**使用機器學習演算法結合**

**蜂巢式網路資訊的移動類型辨識器**

## Transportation Type Identiﬁcation by using Machine

## Learning Algorithms with Cellular Information

林逸豪、曾怡婷、林元偉、何鳳雯、林家弘、林献庭、陳志成

國立交通大學資訊工程學系

Yi-Hao Lin, Yi-Ting Tseng, Yuan-Wei Lin, Fong-Man Ho,

Chia-Hung Lin, Hsien-Ting Lin, Jyh-Cheng Chen

National Chiao Tung University, Hsinchu 300, Taiwan

Email: { yhlin1377, ss920386.cs04, yuanweilin.cs03g, jodyho.cs03,

ljhjimmy87, siamevol41216, jcc}@nctu.edu.tw

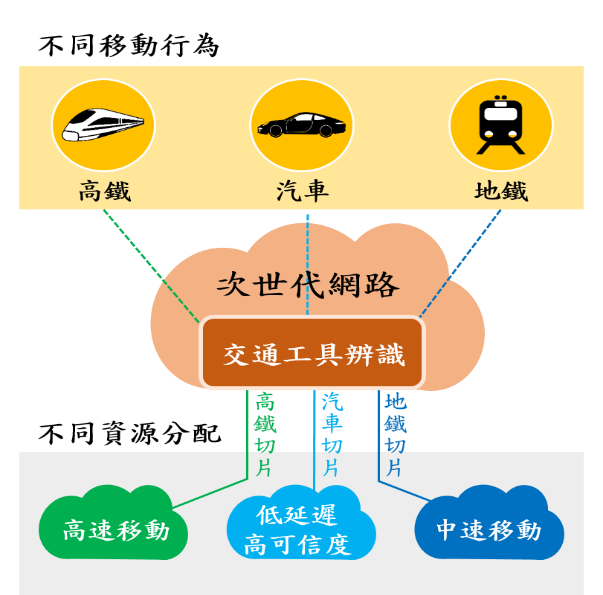


圖 1 次世代網路概覽

**摘 要**

對於第五代 (5G)行動網路而言，了解使用者如何移動以有效地分配網路資源是一個重要的議題。本論文嘗試找出適合的特徵來有效識別四種交通工具，包括地鐵、鐵路、高速鐵路以及汽車 (高速公路)。我們提出了一個只使用蜂巢式網路資訊的系統架構來達成準確且具有即時性和適應性的辨識。由於我們不像大多數的相關其他研究使用GPS，因此包括電力的消耗、紀錄檔案 (log file)的大小和計算的時間都能大幅減少。我們蒐集了大約500小時的資料集來進行表現的評估，實驗結果證明將我們提出的方法應用在知名的機器學習演算法上，能將準確率改善至98%左右，此外，實驗結果也證明我們的方法能減少約8%至16%的耗電量。

**關鍵字：**5G、交通工具識別、機器學習、蜂巢式網路、分類問題。

**一、前言**

網路切片 (network slicing)是5G行動網路裡的其中一項重要特徵。網路切片能讓網路業者在一個共同的網路基礎架構之上，針對不同使用者的不同需求，提供具備不同資源的虛擬網路以滿足這些需求，這些虛擬網路因此能支援5G行動網路預期提供的多種服務。如圖 1所示，被稱作次世代網路 (next generation network)的一個實體網路會被切成多個虛擬網路來支援不同行為模式的使用者，不同的網路資源和特定的移動策略也會透過共同的核心網路被分派給不同類型的使用者。因此，如何自動辨識服務型態、推斷適合的服務提供機制，並且建立需要的網路切片是必要的。

為了要智慧地實現次世代網路，我們必須先探討使用者的服務需求。在本論文中，我們著重在分析使用者的移動行為。我們使用機器學習的演算法來進行移動類型的辨識，且我們主要關注的移動類型為「有引擎」的移動類型，因為無引擎的移動類型 (如行走、慢跑、騎腳踏車等等)其移動速度通常並沒有快到需要網路業者分配額外的特定資源。此外，無引擎的移動類型可以用慣性測量單元 (inertial measurement unit, IMU)的資料來偵測，如智慧型手機裡的加速度計、陀螺儀和旋轉向量感測器[1], [2]等。

在此論文中，我們嘗試擷取出適合的特徵來辨識四種有引擎的交通工具，包括地鐵、鐵路、高速鐵路以及行駛在高速公路上的汽車。我們使用多種不同的機器學習方法來訓練分類器，並將多種知名的機器學習演算法的評估結果呈現在此論文中。這篇論文的主要貢獻包括：

1. 我們提出的系統架構提供移動類型辨識問題一個準確、即時且具適應性的解決方案。此架構只使用了蜂巢式網路資訊，並能被進一步的應用到其他如碳足跡、智慧導航、老人追蹤器、智慧城市和交通分析等應用。
2. 我們從蜂巢式網路資訊擷取位置資訊，並提出蜂巢式資訊結合滑動視窗的特徵擷取方法。此方法並不受限於智慧型手機，而是適用於任何能透過蜂巢式網路連上網的裝置。

**二、問題限制與挑戰**

智慧型手機可以提供包括GPS、加速度計、磁力計、氣壓計、重力感測計、光感測器、蜂窩式/無線信號等量測資訊，這些資訊雖然能協助辨識交通工具，但他們的使用仍存在一些挑戰：

* **環境限制：**由於GPS能夠提供使用者的具體位置資訊，因此能協助辨識交通工具。但是GPS定位需要設備端與衛星保持直視 (line-of-sight)的通訊，所以像是隧道和都市等環境，因為有許多屏障，可能都不適用，或是像在地鐵這種地底環境下，GPS也是沒有辦法定位的。其他感測器，如磁力計、氣壓計或光感測器等，則是容易被環境中如電子產品、濕氣/溫度、陽光等因素影響。

表 1 相關研究概要

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 研究 | 高鐵 | 鐵路 | 地鐵 | 汽車/巴士 | 腳踏車 | 走路 | 跑步 | 靜止 |
| Yu Zheng [11] |  | O |  | O | O | O |  | O |
| Gong. H. [25] |  | O |  | O | O | O |  | O |
| Leon Stenneth [4] |  |  |  | O | O | O | O | O |
| Wenbo Zhu [26] |  |  | O | O | O | O | O |  |
| Threresa Nick [6] |  |  | O | O | O | O |  | O |
| Fang [27] |  |  | O | O |  | O |  |  |
| Arash Jahangiri [7] |  | O | O | O |  | O |  | O |
| Su [2] | O |  |  | O | O | O | O | O |
| Our work | O | O | O | O | O | O | O | O |

**改善方法：**現今不論是在地面上還是地面下，交通路線附近都普遍的部署了基地台，因此我們將如章節五-(二)所述使用蜂巢式網路資訊取代GPS進行定位。

* **耗電問題：**使用智慧型手機進行長時間量測時，計算的資源會是主要的限制。取樣頻率越高，耗電的速度就會越快，若使用GPS更會消耗相當可觀的電量[3]。

**改善方法：**我們使用蜂巢式網路資訊取代基地台，搭配我們在章節五-(三)所提出的一個新演算法，只要有基地台資訊和換手 (handover)頻率，就能根據*事件*進行採樣，可以降低電力的消耗、紀錄檔案的大小和計算時間。

* **隱私問題：**從手機蒐集越多種感測資料，使用者的所在位置和活動狀態就越有可能被偵測及辨識。

**改善方法：**我們透過蜂巢式網路資訊得到的位置精確度比GPS粗略，並且我們的方法不需要使用額外的感測資料，可以大幅降低使用者對於隱私權的憂慮。

**三、背景介紹**

交通工具辨識的問題屬於行為辨識議題的一環，傳統上會使用規則式 (rule-based)演算法解決交通工具辨識的問題，常見的作法是設計一些if-else條件並找到最好的閾值 (threshold)。然而，為了處理多種不同交通工具的狀況，演算法可能會變得很複雜，因此學習式 (learning-based)演算法成為日漸普及的新途徑。近期很多研究利用手機蒐集到的資料來進行機器學習，並且有越來越多的研究著重在*監督式學習*，希望能透過從手機量測到的資料擷取更多能協助判斷的特徵。我們根據欲辨識的交通工具類別將近期的交通工具辨識相關研究整理於表 1，也根據使用的量測資料種類將相關研究分成以下常見的四類：

* GPS/GIS資訊

：這些研究使用地圖匹配技術將使用者的GPS/地理資訊系統 (GIS)軌跡對應到參考資料庫裡記錄的交通工具座標。GIS資料可以被用來產生更突出的特徵，如在[4], [5]的研究。除了GPS/GIS資訊，[6]–[9]的作者們結合了智慧型手機上的加速度計所量測到的資料，表現比單純使用GPS資料要來的好。Zheng等作者[10]使用如速度的平均值、期望值、前三快的速度和加速度等統計值作為特徵，用以辨識四種交通工具 (腳踏車、公車、汽車、走路)。除此之外，Zheng等作者在[11]中提出更進階的特徵，包括方向變化率、停止率和速度改變率，成功達到更高的準確率。

* 慣性測量單元 (inertial measurement unit, IMU)資料：相關演算法使用加速度計、陀螺儀、旋轉向量感測器和磁力計來蒐集手機使用者行為的不同資訊[1], [2]。
* 其他感測器：[12], [13]的作者們使用從其他感測器量得的資訊，包含體溫、心率和光強度等作為特徵來訓練一個能預測使用者活動的模型，預測的活動項目包括走路、跑步、騎腳踏車等。
* 無線訊號：蜂巢式網路、Wi-Fi或藍芽訊號的改變也能用來偵測使用者的行為。例如[13]–[15]的作者們使用蜂巢式網路訊號強度的波動來偵測使用者是處於靜止、走路還是開車中的狀態。

在此論文中，我們著重在使用蜂巢式網路資訊 (即上述的無線訊號一類)，但和其他相關研究的做法不同，我們不使用訊號強度波動作為特徵，而是提出一個全新的演算法來整合位置資訊和蜂巢式網路資料。由於現今交通路線沿線都會廣泛的佈建基地台，蜂巢式網路資訊會比GPS/GIS更可靠，特別是地鐵等交通工具。此外，由於我們的演算法只聚焦在手機連上哪個基地台和換手的頻率為何，因此能減少電力的消耗、紀錄資料的大小以及需要的CPU計算能力。

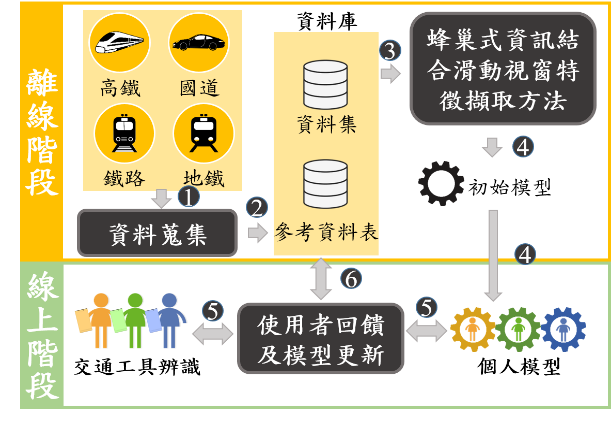


圖 2 系統架構

**四、系統總覽**

表 2 資料集描述

|  |  |  |  |
| --- | --- | --- | --- |
| 種類 | 時間 | 資料筆數 | 可得GPS比例 |
| 高鐵 | 142小時 | 12012 | 69% |
| 地鐵 | 120小時 | 8426 | 0% |
| 鐵路 | 206小時 | 42681 | 68% |
| 國道 | 116小時 | 38160 | 86% |
| 其他 | 56小時 | 20804 | 61% |

(一) 系統架構

我們提出的系統架構如圖 2所示，主要分成兩個階段：離線學習 (offline training)階段和線上學習 (online training)階段。離線學習階段會利用一個初始的資料集訓練出一個分類模型，線上訓練階段則利用使用者的回饋來微調 (fine-tune)此模型以改善辨識的結果，因此最後不同的使用者會擁有各自的模型，這些使用者新產生的資料和其利用個人模型生成的辨識結果，可以存回資料庫做為未來分析使用，這些新的資料也能用於初始分類模型的重新訓練。以下我們將詳細介紹各個步驟：

1. 初始資料蒐集：我們自己蒐集了如GPS和蜂巢式網路資訊等基本資料。值得注意的是我們的演算法實際上並不需要用到GPS的資訊，我們蒐集GPS資訊的原因是使用它來驗證我們資料標籤 (label)的正確性。我們透過自行開發的APP[16]蒐集各種交通工具的資料，資料內的時間資訊可以輔助我們進行標註 (labeling)的工作。到目前為止我們已經蒐集超過500小時的資料，資料集的描述詳見表 2。
2. 上傳資料：在步驟1)蒐集的資料會透過網路上傳至伺服器並儲存存在資料庫中，在本論文研究中我們採用了SQL-based資料庫系統。
3. 模型訓練：在資料清洗和正規化後，我們使用滑動視窗特徵擷取方法擷取出數個特徵進行模型的訓練，並在多次驗證和調整至最佳參數後，挑出表現最好的模型。

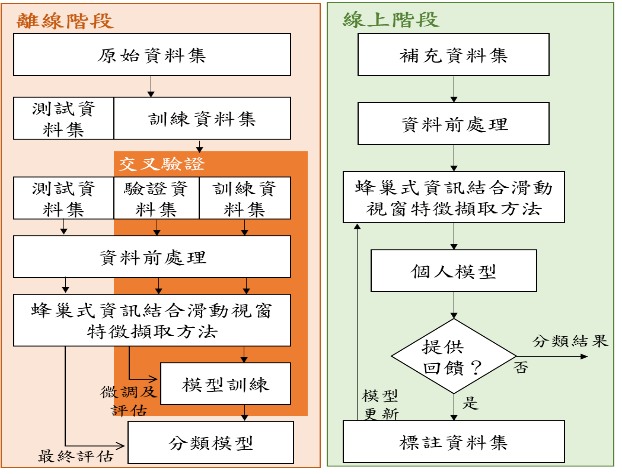


圖 3 系統工作流程圖

1. 模型部署：在步驟3)被訓練好的模型可以部署至手機APP或是透過如RESTful APIs等應用介面供使用者使用。
2. 回饋蒐集和線上訓練：我們採用線上學習演算法，利用使用者回饋進行預訓練模型的微調。使用者回饋可以用一段時間中使用者自行標記的標籤來呈現。由於篇幅有限，我們只評估我們提出之系統架構的可行性，如何驗證使用者的回饋是否正確或是惡意誤報等情況並不在此論文的討論範圍中。使用者的回饋會被輸入至原始的模型中並更新模型，增加對不同使用者的適應性。使用者量測到的資料會在*線上階段*被轉成特徵向量，接著會被輸入至在*離線階段*預訓練出的模型，最後生成個人化的交通工具辨識器。
3. 資料更新： 使用者新產生的資料和個人模型生成的辨識結果可以存回資料庫做為未來分析使用，也可以用於初始分類模型的重新訓練。

(二) 系統工作流程

我們提出的系統工作流程如圖 3所示。在離線階段時，我們先將原始蒐集到的資料 (表 2)以7:3的比例分成訓練和測試資料集，測試資料集將作為最後準確率的評估使用。而其中訓練資料集可以更進一步以4:1被分成訓練資料和驗證資料，此為5折交叉驗證 (5-fold cross validation)的做法，我們將訓練資料集隨機切成五等份，訓練時我們使用其中的四份，剩下的一份資料將作為驗證資料，在訓練完後用來評估模型的性能。在資料前處理階段，我們移除資料中不正常的離群值。接著，我們使用我們提出的蜂巢式資訊結合滑動視窗的特徵擷取方法來從資料中擷取特徵，並利用格點搜尋計算法來找到各種學習演算法的最佳超參數。最後，我們用測試資料集測試並挑選出擁有最高準確率的模型，而這個模型會在線上階段時部署給一般使用者使用。這些使用者會蒐集新的資料集，資料集經過資料前處理和蜂巢式資訊結合滑動視窗的特徵擷取方法等步驟後，會和使用者提供的回饋一同被送進上述預訓練好的模型，預訓練模型將能根據這些資料做更進一步的微調，進而改善個人模型的適應性及準確率。

**五、蜂巢式資訊結合滑動視窗的特徵擷取方法**

在這個章節，我們會先介紹相關術語，接著會介紹我們使用的參考基地台表 (reference cell table)之建立方式，以及我們提出的蜂巢式資訊結合滑動視窗的特徵擷取方法。

1. 蜂巢式網路資訊術語

圖 4呈現了使用者在持續移動時手機拿到的蜂巢式網路資訊，最開始時手機會連上基地台編號 (cell identity，之後都簡寫成Cell ID)為的基地台，此基地台的所屬的細胞 (服務區)標示為。我們可以透過Android系統內一個稱作*onSignalStrengthsChanged()*的應用程式介面 (API)，拿到基地台的資訊，包括： (1)用來辨識基地台所屬電信商的行動國家編碼 (mobile country code，MCC)、行動網路編碼 (mobile network code，MNC)， (2)在3G和4G行動網路裡被使用來辨識位置區或追蹤區 (一組基地台的覆蓋範圍)的位置區碼 (local area code，LAC)和追蹤區碼 (tracking area code，TAC)， (3)用來辨識各個基地台的基地台編號 (Cell ID)，和 (4)在3G和4G行動網路用來識別基地台實體層的主擾論碼 (primary scrambling codes，PSC)和物理小區標識 (physical cell ID)。當使用者移動至和的覆蓋範圍交界時，換手 (handover)可能會發生，也就是服務手機的基地台從換成。代表和間的距離 (cell distance)，細胞停留時間 (cell residence time) 則是指從連上到換手至之間的時間。

1. 參考基地台表 (reference cell table)

現有基於蜂巢式網路的演算法大多只使用換手的次數，或是服務的基地台及其附近其他基地台的訊號強度進行交通工具辨識[15], [17]，然而接收到的訊號強度容易被環境干擾，導致演算法結果不精確。我們另外使用了基地台的位置資訊，讓我們能透過追蹤使用者手機的服務基地台來粗估使用者的所在位置。也就是說，我們能追蹤使用者連上的多個基地台並和交通工具的路線比對，進而辨識使用者所搭乘的交通工具。然而，仍有兩個挑戰需要被克服： (1) 如何拿到基地台的位置資訊，和 (2) 如何知道一個基地台在交通工具的路線附近？我們提出了參考基地台表 (reference cell table) 的方法，它是一個記錄著接近四種交通工具 (高鐵、地鐵、鐵路和國道) 的各種基地台及它們位置資訊的資料庫，以下將會更詳細地解釋如何使用參考基地台表來解決上述的兩個挑戰。

* 基地台的位置資訊：理想上，我們能從電信商取得基地台的位置資訊，然而多數的電信商並不會公開這些資訊。所幸基地台的位置資訊是能透過群眾外包 (crowd-sourcing)來蒐集的，例如：Google[18]和OpenCellId[19]都有蒐集並公開基地台和其位置資訊。我們也招募了一些志願者來蒐集台灣前五大電信商的基地台位置資訊[16]，接著就能透過基於加權質心的計算方式[20]算出基地台的位置，計算方式的核心概念是使用者手機訊號強度良好時，往往代表使用者和基地台的距離較近，基地台的預測位置可以用下面的公式來計算：

其中基地台第次的量測位置為，其權重為，總共有的量測資訊。

我們用自己量測與計算出來的基地台資料和Google及OpenCellID計算平均絕對誤差 (mean absolute error，MAE)並進行比較，和Google的平均距離誤差約300公尺，低於OpenCellID的500公尺，代表Google的資料在台灣的準確度高於OpenCellID的資料。因此，當基地台資訊不存在在我們的reference cell table時，我們選擇使用Google的資料集作為輔助。

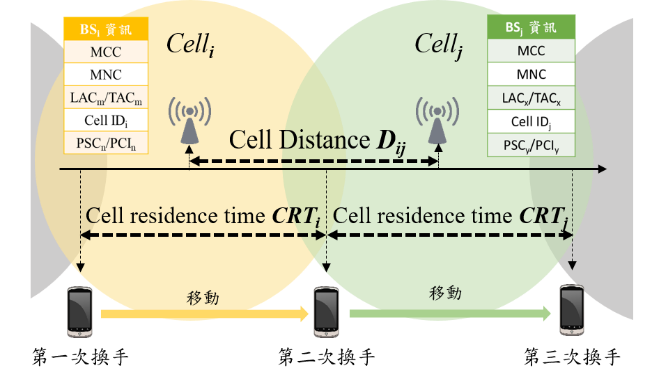


圖 4 手機量到的蜂巢式網路資訊

* Cell-ID：利用我們開發的APP，我們能蒐集使用者在一趟旅程中，搭乘不同交通工具所經過的所有Cell-ID。

簡言之，我們創建了一個reference cell table，它包含了可以用於交通工具辨識的Cell-ID和對應的位置資訊。

1. 蜂巢式資訊結合滑動視窗的特徵擷取方法

如圖 5所示，我們將蜂巢式網路特徵從紀錄檔案的一組樣本點中擷取出來，稱作一個segment，一個segment是由長度為秒的樣本點組合而成的。除了換手的頻率，我們也另外使用了兩種之前沒有被使用過的蜂巢式網路資訊： (1) 一個segment間的cell distance，和 (2) 每種交通工具匹配到的基地台數量。我們使用的蜂巢式網路資訊更詳細的介紹如下：

* 一個segment間的換手頻率：理想上，使用者移動的越快就會發生越多次的換手。在現實世界中，因為無線訊號的傳播本來就有雜訊存在，基地台的負載也隨時在變化，服務使用者的基地台可能會在鄰近的基地台間不停切換，這個現象稱作乒乓效應 (ping-pong effect)[21]。乒乓效應發生時，換手的次數就無法代表使用者的移動速度。為了改善此問題，我們只在每個基地台於segment中第一次被連上使用者手機時將其納入換手次數的計算。
* 一個segment間的cell distance：我們利用segment中紀錄的第一筆和最後一筆基地台的位置距離，來估計使用者的移動速度。基地台的位置資訊能透過先前介紹的參考基地台表來得到。
* 每種交通工具匹配到的基地台數量：這個特徵由五個數字組成，每個數字代表一種交通工具，分別是高鐵、地鐵、鐵路、國道及其他。我們計算一個segment中每個基地台匹配到的交通工具，如果某一種交通工具在這個segment中被匹配到越多次，就代表使用者越有可能在該種交通工具的路線上。

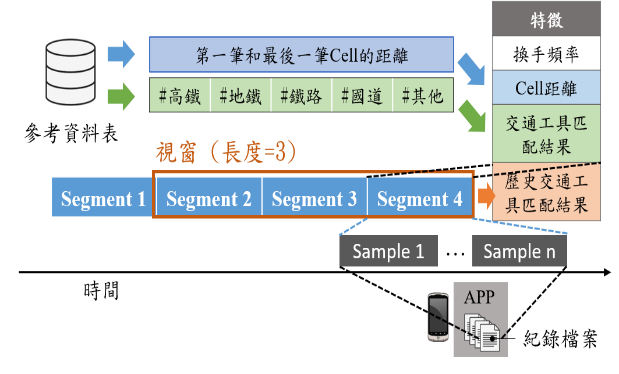


圖 5 蜂巢式資訊結合滑動視窗的特徵擷取方法

然而，我們無法保證參考基地台表包含了所有的基地台的Cell ID，也就是說使用者量測到的Cell ID有可能沒有記錄在我們的參考基地台表內的。為了解決此問題，我們提出了三種解決方法： (1) 忽視沒有記錄的Cell ID， (2) 把沒有記錄的Cell ID當作其他類別 (如果我們的參考資料表包含了四種交通工具沿線上大多數的基地台資訊，沒看過的Cell ID就很可能是其他類別)，和 (3)用歷史資訊預測此Cell ID的匹配交通工具。

上述的第三種解決方法可以被當作一個新的特徵，因為使用者搭乘交通工具的行為具備時間上的連續性，因此將歷史資訊作為一個新的特徵，對於交通工具辨識而言是很重要的。如圖 5所示，如果我們設定滑動視窗的視窗大小為3，就能納入歷史資訊，統計出一共3個segments中所有交通工具匹配數量作為新的特徵，進行交通工具的辨識。

**六、實驗結果**

在此章節中，我們先提出一些問題來尋找適合蜂巢式資訊結合滑動視窗的特徵擷取方法 (為節省篇幅，以下簡稱為*蜂巢視窗方法*)的參數。接著，我們會介紹一些用來評估蜂巢視窗方法的評估指標 (metrics)。最後，我們會用一些知名的機器學習方法來檢視蜂巢視窗方法的表現。

首先我們提出一些問題如下：

1. 在蜂巢視窗方法中，我們提出三種克服基地台沒被記錄在參考基地台表的方法，哪種方法最能提高準確率？
2. 如何設定蜂巢視窗方法的segment時間長度和滑動視窗大小？
3. 利用使用者的回饋來進行線上更新，相較於離線演算法，在準確率、訓練時間、模型大小及預測時間等面向是否能達到更佳的表現？
4. 蜂巢視窗方法是否能比基線 (baseline)演算法減少更多的手機耗電量？
5. 評估

(1) 資料集：我們使用如表 2描述的資料集進行我們的實驗，我們的資料集包含不同種交通工具共超過500小時的資料，且訓練資料包含了以下兩種：

* 基線 (baseline)資料集：這種資料集是用作模型的預訓練並在離線階段驗證模型的準確率。
* 補充資料集：為了驗證線上階段時的模型準確度，我們請線上階段訓練資料的蒐集者蒐集1)同手機 2)同電信商 3)同段路線的資料，並將這些資料稱作補充資料集。

(2) 評估指標：我們使用兩種指標：*精確率 (precision)*和*召回率 (recall)*來評估蜂巢視窗方法的特徵擷取表現。將交通工具的標籤作為真實類別(ground-truth)，精確率為資料辨識結果正確的比例，召回率則為貼有真實類別的資料中，辨識結果和真實類別相同的比例，我們用我們的資料集來算出這兩個指標，兩個指標的值都是越高越好。

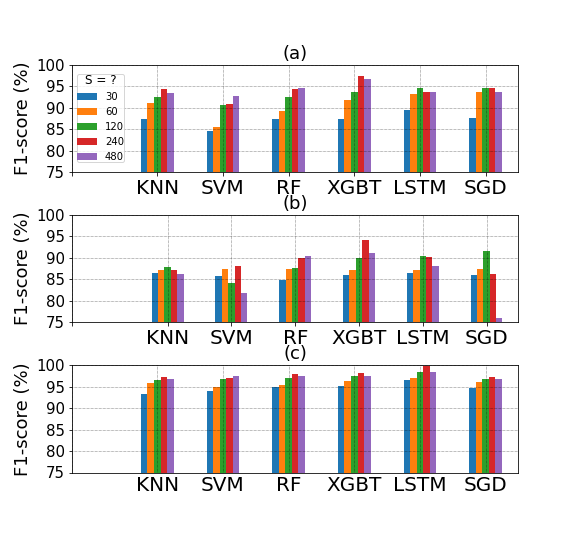


圖 6 比較未知Cell ID的三種解決方式。 (a) 方法1：忽略未知Cell ID。 (b) 方法2：把未知Cell ID當作其他類別。 (c) 方法3：參考歷史資料 (此實驗中視窗大小為4)。

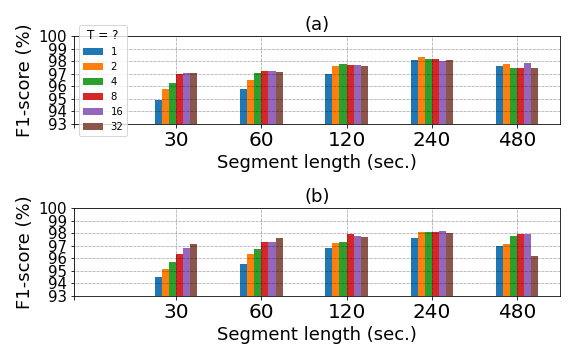


圖 7 比較不同的segment長度和視窗大小。 (a) 離線訓練，XGBT。 (b) 線上訓練，SGD。

為了進一步評估模型的總體表現，我們也考慮F1分數 (F1-score)[22]，即精確率和召回率的調和平均，來進行評估。F1分數 定義如下：

(二) 蜂巢式資訊結合滑動視窗特徵擷取方法表現

(1) 離線階段的分類結果：我們用知名的演算法來進行分類結果評估，如K-近鄰演算法 (k-nearest neighbor，KNN)、支援向量機 (support vector machines，SVM)、隨機森林 (random forest，RF)、XGBoost (XGBT)、多層感知器 (multiple layer perception，MLP)和SVM下的隨機梯度下降法 (stochastic gradient descent，SGD)。我們使用Python的scikit-learn[23]和tensorflow[24]套件來實作學習的演算法。首先，我們比較前述參考基地台表處理未知Cell ID的三種解決方法之表現，並呈現比較結果；接著，我們展示不同segment時間長度和視窗大小的比較結果。由於篇幅有限，我們只呈現F1分數的結果如下：

* 參考基地台表處理未知Cell ID的三種解決方法之準確率：三種方法在不同segment時間長度的F1分數如圖 6所示。方法1 (圖 6-(a))使用XGBT且時，F1分數可以達到97%。總體而言，較大的代表學習演算法能得到更長時間的資訊，因此如果忽視未知的Cell ID且又不夠長時，剩下的資訊可能不足以讓演算法正確的辨識交通工具，也就是說拉長能改善未知Cell ID的問題。然而，如果太長，改善的效果就會開始降低，因為一個segment可能就會包含兩種以上的交通工具，造成辨識結果容易出錯。方法2 (圖 6-(b))的表現並不如方法1，因為直接將未知Cell ID改成其他類別會一定程度地干擾學習演算法的成效 (會降低召回率)。方法3 (圖 6-(c))的F1分數比方法1和2都要高，因為它利用歷史資訊 (滑動視窗)將未知Cell ID的影響降低。綜上所述，使用滑動視窗的方法3更適用於交通工具辨識。
* 不同segment時間長度和滑動視窗大小的準確率：圖 7的(a)和(b)分別呈現了離線 (XGBT)和線上 (SGD)學習演算法在不同segment時間長度和視窗大小的結果。總體而言，越大代表有更多的資訊會被送入學習演算法，所以能達到較高的準確率。相反的，當和的值降低時準確率就會下降，因為學習演算法無法拿到足以用來辨識的資訊。但是當較小但增加時，由於增加代表有較多的歷史資訊會被考慮，所以準確率仍然會增加。利用我們提出的蜂巢視窗方法應用在線上學習的SGD時，如圖 7-(b)所示，即便是在大小只有1和長度只有30秒的狀況下，準確率仍能達到94.5%。當被設置成240秒時能達到最佳的準確率，主要是因為台灣地鐵兩站間的距離通常小於4分鐘，所以能得到較完整的地鐵資料，使辨識地鐵的準確率提升。比較離線和線上訓練的結果，蜂巢視窗方法應用在離線訓練的結果如圖 7-(a)所示，XGBT平均能達到約98%的F1分數，應用在線上訓練的SGD則如圖 7-(b)所示，能達到約97.8%。綜合以上，當segment較長時，系統在擷取蜂巢式網路資訊和將資料輸入至學習模型前的等待時間也會比較長。因此，我們傾向選擇較短但又同時能維持理想準確率的segment和視窗大小。

(2) 線上階段的分類結果：訓練資料集的大小對於傳統學習演算法的成效至關重要，較小的資料集通常會導致較低的辨識準確率。雖然蜂巢視窗方法蒐集了一些有利於分類的特徵，但卻無法確保涵蓋了所有的情境。因此，我們使用線上學習，透過蒐集新的資料集 (即使用者的回饋)更新現有的模型。接下來，我們會用補充資料集在離線和線上學習演算法的表現來評估蜂巢視窗方法。準確率、訓練時間、模型大小和預測時間的評估結果如圖 8。我們發現透過給予更多的資料集，不論是離線還是線上的學習演算法，準確率都能有所提升 (圖 8-(a))。一開始，線上學習的準確率比離線學習要低，但是隨著資料的增加，兩者的準確率也越來越接近。在模型大小的部分，線上學習演算法的模型只有約2 KB，但離線演算法的模型有500 KB (圖 8-(c))。圖 8的(b)和(d)則可以看到不論是模型訓練或是預測，線上學習演算法需要的時間都低於離線學習演算法。線上學習演算法訓練模型只需要約10毫秒，預測模型則只需要約1毫秒的時間。

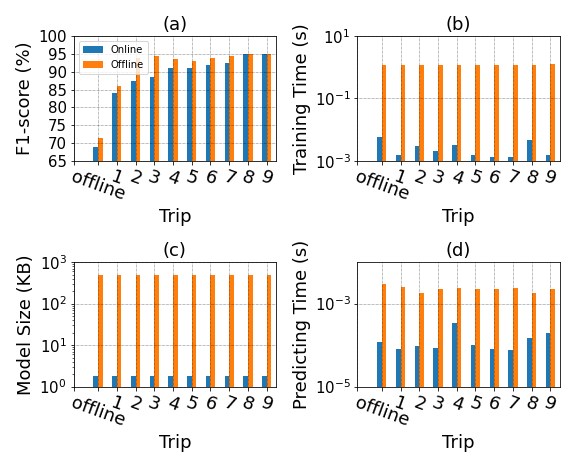


圖 8 補充資料集的表現評估比較。(a)F1分數， (b)訓練時間， (c)模型大小，及 (d)預測時間。

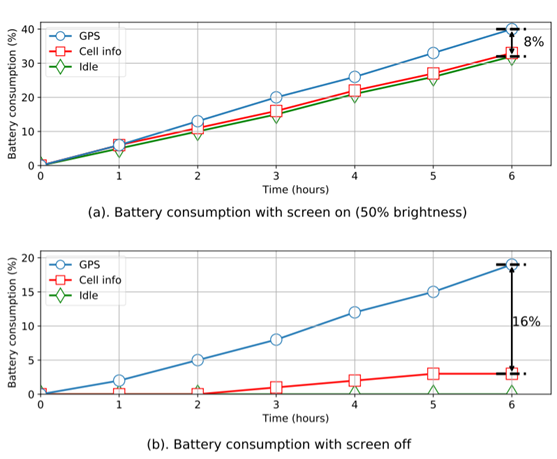


圖 9 耗電量比較

(3)耗電量：我們使用小米Note 4進行耗電量的比較實驗，我們比較在螢幕開和關兩種狀況下，使用蜂巢視窗方法 (蒐集蜂巢式網路資訊)和使用基於GPS的學習方法 (蒐集GPS資訊)的耗電量，並將兩種資訊都不蒐集的耗電量作為基線(圖 9的Idle)，結果如圖 9所示。當螢幕開啟並經過6小時後，使用蜂巢網路方法比使用GPS的方法減少約8%的耗電量，在螢幕關閉時則能減少約16%。

**七、結論**

對於5G網路而言，正確的辨識使用者服務型態，並有智慧的分配資源是一個迫切待解的議題。因此，我們分析使用者在不同交通工具上的移動模式，包括高鐵、地鐵、鐵路和國道。在此論文中，我們提出了一個使用機器學習實現交通工具辨識的方法，以及只使用蜂巢式網路資訊來提供準確的交通工具辨識結果的系統架構。我們蒐集了約500小時的資料集來進行評估，實驗結果顯示我們提出的演算法能有效地被應用在數種知名機器學習方法上，並達到約98%的辨識準確率。我們的實驗結果也顯示我們的方法和使用GPS的方法相比，能減少約8%至16%的耗電量。未來我們希望能更進一步，應用深度學習至蜂巢式資訊結合滑動視窗特徵擷取方法，更加提升準確率等表現。

**八、致謝**

本論文感謝科技部計畫編號MOST 108-2221-E-009-042-MY3、MOST 108-2218-E-009-028、MOST 106-2221-E-009-046-MY3支持。感謝張宏鉦、林雅瑄、許曄律等人支援軟體開發與測試，並感謝吳佳諺、廖紫甯、林詩芸、黃聖芬、黃煒恩、羅靖雅、蔡涵皓、陳紫妍、馬藜珊、林家榆、吳子亦、黃郁宸、楊善涵等人協助資料蒐集。

**參考文獻**

[1] S.-H.Fang *et al.*, “Transportation modes classification using sensors on smartphones,” *Sensors*, vol. 16, no. 8, p. 1324, 2016.

[2] X.Su, H.Caceres, H.Tong, andQ.He, “Online travel mode identification using smartphones with battery saving considerations,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2921–2934, 2016.

[3] S. T.Pasha, “Energy Efficiency in Smartphones: A Survey on GPS energy conservation,” *Imp. J. Interdiscip. Res.*, vol. 3, no. 6, 2017.

[4] L.Stenneth, O.Wolfson, P. S.Yu, andB.Xu, “Transportation mode detection using mobile phones and GIS information,” *Proc. ACM Int. Conf. Adv. Geogr. Inf. Syst.*, pp. 54–63.

[5] F.Biljecki, H.Ledoux, andP.VanOosterom, “Transportation mode-based segmentation and classification of movement trajectories,” *Int. J. Geogr. Inf. Sci.*, vol. 27, no. 2, pp. 385–407.

[6] T.Nick, E.Coersmeier, J.Geldmacher, andJ.Goetze, “Classifying means of transportation using mobile sensor data,” *Proc. IEEE Int. Jt. Conf. Neural Networks*, pp. 1–6.

[7] A.Jahangiri andH. A.Rakha, “Applying machine learning techniques to transportation mode recognition using mobile phone sensor data,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2406–2417, 2015.

\* 本研究接受科技部編號：MOST 108-2221-E-009-042-MY3、MOST 108-2218-E-009-028、MOST 106-2221-E-009-046-MY3研究計畫經費補助

[8] A.Jahangiri andH.Rakha, “Developing a support vector machine (SVM) classifier for transportation mode identification by using mobile phone sensor data,” *Proc. 93rd Annu. Meet. Transp. Res. Board*, no. 14–1442.

[9] P.Widhalm, P.Nitsche, andN.Brändie, “Transport mode detection with realistic smartphone sensor data,” *Proc. IEEE Int. Conf. Pattern Recognit.*, pp. 573–576.

[10] Y.Zheng, L.Liu, L.Wang, andX.Xie, “Learning transportation mode from raw gps data for geographic applications on the web,” *proc. ACM Int. Conf. World Wide Web*, pp. 247–256.

[11] Y.Zheng, Q.Li, Y.Chen, X.Xie, andW.-Y.Ma, “Understanding mobility based on GPS data,” *proc, ACM Int. Conf. Ubiquitous Comput.*, pp. 312–321.

[12] J.Parkka, M.Ermes, P.Korpipaa, J.Mantyjarvi, J.Peltola, andI.Korhonen, “Activity classification using realistic data from wearable sensors,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 119–128.

[13] T.Sohn *et al.*, “Mobility detection using everyday GSM traces,” *Proc. ACM Int. Conf. Ubiquitous Comput.*, pp. 212–224.

[14] I.Anderson andH.Muller, “Practical Activity Recognition using GSM Data,” 2006.

[15] A. M.AbdelAziz andM.Youssef, “The diversity and scale matter: Ubiquitous transportation mode detection using single cell tower information,” in *in Proc. of IEEE Vehicular Technology Conference (VTC’15)*, 2015, pp. 1–5.

[16] Y.-H.Lin, J.-C.Chen, C.-Y.Lin, B.-Y.Su, andP.-Y.Lee, “Poster: SensingGO - Toward Mobile/Cellular Data Measurement with Social and Rewarding Activities,” *Proc. ACM Int. Conf. Mob. Comput. Netw.*, pp. 765–767.

[17] G.Li *et al.*, “Public Transportation Mode Detection from Cellular Data,” *Proc. ACM Conf. Inf. Knowl. Manag.*, pp. 2499–2502.

[18] “Google Maps Geolocation API.” [Online]. Available: https://developers.google.com/maps.

[19] “Unwiredlabs Location API.” [Online]. Available: https://unwiredlabs.com/?ref=ocid.

[20] E.Neidhardt, A.Uzun, U.Bareth, andA.Küpper, “Estimating locations and coverage areas of mobile network cells based on crowdsourced data,” *Proc. IEEE Wirel. Mob. Netw. Conf.*, pp. 1–8.

[21] T.Henderson, D.Kotz, andI.Abyzov, “The changing usage of a mature campus-wide wireless network,” *Comput. Networks*, vol. 52, no. 14, pp. 2690–2712, 2008.

[22] C.Goutte andE.Gaussier, “A probabilistic interpretation of precision, recall and F-score, with implication for evaluation,” *Proc. Eur. Conf. Inf. Retr.*, pp. 345–359, 2005.

[23] “Scikit-learn tool for python.” [Online]. Available: http://scikit-learn.org/stable/.

[24] “TensorFlow framework.” [Online]. Available: https://www.tensorflow.org/.

[25] H.Gong, C.Chen, E.Bialostozky, andC. T.Lawson, “A GPS/GIS method for travel mode detection in New York City,” *Comput. Environ. Urban Syst.*, vol. 36, no. 2, pp. 131–139, 2012.

[26] W.Zhu, J.Ash, Z.Li, Y.Wang, andM.Lowry, “Applying semi-supervised learning method for cellphone-based travel mode classification,” in *Smart Cities Conference (ISC2), 2015 IEEE First International*, 2015, pp. 1–6.

[27] S.-H.Fang, Y.-X.Fei, Z.Xu, andY.Tsao, “Learning Transportation Modes From Smartphone Sensors Based on Deep Neural Network,” *IEEE Sens. J.*, vol. 17, no. 18, pp. 6111–6118, 2017.