較，可發現馬太鞍濕地生態館的支持度又變回小於花蓮文化創意產業園區的狀態，而在表 4.11 具有高支持度的瑞穗牧場一消失於表 4.12，此現象說明單點式關聯規則對於用戶數較少之興趣點在放大其關聯性(也就是有效地提升最小支持度門檻值)的效果是較為顯著的。縱使如此，依然可藉由表 4.12 發現當中多了許多從一般關聯規則的兩個方法(即第 4.3 和第 4.4 節)無法找到的服務缺口興趣點，如：松園別館、花蓮文化中心石雕公園等，而此現象也同樣證實單點式關聯規則於解決雜訊問題的可行性。

表 4.11 單點式關聯規則興趣點支持度結果表(興趣點拜訪人數前 50%)

情境 興趣點	全體- 50%- 1 分鐘	全體- 50%- 2 分鐘	全體- 50%- 5 分鐘	第一類- 50%- 1 分鐘	第一類- 50%- 2 分鐘	第一類- 50%- 5 分鐘	第二類- 50%- 1 分鐘	第二類- 50%- 2 分鐘	第二類- 50%- 5 分鐘
馬太鞍濕地生態館	-	-	-	-	-	-	-	-	-
花蓮文化創意產業	0.4323	0.4062	0.3662	0.4819	0.4704	0.4357	0.4211	0.4270	0.3642

園區									
瑞穗牧場	0.3496	0.3223	0.2943	0.3263	0.2792	0.2719	0.3771	0.3802	0.3342
秀姑巒溪 泛舟中心	NA	NA	NA	0.2370	NA	NA	NA	NA	NA
親不知子 天空步道	NA	NA	NA	0.2113	0.2009	0.2065	NA	NA	NA
九曲洞隧 道	NA	NA	NA	0.2081	NA	NA	NA	NA	NA

表 4.12 單點式關聯規則興趣點支持度結果表(興趣點拜訪人數前 25%)

情境 興趣點	全體- 25%- 1 分鐘	全體- 25%- 2 分鐘	全體- 25%- 5 分鐘	第一類- 25%- 1 分鐘	第一類- 25%- 2 分鐘	第一類- 25%- 5 分鐘	第二類- 25%- 1 分鐘	第二類- 25%- 2 分鐘	第二類- 25%- 5 分鐘
花蓮文化 創意產業 園區	-	-	-	-	-	-	-	-	-
親不知子 天空步道	0.1582	0.1611	0.1570	NA	NA	NA	0.1164	0.1252	0.1025
馬太鞍濕 地生態館	0.1486	0.1523	0.1413	0.1449	0.1524	0.1441	0.1553	0.1522	0.1359
太平洋公 園	NA	NA	NA	0.1820	NA	NA	NA	NA	NA
松園別館	0.1490	0.1363	0.1232	0.1596	0.1396	0.1396	0.1415	0.1448	0.1283
花蓮遠雄 海洋公園	NA	NA	NA	0.1038	NA	NA	0.1246	0.1241	0.1173
花蓮文化 中心石雕 公園	0.1342	0.1128	0.1050	0.16	0.14	0.12	0.1149	0.1187	0.1004
靜埔北迴 歸線界標	NA	NA	NA	NA	NA	NA	0.1167	0.1203	NA

另一方面，從時間的角度上來觀察，在路線中加入花蓮文化創意產業園區同樣是可行之辦法。圖 4-4 為花蓮火車站、慶修院以及花蓮文化創意產業園區任兩者之間的旅行時間示意圖，其中圖 4-4(A)顯示花蓮火車站至慶修院之旅行時間；圖 4-4(B)顯示花蓮火車站至花蓮文化創意產業園區之旅行時間；而圖 4-4(C)則顯示花蓮文化創意產業園區至慶修院之旅行時間。本研究使用以估計旅行時間之系統為 Google Map，而 Google Map 所估計的旅行時間會根據使用者進行路段搜索當下的道路情況進行旅行時間估計，為了避免計算結果隨著不同的搜索時間而有變化，本研究一慮採用 Google Map 所估計出的「交通順暢時」的旅行時間為計算依據(圖 4-4 中紅色方框處所顯示之時間)。從圖 4-4(A)可看到在本研究進行旅行時間搜索的當下，花蓮火車站至慶修院的旅行時間大約為 15 分鐘，而此起彼對在交通順暢時的旅行時間應僅需要 12 分鐘；同理，圖 4-4(B)顯示花蓮火車站至花蓮文化創意產業園區在交通順暢時僅需 8 分鐘即可抵達，而圖 4-4(C)顯示花蓮文化創意產業園區至慶修院在交通順暢時可在 11 分鐘完成旅行。

同理，本研究同樣利用 Google Map 估計圖 4-2 中縱谷花連線各站點之間的旅行時間，並將其平均後得到縱谷花連線的平均站點間旅行時間約為 12 分鐘。再根據第 3.5 節所述的評估原則，可觀察到無論是從花蓮火車站行駛到花蓮文化創意產業園區，亦或是從花蓮文化創意產業園區到慶修院，其旅行時間皆沒有大於站間平均旅行時間，也就是說本研究認為將花蓮文化創意產業園區加入至台灣好型縱谷花連線為可行之辦法。



圖 4-4 花蓮火車站、慶修院、花蓮文化创意產業園區旅行時間示意圖。(A)花蓮火車站至慶修院之旅行時間、(B)花蓮火車站至花蓮文化创意產業園區之旅行時間、(C)花蓮文化创意產業園區至慶修院之旅行時間

表 4.6 也顯示出第一類用戶的旅運行為與花蓮文化创意產業園區有著更密切的關聯。同樣地假設花蓮文化创意園區為台灣好行服務路線中的一個行經站點，可發現可被滿足的樣態數比例幾乎上升至八成以上，而從表 4.7 可看出第二類用戶在加入花蓮文化创意產業園區至台灣好行路線後，其樣態可被滿足比例僅能提升約至七成。因此，相較於第二類用戶，第一類用戶在規劃於花蓮的旅程時有更大的機率將花蓮文化创意產業園區安排進旅程中；也就是說，根據兩群用戶的分類原則來看，有較大可能性使用觀光公共運輸系統的客群前往花蓮文化创意產業園區旅遊的機會較大，故將花蓮文化创意產業園區加入至觀光公共運輸系統本研究認為非常恰當。

同時，根據表 4.8 以及表 4.9 可得知，花蓮文化创意產業園區不僅能改善最多的樣態數，亦是所有「可改善單一興趣點」的樣態中支持度最高的樣態。此現象說明花蓮文化创意產業園區不僅與眾多的台灣好行路線站點有所關聯，亦表達此關聯性是強烈的。不僅如此，同樣根據表 4.8 以及表 4.9，亦可看出馬太鞍濕地生態館以及瑞穗牧場同樣也是遊客前往花蓮時較常去拜訪且無法被台灣好行路線所服務的景點之一。在

樣態支持度的部分無法觀察出此兩興趣點的重要性即是本研究設立多重檢驗門檻的原因，且造成此問題的原因為馬太鞍濕地生態館及瑞穗牧場皆只與花蓮觀光糖廠存有較強烈的關聯性，也就是說在樣態支持度的方面此兩興趣點雖然能改善些許樣態，但改善的量終究不多進而無法被選為最後需要被加入路線的興趣點。

此現象亦可從單點式關聯規則分析的結果看出。從表 4.10 可得知瑞穗牧場以及馬太鞍濕地生態館的拜訪人數介於所有興趣點的前四分之一到二分之一的區間，且不只是在興趣點停留時間門檻值 1 分鐘的情境如此，當興趣點停留時間門檻值為 2 分鐘或 5 分鐘時，此兩興趣點的拜訪人數排序依然介於前 50% 至 25% 之間，故觀察含有此兩興趣點之單點式關聯規則興趣點支持度結果，也就是表 4.11，可發現瑞穗牧場與馬太鞍濕地生態館的樣態支持度皆非常高。也就是說，即便此二興趣點不如花蓮文化創意園區與眾多興趣點有關連性進而能有非常高比例的樣態改善數，但含有此二興趣點的樣態具有高支持度說明此二興趣點的重要性高於其餘被探勘出且同樣為未能被台灣好行服務到之興趣點。

馬太鞍濕地生態館與瑞穗牧場皆是與花蓮觀光糖廠之間存有高支持度的關聯性，圖 4-5 則顯示出此兩興趣點以及花蓮糖廠和台灣好行縱谷花蓮線的最後一個站點，大農大富平地森林園區，之位置分佈。如圖 4-5 的興趣點位置分布圖所示，馬太鞍濕地生態館位於花蓮觀光糖廠西方約 1.5 公里處；而瑞穗牧場則是位於花蓮觀光糖廠南方約 20 公里處。依照兩興趣點的位置以及考量到與此二興趣點具關聯性的興趣點為花蓮觀光糖廠，若要將此兩興趣點安排進台灣好行路線，皆應安排在縱谷花蓮線。而縱使瑞穗牧場與花蓮糖廠的距離不近，但考量到皆是位於台 9 線上的興趣點，在路線規劃上來說也變得相對單純；因此，增加瑞穗牧場一站點雖會使得行駛時間拉長，但也並非是不可行的決策。



圖 4-5 花蓮觀光糖廠、馬太鞍濕地生態館、大農大富平地森林園區以及瑞穗牧場位置分布圖

同樣地本研究亦從旅行時間的角度評量此兩興趣點增加至路線上是否可行。圖 4-6 為花蓮觀光糖廠、馬太鞍濕地生態館以及大農大富平地森林園區任兩者之間的旅行時間示意圖，其中圖 4-6(A)顯示花蓮觀光糖廠至大農大富平地森林園區之旅行時間；圖 4-6(B)顯示花蓮觀光糖廠至馬太鞍濕地生態館之旅行時間；而圖 4-6(C)則顯示馬太鞍濕地生態館至大農大富平地森林園區之旅行時間。從圖 4-6 可看出在從花蓮糖廠至馬太鞍濕地生態館約需 7 分鐘的旅行時間，而從馬太鞍濕地生態館至大農大富平地森林園區則需花費 10 分鐘，兩路段的旅行時間同樣皆沒有超過此路線的平均站間旅行時間 12 分鐘，因此本研究亦認為將馬太鞍濕地生態館加入至路線中為可行。



圖 4-6 花蓮觀光糖廠、馬太鞍濕地生態館以及大農大富平地森林園區旅行時間示意圖。(A)花蓮觀光糖廠至大農大富平地森林園區之旅行時間、(B)花蓮觀光糖廠至馬太鞍濕地生態館之旅行時間、(C)馬太鞍濕地生態館至大農大富平地森林園區之旅行時間

另一方面，圖 4-7 為大農大富平地森林園區至瑞穗牧場之旅行時間示意圖。由於瑞穗牧場在空間上距離較遠，因此從大農大富平地森林園區駛至瑞穗牧場在交通順暢的情況下約需 22 分鐘。本研究認為 20 分鐘以上的額外旅行時間著實對路線的營運造成一定程度的負擔，且遠超過此路線的站間旅行時間。同時又根據表 4.11 興趣點拜訪人數前 50%之單點式關聯規則興趣點支持度結果表，得知瑞穗牧場在被改善的重要性是低於花蓮文化創意產業園區及馬太鞍濕地生態館的，因此本研究認為瑞穗牧場的改善可以從特定班次的面相實施，而非將此景點直接納入常規路線之中，意即可以採用「部分班次延駛至瑞穗牧場」的方式做改善。



圖 4-7 大農大富平地森林園區至瑞穗牧場之旅行時間示意圖。

第五章 結論與建議

本節將針對上述之研究過程及結果做出結論與建議，並分為兩個部分呈現。第一部分為根據研究過程及結果歸納出的結論。而第二部分則為未來相關研究可以精進或嘗試的建議。

5.1 結論

根據第四章的測試結果，可看出本研究研擬之方法應用於找尋觀光用戶的潛在旅運需求是可行的。此外，本研究提出花蓮文化創意產業園區、馬太鞍濕地生態館以及瑞穗牧場三個景點作為改善路線的對象，是根據其各自分別有著最多的樣態改善數以及較高的樣態支持度；然而，根據關聯規則的原理，只要是最終能被探勘出來的樣態都代表著有一定的拜訪人數且該旅遊樣態發生的機率相對高，因此相關單位在考慮新增站點時，除了上述三個景點，亦可將前一章節有列出無法被台灣好行服務到的景點納入考量範圍。

除此之外，研究結果亦顯示出信令資料用於推估民眾的停留點、旅運行為等資訊是有用且可行的。從實際範例測試結果可看出，藉由興趣點資料先一步找出民眾曾到過的地方，再透過信令資料本身即含有的時間資訊推論使用者在到過的地方中哪些是屬於純粹經過，又哪些是屬於有停留行為發生，最後再根據用戶的停留行為進一步將全體用戶劃分成兩個族群。同時，如 Ahas (2007)以及 Zahedi and Shafahi (2017)所言，手機信令資料在蒐集上比起傳統旅運資料有著更有效率、省成本等優點。故根據上述，可得知手機信令資料亦可作為一個優秀的交通調查資料蒐集來源，甚至成為一種主要的交通調查資料來源。

另一方面，本研究提出的樣態支持度亦可補足運研所(2018)只採用改善樣態數一指標可能會造成及果分析上有忽略的問題。從第四章的實測結果可看出，在樣態支持度分析結果的部分，馬太鞍濕地生態館有較高的樣態支持度，說明與台灣好行路線具有強烈的關聯性，為一個較需要被改善的興趣點。但在改善樣態數的方面，可發現馬太鞍濕地生態館可改善的樣態數無論在哪種情境下皆只有零星幾個，在興趣點的改善重要性上便顯得不那麼重要。根據上述，可得知若只採用單一指標，則可能會有忽略部分結果的情況發生，但若兩個指標一併採用，則可以解決此問題。

在興趣點停留時間門檻值的部分，本研究所採用的數值為 1、2 以及 5 分鐘三種，相較於以往的相關研究，此門檻值的設定是相對較低的，而此狀況是本研究由於興趣點資料缺失所採取之應變手段。本研究透過對取得的資料進行檢查後發現，興趣點地理資訊資料中記錄的興趣點包含網格僅含有該興趣點的部分範圍而非全部的空間，也就是興趣點所包含網格的數量有過少的問題，舉例來說，花蓮海洋公園在興趣點地理資訊資料裡，所包含的網格僅顯示售票亭附近的範圍，並非含有整個園區範圍的網格數。因此，為了在資料缺失的情況下能捕捉到用戶於各興趣點停留的行為，本研究將停留時間門檻值設定得較過往研究來的短，以解決此狀況。

5.2 建議

- (1) 位置資訊的解析度會直接影響判斷用戶在各個地點是否有停留行為的準確性。本研究採用的信令資料其位置資訊解析度為 50 乘 50 公尺之範圍，在沒有使用容忍距離之放鬆條件即進行資料比對時，發現由於位置資訊的解析度過高，導致信令資料與興趣點資料的比對率極低，因此才進一步加入容忍距離一放鬆條件。而其餘位置資訊解析度有如 Zahedi and Shafahi (2017) 使用基地站服務範圍，或是如謝雅晴(2018)、施冠毅(2018)以及李泓儒(2018)使用二級發布區，亦或是如 Alexander (2015) 將一個地域相似的區域聚合起來的範圍等。建議往後之研究在位置資訊解析度的選擇上切勿認為定位範圍越小者越好，還得視研究內容、研究範圍等因素來選擇適當的位置資訊解析度。
- (2) 本研究在判斷第一類用戶時所採用的方法為判斷各用戶於花蓮地區的第一個或最後一個停留點是否為台鐵車站。本研究認為此方法是可行的，惟此方法受信令資料的影響非常大，此方法在判定用戶類別時，可能因部份第一類用戶在抵達火車站時因個人因素或外在環境影響導致用戶沒有使用手機的行為，進而使該用戶沒有在火車站的信令資料，造成用戶判斷上的遺漏。此外，從第 4.2 節也可得知第一類用戶的判斷受到興趣點停留時間門檻值的影響，隨著興趣點停留時間門檻值的提升，第一類用戶數有增加的趨勢。因此建議未來的研究可以另外產製一份台鐵軌道的網格資料，並藉由比對用戶軌跡與軌道網格是否符合來避免用戶類別誤判等情況發生。
- (3) 本研究所使用之資料科學方法為關聯規則分析，目的為找出彼此之間具有關聯性的興趣點，並以此作為依據以提出改善台灣好行路線之建議。然而，關聯規則分析僅考量到觀光客彼此之間的旅運行為是否有相似之處，對於觀光客自身的分類則必須先一步完成(如本研究先將觀光客細分為兩族權後再進行關聯規則分析)。因此，建議後續相關研究可以採用其他種資料科學的方法進行分析，如:k-means 分群法、決策樹、簡單貝式分類法等，上述方法在給出台灣好行路線建議的同時，亦能夠針對輸入的各個屬性得到每個興趣點的拜訪遊客特性，即能根據結果給出各個興趣點如何吸引更多遊客之客製化方法。
- (4) 用戶分群建議可有更多種類別。本研究將所取得之觀光用戶透過其旅運資訊再細分成兩群，分別為搭乘公共運輸系統前往(或離開)花蓮的用戶以及全程使用私人運具的用戶。然而，隨著時代進步，現代人旅行時的移動方式也更加多元，如:搭乘公共運輸系統到目的地後，先租車遊玩離市區較遠的景點，旅程的最後一、兩天還車後，再透過市區公車遊玩市區內或市區附近景點這類型使用多種運具之旅程。因此建議未來相關研究可針對用戶分群進行更詳細的探討。
- (5) 本研究進行單點式關聯規則時，為了避免興趣點拜訪人數過低導致些許的旅運樣態便會有過高的支持度進而影響最終分析結果，因此先針對各興趣點進行人數統計，再根據拜訪人數排序為前一半以及前四分之一的興趣點進行單點式關聯規則分析。本研究起初在研擬單點式關聯規則一方法時，並未針對興趣點拜訪人數進行限制，而初步測試結果即顯示拜訪人數少的興趣點會擁有較高的支持度，以第

一類用戶、興趣點停留時間 1 分鐘的部分為例，在沒有人數篩選的情況下，支持度前 2 高的興趣點為「吉蒸牧場(55 個景點中拜訪人數排序第 43，共 118 人拜訪)」和「翡翠谷(55 個景點中拜訪人數排序第 54，共 2 人拜訪)」，因有著此初步測試結果，本研究新增了拜訪人數排序限制。而從結果即可看出，拜訪人數前 50%及拜訪人數前 25%的結果已有明顯差異，便可推論若沒有拜訪人數的限制，最終結果可能會更偏向冷門的興趣點。本研究初步設定能進行單點式關聯規則分析的興趣點其拜訪人數至少為前 50%的原因為考量到路線改善後須吸引到一定數量的觀光客前來使用台灣好行服務，此改善才具有經濟價值，然而本研究認為後續研究可針對此部分進行更細部的探討，如興趣點進行改善後所影響的人數、單點式關聯規則結果與一般關聯規則結果的差異性等，皆可做為評選原始資料庫中何者能進行單點式關聯規則的依據，甚至能探討原始資料庫的資料量、資料間的差異性或雜訊對探勘結果影響的程度等，作為判定該資料庫是否適合於執行單點式關聯規則等相關研究。

- (6) 本研究之目的為透過關聯規則的結果提出台灣好行路線的服務建議，因此在最後評量新增的興趣點加入路線後是否可行的評量依據僅考量興趣點的空間分布以及新增該興趣點後的額外旅行時間。然而，對於一個公車運行路線變化的評估應可以有更詳細之評量方法，如：新增站點後受到影響的人數、運行路線改變後之各項成本變化等，皆是後續相關研究可深入探討之評估各站點的加入是否可行的方向。

參考文獻

1. 聯合行銷研究股份有限公司(2017)。106 年持有手機民眾數位機會調查報告。調查報告。
2. Ahas, R., Aasa, A., Silm, S., Tiru, M. (2007). Mobile Positioning Data in Tourism Studies and Monitoring: Case Study in Tartu, Estonia. *Information and Communication Technologies in Tourism 2007*, pp 119-128.
3. Gur, Y. J., Bekhor, S., Solomon, C., Kheifits, L. (2009). Intercity Person Trip Tables for Nationwide Transportation Planning in Israel Obtained from Massive Cell Phone Data. *Transportation Research Record*, 2121(1), pp 145 - 151.
4. Phithakkitnukoon S., Horanont T., Di Lorenzo G., Shibasaki R., Ratti C. (2010). Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data. *Human Behavior Understanding. HBU 2010. Lecture Notes in Computer Science*, vol 6219. Springer, Berlin, Heidelberg. pp 14-25.
5. Nanni, M., Trasarti, R., Furletti, B., Gabrielli, L., Mede, P. V. D., Bruijn, J. D., Romph, E. D., Bruil, G. (2013). Transportation Planning Based on GSM Traces: A Case Study on Ivory Coast. *CitiSens*, Barcelona, Spain.
6. Alexande, L., Jiang, S., Murga, M., González, M. C. (2015). Origin - destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies Volume 58, Part B*, September 2015, pp 240-250.
7. Trasarti, R., Olteanu-Raimond, A-M., Nanni, M., Couronné, T., Furletti, B., Giannotti, F., Smoreda, Z., Ziemlicki, C. (2015). Discovering urban and country dynamics from mobile phone data with spatial correlation patterns. *Telecommunications Policy*, 39(3), pp 347-362.
8. Zahedi S., Shafahi Y. (2017). Estimating activity patterns using spatio-temporal data of cell phone networks. *International Journal of Urban Sciences*, 22:2, pp 162-179.
9. 謝雅晴(2018)。利用信令資料推估汽車客運旅運起迄之研究。國立交通大學運輸與物流管理學系學位論文，pp 1-48。
10. 李泓儒(2018)。利用信令資料推估軌道運具旅運起訖之研究。國立交通大學運輸與物流管理學系學位論文，pp 1-35。
11. Berlingerio, M., Calabrese, F., Di Lorenzo, G., Nair, R., Pinelli, F., Sbodio, M. L. (2013). ALLAboard: a system for exploring urban mobility and optimizing public transport using cellphone data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp 663-666.
12. Holleczech, T., Anh, D. T., Yin, S., Jin, Y., Antonatos, S., Goh, H. L., Low, S., Shi-Nash, A. (2015). 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*, pp 1859-1868.

13. Demissie, M. G., Phithakkitnukoon, S., Sukhvibul, T., Antunes, F., Gomes, R., Bento C. (2016). 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*, 17(9), pp 2466-2478.
14. 交通部運輸研究所(2017)。運時空資料分析與公共運輸服務應用發展計畫。研究計畫。
15. 交通部運輸研究所(2018)。應用旅次特性大數據精進公共運輸服務計畫。研究計畫。
16. 施冠毅(2018)。應用信令資料於用戶重現性需求之分析。國立交通大學運輸與物流管理學系學位論文，pp 1-46。