

國 立 交 通 大 學
運 輸 與 物 流 管 理 學 系

碩 士 論 文

利用手機信令資料判別觀光旅運使用者之研究

The study of determining tourists based on cellular phone
data

研 究 生：曾 傳 軒

指 導 教 授：王 晉 元

中 華 民 國 一 〇 八 年 七 月

利用手機信令資料判別觀光旅運使用者之研究

The study of determining tourists based on cellular phone data

研究生：曾傳軒

Student：Chuan-Hsuan Tseng

指導教授：王晉元

Advisor：Jin-Yuan Wang



中華民國一〇八年七月

利用手機信令資料判別觀光旅運使用者之研究

學生：曾傳軒

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

國人在從事旅遊活動時，往往會因為私人運具的高機動性而選擇私人運具。為了能夠減緩小客車對觀光景點附近的交通衝擊，政府希望透過推動符合民眾需求觀光公共運輸。一個成功的觀光公共運輸，對於民眾有較大的使用誘因，同時也能夠協助減少自用汽車的使用。因此，在規劃觀光公共運輸時前，應優先深入瞭解民眾的觀光需求。

本研究以手機信令資料為基礎，設計一套從信令資料中判讀出觀光用戶的邏輯。在成功從信令資料中找出觀光遊客後，將能使後續觀光旅運需求分析的結果更為精確，並能根據分析結果設計出貼近民眾需求的觀光公共運輸路線，吸引民眾改以大眾運輸從事觀光旅遊活動。

關鍵字：手機信令資料、旅運需求分析

The study of determining tourists based on cellular phone data

Student : Chuan-Hsuan Tseng

Advisor : Jin-Yuan Wang

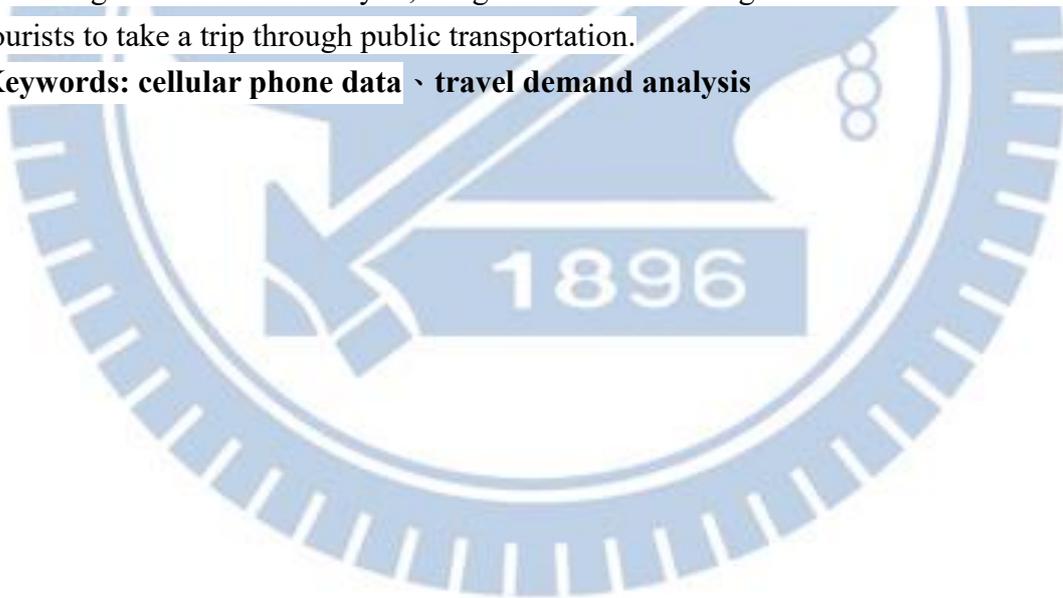
Department of Transportation and Logistics Management
National Chiao Tung University

Abstract

Due to the flexibility of private mode, people often choose private mode when taking a trip. In order to reduce the traffic impact around tourist attractions, the government tries to develop public transportation for tourists. A successful tour public transportation can not only attract tourists but also reduce private car use. Therefore, before planning tour public transportation, the government should know more about people's demands.

This study aims to design a method which can differentiate tourists from cellular phone data. After finding out tourists from cellular phone data, the authority concerned can analyze their cellular phone data to understand their demands. Finally, according to the demand analysis, the government can design better routes to attract tourists to take a trip through public transportation.

Keywords: cellular phone data 、 travel demand analysis



誌謝

碩士生涯的兩年，說長不長，說短不短，正是因為受到許多人的幫助才能夠順利完成碩士學業。首先，最感謝我的指導教授，老師每個禮拜總是會在百忙之中撥空與我們討論論文，討論論文的研​​究方向並點出我的不足之處，同時給予我改進的方向。在與老師見面討論論文後，我都能清楚了解自己的不足之處，並針對這些不足之處深入思考，構想如何解決現有的問題。經由每周不斷地討論以及思考，讓我成長許多，使我在面對問題時能夠更仔細地分析問題，並能思考如何解決問題。從大學部的畢業專題到研究所的碩士論文，都是因為有老師的指導，我才能夠一步一步地成長，並成功地完成學業。

另外，還要感謝從一年級以來就一起奮鬥的同學。在這兩年間，大家都會互相幫忙，一起處理實驗室裡大大小小的事情。而在論文的部分，非常感謝同學能夠在我論文遇到問題的時候能夠一起討論，並在我撰寫程式遇到困境時給予意見，有了同學的建議，我才能夠順利地將程式撰寫完畢，完成自己的碩士論文。

最後，還要感謝整個運管系。感謝系上所有的老師，有了老師們的教導，我才能夠成長茁壯。感謝系上舉辦了各式各樣的演講，透過吸取講者在言談中傳遞的知識，相信對未來的工作一定有實質的幫助。感謝系上提供了豐富的資源，使我踏實地成長。感謝交通大學運管系，豐富了我的閱歷。

曾傳軒 謹誌
2019年7月

目錄

摘要.....	i
Abstract.....	ii
誌謝.....	iii
圖目錄.....	vi
表目錄.....	vii
1.1 研究背景.....	1
1.2 研究目的.....	2
1.3 研究範圍.....	2
1.4 研究流程.....	2
第二章 文獻回顧.....	4
2.1 信令資料應用於推估旅次特性.....	4
2.2 信令資料應用於觀光分析.....	9
2.3 小結.....	12
第三章 研究方法.....	13
3.1 資料蒐集.....	14
3.1.1 信令資料.....	14
3.1.2 興趣點位置資料.....	15
3.2 找出停留之興趣點.....	16
3.2.1 判斷用戶是否位於興趣點.....	16
3.2.2 判斷用戶是否於興趣點停留.....	17
3.3 判斷觀光用戶.....	19
3.3.1 用戶可用旅行時間.....	19
3.3.2 用戶停留於興趣點之時間.....	21
3.3.4 觀光用戶之判讀.....	25
3.4 小結.....	26
第四章 實測分析.....	27
4.1 實測資料介紹.....	27
4.1.1 信令資料.....	27
4.1.2 興趣點.....	27
4.2 實測結果.....	28
4.2.1 用戶於興趣點停留之判斷結果.....	28
4.2.2 用戶可用旅行時間計算結果.....	30
4.2.3 用戶興趣點停留時間計算結果.....	30
4.2.4 用戶興趣點之間移動移動時間計算結果.....	31
4.2.5 觀光用戶判斷結果.....	31
4.2.6 結果說明.....	32

4.3 敏感度分析.....	34
4.3.1 興趣點連續出現時間門檻值之分析.....	34
4.3.2 可用旅行時間之敏感度分析.....	35
4.3.3 可用旅行時間使用比例門檻值之敏感度分析.....	35
4.4 以實際信令資料計算興趣點停留時間之測試.....	36
4.5 小結.....	36
第五章 結論與建議.....	38
5.1 結論.....	38
5.2 未來研究方向之建議.....	38
參考文獻.....	40



圖目錄

圖 1-3 研究流程圖	3
圖 2-1 維也納市中心用戶分佈.....	4
圖 2-2 以奧地利北方之倫巴赫地區為起點之旅次.....	5
圖 2-3 一天不同時段內推估而得的平均旅次長度分佈.....	5
圖 2-4 手機信令資料推估與人口普查結果.....	6
圖 2-5 手機信令資料推估結果.....	7
圖 2-6 信令資料的推估與抽樣家庭訪問結果.....	7
圖 2-7 信令資料推估結果與政府公布數據之關聯性.....	8
圖 2-8 以旅次目的分類之旅次占比.....	8
圖 2-9 愛沙尼亞地區信令資料格式.....	9
圖 2-10 前十大造訪塔爾圖的遊客前往旅遊的地區.....	9
圖 2-11 安道爾地區路網圖.....	10
圖 2-12 三亞市熱門興趣區域.....	11
圖 2-13 三亞市高頻移動樣態.....	11
圖 3-1 研究方法流程圖.....	13
圖 3-2 興趣點比對範例.....	17
圖 3-3 連續時間表計算範例.....	18
圖 3-4 連續時間表過濾一般網格範例.....	18
圖 3-5 興趣點連續時間加總範例.....	19
圖 3-6 可用旅行時間判斷流程圖.....	20
圖 3-7 以旅遊活動時間過濾停留時間表範例.....	21
圖 3-8 用戶 f5320f 信令資料範例.....	22

表目錄

表 1.1 國人旅遊時使用之交通工具比例.....	1
表 3.1 信令資料範例.....	14
表 3.2 觀光景點包含之網格資料範例.....	15
表 3.3 網格中心點經緯度範例.....	15
表 3.4 興趣點網格中心表範例.....	16
表 3.5 用戶可用旅行時間計算範例.....	20
表 3.6 興趣點停留時間表.....	22
表 3.7 用戶連續時間表範例.....	24
表 3.8 OpenStreetMap 與 Google 地圖計算之旅行時間.....	24
表 3.9 花蓮文化創意產業園區之興趣點網格中心資料.....	25
表 4.1 資料相關統計數據.....	27
表 4.2 選定之興趣點.....	28
表 4.3 信令資料與興趣點位置資料比對結果.....	29
表 4.4 三周興趣點連續出現記錄之時間長度分布.....	29
表 4.5 三周可用旅行時間人數分布.....	30
表 4.6 三周興趣點停留時間人數分布.....	30
表 4.7 用戶興趣點移動時間的分布情形.....	31
表 4.8 三周可用旅行時間使用比例人數分布.....	31
表 4.9 三周觀光用戶數統計結果.....	32
表 4.10 觀光用戶判斷結果整理.....	32
表 4.11 用戶 0b44e3 原始信令資料.....	33
表 4.12 用戶 0b44e3 興趣點連續時間表.....	33
表 4.13 用戶 f5320f 原始信令資料.....	34
表 4.14 用戶 f5320f 興趣點連續時間表.....	34
表 4.15 興趣點連續出現時間門檻值敏感度分析結果.....	34
表 4.16 可用旅行時間敏感度分析結果.....	35
表 4.17 可用旅行時間使用比例敏感度分析結果.....	35
表 4.18 以實際信令資料計算興趣點停留時間觀光用戶判讀結果.....	36

第一章 緒論

1.1 研究背景

國人在從事旅遊活動時，往往會因為私人運具的高機動性而選擇透過私人運具以滿足活動過程中的運輸需求。根據交通部觀光局之統計(交通部觀光局，2018)，如表 1.1 所示，於民國 106 年時，國人從事旅遊時最主要利用的交通工具為自用汽車，其比例高達 64.6%。由自用汽車的高使用比例可知，私人運具的便利性仍較吸引民眾。

表 1.1 國人旅遊時使用之交通工具比例

交通工具	利用比例
自用汽車	64.6%
遊覽車	11.7%
公民營客運	9.8%
其他	13.9%

為了能夠減緩小客車對觀光景點附近的交通衝擊，政府希望透過推動符合民眾需求觀光公共運輸，以改變民眾的旅遊習慣。一個滿足民眾需求的觀光公共運輸，對於民眾有較大的使用誘因，同時也能夠協助減少自用汽車的使用。因此，在規劃觀光公共運輸時，應優先深入瞭解民眾的觀光需求。

傳統的旅運需求調查為採用抽樣的方式，瞭解被抽選民眾一整天所有的旅次，但其存在一些缺點。首先，透過抽樣發放問卷或者電話訪問的執行方法，需要耗費大量的金錢以及時間成本。再者，傳統的調查方式約為五至十年進行一次，然而在五至十年中，運輸需求可能發生大幅度的改變，資料的更新速度難以趕上運輸需求的變化。綜合以上的兩項原因，使用傳統的旅運需求調查資料可能使得決策成效不彰。

近年來，由於智慧型手機的普及，手機信令資料成為新興的資料分析對象。當我們使用手機進行上網或者是傳送訊息時，手機會與基地台互動並留下記錄，記錄當下的時間以及使用者大略的位置，這些記錄就是所謂的手機信令資料。信令資料和傳統資料不同，其資料能自動產生，不需要耗費時間以及人力進行蒐集，只需要事後再從資料庫中擷取出所需的資料即可進行分析，故信令資料對於實務規劃具有一定之參考價值。

然而，在瞭解觀光遊客的需求之前，如何分別該手機用戶是否為觀光用戶是一重要的課題。由於在持有手機並且有上網行為的情形下，即會有信令資料的產生，所取得的資料勢必會存在觀光用戶的手機信令資料以及非觀光用戶的手機信令資料。因此，在分析觀光需求前，必須要先從資料中判斷出觀光用戶，以避免找到大量但非觀光運輸需求的結果。

綜合以上所述，本研究將以手機信令資料為基礎，設計一套從信令資料中判讀出觀光用戶的邏輯。若能成功從信令資料中找出觀光遊客，將使得後續觀光旅運需求分析的結果更為精確，並能據分析結果設計出貼近民眾需求的觀光公共運輸路線，吸引民眾改以大眾運輸從事觀光旅遊活動。

1.2 研究目的

本研究之研究目的為建立一套判斷邏輯，從手機信令資料中判讀出觀光用戶，協助後續觀光旅運需求分析。藉由信令資料，可以完整瞭解用戶一天的移動情形，不需要額外進行任何調查。最後，根據用戶的移動情形，例如是否有於觀光景點停留，設計一套判別觀光用戶的邏輯。

1.3 研究範圍

本研究使用之資料為電信公司提供之花蓮地區手機信令資料。資料欄位有加密後的用戶編號、時間、以及所在位置。電信公司會事先將所在位置轉換成對應的網格，而每一網格大小均相同，長寬各為 50 公尺。由於本地居民很少會於當地從事觀光活動，因此信令資料會事先經由電信公司事先過濾花蓮地區住戶，方能針對非花蓮地區之住戶進行判斷邏輯的設計。

1.4 研究流程

本研究研究流程如圖 1-3 所示，總共有以下六個步驟：

(1) 相關文獻回顧

本研究之研究目的為自手機信令資料研擬觀光客判斷邏輯，並從中找出觀光用戶。在此階段，將會針對信令資料應用於用戶移動行為分析的方向，蒐集相關文獻。透過閱讀相關文獻，可以協助瞭解目前的研究成果，並作為後續研究分析的參考依據。

(2) 資料蒐集

本研究最主要分析對象為信令資料，故取得信令資料為此階段之首務。本研究經由中華電信，取得花蓮地區非住戶的手機信令資料，並取得各主要景點之位置資訊。藉由手機信令資料與花蓮地區觀光景點的位置資訊，使本研究得以開始構思觀光用戶的判斷邏輯。

(3) 模式建立

於此階段，將會實際進行觀光用戶判斷邏輯的設計。在構思邏輯過程中，將參考蒐集而得的相關文獻，研擬出一套合適的判斷流程。最後，將設計完成的判斷邏輯以程式語言撰寫，並於下一階段中實際應用於信令資料。

(4) 實測分析

此階段將設計完成之模式，實際應用於手機信令資料。由模式產出之結果，可以分辨每一位用戶是否為觀光用戶，進而得知信令資料中觀光用

戶所佔的比例。在得到結果後，將隨機挑選用戶進行檢查，確認是否有誤判的情形發生，並反覆修正模式，調整相關參數。

(5) 結論與建議

針對每位用戶判斷是否為觀光用戶後，此階段主要進行結果的檢視。根據判斷結果進行相關說明，並闡述判斷結果的合理性。此外，此階段亦將說明研究過程中所遇到的問題，以及探討本研究之研究成果應用，以提供後續研究參考之方向。

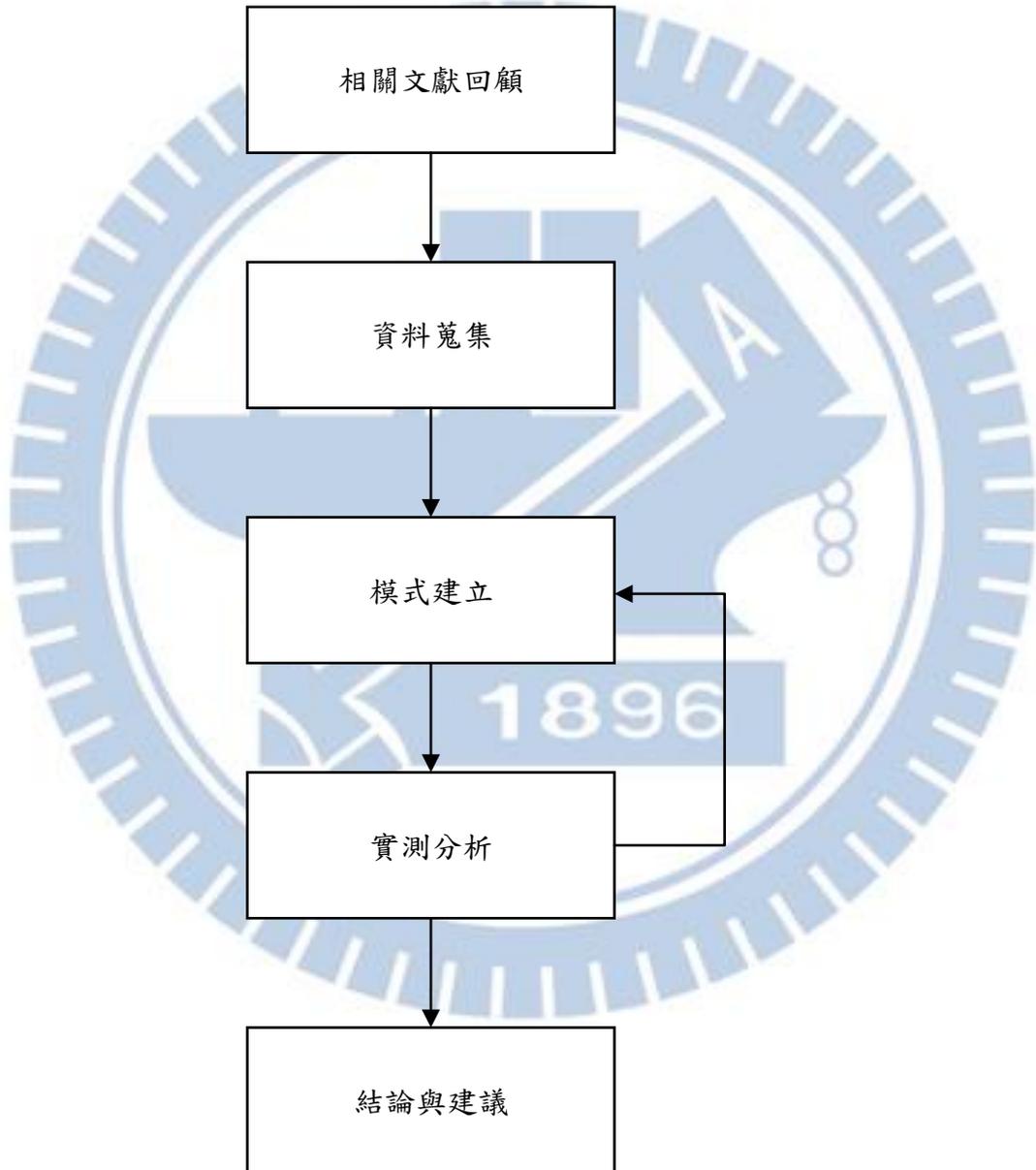


圖 1-3 研究流程圖

第二章 文獻回顧

本研究之研究目的為希望透過信令資料，建立一套判斷邏輯，可以利用信令資料判斷是否為觀光用戶。因此，本章將分成兩部分進行文獻蒐集，分別為應用信令資料於推估旅運型態或目的以及應用信令資料於觀光分析兩部分進行回顧。

2.1 信令資料應用於推估旅次特性

Calabrese 等人(2011)透過信令資料，嘗試推估出起迄矩陣。作者以 8 個位於美國麻薩諸塞州的城鎮，共約 8 億筆信令資料為對象，建置起迄矩陣，以瞭解民眾的旅運型態。信令資料中包含了用戶的編號、時間以及電信公司利用三角定位法所求得之用戶位置。作者以下午六時至隔天早上八時之間，用戶連接最頻繁之基地台為用戶住家，並根據資料中每一位用戶依序造訪的基地台，分別切割成不同旅次，進而產製出起迄矩陣。在與美國聯邦政府公路總署所提供之資料對比後，得到了 0.73 的相關係數，較重力模式的推估結果(相關係數為 0.59)為佳。

Loibl 等人(2012) 以奧地利為研究區域，透過信令資料探討人口的分布以及移動型態。奧地利約有 800 萬人口，但總計卻擁有 1 億 2 千萬隻手機，為人口的 1.5 倍，故作者認為利用信令資料進行研究是可行的。圖 2-1 為維也納市中心的某日上午，作者將位於該處之手機用戶整合至特定觀察點，點越大則代表人數越多，由此可以輕易看出人口分布。圖 2-2 為以奧地利北方之倫巴赫(Lembach)地區為起點之旅次，由圖可以發現大多數的旅次都是通往林茨(Linz)，代表著兩城鎮間有一定量之運輸需求。作者透過兩種範例，說明了應用信令資料的可能性，也提及了信令資料在未來能有更多元的應用。



圖 2-1 維也納市中心用戶分佈

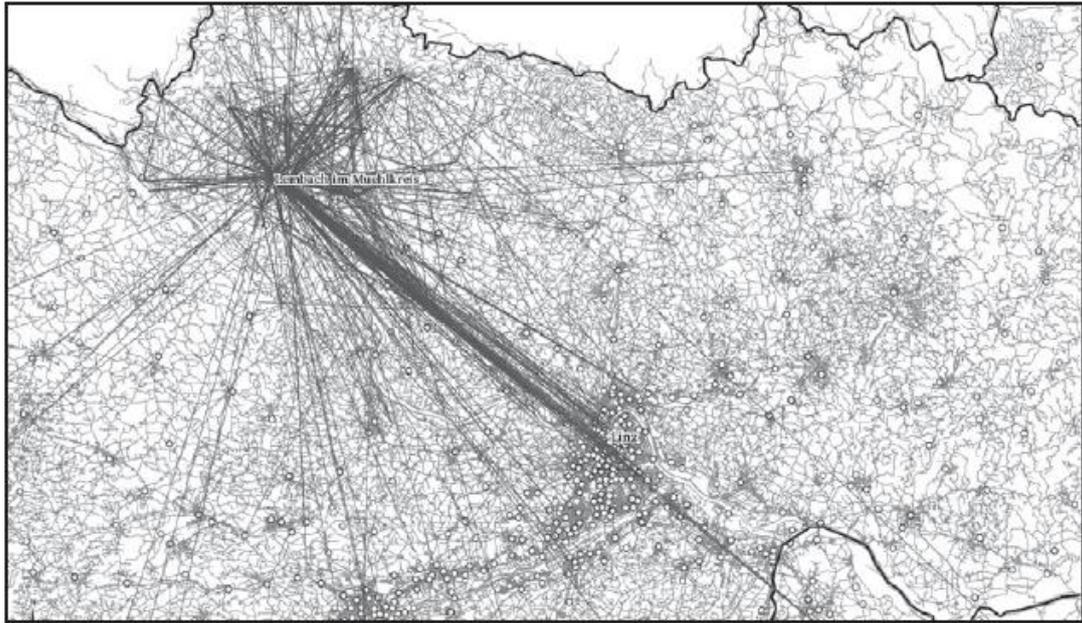


圖 2-2 以奧地利北方之倫巴赫地區為起點之旅次

Bekhor 等人(2013)以以色列地區為例，利用信令資料探討旅次的特性。信令資料之時間為連續的 16 個星期，電信業者已事先在每周抽樣 10,200 位用戶後，再提供抽樣後的信令資料。該研究首先對信令資料進行過濾，過濾掉短時間內與不同位置但很接近之基地台連接的資料，因為這些資料應與旅次無關。該研究對於用戶住家的定義為用戶連續出現最久之處，當用戶有基地台切換的情形時，即形成旅次。圖 2-3 為在一天不同時段內推估而得的平均旅次長度分佈，由圖中可以了解在星期六大部分的時段，平均旅次長度較其他天來的高，也說明了極有可能是因為旅遊活動所導致的。

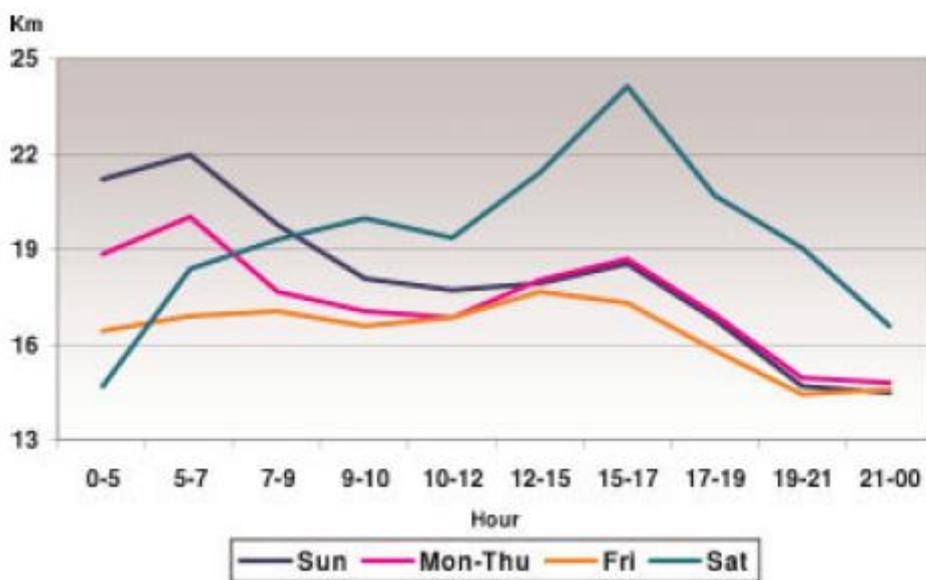


圖 2-3 一天不同時段內推估而得的平均旅次長度分佈

Alexander 等人(2014)利用波士頓地區兩個月之信令資料，根據用戶移動之起迄點將旅次進行劃分。該研究首先以 10 分鐘作為門檻值，篩選用戶的停留點。待篩選完成後，定義每個用戶的住家以及工作場所。住家的定義為於晚上七點至早上八點用戶最常出現之處；工作場所則參考過去文獻，定義為早上八點至晚上七點之間，造訪次數與旅行距離乘積為最大之處。而若用戶於一周內造訪前述之工作場所小於八次或者是該處與用戶之住家距離小於 800 公尺，則將該處轉而定義為其他地點。下一步再根據旅次的起迄點定義旅次的型態，共有家工作旅次、家其他旅次以及非家旅次三種。旅次校估完成後，再將人口放大，得到最後的結果。圖 2-4 中的左圖為美國運輸規劃人口普查的結果，而右邊為此研究利用手機信令資料推估而得的結果。由於兩張圖的旅次分布有一定程度的相似，可以瞭解到信令資料推估的結果是可信的。此研究同時認為若能夠有更深入的土地相關資料，可以將起迄點進行更深入的分類，而非僅有家、工作以及其他共三種類別。

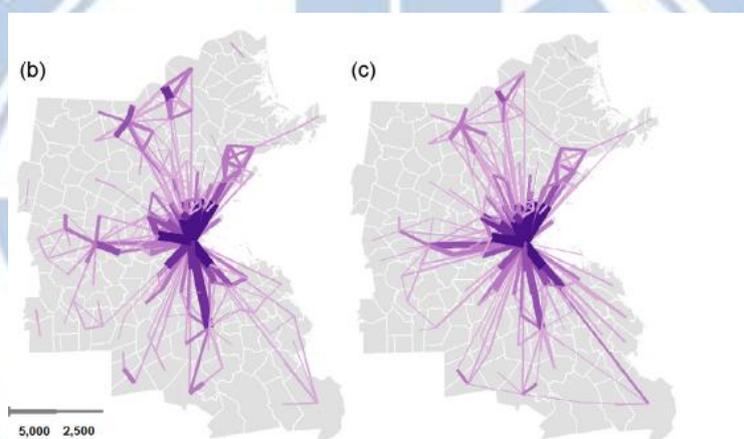


圖 2-4 手機信令資料推估與人口普查結果

Widhalm 等人(2015)利用波士頓以及維也納的手機信令資料進行分析，以土地使用、活動時間帶以及活動時間長度推測用戶的旅次型態。該研究首先進行旅次以及拜訪點的重建，接著透過建立關聯式馬可夫鏈，用以判斷旅次型態。圖 2-5 為推估結果，研究中將旅次分成許多種型態，除了常見的家或工作外，還額外區分出購物以及休閒兩種目的，可以發現加入土地使用資訊之後，能夠更詳盡地判斷出不同旅次目的。

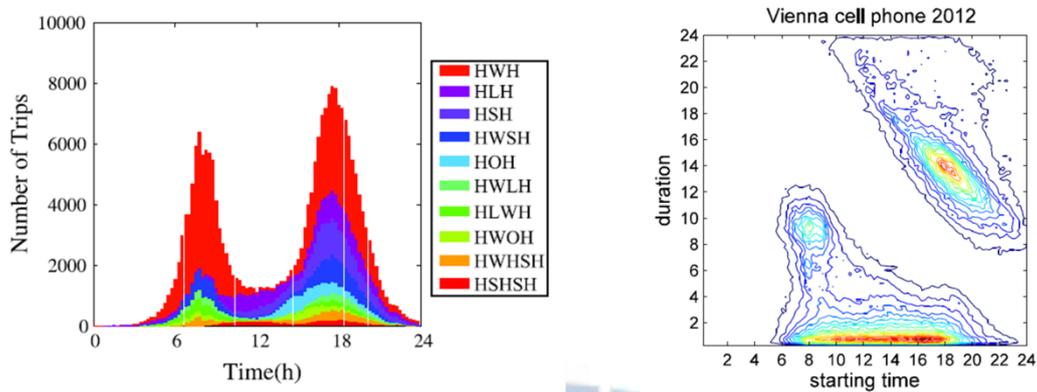


圖 2-5 手機信令資料推估結果

Jiang 等人(2017)應用大數據分析方法，嘗試將龐大的信令資料轉換成民眾的旅次模式(travel pattern)。該研究以新加坡地區連續兩個禮拜的信令資料為分析對象，並將分析結果與人口普查資料及政府部門進行的家庭旅次訪問調查進行比較。首先透過距離以及停留時間的門檻值，判斷出用戶的停留點。由於此研究假設用戶每天一定從家出發，並在最後回到家中，故下一步為找出用戶的家所在位置同時過濾資料筆數過少的用戶。最後，透過整理用戶一天到訪過的停留點，將用戶的移動型態切分成單點旅次、兩點旅次、三點旅次等，並計算人口放大因子，將人口放大至真實數量，推估實際的旅次數。圖 2-6 為信令資料的推估與抽樣家庭訪問結果，雖然兩種分析方式結果不完全一致，但可以瞭解到利用信令資料進行推估能夠取得具有一定水準的結果。此研究亦提及，若能夠取得政府的土地利用資訊，得以針對旅次進行更進一步的探討。

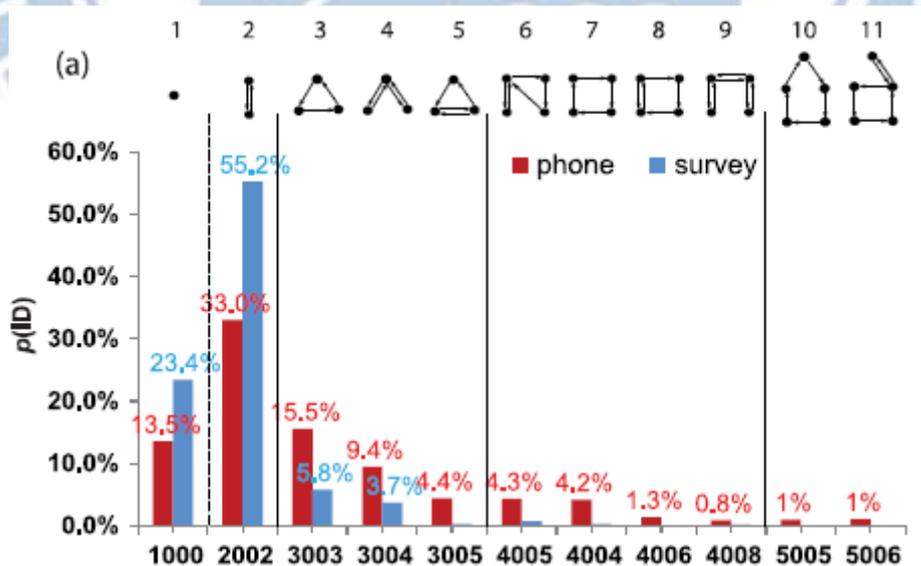


圖 2-6 信令資料的推估與抽樣家庭訪問結果

Mamei 等人(2017)利用信令資料並透過機器學習將用戶進行分類，共分成住戶、通勤人士、經過用戶、遊客以及短途旅行者。由於無法直接從信令資料中瞭解用戶的分類，因此該研究首先以人工的方式，為一些用戶加上標籤，而後再以這些用戶為訓練資料取得模型，並為其他用戶進行標籤分類。完成標籤分類並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，圖 2-7 為研究中推估而得的旅遊人數與政府部門統計旅遊人數之關聯結果，由於皮爾生相關係數高達 0.97，可以瞭解兩者為高度相關。

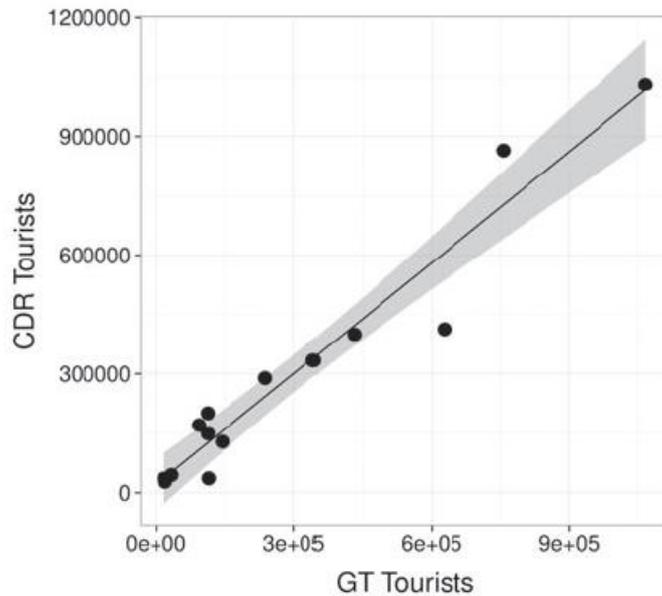


圖 2-7 信令資料推估結果與政府公布數據之關聯性

Zahedi 和 Shafahi(2017)將信令資料應用於辨別用戶的活動型態，以瞭解手機用戶的旅次目的。作者使用之信令資料為以西拉地區 300,000 位市民為對象，並以 5 分鐘為單位，記錄用戶當下所連接的塔台位置，共長達 40 小時。作者認為若用戶處於移動狀態時，不會與同一基地台連接超過 20 分鐘，故作者以 20 分鐘為門檻值，先從信令資料中找出用戶的停留點。在停留點判斷完成之後，作者先以上午八時至下午五時以及晚上十一時至隔天上午八時分別找出用戶的工作場所以及住家位置，而後再結合土地資訊，判斷出非住家亦非工作場所的停留點屬性。在得知所有停留點的屬性後，即可得知用戶一天旅次數量以及各個旅次的目的，作者再進一步與政府所做的訪問調查比較，結果如圖 2-8 所示。圖 2-8 中，W 代表工作旅次、E 代表就學旅次、P 代表接送旅次、C 代表商業旅次、R 代表娛樂旅次、A 代表行政旅次、H 代表家旅次，由圖可以看出出兩種推估方式所得到的結果差異並不大。

Cellphone data in 2016	Type	'W'	'E'	'P'	'C'	'R'	'S'	'A'	'H'
	Percentage	15.35	13.5	4.45	9.5	6.45	3.98	3.85	42.92
Survey data in 2000	Type	'W'	'E'	'P'	'C'	'R'	'S'	'A'	'H'
	Percentage	19.76	13.5	2.59	7.3	3.12	2.8	2.29	48.03

圖 2-8 以旅次目的分類之旅次占比

2.2 信令資料應用於觀光分析

Ahas 等人(2007)利用了信令資料來研究愛沙尼亞地區的旅遊情形，其信令資料是由該地區用戶數最多的電信業者所提供，資料中包含了亂碼的用戶編號、時間、國籍以及用戶所在位置之經緯度，如圖 2-9 所示。作者欲透過手機信令資料，分析外國觀光遊客造訪當地第二大城-塔爾圖的情形，故作者首先透過信令資料中的國籍欄位篩選出國籍非位於愛沙尼亞的用戶。作者接著統計自 2004 年 4 月至 2006 年 7 月的信令資料，瞭解造訪了塔爾圖的外國觀光客尚前往何處旅遊，結果如圖 2-10 所示。圖 2-10 為前十大造訪塔爾圖的遊客前往旅遊的地區，從圖片中除了可以瞭解到較多遊客造訪的旅遊地點，圖中的數字亦代表了該地點被造訪的次數。透過瞭解這樣的資訊，政府單位可以據此規劃旅遊方面的行銷活動，甚至是與鄰近地區進行合作，促進觀光產業的活絡。

Time	Rand. ID	Country	E	N
01/04/2004 00:00:04	43204	FIN	E24-45-34.00	N59-26-36.00
01/04/2004 00:00:23	55861	FIN	E24-44-37.00	N59-25-49.00
01/04/2004 00:00:24	770841	RU	E27-44-38.00	N57-52-12.00
01/04/2004 00:00:31	505262	LAT	E24-45-34.00	N59-26-36.00
01/04/2004 00:00:35	49104	FIN	E24-44-41.00	N59-25-27.00
01/04/2004 00:00:36	586313	UK	E24-43-53.00	N59-26-06.00
01/04/2004 00:00:39	585543	UK	E24-44-37.00	N59-25-49.00

圖 2-9 愛沙尼亞地區信令資料格式

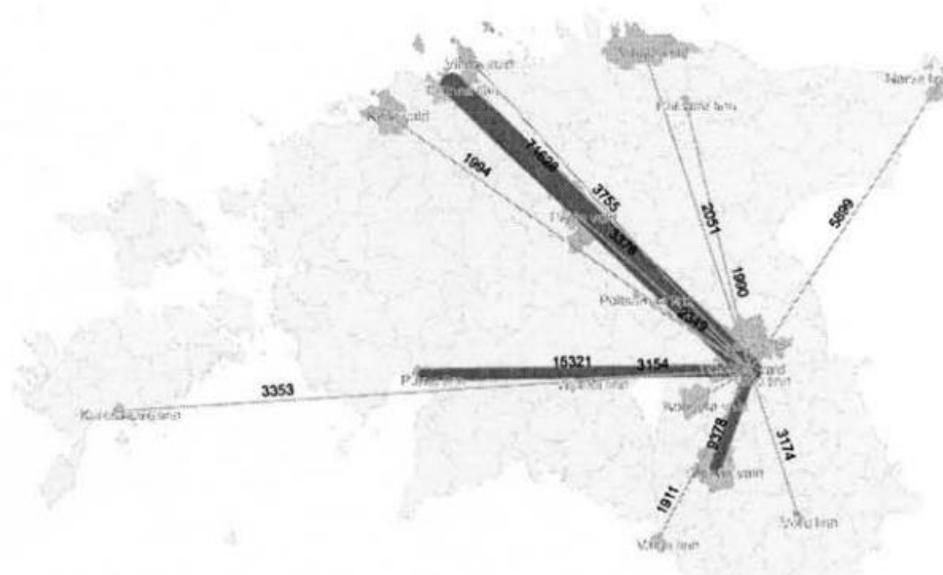


圖 2-10 前十大造訪塔爾圖的遊客前往旅遊的地區

Leng 等人(2016)以歐洲地區之安道爾(Andorra)為對象，研究觀光遊客的活動。信令資料為由當地的電信公司提供，包含用戶編號、時窗、經緯度、註冊國家以及手機型號，再搭配當地的地圖資訊進行分析。作者除了從信令資料中

找出新造訪以及重複造訪的遊客之外，尚分析了道路的路況。圖 2-11 為安道爾地區的路網圖，作者先自信令資料產製出 OD 矩陣，再透過交通量指派將需求指派至各個路徑，即可從中發現較為擁擠的路段。



圖 2-11 安道爾地區路網圖

Chen 等人(2018)使用海南地區於 2015 年 8 月至 12 月之匿名信令資料與興趣點資料(point-of-interests, POI)，透過建構機器學習之演算法，試圖找出觀光熱點地區以及遊客高頻率的移動樣態。作者將機器學習演算法分為三層，分別為資料層、演算法層以及應用層，在資料層中輸入所需使用到之信令資料以及興趣點資料。在演算法層中，主要的工作為將用戶的移動軌跡進行轉換，透過 DBSACN 分群演算法將用戶之停留點進行分群，得到興趣點區域(region-of-interests, ROI)，並以 PrefixSpan 演算法進行循序樣式探勘，找到較高頻率之移動樣態。最後，在應用層中，進行熱門 ROI 分析、住宿地點偏好分析以及旅遊季節性之分析。圖 2-12 為三亞市中，較為熱門的拜訪地點，而其中以亞龍灣以及大東海灣萬豪酒店為最多人拜訪的興趣區域；圖 2-13 為三亞市高頻移動樣態，由圖可以瞭解到亞龍灣為最熱門的觀光景點，遊客幾乎會花一整天的時間在該處遊玩以及住宿。



圖 2-12 三亞市熱門興趣區域



圖 2-13 三亞市高頻移動樣態

交通部運輸研究所(2018)透過對信令資料進行關聯分析，找出高雄市中關聯性較高之觀光景點。該研究取得了高雄市 2017 年 4 月 16 以及 4 月 23 日兩天之信令資料，資料中包含了用戶編號、時間以及用戶所在的二級發布區，而高雄市共有 1,234 個二級發布區。除了信令資料外，尚有觀光景點之位置資訊，可從中得知 480 個觀光景點所對應的二級發布區。該研究首先從信令資料中找出景點停留點，若一個二級發布區在用戶一天的信令資料中出現資料筆數大於 15 筆，則認定為用戶可能的停留點，並找出用戶的停留景點集合。最後，針對停留點集合進行關聯分析，找出兩天共同的關聯規則，並據此向公共運輸提出改善建議。

交通部運輸研究所(2019)利用信令資料，分析了花蓮地區的觀光需求。作者取得了兩份資料，第一份為三個周末中曾被記錄到於花蓮地區活動之用戶的信令資料，資料中包含了用戶編號、時間以及用戶所在位置所對應的網格，每一網格均為長寬皆為 50 公尺之正方形。第二份為當地較為著名的觀光景點與火車站場站，合稱為興趣點，所對應的網格，由於興趣點均有一定之幅員，故每一興趣均會由數個至數十個網格組成。該研究首先以火車站為標的，找到透過搭乘火車來往花蓮的用戶，並將其視為搭乘公共運輸的潛在用戶，並針對這些用戶進行關聯分析，找出發生頻率較高的旅遊樣態。在找到高頻樣態之後，該研究進一步將這些旅遊樣態與既存之觀光公共運輸服務路線進行比對，並找出

服務的缺口，向政府單位提出路線的改善建議。

2.3 小結

由 2.1 小節的文獻可以發現，已有不少利用信令資料進行旅次特性的研究，包含起迄矩陣的推估、旅次目的以及旅次長度。然而，在回顧文獻的過程中，也發現多數文獻均以家旅次以及工作旅次為主進行分析，探討觀光旅次的文獻並不多，又於分析觀光旅次的文獻中，並無討論如何利用信令資料區別用戶是否為觀光用戶的相關研究。因此，本研究欲嘗試以信令資料為基礎，配合景點相關的位置資訊，建立判斷觀光遊客的邏輯，從全部的用戶中找出移動型態屬於觀光遊客之用戶，並期待能將成果提供觀光旅運需求分析，協助得到更為精確的分析結果。



第三章 研究方法

本研究之研究目的，主要是從信令資料中，找出觀光用戶，也就是根據用戶一天的移動型態，判斷該名手機用戶之移動行為是否屬於旅遊活動。對於觀光用戶而言，其一天會花費一定的時間在旅遊上，故本研究將針對此概念進行方法之設計。藉由信令資料的協助，可以大致了解用戶有多少時間可以利用，若用戶花費於興趣點的時間佔可以利用的時間一定比例以上時，則可以推斷該位手機用戶應為觀光用戶。

本研究之研究方法，其流程圖如圖 3.1 所示。第一步為資料蒐集，蒐集本研究所需使用之資料，包含手機信令資料以及興趣點位置資料，而資料提供者將會事先協助將花蓮縣住戶之信令資料刪除，故本研究取得之信令資料僅為非花蓮縣住戶之信令資料。接著則是將信令資料與興趣點位置資料比對，找出用戶曾經停留的興趣點。最後，將個別計算每一用戶之可用旅遊時間、移動時間以及停留於興趣點之時間，在加總移動時間與興趣點停留時間後，將其除以可用旅行時間後即可得到用戶可用旅行時間使用比例，該比例大於或等於 0.5 之用戶為觀光用戶。

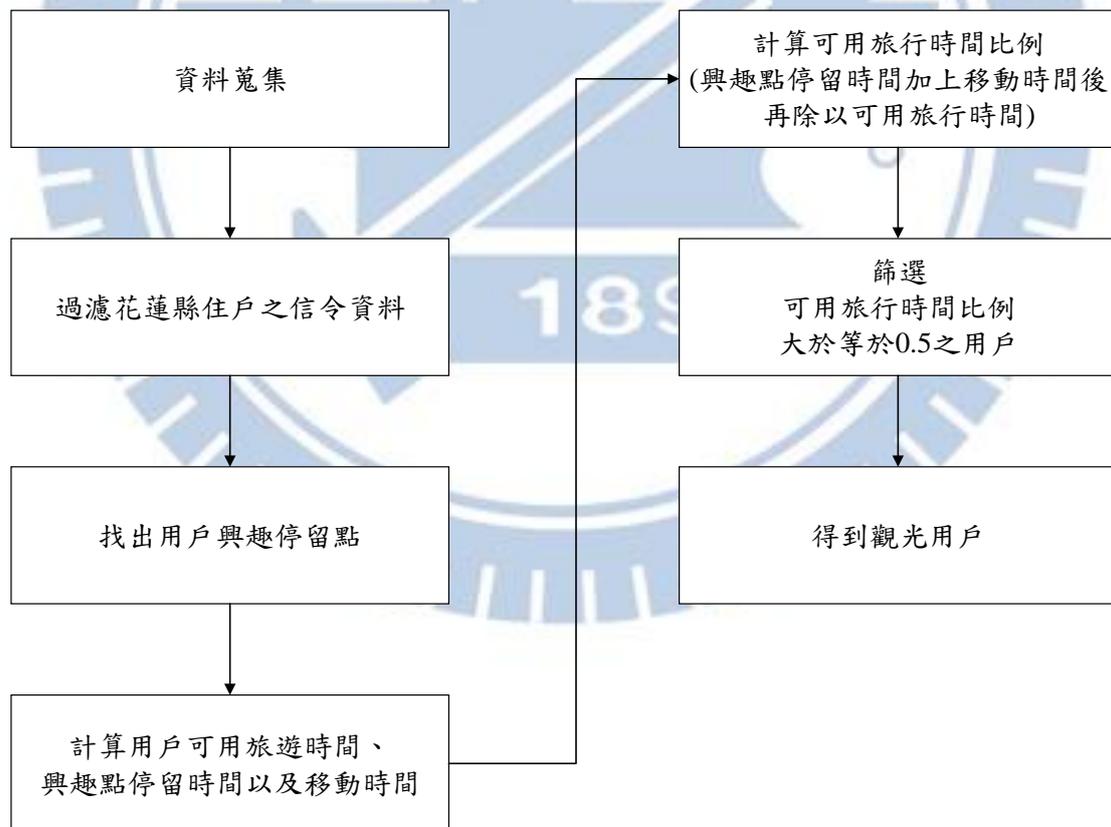


圖 3-1 研究方法流程圖

3.1 資料蒐集

本研究最主要之資料來源為中華電信所提供之手機信令資料以及勤崙國際科技股份有限公司提供之興趣點位置資料。本節將詳細介紹這兩種資料，說明各個資料的產生方式，並以資料的一部分作為例子，闡述資料中的各個欄位。最後，說明如何運用所取得之資料，進行後續的分析。

3.1.1 信令資料

當我們使用手機進行傳送訊息或是瀏覽網頁時，手機與基地台會產生連結，此時基地台會即時記錄相關的資訊，而這些資料就是信令資料。但若手機為待機狀態，用戶沒有使用手機時，手機則會每隔一段時間回傳位置資訊給予基地台。至於當手機處於關機狀態時，則不會產生任何信令資料。

如同上一段所述，信令資料的產生，主要來源為手機用戶的上網行為，當用戶透過手機連結網路時，就會有信令資料。換言之，用戶使用的頻繁程度會影響信令資料量數量的多寡，若用戶長時間未使用手機時，則僅會每隔一段時間產生一筆信令資料。因此，針對手機使用行為較為頻繁的用戶，由於能夠較為完整地瞭解其移動軌跡，故能較準確地判定其是否為觀光用戶；反之，針對信令資料比數很少的用戶，由於信令資料不足，難以瞭解其活動狀況，便不易判定其是否屬於觀光用戶。

本研究取得之花蓮地區信令資料，共有三個欄位，分別為：用戶編號、時間、以及位置，如表 3.1 所示。第一欄用戶編號，由於有個資法的問題，用戶編號有經過加密處理，由 64 個英文字母以及數字組成，表 3.1 中之範例僅以前 8 個字元作為代表。第二個欄位為日期以及時間，即當用戶的手機與基地台有連接時，會記錄當下的日期以及時間。最後一個欄位為用戶之位置訊，本研究取得之信令資料其精確度為長寬均為 50 公尺之網格，也就是 250 平方公尺，當用戶位於某一網格中時，位置欄位即會記錄該網格之編號。

表 3.1 信令資料範例

用戶編號	時間	位置
bb981cf1	2018-05-19 02:21:00	102001013
bb981cf1	2018-05-19 02:25:00	102001013
b8e0daaa	2018-05-19 00:06:00	102020023
b8e0daaa	2018-05-19 00:09:00	102020023
b8e0daaa	2018-05-19 00:11:00	102020023
8c315e96	2018-05-19 02:16:00	243062012

由於本研究是以非花蓮縣之住戶為研究對象，因此信令資料仍需事先進行處理，電信公首先會依據用戶連續一個月的信令資料，判定用戶的居住地。在得知用戶的居住地資訊後，即可依據居住地資訊，篩選居住地不位於花蓮縣之

用戶的信令資料。若有用戶之居住地不位於花蓮縣，但其工作場所位於觀光景點，由於其僅會造訪單一觀光景點，故被判定屬於觀光用戶的機會並不高。

3.1.2 興趣點位置資料

觀光客最常造訪之處莫過於觀光景點，本研究中將觀光景點稱為「興趣點」。為了判斷用戶是否有造訪興趣點，除了用戶信令資料之外，尚須取得景點之位置資訊。本研究自交通部運輸研究所「應用旅次特性大數據精進公共運輸服務計畫」取得兩份資料，第一份為花蓮地區的 55 個興趣景點的位置資訊，第二份為網格中心點經緯度資料。

有關觀光景點的位置資訊如表 3.2 所示，資料中總共包含三個欄位。第一個欄位為該地點之名稱，第二個欄位為該地點的編號，每一個興趣點均有唯一編號。最後一個欄位為興趣點包含之網格的編號，由於興趣點範圍較大，因此會同時涵蓋許多網格。根據表 3.2 的範例，網格編號 259678001、259678002 以及 259678011 之網格均是位於七星潭風景區。

表 3.2 觀光景點包含之網格資料範例

興趣點名稱	興趣點編號	網格編號
七星潭風景區	35	259678001
七星潭風景區	35	259678002
七星潭風景區	35	259678011

第二份資料為網格中心經緯度資料，如表 3.3 所示。此份資料總共有六個欄位，第一個欄位為網格編號；第二個、第三個以及第四個分別為網格座落的縣市、鄉鎮市區以及里鄰，可以知曉網格所在的大略位置；第五個以及第六個欄位為網格中心點的經度以及緯度，代表著網格最精確的位置資訊，經度準確至小數點以下四位數，而緯度準確至小數點以下五位數。

表 3.3 網格中心點經緯度範例

網格編號	縣市	鄉鎮市區	村里	中心經度	中心緯度
156382079	花蓮縣	富里鄉	竹田村	121.3427	23.20073
166719045	花蓮縣	卓溪鄉	古風村	121.1849	23.20821
166266010	花蓮縣	富里鄉	萬寧村	121.3247	23.22830
166266011	花蓮縣	富里鄉	新興村	121.3252	23.22424

由於後續分析需要同時使用到興趣點編號、網格編號以及網格中心點，因此需要先將上述提及的兩份資料進行合併。合併的依據是兩份資料中的「網格編號」，當兩份資料有相同網格編號存在時，便會將網格中心點經緯度中的所有屬性填寫至興趣點資料中，得到初步的整合結果。本研究最後則再進一步過濾出所需的屬性，包含興趣點名稱、興趣點編號、網格編號、網格中心經度以及網格中心緯度五個欄位，並將此表稱為興趣點網格中心表，如表 3.4 所示，此

份資料將會於後續階段分析中使用。

表 3.4 興趣點網格中心表範例

興趣點名稱	興趣點編號	網格編號	網格中心經度	網格中心緯度
九曲洞隧道	10	276923020	121.5293	24.17135
九曲洞隧道	10	276923030	121.5298	24.17135
九曲洞隧道	10	276923040	121.5303	24.17135
大石鼻山步道	27	219463001	121.5514	23.69768
大石鼻山步道	27	219463003	121.5514	23.69858
天祥風景區	43	288810075	121.4437	24.19194
天祥風景區	43	28810076	121.4437	24.19240

3.2 找出停留之興趣點

由於需要瞭解用戶是否有於興趣點停留，故本小節將會說明興趣點位置與信令資料的比對，以及用戶是否停留的判定方式。此階段會確認信令資料中用戶所在的網格，是否為 55 個興趣點涵蓋之網格，或者是位於興趣點附近之網格。最後將計算用戶連續出現於興趣點的時間長度，並藉由門檻值的設定定義停留。

3.2.1 判斷用戶是否位於興趣點

欲瞭解觀光用戶的需求，需要先瞭解用戶是否有曾經於興趣點活動。透過將信令資料的網格欄位與興趣點網格中心表(表 3.4)的比對，可以確認用戶所在的位置是否為屬於興趣點的網格，也就是用戶是否位於興趣點。然而，由於 Alexander 等人(2014)以及 Widhalm 等人(2015)均於文獻中指出信令資料具有飄移的情形，當信號飄移時，將無法精確掌握此使用者實際的時空軌跡，因此本研究將會透過設定一距離容忍值，適當地放寬興趣點的範圍，試圖降低資料飄移所帶來的影響。

以圖 3-2 為例，若有一個興趣點由 8 個網格所組成，且網格正中央的點代表該網格之中心點。當信令資料顯示用戶所處位置於如圖中之 A 點時，由於 A 點位於該興趣點原有之網格中，故可知用戶正位於興趣點。此時會將該筆信令資料的網格欄位以表 3.2 中之興趣點編號取代，表示該筆信令資料是位於特定的興趣點。

若用戶非位於興趣點所包含的網格中，此時會計算該網格與最接近的興趣點網格間的距離，若距離超出所設定容忍值時，則代表用戶非位於興趣點；反之，則將用戶視為位於興趣點，同樣將該筆信令資料的網格欄位以興趣點自身的編號取代原有的網格編號。有關於容忍值設定的部分，將於第四章進行說明。

以圖 3-2 為例，圖中的實線部分為一興趣點，在容忍值設定為 150 公尺的

情境下，由於 B 點與其最接近的興趣點網格距離為 100 公尺，故判定該筆信令資料位於興趣點，將其所在位置之網格編號以興趣點自身的編號取代。至於 C 點，由於 C 點與其最接近的興趣點網格為 158 公尺，超過設定之容忍值 150 公尺，故判定 C 點非位於興趣點。

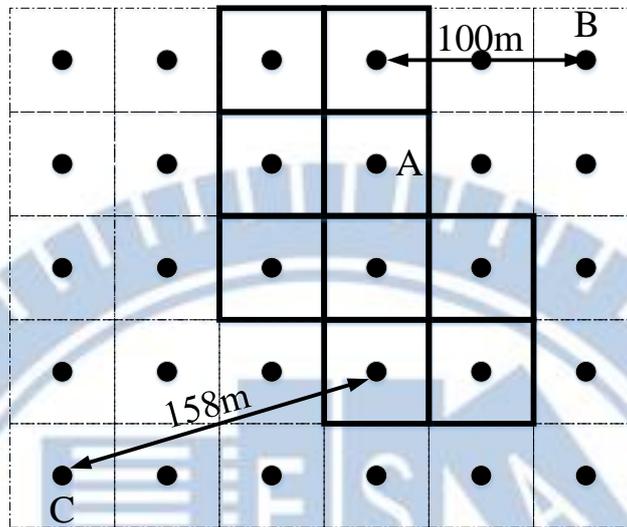


圖 3-2 興趣點比對範例

3.2.2 判斷用戶是否於興趣點停留

在信令資料與興趣點位置比對後，接著為判定用戶是否於興趣點停留。首先，需要計算用戶連續出現在同一地點的時間並產製連續時間表。連續時間的計算，是當用戶連續出現在同一個網格或興趣點時，將連續出現的最後一筆信令資料時間減去第一筆信令資料的時間，得到的時間差值即為用戶在此網格或者是興趣點的連續出現時間。若用戶並未於同一網格或興趣點出現時，代表用戶未於同一地點連續出現，將不會將信令資料列入連續出現表中。

連續時間計算的範例如圖 3-3 所示，在經過是否位於興趣點的比對後，用戶 0cbdcf 於 5/19 日中的 11:40 至 11:46 均出現在同一網格，因此於連續時間表記錄此用戶於網格 2547810214 停留 6 分鐘。另外，該用戶亦於網格 254781082 連續出現兩分鐘，亦記錄於連續時間表中。然而，用戶僅於 11:39 出現於網格 254779021，未有連續出現之情形，故不會將用戶出現於此網格之信令資料記錄於連續時間表中。

用戶編號	日期	網格編號
0cbdcf	2018-05-19 11:39:00	254779021
0cbdcf	2018-05-19 11:40:00	254781024
0cbdcf	2018-05-19 11:41:00	254781024
0cbdcf	2018-05-19 11:42:00	254781024
0cbdcf	2018-05-19 11:43:00	254781024
0cbdcf	2018-05-19 11:44:00	254781024
0cbdcf	2018-05-19 11:45:00	254781024
0cbdcf	2018-05-19 11:46:00	254781024
0cbdcf	2018-05-19 11:48:00	254781082
0cbdcf	2018-05-19 11:50:00	254781082

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0cbdcf	254781024	2018-05-19 11:40:00	2018-05-19 11:46:00	6
0cbdcf	254781082	2018-05-19 11:48:00	2018-05-19 11:50:00	2

圖 3-3 連續時間表計算範例

連續時間表產製完成後，透過將比對連續出現表的網格編號欄位與興趣點網格中心表(表 3.4)的興趣點編號，可以比對出該筆連續出現記錄是否位於興趣點，並僅留下於興趣點連續出現之資訊。由於連續出現表中，依然包含用戶於一般網格(非興趣點)連續出現的記錄，如圖 3-3 中連續時間表的兩筆資料，因此透過與興趣點網格中心表比對，可以準確地將一般網格的連續出現記錄移除，僅留下用戶於興趣點連續出現之記錄。以圖 3-4 為例，圖中上方的表格為初步產製之連續時間表，經過與興趣點網格中心表比對後，可以僅留下該位用戶於興趣點的資料記錄，如圖中下方的表格。

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0cbdcf	42	2018-05-19 10:14:00	2018-05-19 10:43:00	29
0cbdcf	254781024	2018-05-19 11:40:00	2018-05-19 11:46:00	6
0cbdcf	254781082	2018-05-19 11:48:00	2018-05-19 11:50:00	2
0cbdcf	254780040	2018-05-19 12:34:00	2018-05-19 13:00:00	26
0cbdcf	50	2018-05-19 14:40:00	2018-05-19 15:16:00	36

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0cbdcf	42	2018-05-19 10:14:00	2018-05-19 10:43:00	29
0cbdcf	50	2018-05-19 14:40:00	2018-05-19 15:16:00	36

圖 3-4 連續時間表過濾一般網格範例

由於信令資料會有飄移的現象，因此可能發生用戶於同一天有兩筆以上於興趣點連續出現的情形。然而，民眾在旅遊時，並不會安排在同一天中重複前往

相同興趣點，因此本研究會將同一天中連續出現於同一興趣點之時間長度加總，合併至最早出現於該興趣點之連續記錄中。以圖 3-5 為例，由圖中上方表格可以發現，用戶 0cd5b0 於 5 月 19 日當天有兩筆連續出現於編號 31 之興趣點。由於用戶幾乎不會造訪同一興趣點兩次，因此將發生時間較晚之記錄合併至發生時間較早之記錄中，如圖中下方表格所示，

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0cd5b0	31	2018-05-19 13:25:00	2018-05-19 13:30:00	5
0cd5b0	31	2018-05-19 13:37:00	2018-05-19 14:40:00	63

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0cd5b0	31	2018-05-19 13:25:00	2018-05-19 13:30:00	68

圖 3-5 興趣點連續時間加總範例

最後，需要設定興趣點連續時間門檻值，判定用戶是否於興趣點停留。將連續時間表與興趣點網格中心表比對後，其結果可提供用戶於興趣點連續出現的時間長度資訊。透過設定興趣點連續時間門檻值，過濾於興趣點連續出現的時間長度小於門檻值的資料，剩餘之記錄則為用戶當天曾經停留的興趣點。以圖 3-4 中的下方表格為例，若興趣點連續時間門檻值為 30 分鐘，則由表中第一筆記錄的連續時間未達門檻值，將會予以刪除。

3.3 判斷觀光用戶

在得到用戶的停留點資訊後，最後一個步驟就是進行觀光用戶之判定，本小結將介紹本研究如何將用戶判定為觀光用戶。對於觀光用戶，其一天將花費不少時間於興趣點從事旅遊活動，方能稱之為觀光用戶。而透過信令資料，可以瞭解用戶的軌跡，瞭解用戶移動的情形，或者是用戶於興趣點停留的時間。因此，本研究將以時間為基礎，利用從信令資料中能夠得到的資訊，判定用戶是否為觀光用戶。

3.3.1 用戶可用旅行時間

當用戶從事旅遊活動時，實際上能夠運用於觀光景點的時間是有限的，因此瞭解用戶可用的旅行時間，有助於判斷其是否為觀光用戶。本研究認為，絕大多數的用戶以早上至傍晚之前為旅遊時間，且本研究所取得的興趣點多數均較適合於白天前往遊玩，因此本研究將設定用戶從事觀光活動的時間為早上九點至下午五點，最長共 8 小時。

然而，用戶不一定於上述之時間區段間出現於花蓮，故用戶的可用旅行時間的計算需要有相對應的調整。在計算用戶可用旅行時間時，會先檢視用戶一天原始信令資料中第一筆信令資料以及最後一筆信令資料的時間，以更精準得到用戶的可用旅行時間，有關於可用旅行時間的計算流程如圖 3-6 所示。若

用戶當天第一筆信令資料的時間晚於上午九時，則用戶可用旅行時間的起始時間點則會設定成與第一筆資料相同之時間，但若早於上午九時，則以上午九時為起始時間點。同理，若最後一筆信令資料的時間早於下午五時，則會將可用旅行時間的終止時間點設定成與最後一筆資料相同之時間，但若晚於下午五時，則終止點設定為下午五時。最後，用戶的可用旅行時間為起始時間點至終止時間點之時間長度。

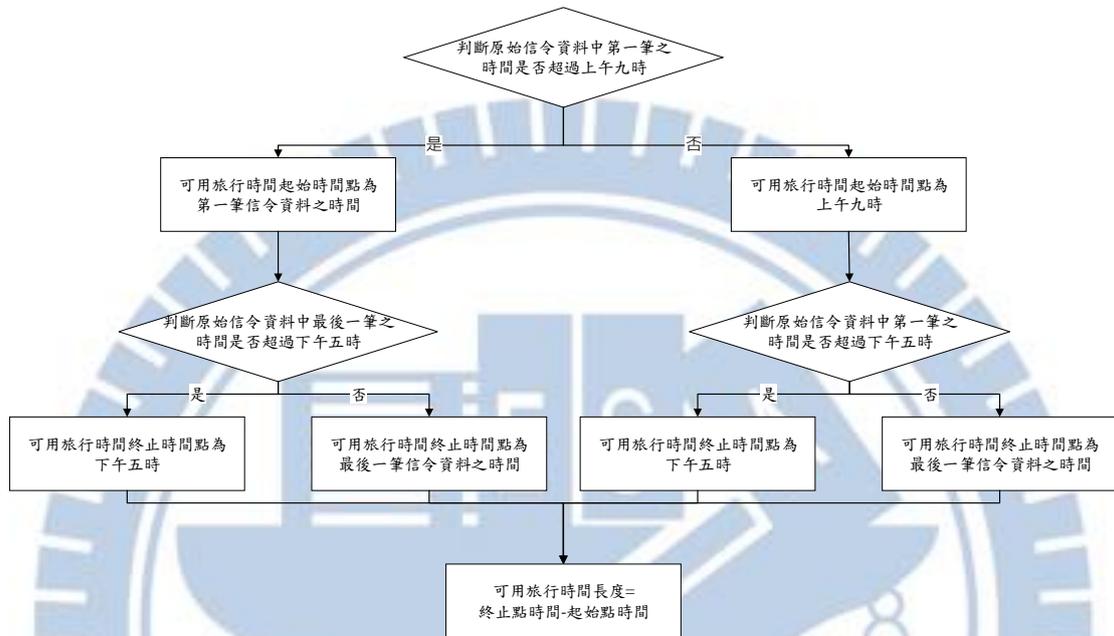


圖 3-6 可用旅行時間判斷流程圖

以表 3.5 為例，用戶 0cbdcf 於 5 月 19 日之原始信令資料中第一筆信令資料的時間為 06:32，由於早於 09:00，故其可用旅行時間的起始點設定為 09:00。該用戶最後一筆信令資料的時間為 22:22，由於超過 17:00，故終止點設定為 17:00。故該名用戶之可用旅行時間即是自上午九點至下午五點，共計 8 小時。

表 3.5 用戶可用旅行時間計算範例

用戶編號	時間	網格
0cbdcf	2018-05-19 06:32:00	242956065
0cbdcf	2018-05-19 07:26:00	242956069
.	.	.
.	.	.
.	.	.
0cbdcf	2018-05-19 22:21:00	254781022
0cbdcf	2018-05-19 22:22:00	254781022

3.3.2 用戶停留於興趣點之時間

在 3.3.1 小節中，說明了本研究如何計算用戶的可用旅遊時間。但自 3.2.2 小節中所產出的連續時間表中可能包含未於旅遊活動時間內的記錄，也就是仍有不在用戶可用旅遊時間內的興趣點停留記錄。因此，尚須先針對連續時間表進行過濾，僅留下於可用旅行時間區段內之興趣點停留記錄。以圖 3-7 為例，圖 3-6 為用戶 0cbdcf 停留時間表的一部分，其中可以發現有一筆停留記錄的起始時間為 17:12:00，超出所設定的最晚旅遊活動時間(17:00:00)，因此該筆停留記錄會從停留時間表中刪除。

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0cbdcf	37	2018-05-19 10:14:00	2018-05-19 10:43:00	29
0cbdcf	24	2018-05-19 14:40:00	2018-05-19 15:16:00	36
.
.
.
0cbdcf	1	2018-05-19 16:17:00	2018-05-19 17:00:00	43
0cbdcf	12	2018-05-19 17:12:00	2018-05-19 17:24:00	12

↓

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0cbdcf	37	2018-05-19 10:14:00	2018-05-19 10:43:00	29
0cbdcf	24	2018-05-19 14:40:00	2018-05-19 15:16:00	36
.
.
.
0cbdcf	1	2018-05-19 16:17:00	2018-05-19 17:00:00	43

圖 3-7 以旅遊活動時間過濾停留時間表範例

透過手機信令資料計算而得的興趣點連續出現時間無法非常準確地與實際停留時間契合，但能得知用戶曾經於該興趣點進行停留。以圖 3-8 為例，圖 3-8 之為將用戶 f5320f 於 5 月 19 日之部分信令資料。根據該圖可知該用戶當天應前往花蓮遠雄海洋公園遊玩，其中第一筆記錄為上午 11 時 43 分至上午 11 時 48 分，而第二筆為下午 2 時 8 分至下午 4 時 24 分。

然而，根據原始信令資料，該用戶於上午 11 時 49 分以及上午 11 時 50 分至下午 1 時 50 分時，均不位於園區內。由於此兩筆資料的關係，中斷了用戶的興趣點連續出現時間的計算，即使將兩筆位於花蓮遠雄海洋公園之連續時間加總，仍舊只有約兩個小時半，不足以讓用戶於園區內活動。因此，本研究將另尋其他方法以計算用戶於興趣點停留的時間長度。



圖 3-8 用戶 f5320f 信令資料範例

關於興趣點停留時間長度，本研究透過旅遊應用程式 TripAdvisor，找尋各個興趣點的建議停留時間，產製興趣點停留時間表，如表 3.6 所示。若無法找到相關資訊時，再至 VZ TAIWAN 或者是 Travel King 等其他應用程式進行蒐集。從停留時間表中，可以得到用戶的興趣停留點，在對照興趣點停留時間表找到個別的建議停留時間，加總後即是用戶停留於興趣的時間。

表 3.6 興趣點停留時間表

編號	興趣點名稱	建議停留時間(小時)
1	花蓮文化創意產業園區	2
2	七星潭風景區	1.5
3	九曲洞隧道	1
4	大石鼻山步道	1.5
5	大農大富平地森林園區	3
6	六十石山	3
7	天祥風景區	1
8	太巴壠文化園區	1
9	太平洋公園(北濱朝日沙灘公園)	2
10	太魯閣國家公園遊客中心	2.5
11	月洞遊憩區	1
12	牛山風景區	1
13	台開心農場(洄瀾灣開心農場)	2
14	玉里神社遺址	1
15	石門遊憩區	0.5
16	石梯坪遊客中心停車場	1.5
17	立川漁場	2.5
18	吉蒸牧場	1.5

19	向日廣場	1
20	池南國家森林遊樂區	3
21	秀姑巒溪泛舟中心(秀姑巒溪遊客中心)	5
22	赤科山	2
23	岳王亭	1.5
24	林田山林業文化園區	2
25	松園別館	1.5
26	知卡宣綠森林親水公園	2
27	芭崎遊憩區瞭望台停車場	0.67
28	花蓮文化中心石雕公園	1
30	花蓮遠雄海洋公園	7
31	花蓮觀光糖廠冰品	2
32	砂卡礑步道(神秘谷步道)	1
33	馬太鞍溼地生態館	2
34	曼波海灘	1.5
35	崇德礫灘	4
36	掃叭石柱(舞鶴石柱)	0.5
37	富源國家森林遊樂區停車場	2.5
38	雲山水自然生態農莊	1.5
39	新光兆豐休閒農場	4
40	新社梯田	0.25
41	瑞穗牧場	1.5
42	翡翠谷	2
43	舞鶴觀光茶園(公主咖啡)	2
44	鳳林環保科技園區 3D 親子水上樂園	7
45	慶修院	1
46	慕谷慕魚遊客中心	2.5
47	燕子口步道	1
48	蕃薯寮休息區	0.5
49	親不知子天空步道	1
50	錐麓古道(合歡越嶺古道錐麓段)	2
51	靜浦北回歸線界標	0.5
52	磯崎海水浴場	2
53	舊長虹橋	1
54	鯉魚潭風景遊憩區潭南停車場入口	2
55	羅山管理站暨遊客中心	1

延續用戶 0cbdcf 的例子，表 3.7 為該用戶在上午九時至下午五時的連續時間表。由表中可知其曾經於立川漁場、松園別館以及花蓮文化創意產業園區連續出現，代表其曾停留於上述之興趣點。再根據表 3.8 可知這三個景點的停留時間分別為 2.5 小時、1.5 小時以及 2 小時，加總後即是此用戶當天於興趣點停留的時間六小時。

表 3.7 用戶連續時間表範例

用戶編號	網格	起始時間	終止時間	連續時間
0cbdcf	17	10:14:00	10:43:00	29
0cbdcf	25	14:40:00	15:16:00	36
0cbdcf	1	16:17:00	17:00:00	43

3.3.3 用戶移動時間

用戶除了將時間花費在興趣點進行旅遊之外，在興趣點間移動的時間亦會於可用旅行時間中占有一定的比例，因此需要額外考量用戶進行移動所需要之時間。將用戶於興趣點停留時間與移動時間加總後，能夠更貼近現實中的情況。

本研究透過 R 語言中的 osrm 套件，取得用戶於興趣點間的移動時間。該套件需要的兩點的經緯度作為輸入資料，當套件取得兩點經緯度後，會將經緯度回傳至 OpenStreetMap，並由 OpenStreetMap 計算兩點間的最短路徑以及最短旅行時間並傳回套件。表 3.8 為 osrm 套件與 Google 地圖所計算之旅行時間，由表可以發現，無論是短距離或者是長距離，該套件所回傳的時間均與 Google 地圖之旅行時間相當接近，因此該套件所求得的旅行時間是值得採用的。

表 3.8 OpenStreetMap 與 Google 地圖計算之旅行時間

起迄對	osrm 套件計算時間	Google 地圖計算時間
花蓮文化創意產業園區 七星潭	15.5 分	15 分
慶修院 立川漁場	19.8 分	20 分
松園別館 花蓮遠雄海洋公園	20.6 分	22 分
慶修院 新光兆豐休閒農場	30.5 分	32 分
七星潭 鯉魚潭	34.1 分	32 分
花蓮文化創意產業園區 太魯閣	35.9 分	34 分
林田山 鯉魚潭	41.4 分	41 分

花蓮觀光糖廠 花蓮文化創意產業園區	59.1 分	55 分
花蓮文化創意產業園區 七星潭	1 小時 10 分	1 小時 4 分

由於套件本身需要兩點之經緯度之值，方能計算兩點間之距離。但興趣點會同時包含許多的網格，因此本研究在計算距離時，會將興趣點之經緯度以其所包含的網格經緯度取平均值作為代表。以表 3.9 為例，表 3.9 為花蓮文化創意產業園區之興趣點網格中心資料，總共包含了 9 個網格，故在計算距離時，會以九個網格中心經緯度的平均值代替，即花蓮文化創意產業園區的中心經度為 121.6046，中心緯度為 23.9765。

表 3.9 花蓮文化創意產業園區之興趣點網格中心資料

興趣點名稱	興趣點編號	網格編號	網格中心點經度	網格中心點緯度
花蓮文化創意產業園區	1	254233058	121.6041	23.97605
花蓮文化創意產業園區	1	254233059	121.6041	23.9765
花蓮文化創意產業園區	1	254233060	121.6041	23.97695
花蓮文化創意產業園區	1	254233068	121.6046	23.97605
花蓮文化創意產業園區	1	254233069	121.6046	23.9765
花蓮文化創意產業園區	1	254233070	121.6046	23.97695
花蓮文化創意產業園區	1	254233078	121.6051	23.97605
花蓮文化創意產業園區	1	254233079	121.6051	23.9765
花蓮文化創意產業園區	1	254233080	121.6051	23.97695

3.3.4 觀光用戶之判讀

對於觀光用戶而言，在觀光景點所耗費之時間，勢必在可用旅行時間佔有一定之比例，因此本研究將利用此概念，設計觀光用戶的判別邏輯。本研究首先將用戶周末兩日於興趣點停留的時間以及用戶所花費之移動時間加總，並除以周末兩日用戶可用的旅行時間，得到用戶從事旅遊活動的時間佔用戶可用旅行時間的比例，公式如下方的式子。

$$\text{可用旅行時間使用比例} = \frac{\text{興趣點停留時間} + \text{移動時間}}{\text{可用旅遊時間}}$$

如同前述，觀光用戶會於興趣點花費一定比例之時間，因此本研究將設定可用旅行時間使用比例之門檻值，將比例大於等於門檻值之用戶判定為觀光用戶。花蓮地區豐富的觀光資源總會吸引民眾從外縣市不遠千里前來旅遊，用戶將花費不少時間停留於興趣點以及於興趣點間移動，故本研究認為用戶應至少耗費一半的時間於興趣點遊玩，方能稱之為觀光用戶。有鑑於此，本研究將以

可用旅行時間使用比例門檻值 0.5 進行篩選，在眾多手機用戶中，找出觀光用戶。

3.4 小結

總括 3.1 小節至 3.3 小節，說明了本研究的研究方法。在 3.1 小節，介紹了信令資料、用戶居住地資料以及興趣點資料。3.2 小節介紹了如何將信令資料與興趣點網格中心表進行比對，確認用戶所在位置是否位於興趣點，並透過計算用戶於興趣點連續出現的時間長度判斷用戶是否曾經於興趣點停留。3.3 小節則說明了如何判定觀光用戶，藉由用戶停留於興趣點的時間以及移動時間與用戶可用旅行時間的比值，以是否大於門檻值判定其是否為觀光用戶。在瞭解用戶是否為觀光遊客後，可以直接針對這些用戶進行需求分析，方能提供民眾更符合其需求的觀光運輸服務。



第四章 實測分析

本章主要將說明將觀光用戶的判斷邏輯實際應用於信令資料之成果。在 4.1 小節中，將針對實測分析所使用的資料進行介紹，而在 4.2 小節將說明應用第三章所提出之觀光用戶判斷邏輯於信令資料的結果。4.3 小節中將針對可能影響判斷結果的參數進行敏感度分析，瞭解參數設定對於結果的影響程度，4.4 小節則接著嘗試以另一種方式計算用戶之興趣點停留時間，並測試在該計算方式下所判讀出之觀光用戶數。本研究所進行之前置資料處理以及判斷邏輯的建構，皆透過 R 語言實踐。

4.1 實測資料介紹

4.1.1 信令資料

本研究進行實測所使用之信令資料，是由電信公司所提供，資料共涵蓋三周周末，分別為 2018 年 5 月 19 日(六)、2018 年 5 月 20 日(日)、2018 年 5 月 26 日(六)、2018 年 5 月 27 日(日) 2018 年 6 月 2 日(六)以及 2018 年 6 月 3 日(日)。本研究取得的信令資料，為於前述日期中曾經到訪花蓮地區之手機用戶所產生之信令資料。由於本研究是以非花蓮縣之住戶為研究對象，故信令資料在取得前已事先將花蓮縣當地的住戶的信令資料剔除，僅留下居住地不位於花蓮縣之手機用戶。此外，信令資料同時亦已經過相關單位事先進行清洗以及插補，故本研究不需再針對取得之信令資料進行額外處理。

表 4.1 為六天的信令資料相關數據。在取得的信令資料中，資料用戶數大約落在 18,000 位左右，僅第二周(5/26 以及 5/27)有 20,000 位用戶。本研究將針對六日之手機用戶，利用於第三章中所提出的觀光用戶判斷邏輯，判斷用戶的行為是否屬於觀光遊客。

表 4.1 資料相關統計數據

	原始信令資料用戶數	原始信令資料筆數
5 月 19 日	18,607 位	7,691,579 筆
5 月 20 日	16,911 位	3,913,119 筆
5 月 26 日	20,005 位	10,919,705 筆
5 月 27 日	19,944 位	10,987,335 筆
6 月 2 日	18,136 位	9,286,832 筆
6 月 3 日	18,104 位	9,447,877 筆

4.1.2 興趣點

除了手機信令資料之外，興趣點位置資訊亦是本研究重要的輸入資料。對於從事觀光活動的民眾，必定會造訪興趣點，故本研究結合信令資料與興趣點位置資訊，進行觀光用戶的判斷邏輯設計。本研究自交通部運輸研究所「應用旅次特性大數據精進公共運輸服務計畫」取得之興趣點資料中，共包含 55 個興

趣點，如表 4.2 所示。

表 4.2 選定之興趣點

花蓮文化創意產業園區	七星潭風景區	九曲洞隧道
九曲洞隧道	大石鼻山步道	大農大富平地森林園區
六十石山	天祥風景區	太巴壠文化園區
太平洋公園	太魯閣國家公園 遊客中心	月洞遊憩區
牛山風景區	洄瀾灣開心農場	玉里神社遺址
石門遊憩區	石梯坪遊客中心停車場	立川漁場
吉蒸牧場	向日廣場	池南國家森林遊樂區
秀姑巒溪泛舟中心	赤科山	岳王亭
林田山林業文化園區	松園別館	知卡宣綠森林親水公園
芭崎遊憩區瞭望台 停車場	花蓮文化中心石雕公園	花蓮遠雄海洋公園
花蓮觀光糖廠冰品	砂卡礑步道	馬太鞍溼地生態館
曼波海灘	崇德礫灘	掃叭石柱
富源國家森林遊樂區 停車場	雲山水自然生態農莊	新光兆豐休閒農場
新社梯田	瑞穗牧場	翡翠谷
舞鶴觀光茶園	鳳林環保科技園區 3D 親子水上樂園	慶修院
慕谷慕魚遊客中心	燕子口步道	蕃薯寮休息區
親不知子天空步道	錐麓古道	靜浦北回歸線界標
磯崎海水浴場	舊長虹橋	鯉魚潭風景遊憩區 潭南停車場入口
羅山管理站暨遊客中心	-	-

4.2 實測結果

本研究將第三章中所設計的判斷邏輯實際以 R 語言撰寫後，將其應用於取得之兩日信令資料，並得到判斷結果。首先將會於 4.2.1 至 4.2.4 說明於第章所提出之研究方法，依序執行的成果，接著於 4.2.5 小節，呈現觀光用戶的判斷成果，最後於 4.2.6 小節針對判斷結果進行更詳盡的分析以及說明。

4.2.1 用戶於興趣點停留之判斷結果

如 3.2 小節，本研究所設計之觀光用戶判斷邏輯的第一步為找出用戶曾經到訪過的興趣點，其判斷結果將於此小節中呈現。首先，本研究將用戶所在位置與興趣點位置比對，透過比對用戶所在網格是否為興趣點所涵蓋之網格，判

斷用戶是否位於興趣點，判斷結果如表 4.3 所示。由表 4.3 可知，在六天的資料中，大約會有 5%的信令資料被比對到位於興趣點內。

除了找到位於興趣點的信令資料外，本研究尚將位於距離興趣點一定距離的信令資料視同位於興趣點。有關距離容忍值的設定，本研究參考交通部運輸研究所「應用旅次特性大數據精進公共運輸服務計畫」，由於該研究是以花蓮地區之信令資料進行觀光需求的分析，故本研究亦參考之將距離容忍值設定為 200 公尺。從表 4.3 可知，在六天的資料中，大約會有 8%的信令資料被比對到位於興趣點附近(不包含位於興趣點內的信令資料)。

表 4.3 信令資料與興趣點位置資料比對結果

	位於興趣點內的 信令資料筆數	位於興趣點附近的 信令資料筆數
5 月 19 日	454,197(5.9%)	721,863(9.4%)
5 月 20 日	242,705(6.2%)	336,104(8.5%)
5 月 26 日	555,115(5.1%)	871,858(8.0%)
5 月 27 日	583,800(5.3%)	926,153(8.4%)
6 月 2 日	471,945(5.1%)	725,466(7.8%)
6 月 3 日	477,161(5.1%)	734,922(7.8%)

完成信令資料與興趣點位置資料的比對後，下一步驟為計算用戶連續出現於同一網格或者是興趣點(包含出現於興趣點附近)的時間長度。表 4.4 為三周興趣點連續出現記錄之時間長度分布，由表 4.4 可以發現，各個組別占總資料筆數的比例大致相近，且 5 分鐘內的停留記錄筆數於六天中均約佔半數。另外，隨著時間長度的增加，符合的記錄筆數亦隨之減少，因此若將連續出現時間長度的門檻值設定得太高，會使得可用資料筆數太少。綜合以上所述，本研究將以連續出現 1 分鐘為門檻值，將用戶於同一興趣點連續出現 1 分鐘以上之記錄判定為停留。

表 4.4 三周興趣點連續出現記錄之時間長度分布

	5/19	5/20	5/26	5/27	6/2	6/3
1分鐘	8625(22.7%)	4981(24.5%)	10355(24.0%)	9535(23.5%)	7715(21.7%)	7595(23.3%)
2~5分鐘	9063(23.9%)	5097(25.1%)	9808(22.7%)	8870(21.9%)	8226(23.1%)	6869(21.1%)
6~10分鐘	4572(12.1%)	2280(11.2%)	4971(11.5%)	4813(11.9%)	4436(12.5%)	4022(12.3%)
11分鐘~20分鐘	4291(11.3%)	2079(10.2%)	4712(10.9%)	4359(10.8%)	4110(11.5%)	3424(10.5%)
21分鐘~30分鐘	2726(7.2%)	1379(6.8%)	3013(7.0%)	2828(7.0%)	2512(7.1%)	2316(7.1%)
31分鐘~45分鐘	2840(7.5%)	1457(7.2%)	3012(7.0%)	3078(7.6%)	2667(7.5%)	2483(7.6%)
45分鐘~1小時	1568(4.1%)	889(4.4%)	2058(4.8%)	1856(4.6%)	1782(5.0%)	1552(4.8%)
1小時~1.5小時	1641(4.3%)	920(4.5%)	2146(5.0%)	1935(4.8%)	1551(4.4%)	1707(5.2%)
1.5小時~2小時	763(2.0%)	394(1.9%)	813(1.9%)	819(2.0%)	702(2.0%)	685(2.1%)
2小時~3小時	680(1.8%)	369(1.8%)	800(1.9%)	812(2.0%)	679(1.9%)	605(1.9%)
3小時以上	1150(3.0%)	490(2.4%)	1426(3.3%)	1609(4.0%)	1220(3.4%)	1349(4.1%)
總和	37,919	20,335	43,114	40,514	35,600	32,607

4.2.2 用戶可用旅行時間計算結果

在判斷用戶是否為觀光用戶時，第一步為需要計算用戶可用的旅行時間，作為可用旅行時間使用比例的分母。表 4.5 為用戶三周用戶可用旅行時間的分布情形。從表中可以發現，周六(5/19、5/26、6/2)的分布情形，以 8 小時為最多；而周日(5/20、5/27、6/3)則較無明顯集中趨勢。本研究認為，由於周六大部分用戶可能會於花蓮過夜，因此可用旅行時間會較長；而由於周一為上班上課日，因此用戶可能會提前自花蓮返家，造成周日的可用旅行時間分布較為平均。

表 4.5 三周可用旅行時間人數分布

	5/19	5/20	5/26	5/27	6/2	6/3
4 小時	699	1082	365	781	269	593
5 小時	2010	3284	679	1719	557	1550
6 小時	2657	3713	834	2422	752	2023
7 小時	3357	653	918	2933	862	2807
8 小時	3650	486	13798	4882	12560	4494

4.2.3 用戶興趣點停留時間計算結果

於 3.2.2 小節中，本研究自信令資料中判斷出用戶曾經停留的興趣點，並利用旅遊應用程式中所建議的停留時間作為用戶停留於興趣點的時間長度。將一個用戶一整天所停留的興趣點，分別對應至建議停留時間並加總後，即是用戶當天於興趣點停留的時間長度，其分布情形如表 4.6。由表 4.6 可以發現有用戶於興趣點的停留時間超過 8 小時，超出本研究所設定的可用旅行時間(上午 9 時至下午 5 時，共 8 小時)，其原因將於 4.2.6 小節中說明。

表 4.6 三周興趣點停留時間人數分布

	5/19	5/20	5/26	5/27	6/2	6/3
0 小時	3,963	3,524	5,125	4,116	5,240	3,991
1 小時	867	674	892	761	883	823
2 小時	2,732	2,263	3,860	2,680	3,301	2,327
3 小時	522	306	659	543	619	468
4 小時	1,819	1,187	2,214	1,770	2,002	1,494
5 小時	463	166	566	377	440	390
6 小時	845	547	1,186	1,128	1,104	905
7 小時	468	348	614	660	423	470
≥8 小時	712	196	1,406	694	936	547

4.2.4 用戶興趣點之間移動時間計算結果

除了用於興趣點的停留時間外，本研究尚將用戶於興趣點之間移動的時間納入計算，詳細說明可參考 3.3.3 小節。表 4.7 為用戶興趣點移動時間的分布情形，從圖中可以發現六天中均有約一半的用戶的移動時間為 0 小時，其成因可能是用戶並未造訪任何興趣點，當用戶並未前往興趣點，其興趣點移動時間自然為 0。

表 4.7 用戶興趣點移動時間的分布情形

	5/19	5/20	5/26	5/27	6/2	6/3
0 小時	5,043	4,654	6,357	5,165	6,603	5,033
1 小時	3,174	2,323	3,855	3,111	2,989	2,722
2 小時	2,699	2,411	3,683	3,059	3,252	2,515
3 小時	1,353	1,090	2,154	1,231	1,748	1,058
>3 小時	122	201	545	171	399	139

4.2.5 觀光用戶判斷結果

在求得每一位用戶的可用旅行時間、興趣點停留時間以及移動時間後，即可計算用戶可用旅行時間的使用比例，其詳細說明可參考 3.3.4 小節。計算方式為將周末兩天的興趣點停留時間以及移動時間加總作為分子，並以兩天之可用旅行時間和作為分母，兩者相除後即是用戶可用旅行時間所使用的比例。

表 4.8 為三周可用旅行時間使用比例人數分布情形。從表 4.8 可以發現，六天中每天的用戶數總和均不與原始用戶數(表 4.1)相同。由於在判定過程中，當用戶於本研究所設定的可用旅行時間中無任何信令資料時，會使得該用戶完全被過濾，無從計算其可用旅行時間使用比例。

表 4.8 三周可用旅行時間使用比例人數分布

比例	第一周	第二周	第三周
0	2,999	3,786	3,892
0.1	264	490	538
0.2	652	1,004	1,108
0.3	837	1,238	1,237
0.4	779	1,163	1,106
0.5	804	1,215	1,110
0.6	848	1,136	1,022
0.7	838	1,171	920
0.8	822	1,182	948
0.9	758	975	895
1.0 以上	2,790	3,234	2,305
總和	12,391	16,594	14,991

表 4.9 三周觀光用戶數統計結果

	第一周	第二周	第三周
觀光用戶數 (所占比例)	6,860 (55.4%)	8,913 (53.7%)	7,200 (48.0%)

在得到每個用戶可用旅行時間使用比例值並將其四捨五入至小數點第一位後，本研究將比例值在 0.5 以上(包含 0.5)的用戶視為觀光用戶，結果如表 4.9 所示。三個周末中，觀光用戶的判斷結果大致相近，觀光用戶數均約占總用戶數的一半。但由表 4.8 可以發現，六天均出現可用旅行時間使用比例在 1.0 以上之用戶，也就是用戶於興趣點停留的時間與移動時間之總和超過了用戶可用旅行時間，其原因將於 4.2.6 小節中探討。

4.2.6 結果說明

在 4.2.5 小節中所呈現的用戶可用旅行時間使用比例(表 4.8)，將其整理後如表 4.10。從表 4.10 可以發現六天均有約 20%的用戶比例在 1.0 以上，但是可用旅行時間使用比例值於邏輯上不應大於 1.0。故在本節中，將深入瞭解這些用戶的信令資料以探究其發生原因。

表 4.10 觀光用戶判斷結果整理

	第一周	第二周	第三周
比例小於 0.5	5,531(44.6%)	7,681(46.3%)	7,791(52.0%)
比例 0.5 至 0.9	4,070(32.9%)	5,679(34.2%)	4,895(32.7%)
比例大於 1.0	2,790(22.5%)	3,234(19.5%)	2,305(15.3%)

以用戶 0b44e3 為例，表 4.11 為該用戶原始信令資料的一部分，表 4.12 為其興趣點連續時間表。由表 4.11 可以得知，該用戶的可用旅行時間為 4 小時 58 分(自上午 9 時 29 分至下午 2 時 29 分)，又由表 4.12，該用戶當天曾於花蓮文化創意產業園區、七星潭風景區以及松園別館停留，而對照興趣點停留時間表(如表 3.6)可知，三個景點所對應之停留時間分別為 2 小時、1.5 小時以及 1.5 小時，共計 5 小時。

由於信令資料的多寡會受到用戶手機使用行為的影響，因此信令資料並不能完全反映用戶實際的活動情形。如用戶 0b44e3，其信令資料之最後一筆為下午 2 時 29 分，但該用戶並不一定是在最後一筆信令資料產生後即沒有任何活動，因此該位用戶的實際可用旅行時間在計算上可能產生低估。而由於用戶可用旅行時間的低估，使得用戶可用旅行時間的使用比例提高，產生了超出 1.0 的情形。

表 4.11 用戶 0b44e3 原始信令資料

用戶編號	時間	網格編號
0b44e3	2018-05-19 09:29:00	254780030
0b44e3	2018-05-19 09:30:00	254780008
...
0b44e3	2018-05-19 14:30:00	300533054
0b44e3	2018-05-19 14:28:00	288715038

表 4.12 用戶 0b44e3 興趣點連續時間表

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
0b44e3	1	2018-05-19 11:36:00	2018-05-19 11:44:00	8
0b44e3	2	2018-05-19 12:04:00	2018-05-19 12:18:00	14
0b44e3	25	2018-05-19 13:12:00	2018-05-19 13:22:00	12

又以另外一位用戶 f5320f 為例，表 4.13 為其原始信令資料的一部分，由表 4.13 可知其可用旅行時間為 7 小時 58 分(自上午 9 時 01 分至下午 4 時 59 分)，將近 8 個小時。表 4.14 為該用戶 5 月 19 日之興趣點連續時間表，其中網格編號 28 代表花蓮文化中心石雕公園，30 則代表花蓮遠雄海洋公園，由表 4.14 可知該用戶當天曾到訪過前述的兩個興趣點。而對照興趣點停留時間表(表 3.6)，花蓮文化中心石雕公園之停留時間為 1 小時，而花蓮遠雄海洋公園則為 7 小時，故在僅計算興趣點停留時間總合與用戶可用旅行時間的比值下，仍已經超過 1.0。

如用戶 f5320f 之情形，由於其曾經造訪花蓮遠雄海洋公園，其建議停留時間為 7 小時，非常接近本研究所設定的最長可用旅行時間 8 小時，故其可用旅行時間使用比例已經高達 87.5%。此外，該用戶當天仍有事先前往其他興趣點，故在同時考慮興趣點移動時間的情形之下，使得其可用旅行時間使用比例超過了 1.0。由該用戶可以瞭解另一個可用旅行時間使用比例超過 1.0 的原因，當用戶造訪了建議停留時間較長的興趣點時，容易導致興趣點停留時間與移動時間超過用戶的可用旅行時間，進而比值超出 1.0。

表 4.13 用戶 f5320f 原始信令資料

用戶編號	時間	網格編號
f5320f	2018-05-19 09:01:00	254238039
f5320f	2018-05-19 09:02:00	254238039
...
f5320f	2018-05-19 16:58:00	254722059
f5320f	2018-05-19 16:59:00	254234086

表 4.14 用戶 f5320f 興趣點連續時間表

用戶編號	網格編號	起始時間	終止時間	連續時間(分鐘)
f5320f	28	2018-05-19 11:21:00	2018-05-19 11:23:00	2
f5320f	30	2018-05-19 14:08:00	2018-05-19 16:24:00	136

4.3 敏感度分析

4.3.1 興趣點連續出現時間門檻值之分析

於 4.2.1 小節中，介紹了興趣點連續出現時間之門檻值，並說明本研究將以 1 分鐘為門檻值，判斷用戶是否於興趣點停留。在本研究中，觀光用戶的判斷邏輯是以用戶是否曾經停留於興趣點為基礎，因此興趣點連續出現時間的門檻值是一個重要的參數。有鑑於此，本節透過針對興趣點連續出現的門檻值進行敏感度分析，以瞭解在不同時間長度門檻值的情境下，利用信令資料所判讀出之觀光用戶人數。

本研究以興趣點連續出現時間門檻值 1 分鐘、3 分鐘與 5 分鐘進行觀光用戶之判定，結果如表 4.15 所示。由表 4.4 之興趣點連續出現之時間長度分布可知，隨著興趣點連續出現的時間門檻值提升，偵測到的興趣點停留記錄數會減少。當興趣點停留記錄減少時，可以預期判讀出之觀光用戶人數將會下降。從表 4.15 中，可以發現三周之判讀結果均呈現如預期之趨勢，觀光用戶數隨著門檻值的上升而下降，5 分鐘之觀光用戶數比例與 1 分鐘相差 10% 左右，但三周均仍有超過 40% 的觀光用戶數。

表 4.15 興趣點連續出現時間門檻值敏感度分析結果

	1 分鐘	3 分鐘	5 分鐘
第一周	6,860(55.4%)	5,771(46.6%)	5,303(42.8%)
第二周	8,913(53.7%)	6,175(45.9%)	5,772(42.9%)
第三周	7,200(48.0%)	5,284(42.3%)	4,945(39.6%)

4.3.2 可用旅行時間之敏感度分析

本研究在計算可用旅行時間使用比例時，其計算公式是以用戶可用旅行時間為分母，而本研究假設上午九時至下午五時為用戶從事旅遊活動的時段，最長為八個小時。然而，每位用戶的旅遊型態並不一定會與本研究之假設相同，可能會有遊客遊玩至晚間，故針對用戶可用之旅行時間，本研究將另以上午九時至晚上九時為旅遊時段，進行觀光用戶人數的判斷。

表 4.16 可用旅行時間敏感度分析結果

	1 分鐘	3 分鐘	5 分鐘
第一周	7,170(49.4%)	5,748(39.6%)	5,210(35.9%)
第二周	7,973(49.5%)	6,541(40.6%)	6,041(37.5%)
第三周	6,259(43.0%)	5,081(34.9%)	4,821(33.1%)

表 4.16 為假設旅遊活動時段為上午九時至晚上九時，根據信令資料所判讀出之旅遊人數。在與表 4.15 對照比較後，可以發現延長旅遊時段至晚上九時後，判讀出之觀光用戶數有減少的現象。雖然信令資料中可能存在遊玩至晚上的遊客，但是由於分母(可用旅遊時間長度)會隨著設定之旅遊活動時段延長而增加，反而使得可用旅遊時間使用比例下降，最終導致從信令資料中判讀出之觀光用戶減少。

4.3.3 可用旅行時間使用比例門檻值之敏感度分析

在完成每位用戶之可用旅行時間使用比例的計算後，本研究以 0.5 為門檻值，將可用旅行時間使用比例在 0.5 以上之用戶判定為觀光用戶。除了興趣點連續出現時間的門檻值以及可用旅行時間之長度外，可用旅行時間門檻值之設定亦會影響觀光用戶判定的結果。

表 4.17 可用旅行時間使用比例敏感度分析結果

門檻值	0.3	0.4	0.5	0.6	0.7
第一周	8,476(68%)	7,639(62%)	6,860(55%)	6,056(49%)	5,208(42%)
第二周	11,314(68%)	10,076(61%)	8,913(64%)	7,698(46%)	6,562(40%)
第三周	9543(64%)	8306(55.4%)	7200(48%)	6090(41%)	5068(34%)

表 4.17 為在興趣點連續現時間為 1 分鐘的情境下，設定不同可用旅行時間使用比例門檻值所判讀出之觀光用戶人數。由表 4.17 可以發現，隨著門檻值的提升，判讀出之觀光用戶人數隨之下降，且均維持著一定的下降幅度，顯示門檻值的選定並不會對於判讀結果有非常大的影響。而由於研究場域為花蓮縣，為台灣地區之觀光重鎮，因此本研究最後選擇了較高之門檻值 0.5，作為觀光用戶之判定依據。

4.4 以實際信令資料計算興趣點停留時間之測試

先前，於第三章說明的觀光用戶判定方法中，曾經提及由於信令資料自身的特性，故本研究將以旅遊應用程式建議之興趣點停留時間作為用戶於興趣點停留的時間長度。在此小節，將嘗試透過信令資料計算用戶停留於興趣點的時間長度，並根據計算之結果，判定用戶是否屬於觀光用戶。

而興趣點停留時間長度的計算方式，為透過尋找用戶當天第一筆出現在興趣點時間與最後一筆出現在興趣點的時間，並計算兩個時間點的時間間隔，作為用戶停留於興趣點的時間長度。以 3.3.2 小節中之用戶 f5320f 為例(圖 3-8)，用戶當天於上午 11 時 43 分首次出現於花蓮遠雄海洋公園，並最後於下午 4 時 24 分出現於花蓮遠雄海洋公園。若依照本節第一段所提出之興趣點停留時間計算方式，可以得到該用戶當天於花蓮遠雄海洋公園的停留時間為 4 小時 41 分。

表 4.18 以實際信令資料計算興趣點停留時間觀光用戶判讀結果

	1 分鐘	3 分鐘	5 分鐘
第一周	5,938(47.9%)	5,648(45.6%)	5,538(44.7%)
第二周	6,592(49.0%)	6,306(46.9%)	6,214(46.2%)
第三周	5,521(44.2%)	5,309(42.5%)	5,220(41.8%)

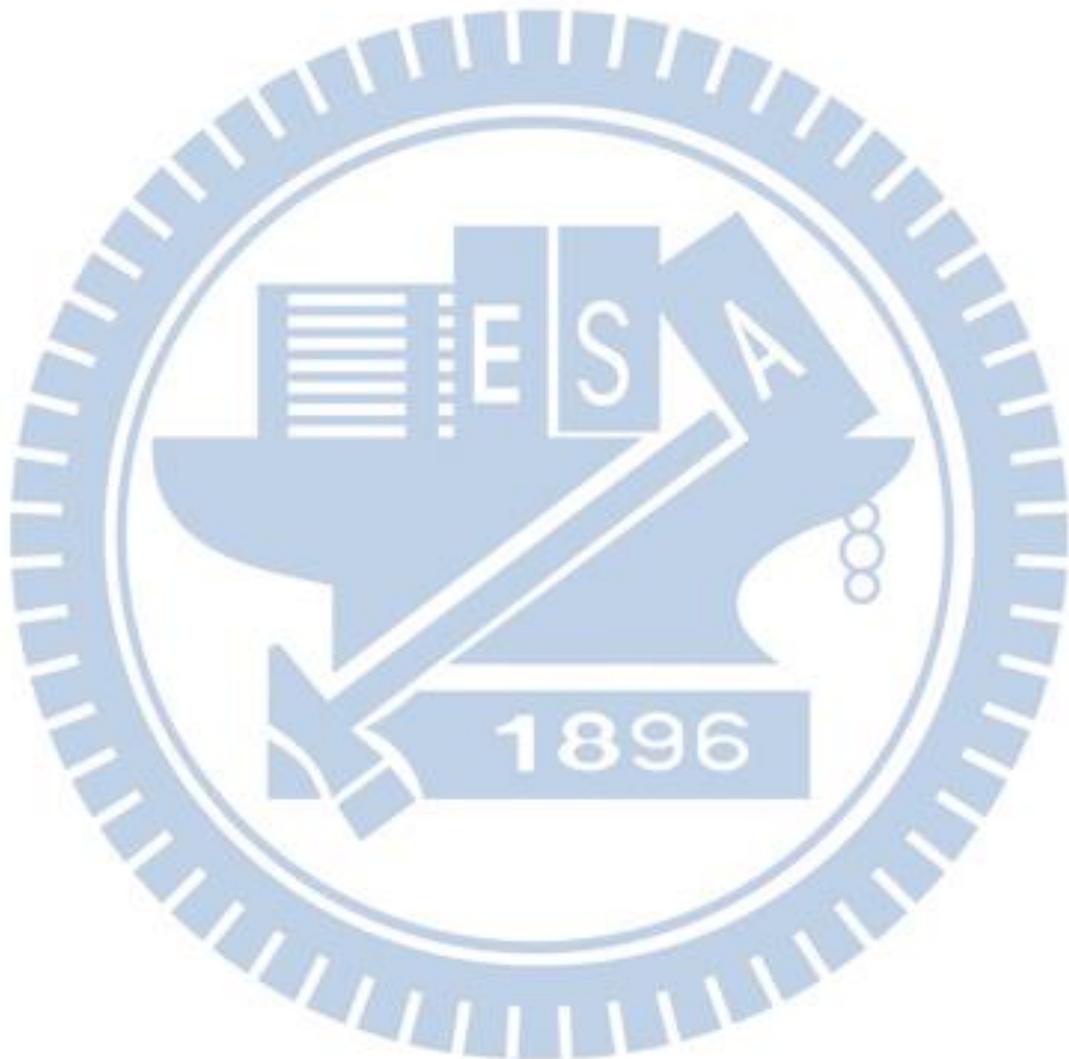
表 4.18 為改以實際信令資料進行興趣點停留時間計算後，觀光用戶人數之判讀結果。由表 4.18 可以發現，以此種興趣點停留時間之計算方式所得到的觀光用戶數均落在總用戶數的 40%至 50%之間，與使用旅遊應用程式之建議停留時間之結果約相差 5%至 10%。雖然此方法是以用戶實際的移動情形計算用戶停留於興趣點的時間，但仍會受因為信令資料自身特性的限制，例如飄移或者是手機持有者的使用行為，而使產生的信令資料受到影響。

4.5 小結

本研究之研究目的為建立一套判斷邏輯，使得觀光遊客能夠從信令資料中被判讀，協助後續觀光旅運需求分析。本研究在取得信令資料後，將於第三章中提出之觀光用戶判斷邏輯以 R 語言進行實踐並應用於信令資料，試圖從中找出觀光用戶。在觀光用戶之判讀成果部分，本研究共自三周之信令資料分別找出 6,860、8,913 以及 7,200 位觀光用戶。

在研究方法設計的過程中，使用了一些本研究自行假設設定的參數。故除了針對信令資料進行研究方法的實測之外，同時也針對可能影響觀光用戶判定結果的因子進行敏感度分析。在興趣點連續出現時間的門檻值以及可用旅行時間使用比例門檻值的部分，觀光用戶數均隨著門檻值的上升而有下降的情形；至於可用旅行時間，若將旅遊時段設定在上午九時至下午九時，縱使將晚間旅遊的情形納入考量，但最終仍因可用旅行時間的增加，而使得觀光用戶數減少。此外，本研究亦嘗試以信令資料計算用戶停留於興趣點的時間長度，進行

觀光用戶之判斷，得到的結果約與原先之實測成果相差 5%至 10%。但由於信令資料會受到自身特性的限制，故本研究最終仍未使用實際信令資料計算用戶於興趣點停留的時間。



第五章 結論與建議

本研究之研究目的，為透過設計一套判斷邏輯，以從信令資料中找出觀光用戶，協助觀光旅運需求分析。於先前的章節中，已詳細說明了判斷觀光用戶的方式，以及觀光用戶的判定成果。本章將針對本研究之研究成果，提出結論以及未來研究方向之建議。

5.1 結論

在研擬觀光公共運輸時，營運路線是否能夠滿足觀光遊客的需求，是影響民眾使用意願的重要因素之一。然而，為了能夠設計出貼近遊客需求的路線，首先需先從瞭解遊客的運輸需求開始著手。本研究之研究目的，即是希望利用手機信令資料，從眾多手機用戶中找出觀光用戶，協助後續觀光需求之分析。

本研究以手機信令資料為對象，透過設計觀光用戶的判斷邏輯，試圖從信令資料中找出觀光用戶。本研究共取得三周之手機信令資料，分別為5月19日、5月20日、5月26日、5月27日、6月2日與6月3日。在參數設定部分，本研究將興趣點容忍值設定為200公尺，若用戶出現於非興趣點之網格但與最近之興趣點距離在200公尺內，視同用戶出現於興趣點。至於興趣點連續出現時間，本研究將門檻值設定為1分鐘，以連續於興趣點出現1分鐘以上視為於興趣點停留。在以上兩個參數的設定條件下，進行觀光用戶之判定。

關於觀光用戶之判定結果，本研究於三周中均約判讀出約半數之觀光用戶，六天的觀光用戶數依序分別為6,860位、8,913位以及7,200位。由於手機信令資料能夠詳細瞭解用戶的移動軌跡，故若針對判讀出的觀光用戶進行移動行為的分析，就能夠瞭解觀光用戶的運輸需求，設計出更貼近其需求的大眾運輸路線。

5.2 未來研究方向之建議

關於未來研究方向之建議，主要可以分成以下兩部分：

1. 研究方法

(1) 資料清洗與插補

信令資料雖然能夠較完整瞭解用戶的移動狀況，但是信令資料存在飄移的問題，即使事先進行清洗以及插補，仍難以完全避免飄移對於資料分析造成的影響。若未來能夠發展出更精確的清洗插補方法，相信能夠更精準地分析信令資料，獲得更佳之研究成果。

(2) 興趣點連續時間

於本研究中，為了保有足夠的資料數量，以1分鐘為門檻值進行興趣點停留與否的判定，若用戶於興趣點連續停留1分鐘以上，則判定用戶曾於該興趣點停留。關於1分鐘之門檻值設定，可能會因為過

短而造成將行經興趣點的用戶誤判為於該興趣點停留，最終誤判成為觀光用戶。建議未來可以針對用戶的移動狀態進行更深入之研究，在先行判定用戶為移動中或者是停留後，方能更準確偵測用戶的停留興趣點。

(3) 興趣點建議停留時間

本研究在計算用戶可用旅行時間使用比例時，以用戶停留於興趣點的時間作為用戶花費於旅遊的一部分。由於信令資料飄移的問題難以完全避免，因此本研究以旅遊應用程式所建議的停留時間作為用戶停留於興趣的時間，而非信令資料實際測得的停留時間。然而，旅遊應用程式所建議的停留時間僅是一個參考值，並非所有的用戶均會分配相同的時間於同一興趣點。因此，若未來能有完整解決訊號飄移的方式，用戶於興趣點的停留時間可採用透過信令資料所計算得到的停留時間，以更接近用戶真實的活動情形。

(4) 天候因素

天氣對於旅遊亦有著莫大的影響，天氣的好壞除了會影響民眾遊玩的時間長度之外，甚至會直接影響用戶是否前往遊玩。若當天天候不理想，例如出現短時間強降雨或者是適逢梅雨季節，都可能導致用戶減少於興趣點停留的時間長度，或者是直接取消前往遊玩的行程。因此，在未來的研究中，建議可以將信令資料結合當天的氣象資料，額外將天氣狀況納入考量。

2. 研究成果

本研究以信令資料為主要研究對象，結合興趣點的位置資訊，並透過本研究所設計的觀光用戶判斷邏輯，從眾多手機用戶中找到觀光用戶。在得到觀光用戶後，可以進一步分析這些用戶的信令資料，進行資料探勘，從中發現觀光用戶的運輸需求。

相關的資料探勘方法，例如針對觀光用戶所拜訪的興趣點進行關聯分析或者是循序樣式探勘。關聯分析可以提供的興趣點之間的關連性，；循序樣式探勘，可以從資料中找到出現頻率較高的樣態，瞭解景點間是否存在特定的拜訪順序。在應用資料探勘方法後，其分析結果中能夠呈現景點間的需求，政府部門亦可據以檢視現有路線是否有不足之處並調整進行路線調整，或者是設計新的營運路線以吸引更多遊客使用。在提升使用率的同時，也能夠降低私人運具對景點周遭所帶來的交通衝擊。

參考文獻

1. 交通部觀光局，「中華民國 106 年國人旅遊狀況調查」，107 年 7 月
2. 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」，研究報告，106 年 12 月。
3. 交通部運輸研究所，「應用旅次特性大數據精進公共運輸服務計畫」，研究報告，107 年 12 月。
4. F. Calabrese, G.D. Lorenzo, L. Liu and C. Ratti, "Estimating origin-destination flows using mobile phone location data", *IEEE Pervasive Computing*, Vol. 10, No. 4, pp 36-44, December 2011.
5. S. Jiang, J. Ferreira, M. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore", *Proc. UrbComp*, pp 1-9, Aug. 2015.
6. L. Alexander, S. Jiang, M. Murga & M. C. Gonzalez, "Origin-destination trips by purpose and time of day inferred from mobile phone data", *Transport. Res. C Emerg. Technol.* 58, Part B, pp 240-250, 2015.
7. Peter Widhalm, Yingxiang Yang, Michael Ulm, Shounak Athavale, Marta C. Gonzalez, "Discovering urban activity patterns in cell phone data", *Transportation* July 2015, Volume 42, Issue 4, pp 597-623.
8. Marco Mamei & Massimo Colonna (2018), "Analysis of tourist classification from cellular network data", *Journal of Location Based Services*, 12:1, 19-39.
9. Qingqing Chen, Zheng Hu, Hang Su, Xiaosheng Tang, Ke Yu, "Understanding travel patterns of tourists from mobile phone data: A case study in Hainan", 2018 IEEE International Conference on Big Data and Smart Computing.
10. Y. Leng, A. Noriega, A.S. Pentland, I. Winder, N. Lutz and L. Alonso: "Analysis of Tourism Dynamics and Special Events through Mobile Phone Metadata". 2016. *Proceedings of Data for Good Exchange (D4GX) 2016*. New York, NY.
11. Seyedmostafa Zahedi and Yousef Shafahi, "Estimating activity patterns using spatio-temporal data of cellphone networks", *INTERNATIONAL JOURNAL OF URBAN SCIENCES* 2018, VOL. 22, NO. 2, pp 162-179
12. S. Bekhor, Y. Cohen, C. Solomon, "Evaluating Long-Distance Travel Patterns in Israel by Tracking Cellular Phone Positions," *Journal of Advanced Transportation*, Vol.47, pp 435-446, Jun. 2013.
13. R. Ahas, A. Aasa, Y. Yuan, M. Raubal, Z. Smoreda, Y. Liu, C. Ziemlicki, M. Tiru & M. Zook (2015) Everyday space-time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn, *International Journal of Geographical Information Science*, 29:11, 2017-2039, DOI: 10.1080/13658816.2015.1063151

14. Wolfgang LOIBL and Jan PETERS-ANDERS, “Mobile Phone Data as Source to Discover SpatialActivity and Motion Patterns”, GI_Forum 2012:
Geovisualization, Society and Learning, Publisher: Wichmann Verlag, pp 524-533
15. Luke Bermingham and Ickjai Lee, “Spatio-temporal Sequential Pattern Mining for Tourism Sciences”, ICCS 2014. 14th International Conference on Computational Science, Volume 29, 2014, pp 379-389

