運輸計劃季刊
第四十卷 第一期
民國一〇〇年三月
頁 1 ～ 頁 30

Transportation Planning Journal
Vo1. 40 No. 1
March 2011
PP. 1 ～ 30

兩步驟類神經網路車輛偵測器遺漏資料之填補及其應用

TWO-STAGE DATA IMPUTATION FOR MISSING VALUE OF VEHICLE DETECTORS AND ITS APPLICATIONS USING ARTIFICIAL NEURAL NETWORKS

吳健生 Jiann-Sheng Wu [1]
廖梓淋 Tzu-Lin Liao [2]
林鈺翔 Yu-Shiang Lin [3]

(99 年 4 月 13 日收稿，99 年 9 月 8 日第一次修改，
99 年 12 月 1 日第二次修改，100 年 3 月 22 日定稿)

摘 要

本研究採用兩步驟資料填補方式，針對雪山隧道路段車輛偵測器遺漏資料之填補進行實證分析，以期找出其中最為適用之填補方法，並發展其可能之應用。於資料填補時，首先採用 K-means 法將資料分群，而後再以最具代表性之三種類神經網路分別進行填補。測試結果發現，將資料分為兩群，並採用回饋式類神經網路進行填補時可獲得最高之填補績效。最後，依據填補績效發展兩種可能之應用，即遺漏資料填補及偵測器佈設間距。在遺漏資料填補方面，速率填補之績效最高，無論是以上、下游任何一對偵測器資料作為輸入，準確度均高達 97.5% 以上。其次為流率，其準確度可達 90% 以上，並可以上、下游 2 或 10 對偵測器資料作為群 1 或群 2 填

1. 國立中央大學土木工程學系副教授 (聯絡住址：32001 桃園縣中壢市中大路 300 號中央大學土木工程學系；聯絡電話：03-4227151 分機 34130；E-mail：wjs@ncu.edu.tw)。
2. 國立中央大學土木工程研究所碩士。
3. 國立中央大學土木工程研究所碩士。

補之輸入。最差者為占有率，僅當準確度門檻降至 80% 時，群 1 資料方能進行填補，群 2 資料則無此限制。在偵測器佈設間距方面，若合併考量流率、速率與占有率三者，則佈設間距由填補績效最差之占有率決定。僅在整體準確度降至 85% 以下時，方可將現行之 350 m 佈設間距擴增至 3,500 m。若僅考慮隨機性較低之群 2 資料，則在準確度高達 90% 以上時，即可將佈設間距增加至 4,200 m。

關鍵詞：兩步驟資料填補；K-means 法；回饋式類神經網路；填補績效；偵測器佈設間距

ABSTRACT

Using a two-stage data imputation method based on artificial neural networks, we carried out, in this study, an empirical analysis of the missing value of vehicle detectors in Hshehshan Tunnel to search for the optimal alternative, and developed its possible applications accordingly. By testing data imputation, we, at first, clustered all the data into groups using K-means, and then chose three typical artificial neural networks to impute the missing data. The result shows that two-group data clustering combined with a recurrent neural network can achieve the highest imputation performance. We, finally, developed two possible applications based on it, including data imputation and installation spacing of vehicle detectors. In respect to data imputation, speed performed the best with an accuracy of greater than 97.5%, and all pairs of vehicle detectors could be input for imputation. Flow performed the second best with an accuracy of over 90%, and the nearest two or ten pairs of detectors up- and downstream could be input for the imputation of data group 1 or 2, respectively. Occupancy performed the worst. Only by an accuracy threshold lowered to 80%, data points in group 1 could be imputed, and those in group 2 were not restricted, nevertheless. In respect to installation spacing, occupancy would dominate due to its relatively poor performance by considering all the three traffic attributes. Only when the overall accuracy decreased to fewer than 85% could we extend the current spacing of 350 m to 3,500 m. If only considering data group 2, we could extend it to 4,200 m with an accuracy of over 90% due to lower randomness.

Key Words: *Two-stage data imputation; K-means; Recurrent neural netwok; Imputation performance; Installation spacing of vehicle detectors*

一、前 言

近年來隨科技的快速發展，智慧型運輸系統 (intelligent transport systems, ITS) 已衍然成為先進交通管理技術不可或缺的一環。欲有效提供 ITS 之各項服務，必須先有準確且可

靠的交通資訊來源，而車輛偵測器 (vehicle detector, VD)，即為蒐集交通資訊最直接、有效的工具之一。通常 VD 多沿道路而佈設，其佈設之位置及密度則視實際需要而定。過去國內因較缺乏相關之應用分析與研究，對於 VD 之佈設在實務上多參考國外相關經驗或依據主觀判斷而定，缺乏一定的標準，致使效果不盡理想。至於相關之學術研究，則因囿於資料取得不易，多改採模擬方式進行。在眾多假設前提下，其結果是否仍具實用性，則有待進一步地驗證 (交通部[1])。

基本上，VD 用於直接蒐集定點之交通資料，至於路段沿線乃至於路網全面性之資料，則須藉助歷史經驗或相關理論模式加以推估。一般而言，VD 佈設數量愈多或愈密集，所蒐集到之資料亦愈為完整與可靠，但相對地成本亦愈高。然而縱使如此，由於受到外在環境與系統運作等因素的影響，VD 資料仍時有發生遺失、錯誤或傳輸失真等情形，不但會導致交通狀況判斷失真，同時亦容易誤導交管人員做出錯誤之決策，反而使交通益形惡化。因此，必須針對此一問題尋求彌補之方法，而資料填補 (data imputation) 即為其中最直接、有效的彌補方法之一。有鑑於此，本研究針對其中路段部分，探討 VD 資料填補之問題，並以雪山隧道為對象，就其 VD 遺漏資料 (missing data) 之填補進行實證分析，找出其中較為適用之填補方法，而後進一步就其可能之應用進行深入探討，以期解決雪隧路段 VD 資料遺漏填補之問題。

本研究在填補技術上延續先前研究之成果 (吳健生與廖梓淋[2])，採用兩階段資料填補之方式，先將資料加以分群，而後分別進行填補。第一階段係採用實證效果不錯之 K-means 法，將同質性高之 VD 資料進行不同群數之分群；第二階段則選擇最具代表性之三種類神經網路模式，即倒傳遞類神經網路 (back-propagation neural network, BPN)、輻狀基底函數類神經網路 (radial basis function, RBF) 以及回饋式類神經網路 (recurrent neural network, RNN) 進行測試與比較分析，並找出其中最適之網路模式與分群數。在填補輸入資料來源方面，有鑑於待填補 VD 資料與鄰近 VD 資料有直接相關性 (Chen 等人[3])，且為簡化實務操作之程序，故選擇同一時段內相鄰上、下游 VD 資料作為輸入進行填補。至於填補之內容，則以目前雪隧路段 VD 所蒐集到的即時資料 (real-time data) 之 5 分鐘流率、速率及占有率三項車流特性資料為主，暫不考慮其延伸之應用，例如事件自動偵測、旅行時間推估等。最後，依據以上分析之結果，規劃其直接可能之應用，包括遺漏資料填補及 VD 佈設間距兩種。希冀藉由繁複嚴謹之實證分析，提出雪隧路段 VD 資料遺漏填補之最適解決方案，並推估 VD 最大可能之佈設間距，使 VD 資源能夠發揮最大的功效。

二、文獻回顧

本研究係以遺漏資料之填補及其應用為主要重點，因此本節首先針對近年來國內、外相關理論與研究成果進行回顧與分析，並將其結果作一總結，供後續研究之參考。

Little 與 Rubin[4]將處理遺漏值的方式區分為忽略處理 (ignorance-based procedures)、

加權處理 (weighting-based procedures)、模式處理 (model-based procedures) 以及插補處理 (imputation-based procedures) 四種，其中插補處理為最典型之資料填補方式，係將遺漏資料透過各種方法，如熱層插補法、迴歸分析法、統計方法等進行填補。

Huang 與 Zhu [5] 利用兩向量加權相關程度發展出擬最鄰近法 (pseudo-nearest-neighbor approach)，將資料視為常態分配產生亂數，但並不確定交通資料是否符合常態分配。Huang 與 Lee[6]則利用灰關聯決定遺漏值最鄰近數列，再透過最鄰近法進行遺漏值填補，最後藉由不同測試範例及填補方法證實其不但可行且精確度高。

Wen 等人 [7] 沿用 Huang 與 Zhu [5] 及 Huang 與 Lee [6] 之概念，使用灰關聯擬最鄰近法 (grey-based nearest neighbor approach) 隨機產生遺漏值進行資料的填補，再使用遞迴式類神經網路 (recurrent neural network, RNN) 進行高速公路旅行時間之預測。結果當遺漏值產生率介於 50%~80% 時，誤差約為 20%，顯示填補後之資料用於預測旅行時間亦可獲得不錯的績效。

Gold 等人 [8] 使用期望最大值演算法 (expectation maximum algorithm，EM 演算法) 先將 VD 資料分群，而後選擇因子分析、內插法、多項式迴歸 (polynomial regression) 以及核迴歸 (kernel regression) 四種方法進行資料填補。結果發現，將期望最大值演算法結合核迴歸可獲得最佳之績效，其次則為與多項式迴歸相結合。

Vanajakshi 與 Rilett [9] 依據 Daganzo [10] 所提出之車輛守恆原則，亦即車輛行駛於道路上不會被創造或消失，檢視路段 VD 上、下游車數是否與此原則相符，並以一般簡化梯度法 (generalized reduced gradient) 求解非線性最佳化條件式，修正不符合邏輯之車流量，以降低 VD 錯誤之機率。然而，此一模式是以單一 VD 資料作為基礎進行推估，若 VD 本身出錯，則會在錯誤的基礎上繼續推估，並無判斷或修正之機制。

Chen 等人 [11] 結合兩種類神經網路填補遺漏值，第一階段使用自組織映射圖網路 (self-organizing map, SOM)，以 15 分鐘為單位輸入 VD 前 1 小時所測得之速率、流率及占有率資料，作為車流型態判別，第二階段則依不同車流型態建立車流量填補預測模式。填補遺漏值時，分別使用 ARIMA (auto-regressive integrated moving average) 與多層感知器 (multi-layer perception, MLP) 兩種模式進行預測及比較，結果發現 SOM/MLP 方法之預測精確度較高。

Zhong 等人 [12] 以通勤、休閒、長途及冬季休閒旅次為對象，針對不同道路型態之小時交通量進行遺漏值填補研究。首先透過遺傳演算法 (genetic algorithms, GAs) 選取 24 個最佳變數，作為迴歸、類神經網路以及 ARIMA 三種填補方法之輸入變數，而後利用前一星期與前後各一星期之小時交通量進行比較分析。結果顯示，使用前後各一星期小時交通量搭配迴歸方法所得到之績效，明顯優於類神經網路及 ARIMA 法。

Chen 等人 [3] 先根據 VD 所測得車流量與占有率兩者之關係，剔除不合邏輯之資料，而後再進行資料填補。於 VD 資料檢測方面，其誤判率僅 2.7%。至於資料填補，則認為與鄰近 VD 資料存在高度相關性，故將同一地點不同車道以及上、下游所有車道之 VD 視

為鄰近 VD，分別求得待填補與各鄰近點之迴歸關係式，最後取其中位數作為填補值。

張堂賢與黃宏仁[13]利用傅立葉轉換 (Transformée de Fourier) 與卡曼濾波器 (Kalman Filter)，發展 VD 遺漏資料之在線插補技術。卡曼濾波器用於推估遺漏資料發生稍早以前之資料走勢，插補作業則依資料本身型態進行插補，並採用大量歷史資料以符合統計學上不偏推定之要求，結果獲得良好之插補績效。

吳健生與廖梓淋[2]採用兩步驟方法進行 VD 遺漏值之填補，首先於第一步驟採用華德法 (Ward's method) 與 K-means 法進行兩階段之資料分群，而後再於第二步驟使用倒傳遞類神經網路 (back-propagation neural network, BPN) 進行資料填補，最後並以雪山隧道 VD 所測得之流率、速率及占有率三種資料進行實證分析，結果獲得不錯之填補績效。由於遺漏之資料係對稱採用上、下游 VD 資料進行填補，因此進一步根據填補所得績效值探討 VD 最大可能之佈設間距。

由以上文獻回顧結果可知，目前 VD 遺漏資料的填補方式可概分為兩個步驟，第一步驟為分群，方法包括灰關聯、EM 演算法、GA 演算法、SOM、華德法以及 K-means 法等，將相似度較高之資料加以群集；第二步驟則為填補遺漏資料，方法很多，包括擬最鄰近法、內插法、多項式迴歸、核迴歸、ARIMA、MLP 以及類神經網路等。上述各種方法均有其特性，且相互組合進行資料填補時，均能獲得一定之填補績效。

三、資料填補方法

綜合以上文獻回顧及總結，本研究決定參照目前發展之趨勢，並延續先前之研究，採用較為普遍且績效較佳之兩步驟方式進行 VD 資料遺漏填補之分析。於第一步驟分群方法上，選擇採用集群分析法進行資料的分群；於第二步驟則使用不同之類神經網路進行資料的填補。國內相關集群分析法曾應用於交通資料的分析並獲得不錯的結果，例如吳冠宏等人[14]將 SOM 與 K-means 兩種分群技術應用於交通事故資料，結果發現 K-means 具有較佳之分群能力。而吳健生與廖梓淋[2]則進一步運用 K-means 進行 VD 資料之分群，並將結果應用於遺漏資料之填補，結果亦證實其效果不錯。分群之優點在於，資料依不同車流狀態分類處理，同群資料之同質性高，變異相對較小，因此可獲得不錯的填補績效。

至於資料填補方面，由於類神經網路具有容錯的能力，且經證實具有不錯之填補績效，因此繼續採用其作為填補之方法，並取其中三種常用且具代表性之網路演算法，即：倒傳遞類神經網路 (back-propagation neural network, BPN)、輻狀基底函數類神經網路 (radial basis function, RBF) 以及回饋式類神經網路 (recurrent neural network, RNN)，進行測試與比較，其中 BPN 為最普遍採用之類神經網路演算模式，而 RBF 與 RNN 則分別代表靜態與動態類神經網路模式。以下分節說明各種方法在本研究之應用。

3.1 聚類演算法

集群分析係根據樣本特性之相似程度，將其化分成數個集群，使同一集群內之樣本具有高度同質性，而不同集群間之樣本則具有高度異質性。集群分析依分類方法的不同可分為階層式集群分析 (hierarchical cluster analysis) 與非階層式集群分析 (nonhierarchical cluster analysis) 兩大類，前者主要利用凝聚 (agglomerative) 與分離 (divisive) 之方式，將樣本個體視為一群，再將相近之個體合成一群，依次結合使群組數漸減，最終集結成一群；或經由相反程序，將所有個體由同一群中逐一分離成個別的群組。而非階層式集群分析則是先決定 k 個集群中心，以其作為起始之分群中心，再依各個體至各中心點距離的遠近，將其移動至最近的群體，並計算各群體之新中心點，然後再繼續移動各個體至最近之群體，直到群聚結果不再改變或滿足某種限制為止。

通常階層式集群分析適用於觀察值較少時，並以樹狀圖來選擇集群數目；而非階層式集群分析則須先選定 k 個集群，再利用疊代法找出最適之分群，一般較適用於觀察值多達 200 個以上時，其中最常用者為 K-means 集群分析法 (周文賢[15])。由於本研究所處理之資料量高達 1,800 筆以上，且經先前之研究 (吳健生與廖梓淋[2]) 亦證實 K-means 分群可獲得不錯的效果，故繼續延用 K-means 法進行資料的分群，其分群之步驟如下：

步驟 1： 由樣本資料中隨機選取 *k* 個資料點 *m1*(0)，*m2*(0)，*m3*(0)，…，*mk*(0)作為初始分群中心，而本研究係以每 5 分鐘之 VD 資料作為一個資料點。

步驟 2： 計算每一資料點與 *k* 個分群中心之歐幾里得距離 (Euclidean distance) 如式 (1)：

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{1}$$

其中 (x_i, y_i) 為資料點 *i* 之座標，(x_j, y_j) 為分群中心 *j* 之座標，而後依距離遠近將每一資料點分配到最近之分群中心。

步驟 3： 重新計算每一集群之重心位置，以其作為新的分群中心。

步驟 4： 重複步驟 2 與 3，直到所有資料點不再重新分配至其他集群為止。

3.2 類神經網路

類神經網路係利用大量人工神經元組合，模擬大腦神經細胞之運作。由高度連結之處理單元 (稱為節點或神經元，neuron) 組成動態運算系統，建構成網路架構來模擬人類之神經功能。類神經網路會透過訓練資料來修正網路參數，使整個網路模式符合訓練資料之特性，並將修正後之參數值儲存至網路中，藉由不斷地自我調整，使輸入資料透過神經元之運算，得到預設之輸出結果 (葉怡成[16])。如前所述，本研究擬採用 BPN、RBF 以及 RNN 三種常用且具代表性之類神經網路進行資料填補測試，並比較分析其填補之績效，以評選出其中何者最為適用。

為檢驗填補之績效，本研究將所蒐集到之 VD 資料分為兩組，其中一組取樣本數之80%作為訓練範例，供模式訓練學習之用；其餘 20%則作為驗證範例，供測試模式填補準確度之用。本研究採用平均誤差百分比絕對值 (mean absolute percent error, MAPE) 作為檢驗模式準確度之指標，其定義如式 (2)。

$$MAPE = \frac{1}{n-1}\sum_{i=2}^{n}\left|\frac{\hat{x}^{(0)}(i)-x^{(0)}(i)}{x^{(0)}(i)}\right|\times 100\% \tag{2}$$

其中 $\hat{x}^{(0)}(i)$ 為第 i 個填補值， $x^{(0)}(i)$ 為第 i 個實際偵測值，n 為測試樣本數。

MAPE 因不受量測值與估計值單位與大小的影響，故可客觀衡量填補值與實際偵測值間差異之程度。其值愈接近 0，代表填補之績效愈佳。資料填補之績效可依據 MAPE 值大小劃分如表 1 之四個等級 (Delurgio [17])。

表 1　填補績效分級

MAPE	填補績效表現
< 10%	高精確度填補
10% ~ 20%	良好填補
20% ~ 50%	合理填補
> 50%	不正確填補

3.2.1 倒傳遞類神經網路

BPN 演算法具有一輸入層、一輸出層以及零到無限多之隱藏層。學習過程中須給定期望輸出值，用以調整網路內之連結鍵值及隱藏層中各處理單元之輸出值。圖 1 所示 BPN 網路架構，係根據本研究之資料標的–流率、速率、占有率及車間距所建構，其內部輸入、輸出及隱藏三層結構之意涵與內容說明如下：

輸入層：表輸入變數，處理單元數視問題而定。本研究所採輸入變數為 VD 之流率、速率、占有率及車間距四種車流屬性，上、下游兩組 VD 共計 8 個處理單元作為輸入進行學習。

隱藏層：表輸入變數間之交互影響，主要作用為連接反映前一層與後一層間之互動關係。其處理單元數以多少為佳或需幾層隱藏層較為合適，目前尚無一定準則可以依循，一般多視問題特性與試驗方法而定，本研究則採常用之 S 形函數作為轉換函數如式 (3)：

$$F(x) = \frac{1}{1+\exp(-\alpha x)} \tag{3}$$

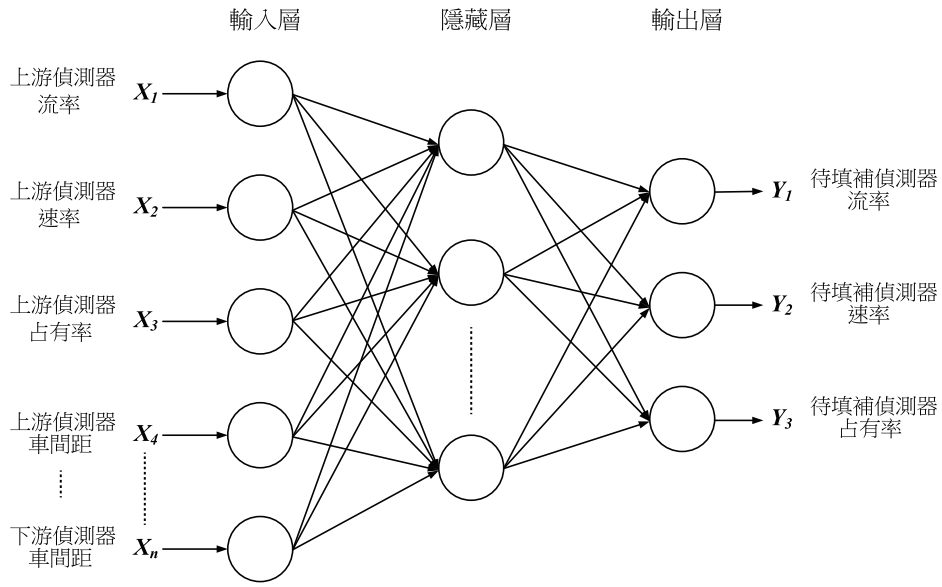


圖 1　倒傳遞類神經網路應用架構

其中 x 為輸入變數，α 則為轉換係數，通常為 0.5、1 或 2，本研究設定其值為 1。

輸出層：表輸出變數，處理單元數依問題而定。於本研究中，輸出層為待填補 VD 之流率、速率及占有率，共計 3 個處理單元。

函數信號：為完全連結前向式架構，每一層只接受前一層之輸出往前傳遞，而後出現在網路上之輸出層。

誤差信號：計算輸出層與對應範例原輸出值之誤差，再以倒傳遞方式將誤差大小進行權重調整。

3.2.2 輻狀基底函數類神經網路

RBF 類神經網路或稱半徑式類神經網路，其架構與多層感知器相同，具有輸入層、一層隱藏層以及輸出層，屬於基本前饋式類神經網路組合，但以函數逼近方式來建構網路(見圖 2)。RBF 之輸入層為輸入資料與網路連結之介面層，唯一隱藏層乃是將輸入資料經過非線性活化函數轉換而來，輸出層則是將隱藏層之輸出進行線性組合而獲得輸出值。

RBF 類神經網路其主要概念，乃是建構許多輻狀基底函數，並以函數逼近法 (curve fitting) 找出輸入與輸出間之映射關係，亦即在隱藏層之各神經元中建立相對應的輻狀基底函數。本研究採用高斯函數作為輻狀基底函數如式 (4)：

$$\Phi\left(\left\|x-c\right\|\right)=\exp\left(-\frac{\left\|x-c\right\|^2}{2\sigma^2}\right) \tag{4}$$

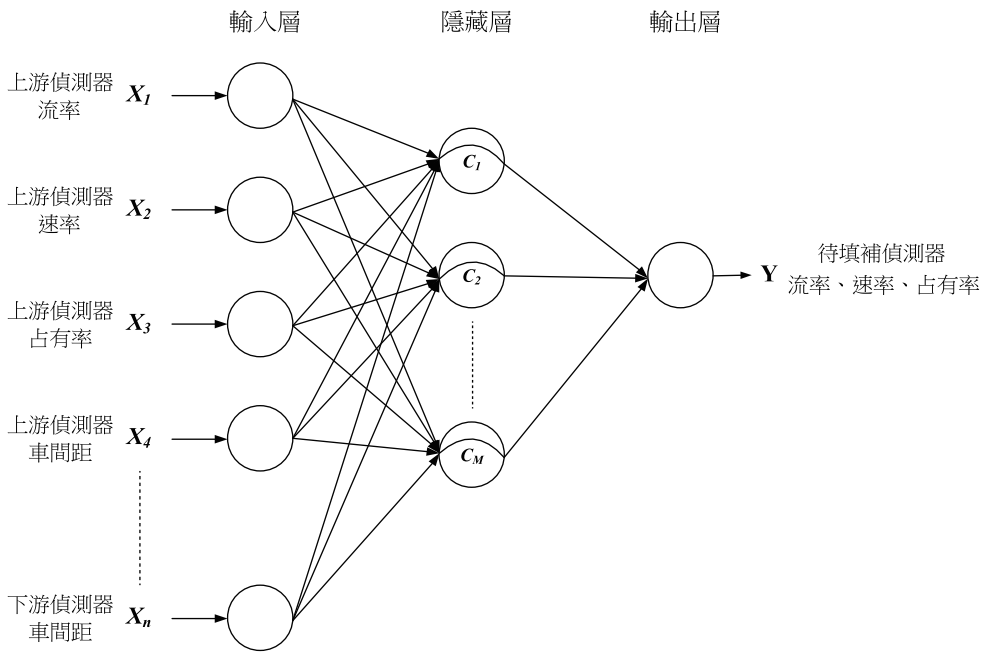


圖 2　輻狀基底函數類神經網路應用架構

其中 c 為群聚之中心點，σ 表兩最近群聚中心之距離。

建構 RBF 網路時，重點在於決定隱藏層大小與輻狀基底函數之中心點，然而並無特定方式決定隱藏層神經元之個數、每一神經元之中心點以及標準差或共變異。因此，一般採用複合式學習法，前階段以非監督式學習選取中心點，後階段則採監督式學習調整連結權重向量。中心點選取法大致可分為隨機性、聚類法 (clustering) 以及監督式選取三種，其中聚類法是將輸入資料依其相似程度聚為一類，例如 K-means 聚類法、fuzzy C-means 聚類法、模糊最小─最大分類法 (fuzzy min-max)、模糊減法聚類法 (fuzzy subtractive clustering)、山形聚類法 (mountain clustering) 等，本研究則是採用 K-means 聚類法進行中心點之選取。

如同 BPN，本方法應用於本研究中時，仍以上、下游兩組 VD 之流率、速率、占有率及車間距四種車流屬性，共計 8 個處理單元作為輸入變數，而輸出變數則為待填補 VD 之流率、速率及占有率，共計 3 個處理單元。

3.2.3 回饋式類神經網路

RNN 屬動態類神經網路 (dynamic neural network)，係以顯性表現方式，將時間因子直接以迴路方式表現在網路結構中。典型的作法為將隱藏層或輸出層神經元之輸出值回傳，作為下一階段本身或其他神經元之輸入訊息，亦即將現階段之訊息處理結果，藉由回傳方式保留在網路結構中，作為處理下一階段的參考。目前最常使用之動態類神經網路為

RNN，有別於傳統靜態 (static) 或前饋 (feed forward) 式網路架構，對人工神經元間之相互連接較無限制，允許神經元間相互回饋，不但更接近生物神經系統，亦因回饋項提供網路局部記憶 (或稱暫時記憶) 功能，故可用於模擬較複雜的系統。

RNN 通常含有一或多個回饋迴圈，其基本架構為：(1) 由輸出層或隱藏層回饋至輸入層；(2) 層內各處理單元間有連結存在；(3) 神經元不分層排列 (僅有一層)，各神經元均可相互連結 (見圖 3)。RNN 最常採用即時回饋學習演算法 (real-time recurrent learning) 作為其演算的方法，其名稱源自於網路連結權重向量的調整為即時進行，亦即輸入一新的資訊後，網路除持續執行訊息程序功能外，並依其誤差情形逐時更新各連結權重向量。

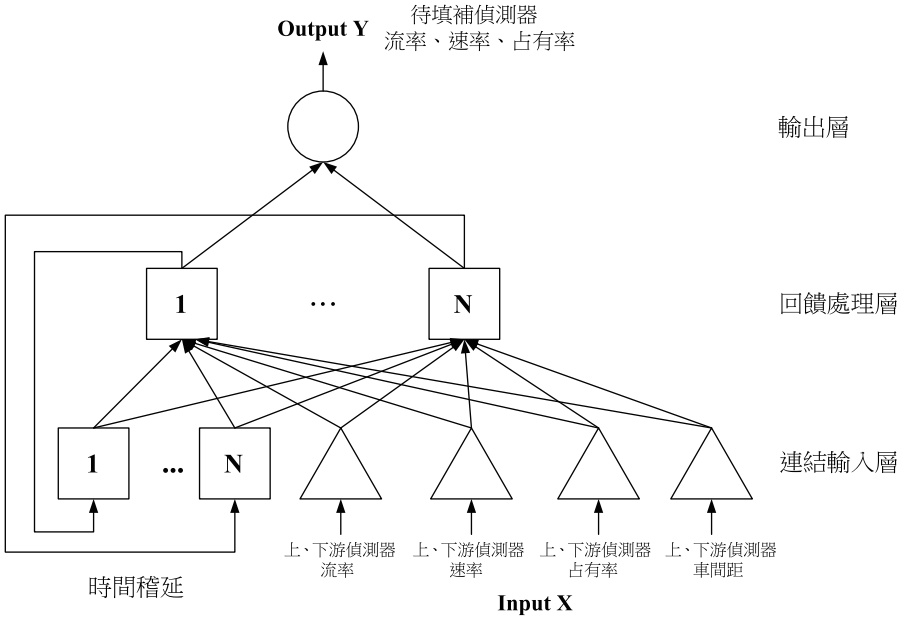


圖 3　回饋式類神經網路應用架構

於本研究中，RNN 雖仍以上、下游兩組 VD 之流率、速率、占有率及車間距四種車流屬性作為輸入變數，但因具有回饋性，多一倍特性供學習，故輸入層共有 16 個處理單元。至於輸出層，則仍維持為待填補 VD 之流率、速率及占有率，共 3 個處理單元。

四、實證分析

4.1 資料蒐集

由於待填補 VD 資料與鄰近 VD 資料具有直接相關性 (Chen 等人[3])，故選擇同一時

段內相鄰上、下游 VD 資料作為輸入進行填補，並逐步向外對稱延伸，以分析上述三種填補方法之績效。經仔細考量之後，決定選擇國道 5 號雪山隧道南下路段作為本研究實證分析之對象，以取其 VD 等距密集佈設可取得完整資料之利。資料內容為隧道內 VD 所測得之交通特性資料，包括流率、平均速率、占有率、車間距、車長以及車種。雪山隧道全長 12.9 km，隧道內約每隔 350 m 佈設感應線圈式車輛偵測器 (Inductive loop detectors) 一組，共計佈設 35 組，表 2 為各 VD 佈設之編號及位置。本系統之 VD 每分鐘登錄交通特性資料一次，累積至 5 分鐘後再回傳至行控中心，故本研究所使用之交通特性資料均以 5 分鐘作為單位，例如流率為 5 分鐘內通過之車輛數，速率為 5 分鐘內通過車輛之平均速率，餘此類推。採用雪隧路段 VD 資料的主要原因為，該路段之 VD 採高密度等間距佈設，資料完整且不受天候與聯結車 (禁止進入) 的影響，分析時可由上、下游兩端等距對稱擷取 VD 資料作為輸入進行填補。

為使填補結果具有周延性，本研究特選取完整之一星期時間作為實證之對象，故資料內容包含平日與假日及其離、尖峰不同之交通型態，且經查證該時段內無意外事故發生。實證資料之範圍為民國 97 年 3 月 25 日至 31 日共 7 天所蒐集到之 VD 資料，以路段雙車道加總之流率、平均速率及占有率作為待填補屬性。因 VD 每 5 分鐘更新資料一次，故 7 天共蒐集到 2,016 筆資料。經事前分析研判刪除異常資料後，共計約有 1,800 筆資料可供模式校估使用，而異常資料之判定係根據統計經驗法則，將 3 個標準差以外之極端值予以刪除。

表 2　雪山隧道車輛偵測器佈設編號及位置

代號	VD 編號	代號	VD 編號	代號	VD 編號
1	VD-5S-SST-15.856	13	VD-5S-SST-20.063	25	VD-5S-SST-24.264
2	VD-5S-SST-16.201	14	VD-5S-SST-20.413	26	VD-5S-SST-24.678
3	VD-5S-SST-16.551	15	VD-5S-SST-20.763	27	VD-5S-SST-24.962
4	VD-5S-SST-P-16.902	16	VD-5S-SST-P-21.063	28	VD-5S-SST-P-25.312
5	VD-5S-SST-17.253	17	VD-5S-SST-21.460	29	VD-5S-SST-25.664
6	VD-5S-SST-17.608	18	VD-5S-SST-21.807	30	VD-5S-SST-26.013
7	VD-5S-SST-17.990	19	VD-5S-SST-22.159	31	VD-5S-SST-26.299
8	VD-5S-SST-P-18.312	20	VD-5S-SST-P-22.506	32	VD-5S-SST-P-26.706
9	VD-5S-SST-18.663	21	VD-5S-SST-22.859	33	VD-5S-SST-27.054
10	VD-5S-SST-19.013	22	VD-5S-SST-23.207	34	VD-5S-SST-27.442
11	VD-5S-SST-19.363	23	VD-5S-SST-23.560	35	VD-5S-SST-27.748
12	VD-5S-SST-P-19.677	24	VD-5S-SST-P-23.910		

實證分析時，為便於上、下游對稱擷取 VD 資料作為填補輸入，故假設隧道正中央編號 18 之 VD (VD-5S-SST-21.807) 其資料遺漏有待填補。填補時，先取上、下游最鄰近兩組 VD (編號 17 與 19) 資料作為輸入，之後續往上、下游各推前一組 VD (編號 16 與 20) 取其資料作為輸入，餘此類推，直至推進至隧道出入口之 VD 為止 (見圖 4)。運算時，先將上、下游 VD 所測得之流率、速率、占有率及車間距四種屬性資料進行分群，其次再將分群後之資料彙集，並分別以上述三種類神經網路進行學習。其輸入層同為上、下游 VD 之流率、速率、占有率及車間距四種屬性，輸出層則為待填補 VD 之流率、速率及占有率，三者均為交通管理與控制所必備之基本交通特性資料。最後，再使用 MAPE 值來檢驗並比較三種填補方法之績效。

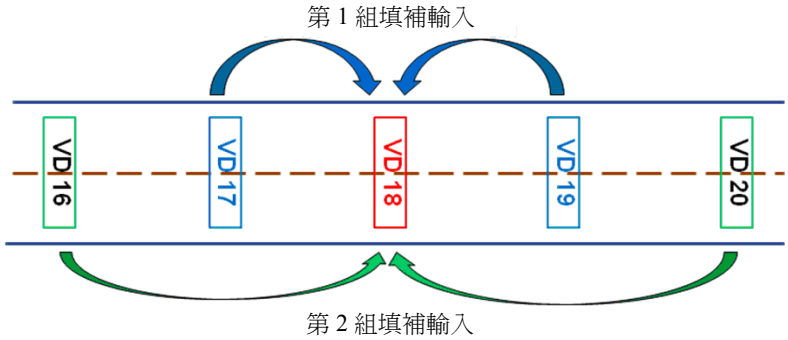


圖 4　資料填補輸入偵測器示意圖

4.2 資料分群

本研究使用 K-means 演算法先將 VD 資料進行分群，並設定群數為 2 至 8 群，分別進行測試。因分群之依據為車流屬性資料，故可視為服務水準之外交通狀況之另一種分類。分群之輸入變數為流率、速率、占有率及車間距四種屬性，表 3 為各屬性資料之相關統計參數值。此四種屬性相互對應可用來共同描述相同之車流狀態，故採用屬性分群即表示依車流狀態加以分類。本研究直接採用完整原始資料進行分群，主要著眼於應用之便利性。原因之一為，資料之蒐集為一時間數列，運作上若再區分平、假日及離、尖峰，將會產生時段切分之問題。原因之二為，正常狀況下車流量之變化均為循序漸進，因此如何界定離、尖峰以擷取資料，恐將產生困難。原因之三則為，離、尖峰轉換間之過渡時期亦須加以處理，易增問題之複雜性。若純依屬性進行分群，不但可避免以上之困擾，同時亦可兼顧離、尖峰及其轉換期間所有時段之各種車流狀態。

由於使用 K-means 進行分群時須事先決定群數，而非藉由資料本身型態加以決定，因此本研究參考 Sgarma [18] 之建議，同時採取階層式與非階層式兩種方式進行集群。第一階

段先以華德法 (Ward's method) 決定群數，第二階段再以 K-means 進行集群。若純以分群的角度來看，2 群為最佳集群數，因其演算次數最少，且收斂最快。然而，本研究之目的為資料填補，應依據填補績效決定最佳分群數，而非以運算績效作為標準，故採強制分群方式，將偵測資料分為 2 至 8 群，分別比較其資料填補之績效。

表 3　雪山隧道車輛偵測器資料統計分析

統　計 參　數	流　率 (veh/5-min)	速　率 (km/h)	占有率 (%)	車間距 (m/veh)
平均數	78.61	75.36	6.17	157.66
眾　數	87	75	1	70
標準差	45.3843	4.0220	4.1829	148.5363
峰　度	0.2113	2.0084	1.9178	3.3093
偏　態	0.6122	-0.7072	1.1858	1.9464
最小值	3	51	1	24
最大值	242	88	29	800

4.3 資料填補

4.3.1 填補績效

K-means 分群完成之後，接著分別採用 BPN、RBF 以及 RNN 三種類神經演算方法進行資料之填補，並比較其填補之績效。填補時輸入層為上、下游 VD 之流率、速率、占有率及車間距四種屬性資料，輸出層則為待填補 VD 之流率、速率及占有率三種屬性。演算時，取總樣本數之 80%作為學習範例，其餘 20%作為驗證範例。為便於相互比較，各群所得 MAPE 值將依其樣本大小予以加權平均。舉分群數 2 為例，假設群 1 與群 2 之樣本數比為 6：4，則分別給予兩者 MAPE 值 0.6 與 0.4 之權重，而後加總得出總 MAPE 值。

每一填補模式經重複測試 5 次之後，取其平均值得如表 4 之結果。根據表 4 之數據，將三種屬性填補結果分別繪如圖 5 至 7。由圖可以看出，就流率而言，除分群數 8 之外，RBF 之 MAPE 值明顯大於 BPN 與 RNN，而 BPN 與 RNN 之間則相差不大。就速率而言，BPN 與 RNN 兩者之 MAPE 值依舊相差不大，而 RBF 之 MAPE 值則明顯大於其他二者。另就占有率而言，三種類神經網路之中，BPN 與 RNN 之 MAPE 值仍相差不大；至於 RBF，則除 6 群以上之高群數外，其餘均較 BPN 與 RNN 為高。整體而言，速率填補之準確性最高，屬高精確度，其次為流率，最差者為占有率 (如圖 8 所示)。

表 4　各填補模式填補 MAPE 值

分群數	BPN			RBF			RNN		
	流 率	速 率	占有率	流 率	速 率	占有率	流 率	速 率	占有率
2	7.23%	1.04%	14.23%	13.30%	3.30%	19.36%	7.39%	1.08%	12.59%
3	7.25%	1.07%	13.39%	11.20%	1.80%	15.94%	7.11%	1.00%	13.80%
4	7.15%	1.09%	13.81%	9.60%	1.90%	14.31%	7.40%	0.96%	12.66%
5	7.50%	1.12%	12.81%	11.00%	1.90%	14.80%	6.82%	1.28%	12.63%
6	7.82%	1.08%	13.10%	9.60%	1.70%	13.70%	7.62%	1.22%	12.12%
7	7.11%	1.12%	15.12%	8.70%	1.70%	13.96%	7.86%	1.06%	13.56%
8	7.28%	1.03%	14.29%	8.50%	1.70%	13.43%	8.22%	1.04%	14.15%

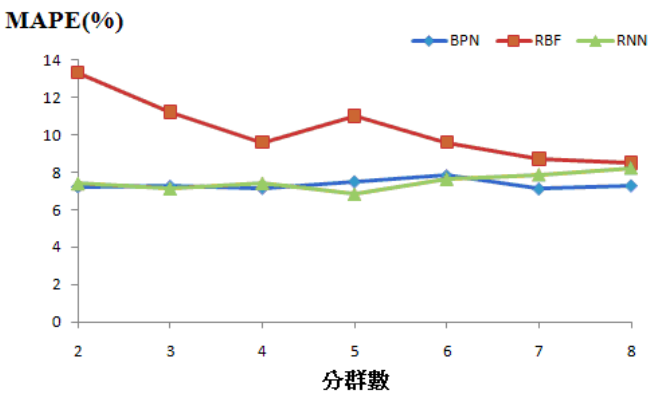


圖 5　流率填補績效

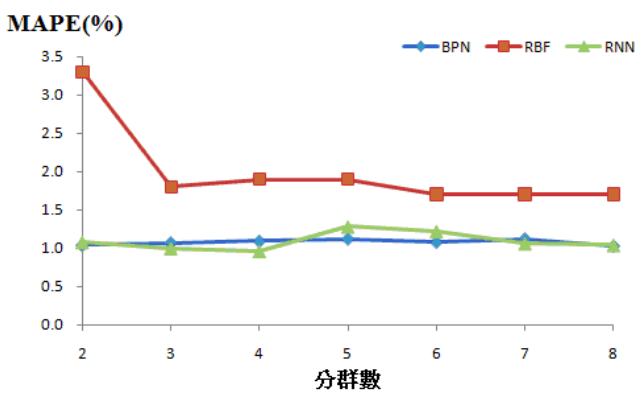


圖 6　速率填補績效

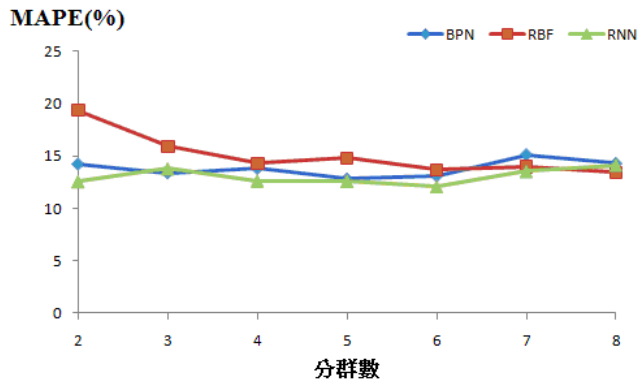


圖 7　占有率填補績效

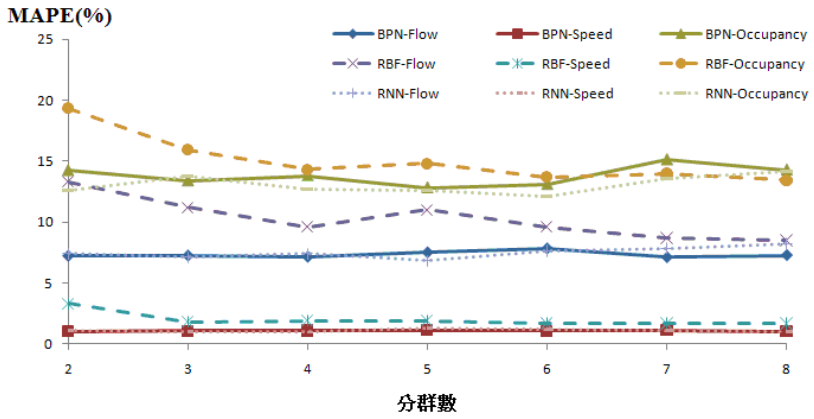


圖 8　各型類神經網路填補績效比較

4.3.2 最適分群數

為了解分群多寡對填補績效之影響，本研究進一步採用變異數分析 (analysis of variance, ANOVA) 針對不同群數之 MAPE 值進行同異檢定，其統計假設如下：

H_0：$\mu_2 = \mu_3 = \mu_4 = \ldots = \mu_8$

H_1：至少兩 μ_i 不相等，$i = 2, 3, \ldots, 8$.

其中 μ_i 表分群數為 i 時之 MAPE 平均值。應用表 4 之數據，分別就流率、速率及占有率進行檢定，檢定結果如表 5 至 7 所示。由表可知，當顯著水準 $\alpha = 0.05$ 時，三者均未呈現顯著性，無法駁斥 H_0 之假設，亦即不同群數間之填補績效無顯著差異。換言之，無論分為幾群，對三種類神經填補模式之績效均無顯著的影響。然而，若就運算績效及實用便利性而言，自以群數愈少愈好，因此於後續應用分析中，決定採用兩群方式進行資料之填補。

4.3.3 最適填補方法

由以上分析得知，分群數之多寡並不影響資料填補之績效。然而，不同類神經網路之間其填補績效是否存在差異，則有待進一步地確認。為比較不同類神經網路間之填補績效，本研究採取兩兩比對的方式，先將三種網路兩兩組合進行統計檢定決定其優劣，而後再將比對結果彙整進行總排序，最後將流率、速率及占有率三種填補績效之排名予以加總，總名次最低者判定其為最適填補模式，並將其應用於後續之分析。由於測試時均採用相同之樣本，因此適合以成對母體平均數差異檢定 (paired t-test) 進行兩兩組合績效之比對，以下為其檢定之方法。

表 5　流率分群填補 MAPE 值變異數分析

變異來源	平方和	自由度	均方和	F 值	P-值	$f_{0.05}(6,14)$
群　間	4.1244	6	0.6874	0.1820	0.9772	2.8477
群　內	52.8713	14	3.7765			
總　和	56.9957	20				

表 6　速率分群填補 MAPE 值變異數分析

變異來源	平方和	自由度	均方和	F 值	P-值	$f_{0.05}(6,14)$
群　間	0.6636	6	0.1106	0.2893	0.9323	2.8477
群　內	5.3530	14	0.3824			
總　和	6.0166	20				

表 7　占有率分群填補 MAPE 值變異數分析

變異來源	平方和	自由度	均方和	F 值	P-值	$f_{0.05}(6,14)$
群　間	25.2458	6	4.2076	0.5760	0.7435	2.8477
群　內	102.2645	14	7.3046			
總　和	127.5103	20				

假設 x_{1i} 與 x_{2i} 分別為填補方法 1 與 2 在第 i 種實驗下，即第 i 種分群數下所得之 MAPE 值。令

$$d_i = x_{1i} - x_{2i}, \quad i = 1, 2, 3, \ldots, n$$

則 d_i 之平均數為：

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} \tag{5}$$

標準差為：

$$s_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \bar{d})^2}{n-1}} \tag{6}$$

其中 n 為分群種類數，於本研究中分群數是由 2 至 8 共 7 種，故 $n = 7$。檢定時，先確認兩種填補方法之 MAPE 值是否相同，其統計假設如下：

H_0：$\mu_d = 0$ 或 $\mu_1 = \mu_2$

H_1：$\mu_d \neq 0$ 或 $\mu_1 \neq \mu_2$

其中μ_d、μ_1及μ_2分別為變數 d_i及填補方法 1 與 2 之母體平均數。檢定時所採用之統計量如下：

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \tag{7}$$

在顯著水準為α之下，若$|t| \le t_{\alpha/2,n-1}$，其中 n-1 為自由度，則接受 H_0，否則進一步檢定何者 MAPE 值較高，MAPE 值較高一方表示預測結果與實際狀況差異較大，即填補績效較差。

檢定所採用之統計量同樣為式 (7) 之 t。若 $t > 0$，表示 μ_1 可能大於 μ_2，故其對立假設為：

H_1：$\mu_d > 0$ 或 $\mu_1 > \mu_2$

若$t > t_{\alpha,n-1}$，則拒絕 H_0，表示 μ_d明顯大於 0，或 μ_1明顯大於 μ_2。反之，若 $t < 0$，表示 μ_1 可能小於 μ_2，故其對立假設為：

H_1： $\mu_d < 0$ 或 $\mu_1 < \mu_2$

若$t < -t_{\alpha,n-1}$，則拒絕 H_0，表示 μ_d明顯小於 0，或 μ_1明顯小於 μ_2。

表 8 為流率填補成對檢定之結果，由表可知，在 $\alpha = 0.05$ 之下，RBF 之 MAPE 值均明顯高於 BPN 與 RNN，而 BPN 與 RNN 之間則無顯著的差異。因此在流率填補績效方面，BPN 與 RNN 兩者相同，且均明顯優於 RBF。進一步成對檢定速率填補之績效，其結果如同流率，即在 $\alpha = 0.05$ 之下，BPN 與 RNN 並無顯著的差異，但均明顯優於 RBF (見表 9)。最後，針對占有率之填補績效進行成對檢定，結果如表 10 所示，在 $\alpha = 0.05$ 之下，僅 BPN 與 RNN 之填補績效有顯著的差異。進一步由單尾檢定得知，$t = 2.52081 > t_{0.05,6} = 1.94318$，表示 BPN 之 MAPE 值明顯大於 RNN，亦即 RNN 占有率之填補績效明顯優於 BPN。綜合而言，RBF 與其他二者均無顯著的差異，RNN 與 RBF 雖無顯著的差異，但明顯優於 BPN；而 BPN 雖與 RBF 無顯著的差異，但較 RNN 為差。因此就占有率之填補績效而言，RNN 最佳，RBF 次之，BPN 則最差。

表 8　BPN、RBF、RNN 流率填補 MAPE 值成對檢定

	流 率					
	BPN	RBF	BPN	RNN	RBF	RNN
平均數	0.07334	0.10271	0.07334	0.07489	0.10271	0.07489
變異數	0.00001	0.00028	0.00001	0.00002	0.00028	0.00002
樣本數	7		7		7	
自由度	6		6		6	
t 統計量	–4.55057		–0.72739		3.64637	
P 值 (雙尾)	0.00389		0.49438		0.01075	
$t_{0.025,6}$ (雙尾)	2.44691					
P 值 (單尾)	0.00194		0.24719		0.00538	
$t_{0.05,6}$ (單尾)	1.94318					

表 9　BPN、RBF、RNN 速率填補 MAPE 值成對檢定

	速 率					
	BPN	RBF	BPN	RNN	RBF	RNN
平均數	0.01079	0.02000	0.01079	0.01091	0.02000	0.01091
變異數	0.00000	0.00003	0.00000	0.00000	0.00003	0.00000
樣本數	7		7		7	
自由度	6		6		6	
t 統計量	–4.09021		–0.31239		4.03124	
P 值 (雙尾)	0.00643		0.76531		0.00687	
$t_{0.025,6}$ (雙尾)	2.44691					
P 值 (單尾)	0.00321		0.38266		0.00344	
$t_{0.05,6}$ (單尾)	1.94318					

表 10　BPN、RBF、RNN 占有率填補 MAPE 值成對檢定

	占 有 率					
	BPN	RBF	BPN	RNN	RBF	RNN
平均數	0.13821	0.15071	0.13821	0.13073	0.15071	0.13073
變異數	0.00006	0.00043	0.00006	0.00006	0.00043	0.00006
樣本數	7		7		7	
自由度	6		6		6	
t 統計量	–1.51645		2.52081		2.25269	
P 值 (雙尾)	0.18019		0.04524		0.06520	
$t_{0.025,6}$ (雙尾)	2.44691					
P 值 (單尾)	0.09010		0.02262		0.03260	
$t_{0.05,6}$ (單尾)	1.94318					

綜合以上檢定之結果，分別給予個別屬性填補績效排序，最優者排名為 1，次優者為 2，最差者則為 3。若排序相同，則給予相同之排名，例如 BPN 與 RNN 相同，且均優於為 RBF，故二者排名同為 1，RBF 則為 3。排名結果如表 11 所示，由表可知，無論是就流率、速率或占有率而言，RNN 均為最優排名第 1，明顯優於 BPN 與 RBF，因此以下繼續採用其作為應用分析之方法。

表 11　填補績效排序

填補屬性 / 填補方法	流 率	速 率	占有率
BPN	1	1	3
RBF	3	3	2
RNN	1	1	1

五、應用分析

5.1 實驗設計

由以上實證分析結果得知，無論是分群數或填補方法均可先行收斂，以簡化後續應用分析之程序。首先於第一步驟資料分群時，考量運算之速度與應用之便利性，選擇最少之兩群作為分群數。其次於第二步驟資料填補時，採用三種類神經網路中最具績效之 RNN 作為填補方法。資料分群係採用 K-means 法將資料分為兩群，群 1 之流率分群中心為 50 veh/5-min，群 2 則為 122 veh/5-min，兩者約以 75 veh/5-min 作為分界點，故群 1 可視為中低流量狀態，群 2 則代表中高流量狀態。

與實證分析相同，實驗所採資料為民國 97 年 3 月 25 日至 31 日雪山隧道內 7 天所蒐集到之 1,800 筆 VD 資料，並假設隧道正中央編號 18 之 VD 資料發生遺漏有待填補。填補時先擷取同一時段內相鄰上、下游 VD 資料作為輸入，並逐步向外對稱延伸，直至隧道出入口為止，同時並分析流率、速率及占有率三種屬性之填補績效，最後再依據填補績效進行遺漏資料填補與 VD 佈設間距兩種應用分析。

5.2 績效分析

依據上述原則及步驟，分別就雪山隧道路段之流率、速率及占有率三種車流屬性進行訓練與測試，以下依序探討其填補之績效：

(一) 流率

流率填補結果其績效如圖 9 所示，圖中橫座標表不同 VD 之輸入組合，$(i-j)+(i+j)$ 表

示 VD*i* 所遺漏之資料由其上、下游第 *j* 個 VD 資料組合輸入進行填補。為了解分群與否對填補績效之影響，本研究另行將原始資料不經分群直接進行訓練，而後再以相同的樣本進行測試，以比較兩者是否有差異。整體而言，MAPE 值隨輸入 VD 距離之增長而有逐漸增加之趨勢，符合距離愈近車流屬性愈為接近之一般通則。若就分群資料來看，群 1 之填補績效明顯低於群 2，其主要原因在於，群 1 屬中低流量狀態，車輛到達隨機性高，較難掌控；群 2 則屬中高流量，隨機性較低，相對較易掌控。若以 95% 準確度 (MAPE 值 5%) 為門檻，則群 2 資料可由其上、下游 3 組 VD 中之任一對組合進行填補，而群 1 資料之填補則完全無法達到此一要求。若以 90%準確度 (MAPE 值 10%) 作為門檻，則用以填補群 2 資料之 VD 更可溯至上、下游 10 組 VD。反觀群 1，如欲維持 90%之準確度，則僅能取上、下游 2 組 VD 中之任一對組合進行填補；若將準確度降至 85% (MAPE 值 15%)，則可延伸至上、下游 10 組 VD；而依據表 1 所列標準，85%準確度仍屬良好之填補。

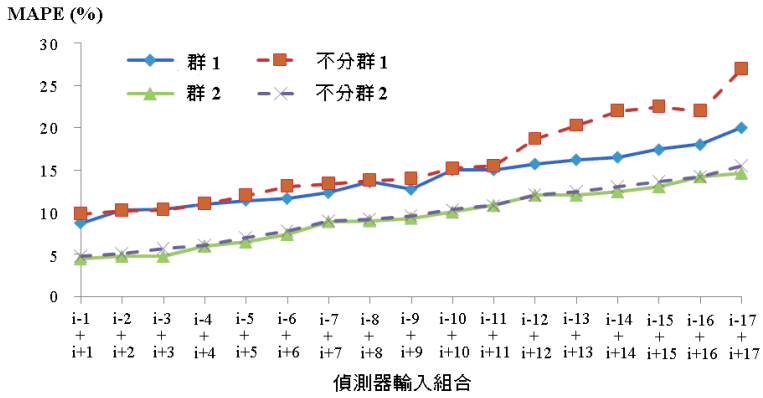


圖 9　兩群流率填補績效

除群體間相互比較之外，本研究並將上述群 1 與群 2 資料，分別代入未分群之類神經網路中進行測試，並比較分群與不分群間之差異。經成對母體平均數差異檢定後得知，在 $\alpha = 0.05$ 之下，兩者均呈現顯著的差異，顯示資料分群後在流率填補方面可獲得較佳的績效 (見表 12)。

(二) 速率

速率填補之結果如圖 10 所示，MAPE 值同樣隨輸入 VD 距離之增長而增加，但結果均相當準確，僅介於 1% 與 2.5% 之間，已達表 1 高精確度填補之等級。由於速率填補之績效均相當高，且上下變動之幅度亦不大，因此無論是群與群之間，抑或是分群與不分群之間，皆難以顯著區別其差異 (見表 13)，故速率遺漏時，縱使不經分群亦可獲得不錯之填補績效。

表 12　分群與不分群流率填補績效成對檢定

	群 1	不分群 1	群 2	不分群 2
平均數	13.8706%	15.9294%	9.4235%	9.7706%
變異數	0.001002	0.002696	0.001116	0.001124
樣本數	17		17	
自由度	16		16	
t 統計量	–3.747846		–5.054565	
P 值 (雙尾)	0.001756*		0.000117*	
$t_{0.025,16}$ (雙尾)	2.119905			
P 值 (單尾)	0.000878*		0.000059*	
$t_{0.05,16}$ (單尾)	1.745884			

註：*表示呈現顯著差異。

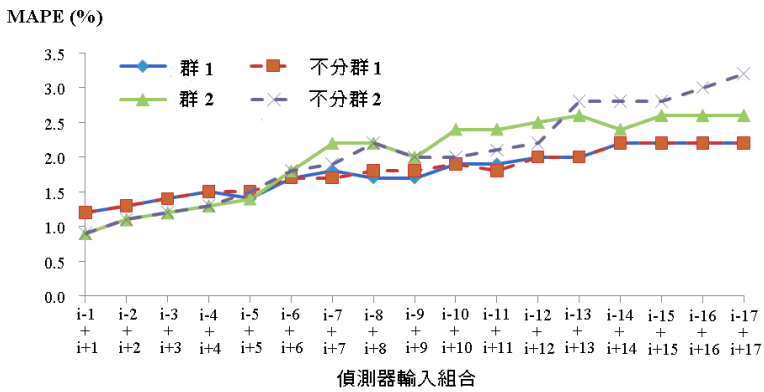


圖 10　兩群速率填補績效

表 13　分群與不分群速率填補績效成對檢定

	群 1	不分群 1	群 2	不分群 2
平均數	1.7824%	1.7882%	2.0118%	2.0471%
變異數	0.000011	0.000011	0.000036	0.000049
樣本數	17		17	
自由度	16		16	
t 統計量	–0.436436		–0.536120	
P (雙尾)	0.668353		0.599250	
$t_{0.025,16}$ (雙尾)	2.119905			
P (單尾)	0.334176		0.299625	
$t_{0.05,16}$ (單尾)	1.745884			

(三) 占有率

在占有率填補方面，其結果如圖 11 所示，就分群資料來看，群 1 同樣因屬中低流量狀態，隨機性高，故其填補績效明顯低於群 2，且準確度均在 85%以下 (MAPE 值 15%以上)，而群 2 則可高達 95% (MAPE 值 5%)。若以 80%準確度 (良好填補，見表 1) 作為門檻，則群 1 資料可由其上、下游 5 組 VD，而群 2 資料更可溯至上、下游所有 VD 進行填補。若提高準確度門檻至 85%及 90% (MAPE 值 15%及 10%)，則群 2 資料仍可由上、下游 13 組及 6 組 VD 進行填補。若再提高門檻至 95% (MAPE 值 5%)，則僅剩上、下游最鄰近 VD 符合要求。進一步比較分群與不分群間填補之績效，經成對母體平均數差異檢定後發現，在 α=0.05 之下，資料分群後在占有率填補方面可得到較佳之績效，如表 14 所示。

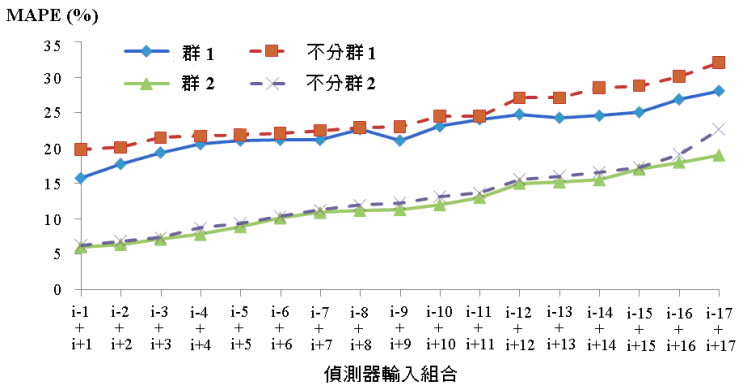


圖 11　兩群占有率填補績效

表 14　分群與不分群占有率填補績效成對檢定

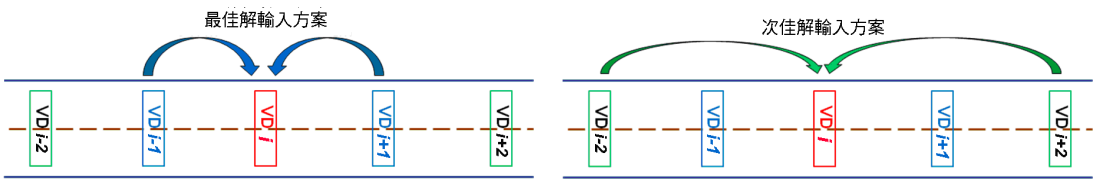
	群 1	不分群 1	群 2	不分群 2
平均數	22.3647%	24.5000%	12.0235%	12.7882%
變異數	0.001005	0.001354	0.001669	0.002091
樣本數	17		17	
自由度	16		16	
t 統計量	−6.830843		−3.956919	
P (雙尾)	0.000004*		0.001130*	
$t_{0.025,16}$ (雙尾)	2.119905			
P (單尾)	0.000002*		0.000565*	
$t_{0.05,16}$ (單尾)	1.745884			

註：*表示呈現顯著差異。

5.3 遺漏資料填補

本應用為資料填補最基本之目的，即將遺漏之資料透過演算方法，在一定誤差範圍內將其回復。選擇最佳填補輸入方案時，其基本程序如下：首先測試各輸入資料方案之填補績效 (MAPE 值)，其次界定容許之 MAPE 值範圍，並據以找出上述方案中何者可以接受，亦即何種方案為可行解 (feasible solution)，最後將可行解依 MAPE 值大小進行優劣排序。當資料發生遺漏時，優先選擇最佳解，即 MAPE 值最小之方案進行填補。若最佳輸入方案中之 VD 亦發生資料遺漏，則退而求其次選擇次佳解替代，餘此類推。

以圖 12 為例，假設任一 VD_i 發生資料遺漏，經測試分析後得知，在容許誤差範圍內共可求得 2 組可行解，其中最佳解為上、下游最鄰近 VD 之輸入組合 $(VD_{i-1}+VD_{i+1})$，如圖 12(a) 所示；其次則為上、下游次鄰近 VD 之輸入組合 $(VD_{i-2}+VD_{i+2})$，如圖 12(b) 所示。當 VD_i 資料實際發生遺漏時，優先選擇 MAPE 值最小之最佳輸入方案，即 $(VD_{i-1}+VD_{i+1})$ 組合進行填補。若 VD_{i-1} 或 VD_{i+1} 中有任何一組資料亦發生遺漏，則退而求其次改採次佳解，即上、下游次鄰近二組 VD 之輸入組合 $(VD_{i-2}+VD_{i+2})$ 進行填補。



(a) 可行解一：最佳輸入方案　　(b) 可行解二：次佳輸入方案

圖 12　可行解輸入方案

以下依據上述原則及方法，分別就雪山隧道路段流率、速率及占有率三項車流屬性之填補績效及其填補應用依序探討之。

1. 流率：如前所述，流率填補績效隨輸入 VD 距離之增加而漸減。若以 95% 準確度作門檻，則群 1 資料完全無法達到此一要求，故無填補可行解；而群 2 資料則可上、下溯至 3 組 VD 作為可行解。若將準確度門檻設定為 90%，則群 1 與群 2 之填補可行解可分別上、下溯至 2 及 10 組 VD；若調降準確度門檻至 85%，則所有上、下游 VD 均可滿足群 2 資料填補之需要，而群 1 之填補可行解則可延伸至上、下游 10 組 VD；若進一步調降門檻至 80%，則無論群 1 或群 2 資料，均可由所有上、下游 VD 之組合進行填補。
2. 速率：由於速率填補之準確度無論群 1 或群 2 均高達 97.5% 以上，能完全滿足上述門檻之要求，故其可行解為上、下游所有 VD 之組合。
3. 占有率：占有率填補之績效仍隨輸入 VD 距離之增加而漸減，但就分群資料來看，群 1 填補之準確度均在 85% 以下。若降至 80%，則可上、下溯至 5 組 VD 作為可行解。群 2

填補則有較高之績效，準確度門檻設為 95% 時，其填補可行解為上、下游 1 組 VD；若降低門檻至 90 及 85%，則填補可行解可分別上、下溯至 6 及 13 組 VD；若進一步調降至 80%，則上、下游所有 VD 組合均可作為其可行解。

綜合以上之分析，將其結果彙整如表 15，應用時，可依各屬性之需求分別填補之。例如填補流率時，群 1 資料無法達到 95% 準確度之要求，必須降低門檻至 90%，並優先選取上、下游最鄰近 1 組 VD 資料作為輸入。若此組 VD 資料亦發生遺漏，則退而求其次，取上、下游次鄰近 1 組之 VD 資料作為輸入。萬一兩組 VD 資料均發生遺漏，則必須再降低門檻至 85%，並依上述方式逐步向外推延尋解，直至找到最佳可行解為止。除非所有 VD 資料均發生遺漏，否則群 1 資料之填補平均至少可達到 80% 之準確度。至於群 2 資料，若上、下游鄰近 3 組 VD 中有任何 1 組之資料完整，其填補準確度即可達到 95% 以上。除非所有 VD 資料均發生遺漏，否則平均至少可達到 85%之準確度。

此外，亦可先設定最低門檻要求，而後再逐一檢視表 16 中可行方案之輸入 VD 資料是否完整，最後再由其中選擇最鄰近一組 VD 作為最佳解。例如同樣填補流率時，若準確度之最低門檻要求為 90%，則群 1 至多能夠選取上、下游 2 組 VD 作為其可行解，而群 2 則可擴增至上、下游 10 組 VD。

表 15　遺漏資料填補可行方案

車流屬性 / 準確度門檻		流　　率	速　　率	占　有　率
95%	群 1	無可行解	上、下游所有 VD	無可行解
	群 2	上、下游 3 組 VD	上、下游所有 VD	上、下游 1 組 VD
90%	群 1	上、下游 2 組 VD	上、下游所有 VD	無可行解
	群 2	上、下游 10 組 VD	上、下游所有 VD	上、下游 6 組 VD
85%	群 1	上、下游 10 組 VD	上、下游所有 VD	無可行解
	群 2	上、下游所有 VD	上、下游所有 VD	上、下游 13 組 VD
80%	群 1	上、下游所有 VD	上、下游所有 VD	上、下游 5 組 VD
	群 2	上、下游所有 VD	上、下游所有 VD	上、下游所有 VD

5.4 偵測器佈設間距

資料填補應用於 VD 佈設間距之基本概念為，當 VD 資料能夠由其上、下游 VD 資料進行填補或推估，並達一定準確度要求時，即可將此一 VD 視為多餘 (redundant)，故佈設間距可擴增為上、下游填補輸入 VD 之距離。此處延續應用表 15 之結果，分析雪山隧道中 VD 最大可能之佈設間距。分析時，可依單一屬性及多重屬性兩種角度進行。當僅考慮單一屬性，例如流率時，若設定準確度門檻值為 95%，則因群 1 資料無法達到此一要求，故 VD 佈設間距無法擴增，仍須維持在既有之 350 m。若降低門檻至 90%，則因群 1 與群

2 資料必須同時達到此一要求，故取兩者之交集，至多選取上、下游 2 組 VD 作為填補輸入。依此結果推算，每組 VD 約相距 350 m，上、下游各 2 組共 4 組，故最大可能之佈設間距為 1,400 m。依此方式，推估不同屬性在不同準確度之要求下，雪山隧道 VD 最大可能之佈設間距如表 16 所示。

表 16　單一屬性車輛偵測器最大可能佈設間距

準確度門檻 \ 車流屬性		流　率	速　　率	占　有　率
95%	群 1	350 m	11,900 m	350 m
	群 2	2,100 m	11,900 m	700 m
	全體	350 m	11,900 m	350 m
90%	群 1	1,400 m	11,900 m	350 m
	群 2	7,000 m	11,900 m	4,200 m
	全體	1,400 m	11,900 m	350 m
85%	群 1	7,000 m	11,900 m	350 m
	群 2	11,900 m	11,900 m	9,100 m
	全體	7,000 m	11,900 m	350 m
80%	群 1	11,900 m	11,900 m	3,500 m
	群 2	11,900 m	11,900 m	11,900 m
	全體	11,900 m	11,900 m	3,500 m

當同時考慮多種屬性，例如流率與占有率時，則填補準確度必須同時滿足各屬性之要求。舉流率與速率為例，因流率之填補績效較速率為差，故 VD 之佈設間距均依流率而定。在準確度為 95% 之要求下，佈設間距維持現況不變為 350 m。當準確度調降至 90% 時，依流率之填補績效可延長至 1,400 m，餘此類推。因占有率之填補績效為三者中之最差者，故只要將其納入考慮，即會對 VD 佈設間距產生決定性的影響，僅在 80% 之準確度要求下，可由現有之 350 m 擴增為 3,500 m。依此，推算出各種屬性組合下 VD 之最大可能佈設間距如表 17 所示。

六、結論與建議

6.1 結論

本研究採用兩步驟類神經網路方式，針對雪山隧道路段車輛偵測器遺漏資料之填補進行實證分析，以找出其中較為適用之填補方法，並規劃其可能之應用。經研究結果得到以下之結論：

表 17　多重屬性車輛偵測器最大可能佈設間距

車流屬性 / 準確度門檻		流率&速率	流率&占有率	速率&占有率	流率&速率&占有率
95%	群 1 群 2	350 m 2,100 m	350 m 700 m	350 m 700 m	350 m 700 m
	全體	350 m	350 m	350 m	350 m
90%	群 1 群 2	1,400 m 7,000 m	350 m 4,200 m	350 m 4,200 m	350 m 4,200 m
	全體	1,400 m	350 m	350 m	350 m
85%	群 1 群 2	7,000 m 11,900 m	350 m 9,100 m	350 m 9,100 m	350 m 9,100 m
	全體	7,000 m	350 m	350 m	350 m
80%	群 1 群 2	11,900 m 11,900 m	3,500 m 11,900 m	3,500 m 11,900 m	3,500 m 11,900 m
	全體	11,900 m	3,500 m	3,500 m	3,500 m

1. 經採用雪山隧道車輛偵測器資料測試結果發現，先將資料分群，而後再以類神經網路進行填補之兩步驟資料填補方式，確實可以獲得不錯之填補績效，且流率、速率及占有率三項基本交通屬性資料，均可達到準確度 80%以上之良好填補績效。
2. 透過 K-means 分群方式，將同質性高之交通屬性資料予以分類處理，確實可以提高資料填補之準確度，且資料隨機變異性愈小，準確度愈高。經進一步檢定結果發現，分群數多寡對填補績效並無顯著之影響，故基於運算績效及實用便利性之考量，可將資料分為兩群分別進行填補。經實際測試結果發現，兩群分類方式不但樣本數最為平均，且亦最容易收斂，其中群 1 資料屬中低流量狀態，而群 2 資料則屬中高流量狀態。
3. 於資料填補方面，分別採用倒傳遞、輻狀基底函數以及回饋式三種最具代表性之類神經網路進行測試與比較分析。結果發現，具有動態特性之回饋式類神經網路，無論是流率、速率抑或是占有率，均可獲得最高之填補績效。
4. 就整體填補績效而言，速率填補之準確性最高屬高精確度，其次為流率，最差者為占有率；且 MAPE 值隨輸入偵測器距離之增長而有逐漸增加之趨勢，符合距離愈近車流屬性愈為接近之一般通則。另就不同群組來看，群 1 因屬中低流量狀態，車流隨機性較高，故其填補績效普遍低於群 2。
5. 以流率填補績效來看，若以 95% 準確度為門檻，則群 2 資料可由其上、下游 3 組偵測器中之任一對組合進行填補，而群 1 資料則完全無法達到此一要求。若準確度門檻降至 90%，則用以填補群 2 資料之偵測器可溯至上、下游 10 組。反觀群 1，如欲維持 90% 之準確度，則僅能取上、下游 2 組偵測器中之任一對組合進行填補；若將準確度降至

85%，則可延伸至上、下游 10 組偵測器。

6. 以速率填補績效來看，其結果均可達到高精確度填補之等級，MAPE 值上下變動幅度不大，僅介於 1% 與 2.5% 之間，且無論分群與否，抑或是分為幾群，均可獲得不錯之填補績效。
7. 以占有率填補績效來看，群 1 之準確度均在 85% 以下，而群 2 則可高達 95%。若以 80% 準確度作為門檻，群 1 資料可由其上、下游 5 組偵測器資料，而群 2 資料則可溯至上、下游所有車測器資料進行填補。若提高門檻至 85% 及 90%，則群 2 資料仍可由上、下游 13 組及 6 組車測器資料進行填補。若再提高門檻至 95%，則僅剩上、下游最鄰近車測器資料符合要求。
8. 在遺漏資料填補應用方面，速率填補之績效最好，無論是以上、下游任何一對偵測器資料作為輸入，準確度均高達 97.5% 以上；其次為流率，其準確度亦可達 90% 以上，並可以上、下游 2 至 10 對偵測器資料作為填補輸入。最差者為占有率，僅當準確度門檻降至 80% 時，群 1 資料方能進行填補，群 2 資料則無此限制。
9. 在偵測器佈設間距應用方面，若合併考量流率、速率與占有率三者，則佈設間距是由填補績效最差之占有率來決定。僅在整體準確度降至 85% 以下時，方可將現行之 350 m 佈設間距擴增至 3,500 m。若僅考慮隨機性較低之群 2 資料，則在準確度高達 90% 以上時，即可將佈設間距增加至 4,200 m。

6.2 建議

經以上分析結果得知，兩步驟回饋式類神經網路方法，確實可用於路段交通資料遺漏之填補，且可達到一定的準確度要求。然而，資料填補所涉及的層面相當廣泛，為期後續相關研究更臻完整，考慮更為周詳，特提出以下之建議供參考：

1. 欲進行完整之資料填補研究，基本上必須考量輸入資料來源、資料前置處理、資料填補技術以及填補資料應用四個面向，其中每一面向又可再細分為若干維度，組合起來即會產生極多的填補可能方式。實際研究時，建議先針對填補之對象及其特性加以界定，以簡化繁瑣之測試分析程序，快速找尋出一準確而且可靠的填補方法。
2. 以輸入資料來源來看，建議分由時、空兩個維度加以界定，其中時間方面又可分為即時性 (real-time) 與非即時性 (non-real-time) 兩類，而空間方面則可分為自我 (self) 與鄰近 (near) 車輛偵測器兩類。本研究在時間維度上係採用即時性資料進行填補，空間維度上則對稱選取上、下游各一組偵測器資料作為輸入，因此未來可繼續將研究範圍延伸至尚未觸及之時、空維度上，以使成果更為周延。
3. 以資料前置處理來看，為避免誤用異常之資料，提高填補之準確度，通常會將填補輸入之資料進行前置處理。處理的方式主要包括資料過濾 (filtering)，找尋資料變化之趨勢，以及資料分類 (data classification)，將同質性高之資料予以類聚，使其變異降至最低，常見方法有灰關聯、EM 演算法、GA 演算法、SOM、華德法以及 K-means 法等。本研

究係採過濾方式剔除異常之資料，未來則可進一步嘗試資料分類之各種方法。

4. 以資料填補技術來看，目前已知者主要包括擬最鄰近法、內插法、多項式迴歸、核迴歸、ARIMA、MLP 以及類神經網路等，而各種技術均有其適用的條件與資料型態。本研究主要是以三種典型之類神經網路作為資料填補技術，未來則建議繼續嘗試其他各種方法，或能找出填補準確度更高之演算技術。
5. 以填補資料應用來看，可由直接與間接，以及即時與事後幾個角度觀之。直接填補係純粹將遺漏值回復，供後續處理應用，而間接填補則是根據應用所需，例如旅行時間預測，填補相關屬性資料。至於即時填補，係將即時遺漏之資料立即填補，而事後填補則較無填補急迫性，可利用遺漏時間點之前或之後資料進行填補，多用於長期性資料分析或資料庫 (database) 之建立。由於資料應用之差異性極大，且牽連到其他面向維度之決定，因此建議根據研究所需，優先將其界定，以避免研究過於發散。
6. 本研究為取資料完整之利，選擇雪山隧道作為實證分析之對象。因屬封閉式空間，車流特性自有其限制性，因此未來如資料來源充足，建議進一步針對高速公路其他開放路段，甚至一般道路路段進行實證分析，使其實用範圍更為擴大。
7. 若純由資料蒐集之觀點觀之，不同時距所蒐集之資料，可能會對填補績效產生不同的影響，因此建議未來可針對此一課題進行更進一步之分析，以了解此一影響是否顯著。

參考文獻

1. 交通部，**智慧型交通資訊蒐集、處理、傳播與旅行者行為系列之研究－號誌化道路路況資訊偵測方法與省道路段固定式偵測器佈設規劃**，民國 95 年 3 月。
2. 吳健生、廖梓淋，「利用資料填補概念探討車輛偵測器佈設間距」，**運輸學刊**，第 22 卷，第 3 期，民國 99 年 9 月，頁 307-326。
3. Chen, C., Kwon, J., Rice, J., Skabardonis, A., and Varaiya, P., “Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems”, *Transportation Research Record*, Vol. 1855, Transportation Research Board, 2003, pp. 160-167.
4. Little, R. J. A. and Rubin, D. B., *Statistical Analysis with Missing Data*, John Wiley & Sons, New York, 1987.
5. Huang, X. L. and Zhu, Q. M., “A Pseudo-Nearest-Neighbor Approach for Missing Data Recovery on Gaussian Random Data Sets”, *Pattern Recognition Letters* , Vol. 23, 2002, pp. 1613-1622.
6. Huang, C. C. and Lee, H. M., “A Grey-Based Nearest Neighbor Approach for Missing Attribute Value Prediction”, *Applied Intelligent,* Vol. 20, No. 3, 2004, pp. 239-252.
7. Wen, Y. H., Lee, T. T., and Cho, H. T., “Missing Data Treatment and Data Fusion toward Travel Time Estimation for ATIS”, *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, 2005, pp. 2546-2560.

8. Gold, D. L., Turner, S. M., Gajewski, B. J., and Spiegelman, C., "Imputing Missing Values in ITS Data Archives for Intervals under 5 Minutes", the 80th Annual Meeting, Transportation Research Board, January 7-11, Washington, D.C., 2001, Paper No. 01-2760, pp.1-7.

9. Vanajakshi, L. and Rilett, L. R., "System Wide Data Quality Control of Inductance Loop Data Using Nonlinear Optimization", *Journal of Computing In Civil Engineering*, Vol. 20, No. 3, 2006, pp. 187-197.

10. Daganzo, C., *Fundamentals of Transportation and Traffic Operation*, Pergamon Elsevier, Oxford, U.K., 1997.

11. Chen, D., Muller, S. G., Mussone, L., and Montgomey, F., "A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting", *Neural Computing & Applications*, Vol. 10, 2001, pp. 277-286.

12. Zhong, M., Lingras, P., and Sharma, S., "Estimation of Missing Traffic Counts Using Factor, Genetic, Neural, and Regression Techniques", *Transportation Research Part C*, No. 12, 2004, pp. 139-166.

13. 張堂賢、黃宏仁，「車輛偵測器資料遺失之在線插補技術研究」，**運輸學刊**，第 20 卷，第 4 期，民國 97 年，頁 377-404。

14. 吳冠宏、吳信宏、郭廣洋，「應用分群技術於交通事故資料分析」，**品質學報**，第 13 卷，第 3 期，民國 95 年，頁 305-312。

15. 周文賢，**多變量統計分析**，初版，智勝文化，臺灣，民國 91 年。

16. 葉怡成，**類神經網路模式應用與實作**，第 7 版，儒林圖書有限公司，臺北，民國 89 年。

17. Delurgio, S. A., *Forecasting Principles and Applications*, McGraw-Hill, New York, 1998.

18. Sgarma, S., *Applied Multivariate Techniques, Strategies and Case Studies*, John Wiley & Sons, New York, 1995.