



電信與車輛偵測 大數據應用 — 高雄市 旅運特性 分析與易壅塞 道路預報

周諺鴻／鼎漢國際工程顧問公司 資訊長／副總經理

顏郁航／鼎漢國際工程顧問公司 數據師

王志綱／高雄市政府交通局智慧運輸中心 主任

高雄市已實施多項交通管理或工程手段來解決塞車問題，然市區車流與相鄰縣市間及過境性城際車流仍需進一步透過整合式運輸走廊智慧交通管理的手段，引入先進交通預報管理解決方案，藉由科技管理、大數據分析與交通預警運作機制及跨單位複合運輸與交通協調構建交通預報與管理功能。先進技術上可結合道路車輛偵測數據及大樣本電信數據來精細掌握市民之旅運特性，包括分時段別旅次起迄、旅次目的、運具別等特性。因此，藉由建立大數據旅次起迄分析與短期交通預報資訊服務，持續性提供熱門景點及易壅塞道路車流特性分析與預報服務，作為管控與疏散運輸走廊壅塞車流之基礎。

電信大數據應用 — 旅運特性與飽和點分析

高雄市府與電信業者合作，將電信源頭端進行資料去識別化處理，再依據各地區基地台分佈位置、密度、訊號投射方向等，透過三角定位法，將基地台資料正規化轉換為空間網格維度資料，建立各網格對應之空間屬性資料，高雄市共劃分為 51,658 個分析網格。由於原始信令資料為行動通訊傳遞紀錄，需透過邏輯定義與轉換，依據不同分析維度將其轉換為旅運行為資料、路段速率、空間飽和與熱度等類型應用資料，如圖 1。

電信數據可擷取之旅運資料

1. 旅次行為資料：以用戶為基礎 (imsi-based) 之行動數據資料，透過用戶手機移動軌跡與基地台傳輸紀錄，定義各基站 POI 之對應關係 (例如：家、工作、學校等)，並以訊號移動的速度與停留時間，判

斷該旅次移動與停留特徵，轉換為旅次行為資料。

2. 路段速率資料：以 link-based 為基礎之行動數據資料，將用戶移動軌跡，以距離與時間進行行駛速率之運算，匹配於道路路網、路段上，轉換為不同時段各道路路段速率資料。

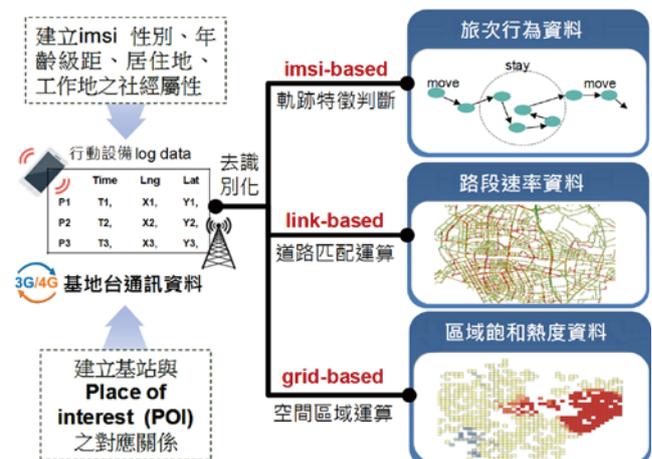


圖 1 電信數據資料轉換概念

3. 區域飽和度資料：以網格為基礎 (grid-based) 之行動數據資料，以不同時間斷面各基地台通訊傳輸之用戶數變化，轉換為以每 5 ~ 15 分鐘為一時間單位之空間網格飽和度熱度資料。

電信數據特性

電路控制需時刻進行，因此可以針對個體行動裝置活動，包含通話、通信、3G/4G 網路等進行追蹤。透過行動裝置回傳的信號資料與電信基地台間的相對關係，界定該信號用戶所在位置 (或所屬基站服務範圍)，並利用訊號於基站與基站間之移動距離、時間、速度等資訊，得知用戶於特定時間下的移動行為與軌跡。

根據不同信令資料類型，能應用於分析不同類型之議題，如 Call Detail Record (CDR) 或 Subscriber Dr. Data 由於資料密集度相對較低，因此較適用於判斷手機使用者 OD 起迄與旅次目的，Signal Data 為主動式資料，資料擷取時間頻率較高，因此除 OD 起迄與旅次目的外，亦可用於觀察使用者軌跡路徑，相對而言，其設備要求、資料量以及運算作業的硬體需求亦相對要求較高，各類型資料回傳密度示意如圖 2 所示，相關特性彙整如表 1 所示。

電信數據擷取旅運資料作法

基地台接收到用戶行動裝置的資料後，將其回傳至電信業者後端硬體設備中，再透過演算分析初步辨識和解析，轉換為以地區或是人為本位的資料庫進行存放。基於個人資料保護法，各用戶之個人資訊與所在位置為機密資訊，需將用戶所屬之國際行動用戶辨識碼 (International Mobile Subscriber Identification Number, IMSI) 轉編成臨時性 ID 資料，經由判讀用戶之每日各時段的旅次活動行為，先依據信號移動特性區分為停留與移動旅次，再針對有意義的停留旅次判斷活動目的、使用運具。

表 1 信令資料類型特性比較

資料類型	資料特性	數據密集度	限制
Call Detail Record (CDR)	提供通信、SMS 傳輸期間之基地台位置、時間	低	為被動式傳輸資料，資料間隔期長，較不適合做精細之行為分析
Subscriber Dr. Data	被動個體行動裝置活動追蹤 (包含：通信、SMS、3G/4G 網路等)	中	通信、上網等資料上傳期間才有資料
Signal	主動個體行動裝置活動追蹤 (包含：通信、SMS、3G/4G 網路等)	高	資料量龐大，資料倉儲費用高，需經過壓所萃取出具有特性意義之資料

電信數據分析高雄熱門活動據點

以一個月份的電信數據 (106 年 5 月份) 轉換為網格維度計算停留該區超過 15 分鐘之用戶數，透過空間運算克利金 (Kriging) 內插法，將高雄市各地區活動熱度區分為 1 ~ 10 分，指標分數越高代表活動熱度越強，藉此找出高雄熱門活動據點，參見圖 4。

以活動熱點最高之高鐵新左營站為例，電信數據飽和度為 10 分，該空間單元因涵蓋高鐵新左營站、臺鐵新左營站以及高雄捷運左營站，高鐵站之平均每日進出站量約 4.7 萬人次、臺鐵站之進出站量約 1.6 萬人次、捷運站之進出站量約 3.3 萬人次，三個車站合計之每日進出站量近 10 萬人次，加上車站 2 至 4 樓為環球購物中心，並緊鄰 10 層樓之新光三越百貨與彩虹市集，因此活動熱度高。進一步分析此地區之活動相對熱度，主要於週五、日以及連假收假日最高，並集中於晚上 17 點至 22 點，參見圖 5。

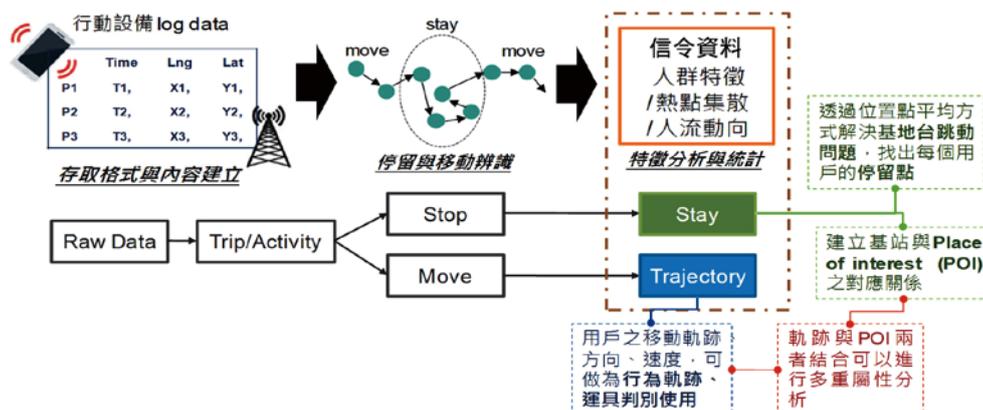


圖 2 電信數據移動停留特徵判斷處理

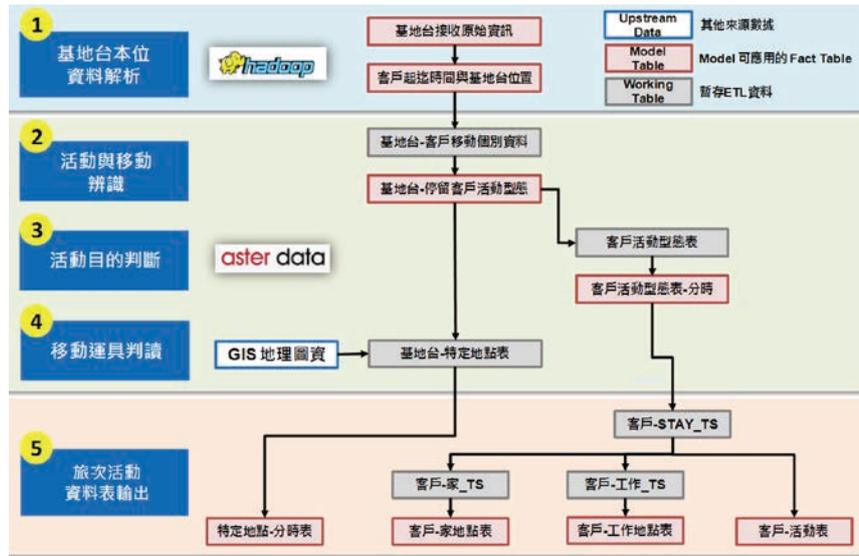


圖 3 旅運特性資料擷取作業

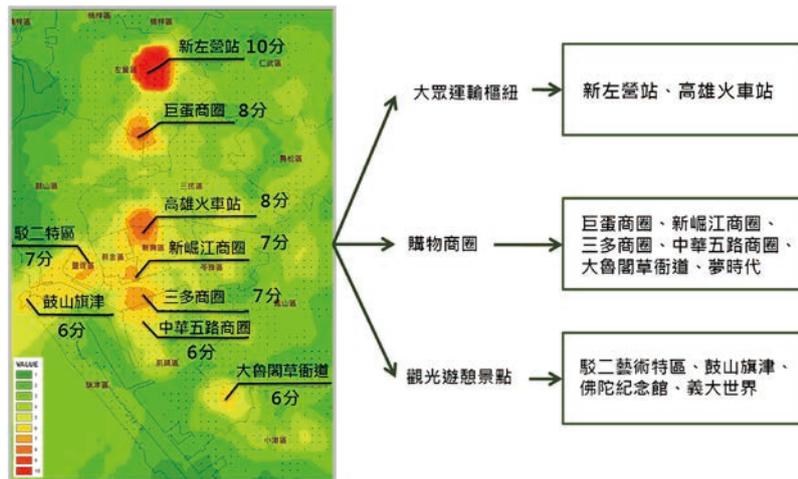


圖 4 電信數據分析高雄市熱門活動據點

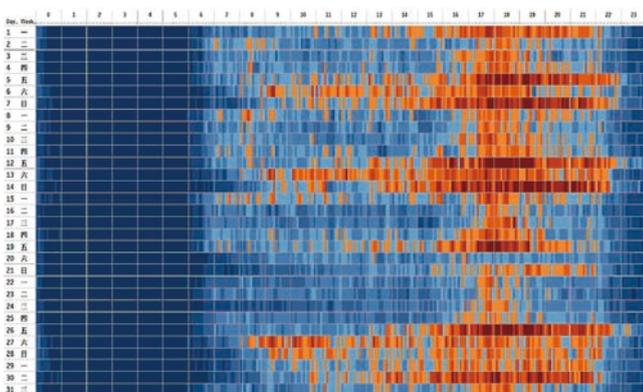


圖 5 電信數據分析高鐵新左營站地區分時活動熱度變化

電信數據分析高雄旅運特性

1. 平常日：以住宅區內旅次以及往來工業區旅次為大宗，主要起迄以左營區、前鎮區、小港區、楠

梓區、苓雅區、三民區、岡山區各自區內活動，以及跨行政區旅次之三民區至左營區、前鎮區至苓雅區、鼓山區至左營區、苓雅區至前鎮區、左營區至三民區、前鎮區至左營區等較高。

2. 週末假日：以往來商圈與熱門景點區域為最大宗，主要起迄分佈分別以左營區、三民區、前鎮區、苓雅區、鳳山區各自區內活動，以及跨行政區旅次之三民區往返於左營區、鼓山區往返於左營區、苓雅區往返前鎮區等較高。

3. 連續假期：跨區域活動之特性比平日、假日明顯，以分析當月份之端午節連假可看出左營區、前鎮區、三民區、小港區、鳳山區之各自區內活動以及跨行政區旅次之三民區至左營區、前鎮區往返苓雅區、鼓山區至左營區、鳳山區至前鎮區等較高。



圖 6 電信數據分析高雄不同日期主要旅次起迄

車輛偵測數據應用 — 道路壅塞預警門檻

一般道路偵測設備包括車輛偵測器 (Vehicle detector, VD)、車輛自動辨識 (Auto Vehicle Identify, AVI)、路側無線射頻感測器 (RFID Reader)、路況監視器 (Close Circuit television, CCTV) 等。其中，高雄市以車輛偵測器之建置數量 (約 220 處) 與分布最多最廣，車輛偵測器可記錄道路上車輛的平均速度、車流量和時間佔有率等數據，如圖 7，本文選擇以此作為道路績效數據蒐集與分析應用之基礎。

道路壅塞門檻值分析方法

將道路偵測回傳數據透過 K-means 法，以最佳分群數法進行分群，即群內差異最小、群間差異最大，分析道路壅塞臨界值。K-means 是一個分割式分群 (Partitional Clustering) 的演算法，主要的目的是在大

量資料中快速找出具有代表性的資料點，這些資料點稱為群中心 (center)，再依據相近的特徵邏輯分成 K 群。分割式分群法目的是希望盡量減小每個群聚中，每一資料點與群中心的距離平方誤差 (square error)，使其群間誤差極大化。亦即使同一分群裡需要具有類似特徵，不同分群必須具有相異之區別特性方能達到分群分類之用意。觀測值間之距離計算式如下：

$$\arg \min \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

μ_i : S_i 群體的平均數；

x_j : 第 j 個資料點；

k : 分群的總群數；

j = 1, 2, 3, ..., 資料總筆數 ; i = 1, 2, 3, ..., k 。



圖 7 高雄市車輛偵測器區位與數據查詢資訊

於 K- 平均數分群法中，最佳分群數之決定可由 CH 指標 (Calinski-Harabasz criterion) 評估而得，CH 指標透過計算各群間變異數與群內變異數之比例，以群間離散程度與群內緊密程度比例做為依據，以取得最佳分群數，公式如下：

$$\frac{SS_B}{SS_W} \times \frac{(N - k)}{(k - 1)}$$

SS_B：群間變異數；

SS_W：群內變異數；

N：資料總筆數；

k：分群的總群數；

而各分群之起始中心點則透過 Lloyd 演算法 (Lloyd's algorithm) 設定，Lloyd 演算法根據 CH 指標取得之分群數，將所有資料點隨機切分為初始化分組，並計算每組之中心點做為各分群之起始中心點。接著，計算各資料點與各分群中心點之歐氏距離 (Euclidean Distance)，並將資料點重新切分至距離最近之中心點所在分群後，再次計算各群中心點並重新分組，直到中心點位置不再改變。

易壅塞道路之壅塞預警門檻建立

本文以高雄市境內之道路車輛偵測器回傳之 105 年 1 月至 106 年 9 月間數據，透過分時流量、平均速率等兩項指標篩選高雄市易壅塞之路段，篩選出九如路、中正路、大中路、中華三路、自由二路、以及博愛路等 6 處為市區重要之易壅塞道路。將此 6 處壅塞道路以各路段順逆向分時之交通量與平均速率為一個資料分析點，共有 1.4 萬個數據點 (已剔除異常或缺漏數據點) 進行壅塞預警門檻值分析，如圖 8。

藉由分析找出各路段之壅塞臨界值，若車流量超過此門檻值時，速率會逐漸降低出現交通壅塞之情形，各路段之門檻值再納入交控中心，結合資訊可變標誌發佈排程設定，提供用路人於上游路段事先壅塞預警資訊，以避開壅塞時段或改道，以疏緩塞車之現象。

市區易壅塞道路交通預報

交通預報技術

1980 年代交通預測主要利用統計概率論建立數學模型，收集所觀察系統的數據，主要以迴歸分析、時間序列分析等進行量化的分析、總結，併進而推斷和預測。2010 年機器學習方法崛起，以大量特徵變數

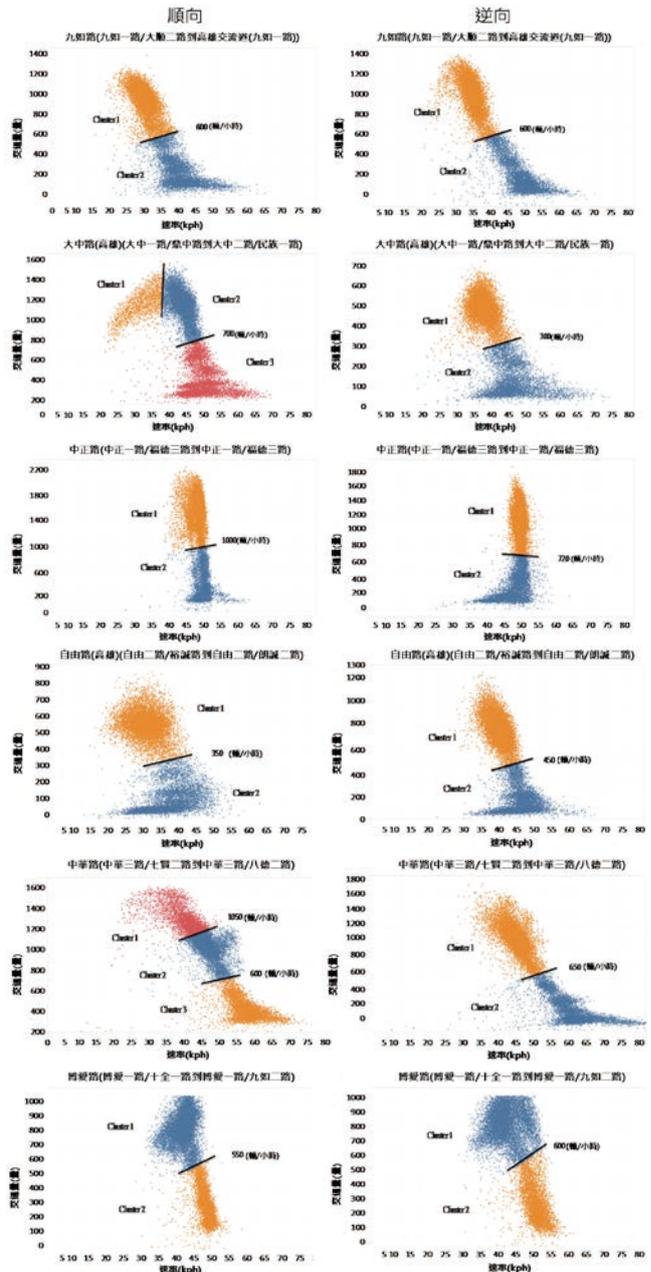


圖 8 高雄市區易壅塞道路壅塞預警門檻值界定

數據建立訓練集與測試集建立模式進行預測，機器學習方法發展相當快速且多元，包含決策樹 (Decision Tree)、隨機森林 (Random Forest)、類神經網路 (neural network)、最近鄰居法 (KNN)、貝葉斯算法 (Bayesian Algorithms) 等，國內交通部運輸研究所、高公局相關研究與創意競賽，亦導入機器學習法進行交通速率、旅行時間預報。2017 年隨著人工智慧 AI 時代的來臨，Google DeepMind 在人工智慧方面取得重大突破，AlphaGo 透過深度學習法 (Deep Learning) 戰勝棋王而聲名大噪，北京交通大學也導入深度學習法預測北京道路速率，亦獲得不錯效果。

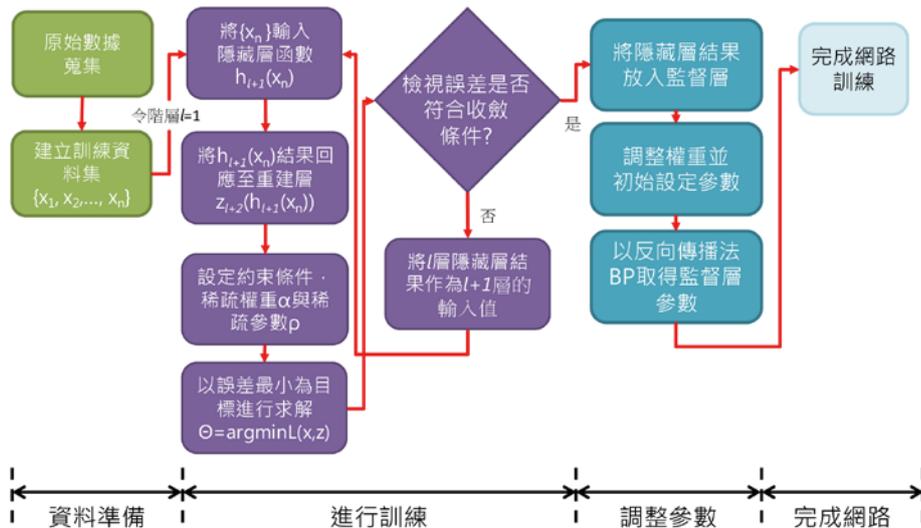


圖 9 深度學習模式訓練邏輯

本文針對此項機器學習法評估後，採用深度學習作為交通預報之方法，此演算法是基於最陡坡降法 (gradient steepest descent method)，透過疊代使資料輸出值實際值之誤差達到最小，並將過程建立成模式。深度學習的概念以目前公認的理解上系屬機器學習 (Machine Learning, ML) 的分支，並歸類為特徵學習 (Feature Learning) 之類別。簡略概之是期望可從眾多函數集 (現多採用線性、非線性等函數) 中選取最好的函數，然而函數會以不同的組合產生出不同等級的階層，產生出多個處理層對資料進行分析的演算法，並從中取得最佳績效函數與階層。

交通預報模式建立

1. 交通預報模式變數選擇，模式所需投入變數應具有可事前掌握之特性，並具有足夠的數據量，方能透過機器學習演算法建立預測模式，高雄交通預報模式變數可歸納為假期、氣象、交通、活動四大面向。

- (1) 假期：根據人事行政局每年公告行事曆，建立不同年月、星期別、假期名稱 (例如：春節、中秋節、清明節、端午節、雙十節、元旦等連續 3 天以上之假期)、假期天數 (例如：3 天、4 天等) 建立對照模式變數所需屬性變數。
- (2) 氣象：介接交通部中央氣象局氣候統計觀測資料查詢系統，透過 ETL 資料萃取轉換程序，處理成高雄交通預報模式所需變數，例如日雨量是否超過 80 mm，有則 heavyrain_yn 變數數值為 1，沒有超過 80 mm heavyrain_yn 變數數值為 0。

- (3) 交通：以道路偵測設備 (VD) 數據資料為基礎，收納 2016/1 ~ 2017/9 每 5 分鐘 VD 歷史數據資料，並透過 ETL 資料萃取轉換程序，剔除異常數據，並針對各 VD 進行分時流量與分時平均速率之運算，轉換為模式之訓練變數。
- (4) 活動：根據體育署與各縣市政府核可之路跑活動日期與範圍，將其建立為觀察變數。根據交通部觀光局、高雄旅遊網每年公告之活動行事曆、大型燈會、大型展覽、年度盛事之舉辦地點與日期建立模式所需之變數，及結合電信大數據分析之活動飽和度資料。



圖 10 高雄市區幹道交通預報模式建立流程

2. 深度學習模式選擇參數：參數值包括層數 (Layer)、疊代次數 (Epochs)、神經元數 (Neuron)、學習率 (Learning Rate)、活化函數 (Activation Function)、分配函數 (Distribution Function)、輸入層刪除率 (Input Dropout Ratio)。高雄交通預報模式經反覆測試訓練，最佳參數如表 2。

表 2 深度學習法選擇參數

模式參數	參數值
層數 Layer	3
疊代次數 Epochs	60
神經元數 Neuron	400
學習率 Learning Rate	Adaptive
活化函數 Activation Function	Rectifier
分配函數 Distribution Function	Auto
輸入層刪除率 Input Dropout Ratio	0.1

(3) 交通預報模式應用範圍：以兩處熱門據點之聯外道路，以及三條市區易壅塞道路進行預報模式驗證應用，模式預報可產出各路段不同日期每 5 或 30 分鐘為單位之每日時空流量與時空速率資料，作為交控中心事前掌握交通壅塞程度與發佈交管措施之基礎。

易壅塞道路交通預報應用與驗證

以雙十連續假期 (2017/10/7 ~ 10/10) 進行交通預報服務與驗證，於假期前 14 天啟動預報資訊提供，並於假期後進行交通預報驗證，先以三條市區易壅塞道路進行試辦。假期後透過平均絕對值誤差率 (mean absolute percentage error, MAPE) 檢驗預報與實際值之誤差，如表 3，驗證結果 MAPE 值多小於 15%，預報能力良好，可持續透過模式之滾動學習精進與持續將交通預報資料提供交通管理權責單位與用路民眾參考。

表 3 三條市區幹道壅塞路段交通預報 (2017 雙十連假) MAPE 值

日期 / 方向	大中路		九如路		中正路		
	速率	流量	速率	流量	速率	流量	
10/7	順向	9.11%	13.25%	6.72%	3.34%	2.50%	7.75%
	逆向	12.00%	6.58%	3.61%	1.80%	1.39%	2.22%
10/8	順向	6.28%	10.18%	8.11%	7.21%	2.83%	5.63%
	逆向	10.56%	6.50%	4.28%	0.19%	1.33%	2.79%
10/9	順向	6.78%	9.17%	5.61%	0.58%	1.83%	5.08%
	逆向	10.50%	3.44%	3.94%	0.10%	2.28%	12.92%
10/10	順向	5.11%	8.37%	5.11%	5.57%	0.67%	4.35%
	逆向	9.61%	0.43%	5.22%	2.99%	1.72%	11.49%

參考文獻

- 應用大數據技術建置重要瓶頸路段及運輸走廊之交通預警機制，交通部運輸研究所 (2016)。
- 106 年多元偵測技術整合應用計畫，高雄市政府交通局 (2018)。
- Mark K. Y. Mak, George T. S. Ho and S. L. Ting (2011), A Financial Data Mining Model for Extracting Customer Behavior.
- Miao Chong, Ajith Abraham and Marcin Paprzycki (2011), Accident Data Mining Using Machine Learning Paradigms.

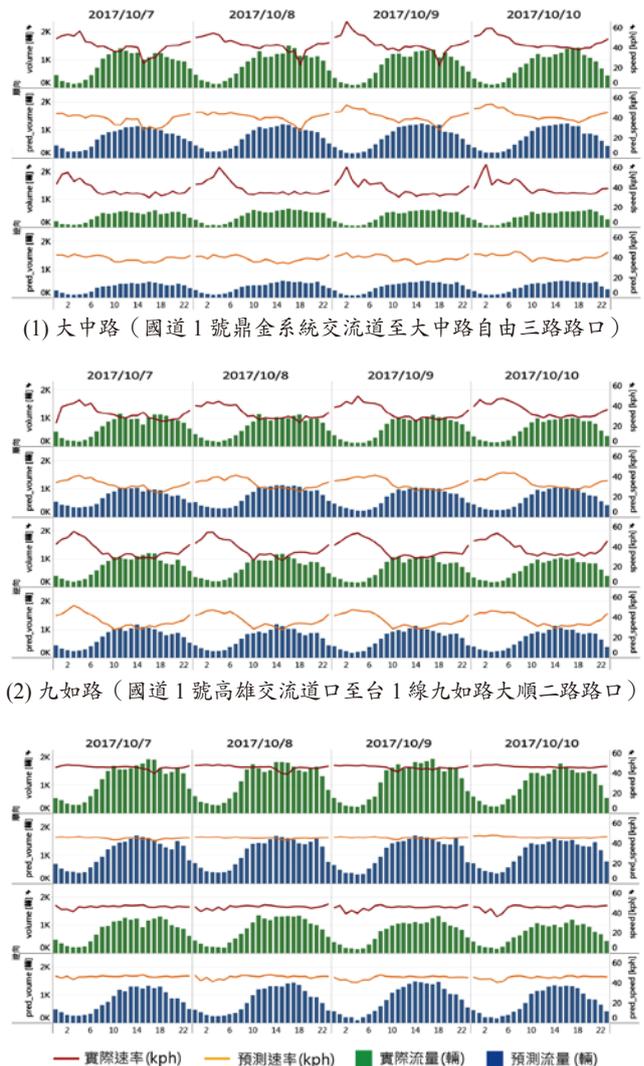


圖 11 三條市區幹道壅塞路段交通預報 (2017 雙十連假) 驗證

- Yann LeCun, Yoshua Bengio and Geoffrey Hinton (2015), Deep Learning.
- Yisheng Lv, Yanjie Duan, Wenwen Keng, Zengxi Li, Fei-Yue Wang (2015), Traffic Flow Prediction With Big Data: A Deep Learning Approach. IEEE. Vol. 16. No. 2. pp. 865-873.
- Manuel Fernandez-Delgado, Eva Cernadas, and Barro (2014), Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? Journal of Machine Learning Research.
- Xiaolei Ma, Zhuang Dai, et al. (2017), Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction. Sensors. Vol. 17. Issue 4.
- Atos Group. Real Time Traffic Forecast, Ascent white paper. November 2013.
- Nicholas Polson and Vadim Sokolov (2016), Deep Learning Predictors for Traffic Flows. Cornell University Library, arXiv: 1604.04527.
- Xiaolei Ma and Haiyang Yu (2015), Large-Scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory.
- Zilu Liang and Yasushi Wakahara (2014), Real-time urban traffic amount prediction models for dynamic route guidance systems.