

二、研究計畫內容（以 10 頁為限）：

（一）摘要

近年來智慧型裝置隨著科技的進步，各式各樣的行動裝置也有越來越多的功能也更加便利。智慧型裝置的普及化，讓行動裝置在大眾的生活作息中更加形影不離。而智慧型裝置就另一層面而言，其不僅僅只是一隨身電腦，事實上其為一個行動裝置持有者之個人資料蒐集器，能夠貼身且密集地收集到大量的使用者資訊，我們認為能夠藉由這些資料來分析使用者、瞭解使用者行為，並以其為基礎設計出更智慧之服務。

在過去范耀中教授實驗室所執行的計畫中，累積收集到許多使用者的 Wi-Fi log 資料，而這些資料，伴隨著裝置持有者的使用所產生，隱含著許多有趣且具有高應用價值的使用者資訊。我們可以藉由這些 Wi-Fi 資訊，分析出使用者可能的偏好、年齡或是職業等等。具體來說，由於 Wi-Fi log 資料帶有語意資訊(Semantics)與時間序列資訊。例如中興大學之學生將會頻繁觀察到 nchu-wifi 這樣的 Wi-Fi log。又例如早上九點會固定觀察到在學校實驗室所安裝之 Wi-Fi 存取點之規律行為。因此，本研究預計以范耀中教授實驗室的所累積之使用者 Wi-Fi 資訊為基礎出發點，來進行本研究計畫申請人之研究，希望利用使用者 Wi-Fi 資訊、時間特徵，以及 Wi-Fi 存取點的地理位置特徵等等觀測資料，探索分析出使用者身分、興趣、職業等資訊。並具體以探勘 Wi-Fi 資料分析使用者之個人資料 (Inferring Smartphone User Demographics from Wi-Fi trace Logs)為目標，嘗試應用或發展資料探勘演算法來取得行動裝置持有者之個人資料，並進行實驗之設計來驗證所發展之技術。

（二）研究動機與研究問題

隨著智慧型裝置的科技越趨發達，行動裝置與大眾的生活作息也更加密不可分，而這些智慧型裝置的功能也越來越多樣化。本研究在計畫執行的過程中，收集了大量的行動裝置使用資料以及觀察過去范耀中教授實驗室所收集的 Wi-Fi 資訊，我們發現了許多有趣且值得深入研究的使用者行為。而這些寶貴的巨量資料，皆由使用者之行動裝置所蒐集，這使我們對智慧型裝置有另一個不同的觀點：我們可以把智慧型裝置視為一個使用者之資料收集器。利用這些資料收集器所得到的觀測資料，進一步分析探索出具有商業價值的使用者行為之側寫。

在過去范耀中教授實驗室所執行的 Wi-Fi 計畫研究中，開發建置伺服器與撰寫 App 應用程式並邀請中興大學 90 位使用者安裝 App 應用程式。在每一部行動裝置安裝此應用程式時取得載具上唯一辨識碼，並以每 30 秒唯一區間取得該行動裝置所觀測到之所有 Wi-Fi 存取點名稱(SSID)與 Wi-Fi 存取點之唯一辨別碼(BSSID)，上傳至所建置的伺服器，讓每一個使用者成為資料提供者。針對每一個 I-MEI 碼，隨著時間記錄該行動裝置所觀測到之 Wi-Fi 訊號軌跡值。如下圖一為伺服器所收集到之計資料例。如下圖一所示，每一行 Wi-Fi 乃為使用者之智慧型裝置所在偵測到 Wi-Fi 存取點之資訊，包含存取點名稱(SSID)、BSSID 辨識碼、存取點觀測訊號之強度(Level)以及時間(Timestamp)。

約四個月的時間中，累積 90 位使用者，產生約一百六十萬筆 Wi-Fi 資料，目前並持續累積中。這樣的大量資料累積後，我們偶然間發現其中隱含著許多有用的資訊

可以幫助我們瞭解行動裝置的使用者之行為與相關資訊，以下將簡述我們所觀察的現象。

IMEI:77132194f7a8fdc0			
SSID	BSSID	Level	TIME
Fami-WiFi	c8:be:19:72:11:92	-51	2016-02-01 22:13:47
Fami-WiFi	c8:be:19:72:11:92	-82	2016-02-01 22:13:48
Fami-WiFi	c8:be:19:72:11:92	-83	2016-02-01 22:13:49
Fami-WiFi	c8:be:19:72:11:92	-50	2016-02-01 22:13:54
hikaru	f0:79:59:d1:9b:40	-83	2016-02-01 22:13:54
NCHU-WiFi	bc:4d:fb:3c:b5:98	-50	2016-02-01 22:13:56
hikaru	f0:79:59:d1:9b:40	-83	2016-02-01 22:13:56
NCHU-WiFi	bc:4d:fb:3c:b5:98	-50	2016-02-01 22:13:59
hikaru	f0:79:59:d1:9b:40	-83	2016-02-01 22:13:59
Fami-WiFi	c8:be:19:72:11:92	-49	2016-02-01 22:14:01
Fami-WiFi	c8:be:19:72:11:92	-83	2016-02-01 22:14:02
GNS	c4:6e:1f:2f:37:c6	-79	2016-02-01 22:14:03
Fami-WiFi	c8:be:19:72:11:92	-87	2016-02-01 22:14:04
Fami-WiFi	c8:be:19:72:11:92	-49	2016-02-01 22:14:05
hikaru	f0:79:59:d1:9b:40	-83	2016-02-01 22:14:05
GNS	c4:6e:1f:2f:37:c6	-79	2016-02-01 22:14:06
RT-N12HP	e0:3f:49:92:6b:fc	-87	2016-02-01 22:14:06
NCHU-WiFi	bc:4d:fb:3c:b5:98	-49	2016-02-01 22:14:08

圖一：Wi-Fi存取點觀測資料

使用者 Wi-Fi 資料觀察分析 首先，大部分 Wi-Fi 存取點 SSID 名稱有其意義，並非隨意亂取，其命名多與該 Wi-Fi 存取點所裝置之地點有所關聯，例如全家便利商店內所佈建的 Wi-Fi 存取點，即取名為 Fami-WIFI，再例國立中興大學校園中所佈建 Wi-Fi 存取點，即取名為 NCHU-WiFi。由於研究計畫中所開發的 App 應用程式，乃是藉由智慧型裝置掃描附近可偵測到的 Wi-Fi 存取點之資訊，並上傳至伺服器。而每一筆 Wi-Fi 觀測值所對應即是使用者之智慧型裝置所能偵測到之 Wi-Fi 存取點。我們觀察出每一筆 Wi-Fi 觀測資料之產生是來自於使用者曾經拜訪過此 Wi-Fi 存取點之相關地理位置。

在每一個行動裝置的資料集中，我們發現同一地點之 Wi-Fi 觀測值，即 SSID 所對應之存取點名稱，皆為相同時，其原因來自於使用者長時間停留或者是經常拜訪此地點。由於資料收集的應用程式每三十秒將會取樣一次行動裝置附近所有可觀測到之 Wi-Fi 存取點訊號，因此若該行動裝置持有者停留於某一 Wi-Fi 地點約三百秒，則我們將獲得十筆相同的 Wi-Fi 訊號觀測值。

我們觀察使用者長時間停留於同一 Wi-Fi 存取點之地理位置，並加以分析其地理位置之特徵，進而可以探索出使用者之可能相關之偏好或是身份。例如，我們所收到的 90 位資料提供者中，有使用者的智慧型裝置在每天中午 12 點至 1 點之間收集到大量的 Fami-WIFI(便利超商)存取點之觀測值，或是在早上 9 點會收集到資料庫實驗室 Wi-Fi 存取點之資訊等等頻繁規則。我們可以進一步的推論，此使用者可能在中午時段喜愛在便利超商購買午餐，而且此使用者可能是理工學院的研究生。本研究最基本的想法為我們認為頻繁觀察到之 SSID 相對隱含著使用者之身分或職業類別之資訊。

使用者 Wi-Fi 資料與時間特徵之分析 如上資料觀察分析所述，我們是針對 Wi-Fi 資料集中高頻率 SSID 之存取點資訊來分析使用者。若我們能以時間觀點的不同來探

索 Wi-Fi 資料，就能觀察出不一樣的使用者行為之側寫。具體來說，我們可以用巨觀與微觀這兩種不同的時間觀點來分析 Wi-Fi 資料。首先，以巨觀時間來說，我們觀察一 I-MEI 碼之全域頻繁 Wi-Fi 資料集，即一位使用者之所有時間點 Wi-Fi 觀測資料，發現如果同一地理位置之 Wi-Fi 存取點名稱高頻率出現者，代表使用者經常拜訪或是停留該地點，此存取點之相關場所可能與使用者的身分職位有關。而以微觀時間來說，我們以區域頻繁為基準來分析其 Wi-Fi 資料集，則能夠分析出使用者之興趣與偏好等等資訊，如上資料觀察分析我們在中午時段收到大量的便利商店 Wi-Fi 存取點名稱，便能夠進一步分析使用者午餐選擇之偏好。

上述的觀察看似有道理並且可行，然而在我們著手整理資料後發現在我們所收集的資料中，雖然有些資料極具價值，然而我們發現多數的資料皆為雜訊或者具意義但不完整之資料。簡言之，個人認為本研究計畫主要的挑戰將為如何從充滿大量雜訊的資料中，過濾出我們所需要的資訊，若無法有效適當地處理這些雜訊，真正有用的訊息將淹沒於大量的雜訊之中。綜合上述研究動機，本研究計畫擬探討三個研究問題：首先，若是以一個長期停留或多次拜訪之地點應對行動裝置持有者有其重要性為基礎假設，則假設使用者短暫且相對低頻率拜訪重要之地點，是否會過濾部分重要之 Wi-Fi 存取點字集？再者，若是使用者所探測之 Wi-Fi 觀測值甚少或是有其他因素造成資料不齊全，是否能夠準確判斷使用者之相關偏好？最後，在智慧型裝置所收集到的 Wi-Fi 資料及當中，各筆 Wi-Fi 之存取點名稱是否為具有訊息(Informative)之資料，若是為非有用資料該如何剔除過濾？

(三) 文獻回顧與探討

本研究所預計探討之內容，主要是以大量使用者行動裝置累積之 Wi-Fi 資訊所衍生之應用與利用智慧型裝置所收集之觀測資料以推測使用者的偏好興趣。以我們目前所知，尚未有相關研究涉及其核心之關鍵技術，而據我們所知，目前利用智慧型裝置來探索使用者行為的有兩個主要的計畫研究：分別為微軟亞洲研究院的 GeoLifeProject[1][2][3][4][5][6][7][8][9][10][11][12][13] 與麻省理工學院 MIT Media Lab 所從事之 RealityMiningProject[15][16][17][18][19][20]。以下我們分別介紹此兩專題研究計畫之進展與其優缺點以及本研究計畫所擬開發技術之優劣比較。

在 GeoLife Project 中，其研究之主軸是建立使用者行動裝置內的 GPS(Global Positioning System)軌跡資料，並大量收集資料後[9][11]，透過使用者交友系統[2][8][10]，例如使用者經常拜訪相同之地點[8]或是經常前往地點屬性相同或者一個常常於區域地點移動的使用者(Local Expert)，所前往的商店，對使用者進行商店推薦[3]、拜訪的餐廳，對使用者進行餐廳推薦[3]、景點與活動推薦系統[4]、使用者移動模式分析系統[6][13]等等技術便基於使用者之智慧型行動裝置所取得的 GPS 軌跡資料。而由於軌跡資料量過於龐大，相關研究也於[1][14]中探索如何快速取出並分析使用者的軌跡資料。

在另一 MIT Media Lab Reality Mining Project 中，也是希望能從比較貼近使用者的智慧型行動裝置所產生的資料進行著手[16][18]，然而相關的研究發展主題則較為分散，部份之研究[16][17]也遭逢使用 GPS 定位系統所帶來的問題。其中於研究[16]中，

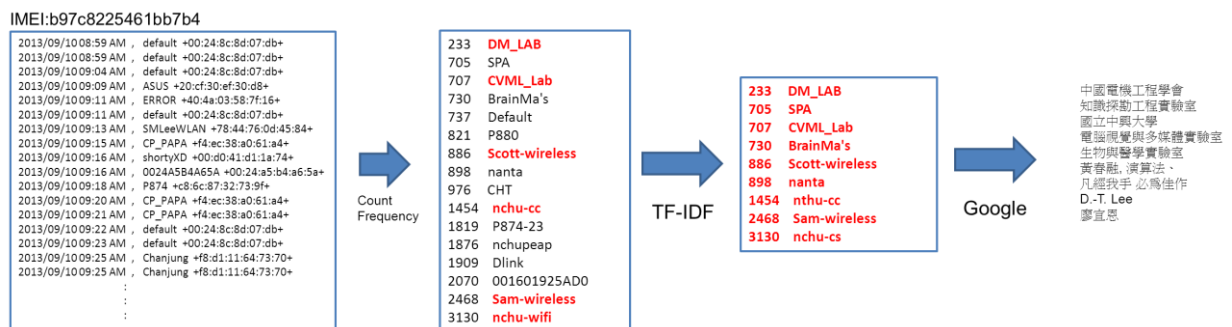
作者們也想利用 GSM 定位系統來推測使用者可能的身份，然而相較於我們除了推測使用者之身分外，能進一步地推測出使用者之可能喜好與社群關係等等訊息。而於研究[15][18]中，作者們則提出利用藍芽裝置與部份行動通訊的資料來推測可能的使用者彼此關係，使用者的藍芽裝置若常常彼此觀測到，則我們可推論使用者是朋友，然而此一做法之前提為使用者必須無實無刻開啟藍芽裝置，有實際推廣上之考量。而於[17]中，作者們則探討利用使用者所在情境中來進行行動裝置電量之節省，此一研究也可為本研究計畫之動機，我們也可協助偵測使用者行動裝置現所在之情境資訊。

上述的所有研究結果皆奠基於透過使用者行動裝置之 GPS 軌跡資料取得。然而我們發現使用 GPS 軌跡資訊將有下列三大缺點，使得所有基於 GPS 軌跡資料分析之系統於實際運作上之不可行。首先，GPS 感測器之於行動裝置為一非常耗電之裝置，使用者不可能長期地將 GPS 感測器開啟。第二，使用 GPS 軌跡資料將有室內定位之問題，若我們無法有效地定位到使用者所在地，我們將無從分析起使用者之軌跡。Google Wi-Fi 定位技術也許可能可提供定位部份的室內環境，然而對於建築物內之詳細地點資訊，Google Wi-Fi 定位將無法提供協助。第三，我們主要想法為利用使用者所在地點來瞭解使用者，所以能夠知道使用者所處地點之類型或者屬性將是最關鍵一步驟，然而 GPS 定位系統，我們必須使用 Google Place API 地點轉譯技術，而 Google Place API 有其存取限制並且僅能告知我們給定範圍內的景點類型，並無法精準提供使用者所在地點之精確資訊。相較於本研究計畫所提出的想法，我們主要利用 Wi-Fi 存取點之 SSID 名稱來推測可能使用者可能的所在地，並無使用 GPS 系統所帶來的缺點。

(四) 研究方法及步驟

根據研究動機所述，本研究乃是依據智慧型行動裝置所蒐集之 Wi-Fi 存取點資訊來進一步探討及分析使用者。我們所擬提出之方法將基於下列之四項觀察：

首先，部分 Wi-Fi 存取點名稱有其意義，並非隨意亂取，其取名多與該 Wi-Fi 存取點所裝置之地點有所關聯。第二、Wi-Fi 觀測之資料之產生來自與使用者曾經拜訪過安裝該 Wi-Fi 存取點之裝置地點。第三、同一 Wi-Fi 地點擁有多筆觀測值，其因來自於使用者多次拜訪該地點或長期停留於該地點。第四、一個長期停留或多次拜訪之地點應對行動裝置持有者有其重要性。綜合以上四點，我們發現適當分析此每一具智慧型行動裝置所提供之 Wi-Fi 觀測資料，將有助於瞭解對應之行動裝置持有人。

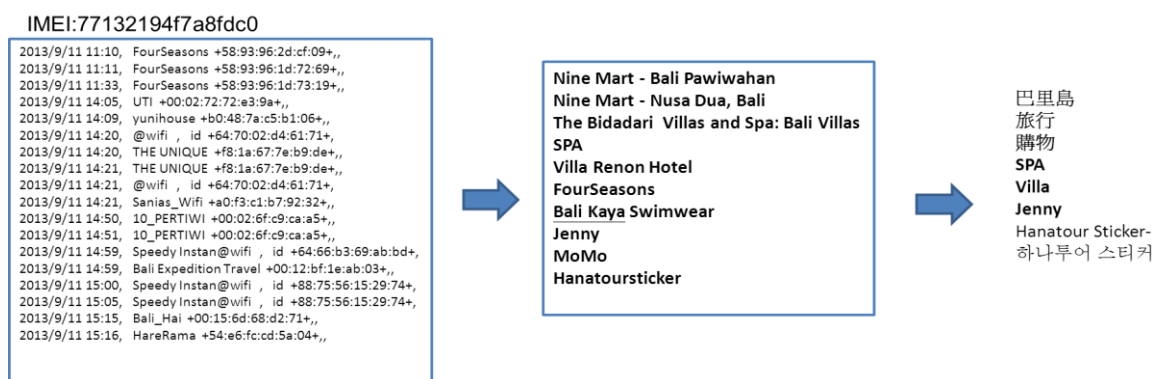


圖二：以 Wi-Fi 觀測資料為基礎進行使用者側寫

我們的基本分析想法如下，首先一個地點停留越多次該地點應對行動裝置持有者擁

有較高之重要性。因此我們可先針對每份 IMEI 中，統計所有 Wi-Fi 存取點名稱所出現的次數並進行排序。經排序後我們發現確有部分 Wi-Fi 存取點名稱擁有極高的出現頻率可能對使用者有意義，但也出現部分出現頻率較高但不具意義之 Wi-Fi 存取點名稱，如 Dlink 與 Default。針對此一因素我們可先利用 TF-IDF(Term Frequency and Inverse Document Frequency)來去除頻率高但常見於不同 IMEI 資料檔中 Wi-Fi 存取點名稱，並對所有觀測到之 Wi-Fi 存取點名稱進行重新排序。藉此，如下圖所示，我們可取得較有意義之 Wi-Fi 存取點名稱。我們進一步地將此整理過後之 Wi-Fi 存取點名稱集合，將其投入 Google 搜尋引擎中，取得所有搜尋結果之關鍵字，可得下列最右子圖之結果。我們可以利用最右子圖之結果推敲而知，該行動裝置之持有者可能為一中興大學資工系之研究生、研究專長為演算法、電腦視覺、資料探勘。

我們可再一 IMEI 所收集部分資料為相同示範例，如下圖所示。我們可從資料中整理出 FourSeasons, Bali, Villas, SPA 等有趣 Wi-Fi 存取點名稱，並將之投入 Google 搜尋引擎中，我們可以進一步衍生更多訊息，如旅行、購物等等關鍵字。由於安裝我們程式者皆為中興大學之學生，因此可推測該行動裝置之持有者可能為一位中興大學大學生之女同學、於受測期間到過巴里島、喜好為購物、經濟能力佳。透過上述之兩真實資料之簡易分析，我們將發現針對行動裝置之持有人之大概輪廓進行一概約認識將為可能。



圖三：以 Wi-Fi 觀測資料為基礎進行使用者側寫

由於 SSID 名稱多為一縮寫，字面上涵意可能不足，然而我們若妥善利用現有開放的搜尋引擎 API，例如我們利用 Google Web Search API，將 SSID 投入搜尋引擎中，我們便可衍生其字義；Google 搜尋引擎將回傳我們若干網頁內容描述我們所投入之 SSID 背後可能代表的內容，例如將 Nchu-cs-WiFi 投入 Google Web Search API 後，如圖四所示，我們將得到中興大學資訊科學與工程學系之搜尋結果，因此透由此舉我們便可擴增一個 SSID 背後所代表的意涵。將 SSID 投入搜尋引擎中帶來了擴增 SSID 意涵的好處，然而搜尋引擎所回傳的資料皆為半結構化(semi-structured)或非結構化(unstructured)的資料，其中包含了更多的文字雜訊，因此如何分析搜尋引擎所回傳的非結構化資料也將成為預期之難題。

```

1 sn_wap -> sn wap 41118
2 DMLAB-10F -> DMLAB 37344
3 NCHU-CSIE-wifi -> NCHU CSIE 36689
4 Condor -> Condor 27978
5 BiVihome -> BiVihome 26914
6 NCHU-WIFI -> NCHU 23974
7 dd-wrt -> dd wrt 23073
8 CVML8F -> CVML8F 22137
9 Jacky Lu -> Jacky Lu 21612
10 yfan -> yfan 19421
11 ATAlab -> ATAlab 18760
12 FINE-ART -> FINE ART 18214
13 love520 -> love 14321
14 Grasshopper Eggs Inc. -> Grasshopper Eggs Inc 14030
15 ArroGaNt -> ArroGaNt 13730
16 yaosn -> yaosn 13298
17 happyhappy -> happyhappy 12305
18 E-SPEED -> E SPEED 8586
19 harryl -> harry 7955
20 Tenda 53F8E8 -> Tenda 53F8E 802
21 LIN -> LIN 7629
22 OSNET -> OSNET 7484
23 HTC Portable Hotspot 71FE -> 71FE 7086
24 LAB OF MICROBIOLOGY -> LAB OF MICROBIOLOGY 6010
25 real66 -> real 5486
26 derek -> derek 4680
27 asua-1 -> asua 4505
28 FM -> FM 4476
29 linjay -> linjay 4216
30 hauman_chao -> hauman chao 3534

```

Google Web
Search API

Google

Google Search I'm Feeling Lucky

```

NCHU CSIE 36689 首頁國立中興大學資訊科學與工程學系
工所 網路使用 疑難排解 http network cs nchu edu tw new 隱私權 政策 本系 位置 國立 中興 大學 理學 大樓 電話
使用 csie cloud03 機器 使用 csie cloud04 機器 vm 注意 ip hostname tap no 是否重複 分配原則 如下 csie20 為例
ssh cloud ssh ctsp tn csie cloud01 replace csie cloud01 by ip check whether you got linux uml and its related module
jun 中興 系統 網路 管理 實驗室 nchu csie smmlab 中興 工資 挖掘 知識工程 實驗室 nchu csie dm amp ke lab oral presentation
tronelish hwang and yi ping luo quot 簽署 量子 件數 量子 簽署 quot cryptography and information security conference may nchu taichung taiwan
中興 大學 http ww nchu edu tw nchu campus farm foodscie food html 搜尋引擎 自然科學 搜尋 http taiwan csie ntu edu tw b5 yam 國外
origin 中興 大學 資訊網 it bbs bbs nchu edu tw qt from prt bbs ee ntu freebed ntu news neto net db csie ncu news ncu news2 cs
發信 交大 鳳凰 資訊 sat dec ntlc news nchu conews nchu ctu gate news nctu spring ee ttit news csie 一般 都是 mallo

```

```

NCHU 23974 國立中興大學(National Chung Hsing University)
立為 農林專門學校 改為 附屬農林 獨立 遷至 台中 改為 國立大學
諮詢 建議 圖書館 關於 概況 歷史 概況 歷史 歷任 館長
說明 使用 對象 目前 可以使用 單一 系統 對象 在職 教職員 在學
招生 資訊 網站 暫停 網路 伺服器 服務 公告 image 年度
國際 學者 講座 samuel weber 學年 碩士
中興 大學 教務處
本處 設立 旨在 積極 推動 本校 邁向 國際化 掌握 國際 脈動 學校 擁有 國際化
首頁 工學院 english 通訊 光電 晶片 中心 光電 半導體

```

圖四：SSID 語意擴增

根據上述 Wi-Fi 資料的基礎分析之後，以下為我們所規劃之研究分析：

〔以全域高頻率為基準之使用者身分分析技術研究〕

在本研究中，我們最初的想法是：每一 I-MEI 碼之全域高頻率 Wi-Fi 資料集中的 SSID 應該對使用者有意義。因為使用者在全域時間資料集中，長時間停留或是拜訪該對應之 Wi-Fi 存取點，若我們能夠瞭解其 SSID 所代表之意涵，便能夠分析出此使用者可能之身分。我們最基礎的出發點為妥善利用現有開放的搜尋引擎 API，將 SSID 投入搜尋引擎，我們便能夠衍生其字義。然而搜尋引擎所回傳之資料仍需要結構化分析，且賦予有效的權重，才能讓每一個 SSID 表示該真正意涵，因此我們擬定了以下的步驟，並將逐一說明其步驟：

步驟一：SSID 權重分配機制

在本研究中，最基本的想法為針對每一 I-MEI 碼之全域 Wi-Fi 資料集中的 Wi-Fi 存取點名稱進行統計，並排序來取得使用者之長時間停留之地點。具體來說，我們以出現頻率最高的為統計分析之準則，權重最高的乃為拜訪最多次或是停留最久之地點，我們認為該地點與使用者工作單位有高相關性。我們也將觀測時間納入權重考量之中，舉例來說，在一使用者全域 Wi-Fi 資料集中的白天工作時間，出現的高頻率 SSID 應偏向於工作地點之存取點，其相對應之權重應當較為夜晚時間所收集的高頻率 SSID 來的高。此研究步驟，將針對其 SSID 的有效程度進行適度的衡量分配，以篩選出對使用者更有意義之資訊。

步驟二：SSID 意涵之分析

我們將步驟一中權重較高的 SSID 資訊投入搜尋引擎當中，並將回傳之相關資料進行適當斷詞分析，並將所有字詞當作一關鍵字集合。我們發現帶有訊息之 SSID 有不同的資訊意義，部分的 SSID 名稱背後代表的訊息可能是一個使用者姓名、是一個單

位名稱,例如圖一中 NCHU-WiFi。因此一個字詞可能代表使用者的姓名或者使用者的工作地點。在本步驟中,我們擬設計一階層式之模型,分門別類地來描述使用者 SSID 所帶來的各種訊息。

步驟三：多重 SSID 訊息分類架構

本研究步驟中,在 SSID 名稱排序後我們將建構不同的分類器來進行資訊之多層分類,我們將層層過濾來辨識 SSID 名稱後所代表之意涵,並分類整理。具體而言,我們的初步想法為建構一連串之二元分類器,來使資訊的分類與過濾將為可能。我們預計實驗現有各式各樣設計良善之分類器方法來建構我們的多層資訊分類架構,使取得使用者之側錄資訊更加完整。

以下為本研究之預期進度表

預期進度	1	2	3	4	5	6	7	8
關鍵技術論文研讀								
研究計畫架構設計								
演算法設計及程式撰寫								
SSID 權重分配機制之研究								
SSID 意涵之分析研究								
多重 SSID 訊息分類架構								
SSID 訊息雜訊過濾技術研究								
計畫研究報告撰寫								

(五) 預期結果

本研究是以智慧型行動裝置所蒐集之 Wi-Fi 存取點來進行使用者之探討,利用 SSID 觀測資料來側寫使用者之身分、職業,或是偏好興趣等等個人資料。本研究之主要依據 SSID 具有可能揭露使用者之職業,或是興趣偏好等等資訊來進行研究。

在全域高頻率分析當中,預期能夠把搜尋引擎所回傳的關鍵字集,進行妥善斷詞分析,將所有的關鍵詞找出與使用者有相關的集合,將該集合視為使用者側錄分析之依據,以完整表達使用者之身分。

在另一方面區域高頻率分析中,我們利用滑動視窗為基礎在特定時間範圍內之高頻率而在整體為低頻率之 SSID 中,尋找出可能之使用者相關喜好資訊,預期能夠架構出階層式使用者偏好資訊,並能夠妥善的整合出使用者多重的興趣喜好,讓使用者的個人資料更加完善。

本研究預計能夠將使用者的個人資料,透過以上研究之分析之後,拼湊出完整之身分、職業,興趣等等個人資料。我們預計持續地對資料提供者進行長時間的關測,讓使用者之 Wi-Fi 觀測資料更加具有可靠性,以增加使用者資訊側錄分析之重要依據。

(六) 參考文獻

[1] Longhao Wang, Yu Zheng, Xing Xie, Wei-Ying Ma. A Flexible Spatio-Temporal Indexing Scheme for Large-Scale GPS Track Retrieval, In Proceedings. of International

conference on Mobile Data Management, MDM 2008.

[2] Quannan Li, Yu Zheng, Yukun Chen, Xing Xie. Mining user similarity based on location history. In Proceedings. of ACM SIGSPATIAL conference on Geographical Information Systems, ACM GIS 2008.

[3] Wenchen Zheng, Bin Cao, Yu Zheng, Xing Xie, Qiang Yang. Collaborative Filtering Meets Mobile Recommendation: A User-centered Approach, In Proceedings. of AAAI conference on Artificial Intelligence (AAAI 2010).

[4] Wencheng Zheng, Yu Zheng, Xing Xie, Qiang Yang. Collaborative Location and Activity Recommendations With GPS History Data. In Proceedings. of International conference on World Wide Web (WWW 2010).

[5] Yu Zheng, Xing Xie. Learning Location Correlation from GPS trajectories. In Proceedings of the International Conference on Mobile Data Management 2010, MDM 2010.

[6] Yu Zheng, Like Liu, Longhao Wang, Xing Xie. Learning Transportation Modes from Raw GPS Data for Geographic Application on the Web, In Proceedings. of International conference on World Wide Web.

[7] Yu Zheng, Lizhu Zhang, Xing Xie, Wei-Ying Ma. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of International conference on World Wide Web, WWW 2009.

[8] Yu Zheng, Lizhu Zhang, Xing Xie. Recommending friends and locations based on individual location history. ACM Transaction on the Web, 2010.

[9] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie. Understanding Mobility Based on GPS Data. In Proceedings of ACM conference on Ubiquitous Computing, UbiComp 2008.

[10] Yu Zheng, Yukun Chen, Xing Xie, Wei-Ying Ma. GoLife2.0: A Location-Based Social Networking Service. In Proceedings of International Conference on Mobile Data Management 2009, MDM 2009.

[11] Yu Zheng, Longhao Wang, Xing Xie, Wei-Ying Ma. GeoLife-Managing and understanding your past life over maps, In Proceedings of International conference on Mobile Data Management, MDM 2008.

[12] Yu Zheng, Xing Xie. Learning travel recommendation from user-generated GPS trajectories. ACM Transaction on Intelligent Systems and Technologies 2012

[13] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie, Wei-Ying Ma. Understanding transportation modes based on GPS data for Web applications. ACM Transaction on the Web. 2010.

[14] Zaiben Chen, Heng Tao Shen, Xiaofang Zhou, Yu Zheng, Xing Xie. Searching Trajectories by Locations – An Efficiency Study, In Proceedings of ACM SIGMOD International Conference on Management of Data, SIGMOD 2010.

[15] N. Eagle, A. Pentland, and D. Lazer, Inferring Social Network Structure using Mobile Phone Data, Proceedings of the National Academy of Sciences (PNAS) Vol 106(36), pp. 15274-15278.

[16] Y. De Mulder et al., Identification via location-profiling in GSM networks, In Proceedings of 7th ACM Conference on Computer and Communications Security, ACS 2008

[17] N. Ravi et al., Context-aware Battery Management for Mobile Phones, In Proceedings of IEEE

(一) 需要指導教授指導內容

(如篇幅不足，另紙繕附)

在研究開始時，需要與指導教授討論研究方法及設計實驗的步驟，並探討在 SSID 權重分析機制當中，如何妥善選取該項對應之演算法以期符合使用者側錄依據之權重排序，並且在斷詞系統的架構內分析過濾出有效之字詞，以及在滑動視窗為基礎之區域高頻 SSID 擷取技術中，如何妥善的定義滑動視窗的大小等等相關關鍵技術。探討過程當中向教授請教論文相關問題，包括研讀及技術方面的指導，最後需要指導教授指引研究最終報告書。

表 C802