

國立中央大學

資訊工程學系軟體工程碩士班  
碩士論文

預測交通需求之分佈與數量—基於多重式注意力  
機制之 AR-LSTMs 模型

Predicting Transportation Demand based on  
AR-LSTMs Model with Multi-Head Attention



研究生：徐志榮

指導教授：陳弘軒 博士

中華民國一百零八年六月



# 國立中央大學圖書館 碩博士論文電子檔授權書

(104 年 5 月最新修正版)

本授權書授權本人撰寫之碩/博士學位論文全文電子檔(不包含紙本、詳備註 1 說明)，在「國立中央大學圖書館博碩士論文系統」。(以下請擇一勾選)

☒同意 (立即開放)

☐同意 (請於西元 \_\_\_\_\_ 年 \_\_\_\_\_ 月 \_\_\_\_\_ 日開放)

☐不同意，原因是：\_\_\_\_\_

在國家圖書館「臺灣博碩士論文知識加值系統」

☒同意 (立即開放)

☐同意 (請於西元 \_\_\_\_\_ 年 \_\_\_\_\_ 月 \_\_\_\_\_ 日開放)

☐不同意，原因是：\_\_\_\_\_

以非專屬、無償授權國立中央大學、台灣聯合大學系統圖書館與國家圖書館，基於推動「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名： 徐志榮

學號： 106525009

論文名稱： 預測交通需求之分佈與數量—基於多重式注意力機制之 AR-LSTMs 模型

指導教授姓名： 陳弘軒

系所： 資訊工程學系軟體工程碩士班 所 ☐博士班 ☒碩士班

填單日期： 108.7.2

備註：

1. 本授權書之授權範圍僅限電子檔，紙本論文部分依著作權法第 15 條第 3 款之規定，採推定原則即預設同意圖書館得公開上架閱覽，如您有申請專利或投稿等考量，不同意紙本上架陳列，須另行加填申請書，詳細說明與紙本申請書下載請至本館數位博碩論文網頁。
2. 本授權書請填寫並親筆簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
3. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應遵守著作權法規定。

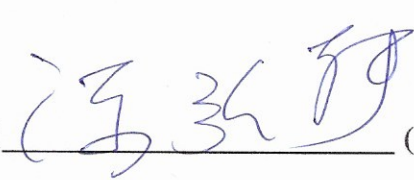
國立中央大學碩士班研究生

論文指導教授推薦書

資訊工程學系軟體工程碩士班學系/研究所 徐志榮 研究生所提之論

文 預測交通需求之分佈與數量—基於多重式注意力機制之 AR-LSTMs 模型

係由本人指導撰述，同意提付審查。

指導教授  (簽章)

108 年 7 月 2 日

國立中央大學碩士班研究生  
論文口試委員審定書

資訊工程學系軟體工程碩士班 學系/研究所 徐志榮 研究生

所提之論文

預測交通需求之分佈與數量—基於多重式注意力機制之AR-LSTM模型

經本委員會審議，認定符合碩士資格標準。

學位考試委員會召集人

委

員

陳忠國  
陳弘軒  
洪炯宇

中華民國 108 年 6 月 24 日

# 預測交通需求之分佈與數量—基於多重式注意力 機制之 AR-LSTMs 模型

## 摘要

智慧交通儼然成為智慧城市的重要一環，計程車需求預測是智慧交通中一項重要課題。有效地預測下個時間點載客需求的分布可以減少司機空車時間、降低乘客等待時間及增加獲利載客次數，將計程車產業獲利最大化並解決車輛巡迴攬客所造成的能源消耗及汙染。

本文利用計程車行車紀錄資料結合深度學習的架構提出有效的計程車載客需求預測模型，使用善於處理時間序列架構的短中長期記憶模型 (LSTM) 為基礎，交通議題的資料與長時間周期變化有關，過去的方式難以克服於尖峰與離峰間變化的預測，因此我們使用注意力機制 (Attention) 加強長時間週期的交通問題資訊處理，並設計多層的深度學習網路架構來提高預測準確率。我們並自訂了一個同時考慮均方損失誤差及平均百分比誤差的損失函數，因為均方損失誤差通常會低估低需求區域的叫車數，而平均百分比誤差則容易錯估高需求區域的叫車數。

為驗證模型的一般性，我們使用兩組資料集，分別為紐約市計程車的行車紀錄資料與台灣大車隊在台北的計程車叫車資料進行驗證。在實驗中我們比較了傳統的預測方式、淺層機器學習、及深度學習模型等方式預測計程車需求分佈，實驗結果顯示我們提出的多重式 AR-LSTMs 預測模型能有效的提高預測的準確度。

**關鍵字：**計程車需求預測, 深度學習, 遞歸神經網絡, 長短期記憶模型, 注

## 意力模型

# Predicting Transportation Demand based on AR-LSTMs Model with Multi-Head Attention

## Abstract

Smart transportation is a crucial issue for a smart city, and the forecast for taxi demand is one of the important topics in smart transportation. If we can effectively predict the taxi demand in the near future, we may be able to reduce the taxi vacancy rate, reduce the waiting time of the passengers, increase the number of trip counts for a taxi, expand driver's income, and diminish the power consumption and pollution caused by vehicle dispatches.

This paper proposes an efficient taxi demand prediction model based on state-of-the-art deep learning architecture. Specifically, we use the LSTM model as the foundation, because the LSTM model is effective in predicting time-series datasets. We enhance the LSTM model by introducing the attention mechanism such that the traffic during the peak hour and the off-peak period can better be predicted. We leverage a multi-layer architecture to increase the predicting accuracy. Additionally, we design a loss function that incorporates both the absolute mean-square-error (which tends under-estimate the low taxi demand areas) and the relative mean-square-error (which tends to misestimate the high taxi demand areas).

To validate our model, we conduct experiments on two real datasets — the NYC taxi demand dataset and the Taiwan Taxi's taxi demand dataset

in Taipei City. We compare the proposed model with non-machine learning based models, traditional machine learning models, and deep learning models. Experimental results show that the proposed model outperforms the baseline models.

**Keywords:** Taxi Demand Prediction, Deep Learning, Recurrent Neural Networks, Long Short-Term Memory Work, Attention



# 目錄

	頁次
摘要	iv
Abstract	vi
目錄	viii
圖目錄	ix
表目錄	x
一、緒論	1
1.1 研究動機 .....	1
1.2 研究目標 .....	2
1.3 研究貢獻 .....	2
1.4 論文架構 .....	3
二、預前工作	4
2.1 遞歸神經網路 (Recurrent Neural Network, RNN).....	4
2.2 長短期記憶模型 (Long Short-Term Memory, LSTM) .....	6
三、模型及方法	8
3.1 原始座標資料編碼 .....	8
3.2 保留連結 (Residual Connections) .....	9
3.3 注意力機制 (Attention).....	10
3.3.1 比例內積式注意力 (Scaled Dot-Product Attention) ...	10

3.3.2	多頭式注意力 (Multi-Head Attention) .....	11
3.4	損失函數 .....	11
3.5	Residual-LSTMs 預測模型 .....	12
3.6	Attention-Residual-LSTMs 預測模型 .....	13
<b>四、</b>	<b>實驗結果與分析</b>	<b>15</b>
4.1	資料集介紹 .....	15
4.1.1	NYC Taxi .....	16
4.1.2	台灣大車隊 .....	16
4.2	實驗環境 .....	16
4.3	實驗對照模型與方法介紹 .....	17
4.3.1	歷史平均 .....	17
4.3.2	ARIMA .....	17
4.3.3	XGboost .....	18
4.3.4	線性回歸 (Linear Regression) .....	18
4.3.5	DMVST-Net .....	18
4.4	評量指標 .....	18
4.5	實驗結果 .....	19
4.5.1	實驗結果分析 .....	20
4.5.2	不同時間區段比較 .....	27
<b>五、</b>	<b>相關研究</b>	<b>32</b>
5.1	LSTM-MDN-Conditional .....	33
5.2	DMVST-Net .....	33
<b>六、</b>	<b>結論及未來展望</b>	<b>35</b>
6.1	結論 .....	35
6.2	未來展望 .....	35

參考文獻	37
附錄一	39

# 圖目錄

	頁次
2.1 基本 RNN 的形式 . . . . .	4
2.2 RNN 單元的架構圖 . . . . .	5
2.3 LSTM 單元的架構圖 . . . . .	6
3.1 GIS 編碼處理過程 . . . . .	8
3.2 保留連結 (Residual Connections) . . . . .	9
3.3 Residual-LSTMs 架構圖 . . . . .	12
3.4 Attention-Residual-LSTMs 架構圖 . . . . .	14
4.1 Loss 變化圖 . . . . .	20
4.2 第一組資料集：2016/6/15(平日) 預測結果 . . . . .	23
4.3 第一組資料集：2016/6/5(假日) 預測結果。 . . . .	24
4.4 第二組資料集：2017/1/24(平日) 預測結果 . . . . .	25
4.5 第二組資料集：2017/1/15(假日) 預測結果 . . . . .	26
4.6 第一組資料集：24 小時區段之 RMSE 變化比較 . . . . .	28
4.7 第一組資料集：24 小時區段之 MAPE 變化比較 . . . . .	29
4.8 第二組資料集：24 小時區段之 RMSE 變化 . . . . .	30
4.9 第二組資料集：24 小時區段之 MAPE 變化 . . . . .	31
5.1 LSTM-MDN-Conditional 架構圖 . . . . .	33
5.2 DMVST-Net 架構圖 . . . . .	34

# 表目錄

	頁次
4.1 資料集介紹 . . . . .	15
4.2 不同模型間的效能 (第一組資料集) . . . . .	19
4.3 不同模型間的效能 (第二組資料集) . . . . .	20

# 一、緒論

## 1.1 研究動機

交通和人們日常生活息息相關，也是智慧城市的核心，然而城市的交通建設規劃牽涉到人口的分布數量、龐大的通勤變化及不同運輸交通工具等情況，如何將龐大的變動資訊整合出有效的預測需求模型是一個很複雜且困難的問題。

在運輸交通工具上，計程車產業扮演重要的一環，當計程車企業進行車輛調度依據時，獲得下個時間的交通需求分佈的資訊才得以進行調度車輛。此時建立精準的計程車即時預測模型儼然是十分重要的課題。

近期基於深度學習的技術日漸成熟，在處理圖學影像上已經成功達到超越人類觀測的結果 [1][2][3]。我們認為有機會將深度學習預測技術衍伸至交通議題上，然而交通議題因牽涉層面涉及不同面向，故如何以深度學習的技術達到有效解決交通上的問題便是困難的挑戰更是很有價值的議題。

在進行任何交通分析決策前，對於目前以及未來的車輛掌握，是決策依據極為重要的關鍵資訊，尤其在計程車產業上，對於車輛派遣依據決定於車輛需求前提。因此我們選定以交通議題中計程車需求問題為出發，進而預測需求分布。

## 1.2 研究目標

本論文目的是建立一個精準的計程車叫車預測需求的模型。我們使用於真實社會中計程車行車裝置所紀錄的車用資訊及 GPS 位置訊號進行處理，將一個地區的圖層資訊切成數個小網格，並將載客的座標資料分布結合於網格之中，利用深度學習的方式進行預測，希望提供有價值的預測需求分布以達到精準派遣車輛之目標。

因交通上的問題與時間的因素息息相關，將計程車的數據進一步分析中呈現需求的變動有明顯週期性。因此我們選擇適合處理時間序列的長短期記憶模型 (LSTM) [4]，將當前的需求分佈的圖層資訊結合相關的資訊 (天氣、日期…等) 整合後輸入模型，藉由深度學習的運算後得知下個時間點的預測需求分佈及數量。

在深層的 LSTMs 通常會得到比單層的 LSTM 更佳的準確度 [5]，在本篇論文中提出多層 LSTMs 的網路架構計程車需求數量預測模型，並加上注意力 (Attention) 機制加強長時間的資訊保留，最後實驗證明本模型優於傳統的 ARIMA 模型、機器學習的線性回歸模型、基於決策樹的 XGBoost 以及已發表的深度學習的 DMVS-Net 模型。

## 1.3 研究貢獻

本論文利用深度學習建立預測模型，在接下來章節實驗結果中說明我們研究驗證深度學習得以應用於計程車需求預測上並且得到極佳的結果。我們透過過去的資料成功預測下個時間的階段需求分佈，將這些資訊提供車輛派遣之決策者得以進行更有效率的車輛調度，以達到減少司機空車時間、降低乘客等待時間及增加獲利載客次數之目的，最終提升計程車相關產業之最大化獲利。

本論文之研究將可運用於其他交通運輸工具預測需求的議題上。在交通議題上資料與時間的關係是密不可分，對於特地週期有明顯相關聯

性，我們數據分析結果也顯示人們上下班通勤的時間與交通需求數量變化相關。我們提出的模型對於長時間區段的資訊保留有極佳的效果，因此未來進行其他交通工具需求預測上可以藉由我們模型得出極佳的預測結果。

## 1.4 論文架構

本篇論文共分為六個章節，其架構如下：

第一章、說明本篇論文之研究動機、研究目標、研究貢獻。

第二章、介紹本論文所需的知識。

第三章、說明本篇論文資料前處理過程、模型所使用的機制及模型架構。

第四章、說明實驗之資料集、環境設定、比較模型及結果。

第五章、介紹與本篇論文之相關主題研究與其研究成果。

第六章、本篇論文之結論與未來展望。



## 二、 預前工作

### 2.1 遞歸神經網路 (Recurrent Neural Network, RNN)

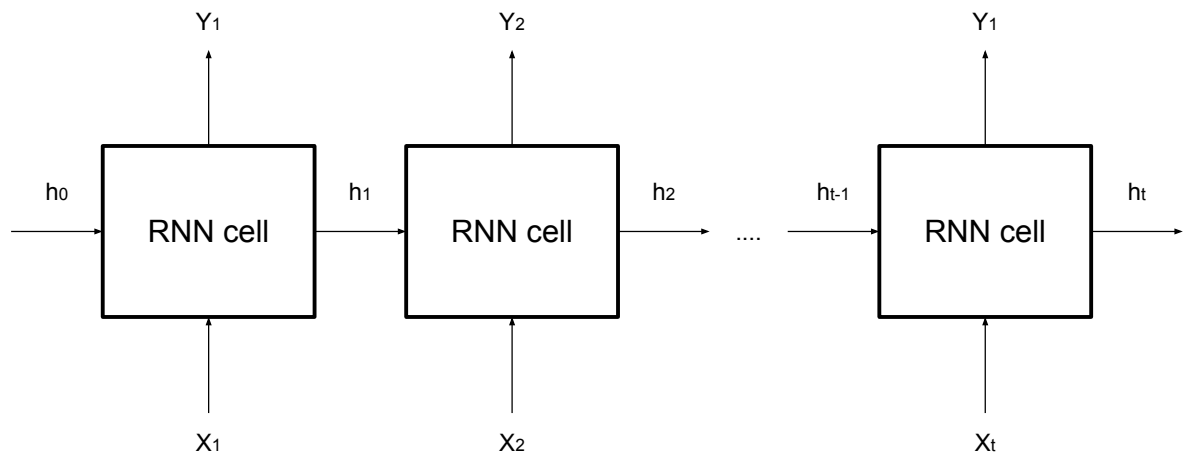


圖 2.1: 基本 RNN 的形式

RNN 是類神經網路的一種，將自身單元資訊在網路中循環傳遞，故相較於一般前饋式神經網路 (Feedforward Neural Network, FNN) 來說更適合用於時間序列資料的處理。

我們交通需求預測的任務屬於連續時間上的問題，因此我們設計深度學習模型的類神經網路選用 RNN 的架構，藉此希望達到保留過去時

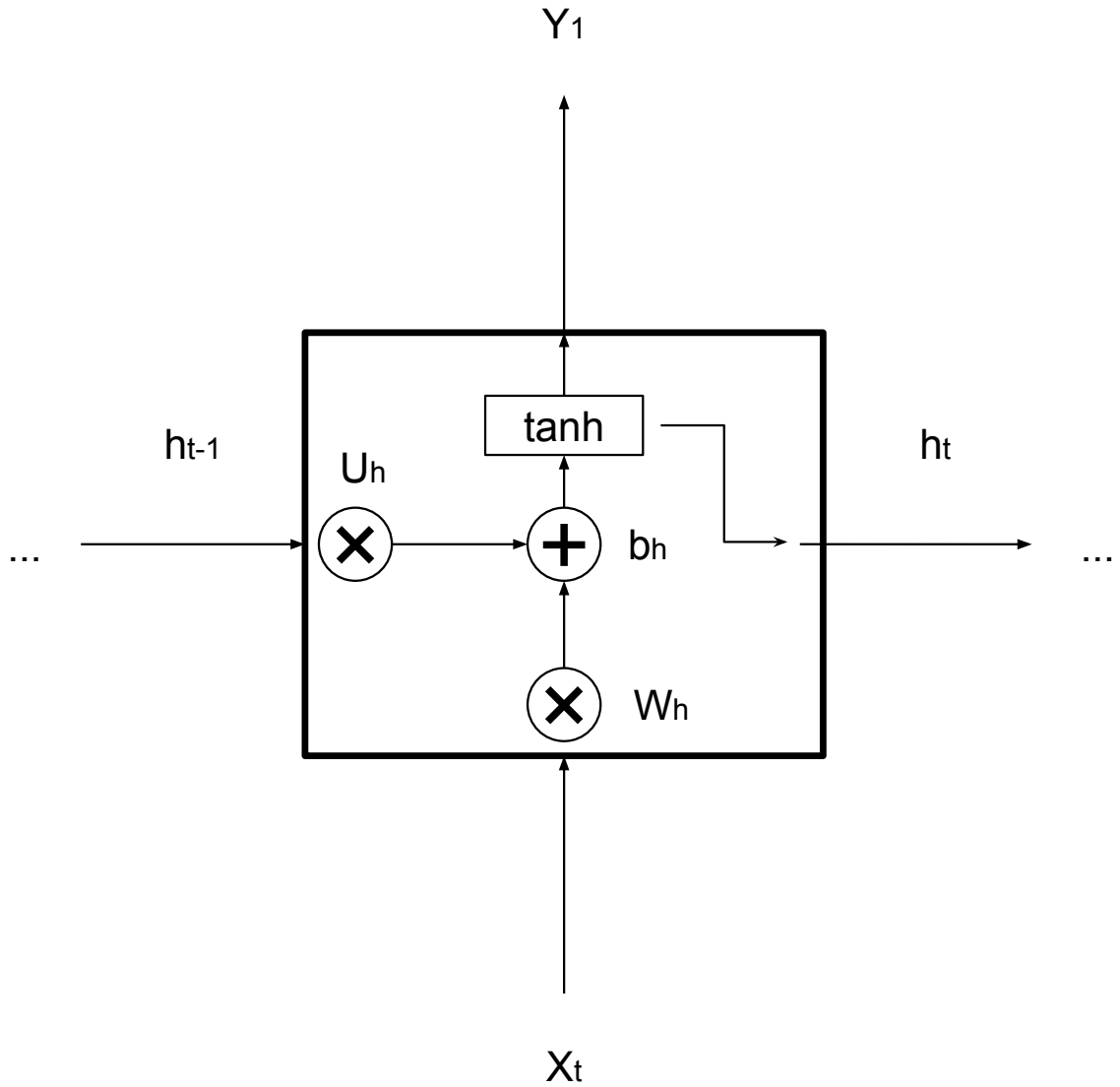


圖 2.2: RNN 單元的架構圖

間點上的資訊。其定義如下：

$$\begin{aligned} h_t &= \sigma(W_h x_t + U_h h_{t-1} + b_h) \\ y_t &= \sigma(W_y h_t + b_y) \end{aligned} \quad (2.1)$$

其中， $x_t$  及  $y_t$  分別代表輸入及輸出的向量。 $h_t$  為隱藏層的向量。 $\sigma$  為激勵函數。 $W_h, W_y, U_h$  代表權重，透過訓練過程中更新。 $b_h, b_y$  為偏差值參數。

然而在一般 RNN 架構中常見於資訊在遞迴的過程中，如果權重值

大於 1 在多次相乘後得到極大的值時，此資訊變極大則發生梯度爆炸的問題；如果權重值小於 1 在多次相乘後得到極小的值時，此資訊變消失發生梯度消失的問題。資訊很難傳遞到時間步數越大的網路中，因此很難得以訓練出有效的預測結果 [6]。

## 2.2 長短期記憶模型 (Long Short-Term Memory, LSTM)

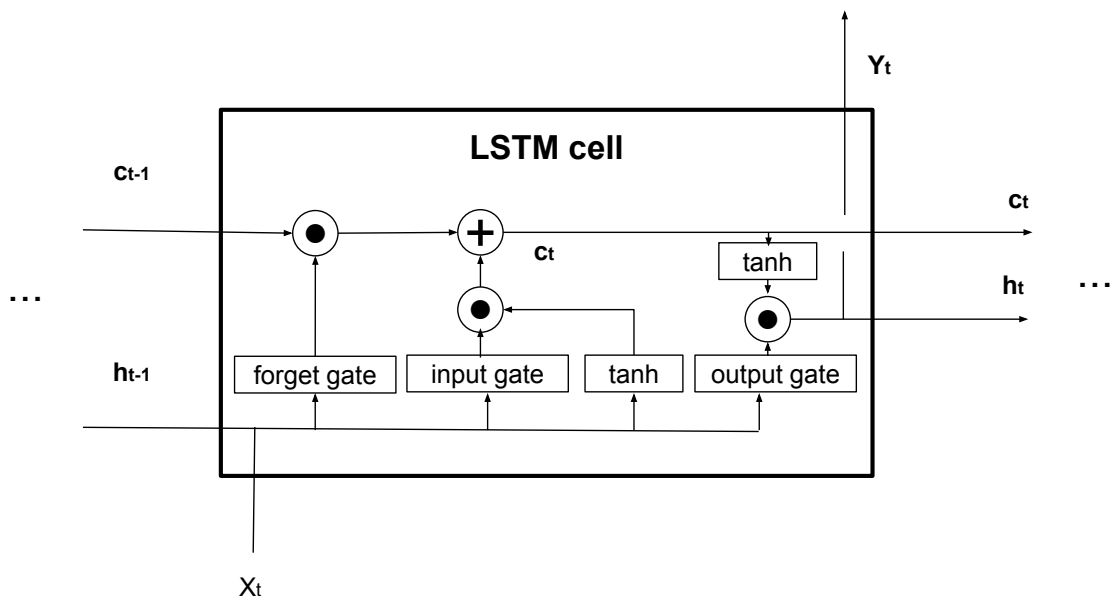


圖 2.3: LSTM 單元的架構圖

LSTM 相較於一般的 RNN 架構中多三種控制閥來增強對應時間的決策。利用閥來決定過去及當前那些資訊需要被儲存及使用，進而使資訊有效的傳遞於遞迴神經的末端，以解決上節所提及梯度爆炸或梯度消失的問題。三種閥分別為：

1. 輸入閥 (Input Gate)：用以選擇當前的輸入及新的記憶候選單元是否需要被保留，其中以 Sigmoid 函數作為激勵函數。

2. 輸出閥 (Output Gate)：用以選擇當前的記憶是否要被保留至隱藏單元，其中以 Sigmoid 函數作為激勵函數。
3. 遺忘閥 (Forget Gate)：如果當前的資訊與新的記憶或前者資訊相反時則被過濾，其中以 Sigmoid 函數作為激勵函數。

其定義如下：

$$\begin{aligned}
i_t &= \sigma(W_i h_{t-1} + U_i x_t + b_i) \\
f_t &= \sigma(W_f h_{t-1} + U_f x_t + b_f) \\
o_t &= \sigma(W_o h_{t-1} + U_o x_t + b_o) \\
\tilde{c}_t &= \tanh(W_c h_{t-1} + U_c x_t + b_c) \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned} \tag{2.2}$$

其中， $i_t, f_t, o_t$  分別代表輸入閥、遺忘閥及輸出閥。 $h_t$  為隱藏層的向量。 $\sigma, \tanh$  為激勵函數。 $W_i, W_f, W_o, W_c, U_i, U_f, U_o, U_c$  代表權重，透過訓練過程中更新。 $b_i, b_f, b_o, b_c$  為偏差值參數。 $h_t, c_t, \tilde{c}_t$  分別為  $LSTM_i$  隱藏單元、記憶單元及候選記憶單元。

### 三、 模型及方法

在這章節中，我們首先說明如何將原始座標資料處理成深度學習網路可接受運算格式。依序介紹我們提出的模型中所使用的深度學習架構，包含遞歸神經網路、長短期記憶模型、保留連結及注意力機制。接著說明我們在更新權重參數時所使用損失函數，最後說明我們提出的 AR-LSTMs 預測模型設計架構。

#### 3.1 原始座標資料編碼

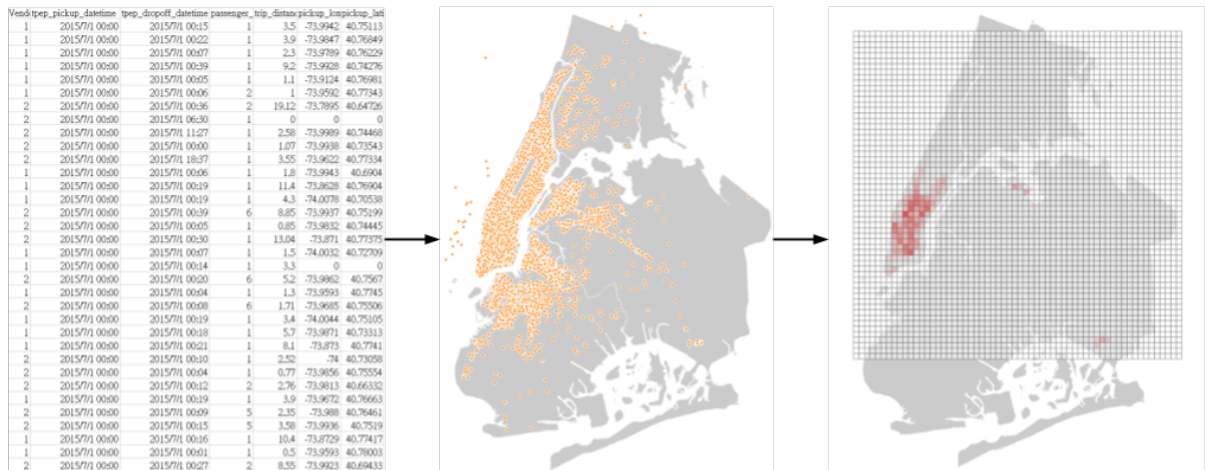


圖 3.1: GIS 編碼處理過程

我們使用地理資訊系統 (Geographic Information System, GIS) 將原始計程車行車裝置所紀錄的全球定位系統 (Global Positioning System, GPS) 之地點座標資料進行地理資訊處理。

將座標資料從世界大地測量系統 (World Geodetic System, WGS84)

座標編碼轉換成對應地區所合適的投影系統，透過這種方式轉換後的位置可以得到較高的精準度，使得計程車需求分佈分析得到更佳符合真實位置的結果。

以我們實驗資料說明，我們利用 GIS 的系統建立於紐約市曼哈頓區域  $0.5\text{km} \times 0.5\text{km}$  大小地理區塊 (共 3600 格)，將原始計程車需求座標資料轉換後的位置根據其分布的數量匹配進相對應的地理區塊位置中。

## 3.2 保留連結 (Residual Connections)

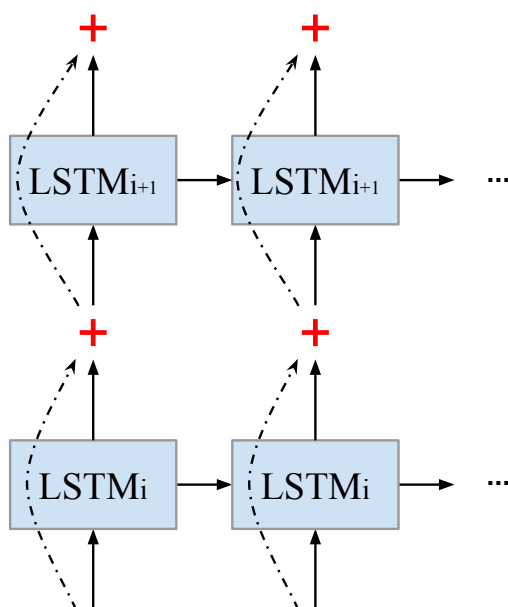


圖 3.2: 保留連結 (Residual Connections)

多層的 LSTMs 模型比單層可以學習到更多且相對關係的資料特性，然而多層的 LSTMs 資料容易在訓練的過程中發生梯度爆炸或梯度消失的問題 [3]。在我們後續實驗章節中的結果會發現如果使用一般堆疊方式達到 4 層時，得到的預測結果不佳。

因此我們設計模型時加入保留連結部分，和一般堆疊的 LSTMs 差別於當其中一層 ( $LSTM_i$ ) 輸出到下一層 ( $LSTM_{i+1}$ ) 資訊時會結合當層

( $LSTM_i$ ) 輸入資料進行傳遞，這樣可以有效改善後向傳遞時的梯度變化 (圖 3.2)。在過去研究中說明在深層 LSTMs 中使用保留連結會得到很好的效果 [3]。其定義如下：

$$\begin{aligned} c_t^i, m_t^i &= LSTM_i(c_{t-1}^i, m_{t-1}^i, x_t^{i-1}; W^i) \\ x_t^i &= m_t^i + x_t^{i-1} \\ c_t^{i+1}, m_t^{i+1} &= LSTM_{i+1}(c_{t-1}^{i+1}, m_{t-1}^{i+1}, x_t^i; W^{i+1}) \end{aligned} \quad (3.1)$$

其中， $x_t^i$  為在  $t$  時間於  $LSTM_{i+1}$  輸入的資訊。 $c_t^i, m_t^i$  分別為  $LSTM_i$  隱藏狀態及記憶狀態。 $W^i$  為在  $LSTM_i$  的權重。

### 3.3 注意力機制 (Attention)

Attention 是使用一組向量 queries 及一組向量 keys 對應一組向量 outputs，每一組 queries 與其對應的每一組 keys 計算出權重總和，並將權重總和與一組向量 values 計算後輸出向量。經由額外的向量進行全部時間點的運算後，可以加強不同時間之資訊處理，進而保留該時間點所相關對應的時間點資訊。

#### 3.3.1 比例內積式注意力 (Scaled Dot-Product Attention)

常見的 Attention 運算方式有兩種類別分別為加法式注意力 (additive attention) 及內積式注意力 (dot-product attention)，兩者運算方式類似，然而在實務上加法式使用一層前饋式類神經網路實作需要耗費的運算成本大於內積式，故在此選擇以內積式注意力進行運算。為了減緩當內積運算後成為一個很大的值又將其進行 softmax 的函數運算導致梯度為變成極小，故在分母加加上  $\sqrt{\frac{1}{d_k}}$ 。其定義如下：

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3.2)$$

其中，Q, K, V 為對應 queries, keys, values 所組合的矩陣。\$d\_k\$ 為 queries, keys, values 的維度。

### 3.3.2 多頭式注意力 (Multi-Head Attention)

一般在建立類神經網路時是使用單一 Attention。過去相關研究 [7] 將 queries, keys 和 values 分為等分的大小後進行線性投影，並將其值運用平行運算。此架構透過不同線性投影的結果，能萃取更完整的資訊，提升最後結果的準確度。將原本高維等分多頭式，能使每組的維度被降低，所以運算成本和單一 Attention 相近。其定義如下：

$$\begin{aligned} MultiHead(Q, K, V) &= Concat(head_1, \dots, head_n)W^o \\ \text{where } head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V) \end{aligned} \quad (3.3)$$

其中，\$W\_i^Q, W\_i^K, W\_i^V, W^o\$ 皆為線性投影的參數。

## 3.4 損失函數

我們使用的損失函數其定義如下：

$$Loss = \sum_{t=1}^T \sum_{r=1}^R ((Y_t^r - \hat{Y}_t^r)^2 + \gamma (\frac{(Y_t^r - \hat{Y}_t^r)^2}{Y_t^r + 1})) \quad (3.4)$$

其中，\$Y\_t^r\$ 和 \$\hat{Y}\_t^r\$ 分別代表實際分布於 \$t\$ 時間時 \$r\$ 地理區塊的真實與預測之載客需求，\$T\$ 代表 LSTM 時間單位數，\$R\$ 代表網格總數。

本論文使用的損失函數包含均方損失誤差及平均百分比誤差，在模型學習的過程中均方誤差會偏向預測保留值大的部分，然而交通需求量有很多地方是少量需求但穩定的地方，為了避免損失掉長時間有載客需求地區但數量較小的區塊，加入平均百分比誤差來適當處理其部分 [8]。



### 3.5 Residual-LSTMs 預測模型

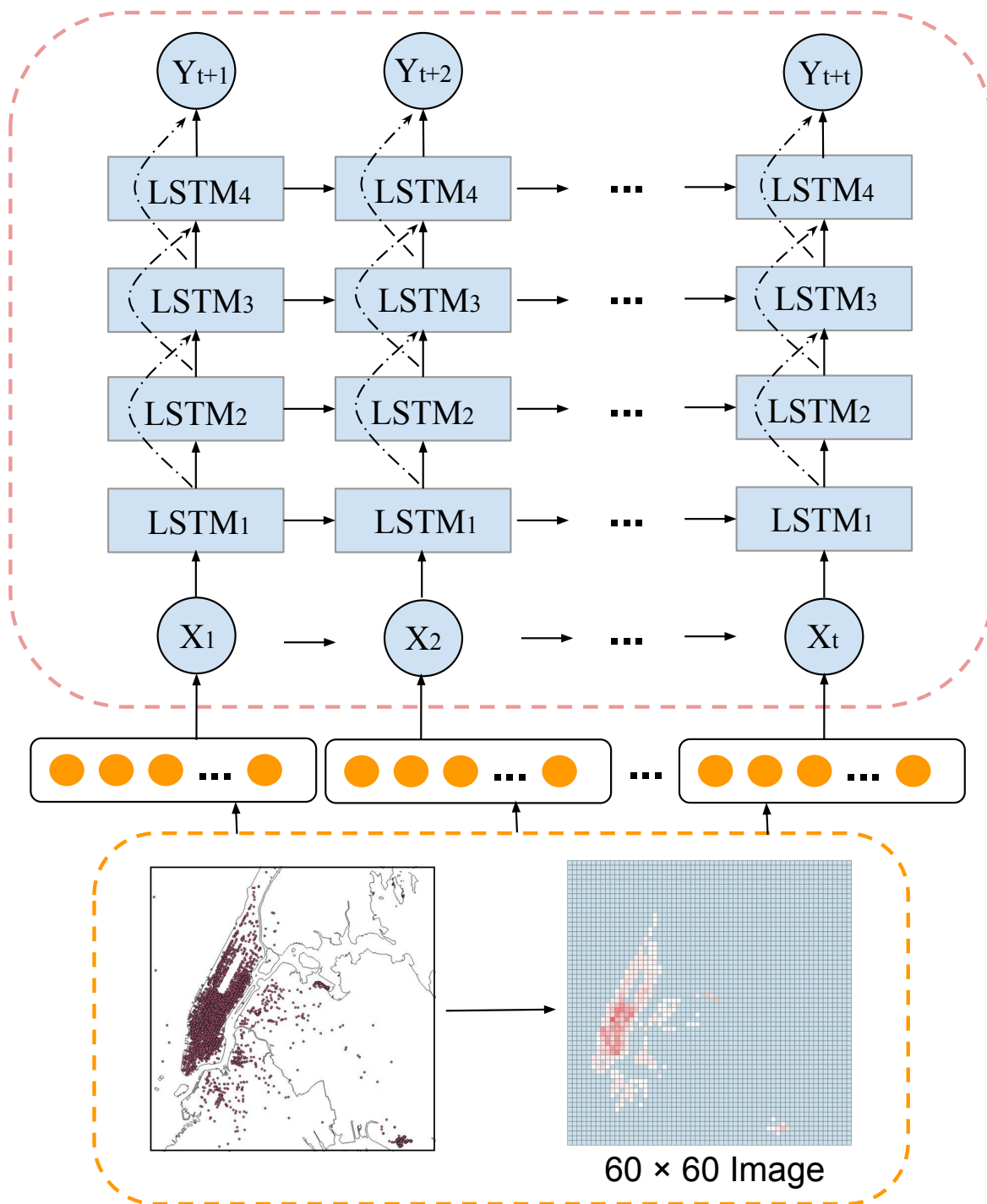


圖 3.3: Residual-LSTMs 架構圖

我們於過去研究中 [9] 提出的 Residual-LSTMs 預測模型 (圖3.3)，使用處理時間序列的 LSTM 架構，將層數堆疊數層，其中  $X_t, Y_t$  分別為真

實載客需求及預測的下個時段載客需求。

利用足夠深層的類神經網路架構使模型能有效地學習地理區塊間的相對應空間關聯關係。然而在建構深層的 LSTMs 時，常會發生資訊無法傳遞到最後一層 (梯度消失或梯度爆炸)，因此我們在 LSTM 之間加入保留連結 (圖3.2)，使得梯度流動 (Gradient Flow) 順利進行。

原始座標資料處理後成地理區塊資料後透過全連接層的方式輸入 LSTMs 層。經由網路後輸出將預測需求數量與真實資料進行損失函數計算。

### 3.6 Attention-Residual-LSTMs 預測模型

計程車的需求預測中，交通議題上必需要考量的主要關鍵點為空間及時間的資訊處理。近幾年使用深度學習進行預測的常見使用 LSTM 模型進行時間處理，在時間上的處理都可以得到極佳的結果。考量到空間資訊時，有研究提出使用 LSTM 方式將地理網格的資訊每一格依序視為 LSTM 輸入 [10]；另外有研究提出使用 CNN 方式 [8] 進行處理地理網格，利用卷積層將網格資訊呈現不同形式到 LSTM 輸入。然而前者的方式將同一個時間上的空間資訊使用 LSTM 進行運算會導致連續時空處理方式處理同一時間上的空間，且空間的關聯並非單純連續的地理區塊的資訊整合，現實中往往需要同時間完整城市地理區塊資訊；後者的方式會使很多空間資訊消失，尤其交通問題上牽涉整個城市地理區塊，有時而非單看鄰近地理區塊就能得知。

我們在架構 (圖3.4) 中使用注意力機制加強交通資訊長時間變化偵測，藉此解決在預測結果中尖峰時段預測困難問題。原始座標資料以我們設計格式處理後，經由多層的 LSTMs 類神經網路傳遞至下一 Attention 層，所有資訊進入 Attention 層分至多頭式進行運算，將資訊進行線性映射後再做結合，並將最後資訊再做一次線性映射後傳遞至全連接層與原先多層的 LSTMs 類神經網路輸出結合進行最終預測。

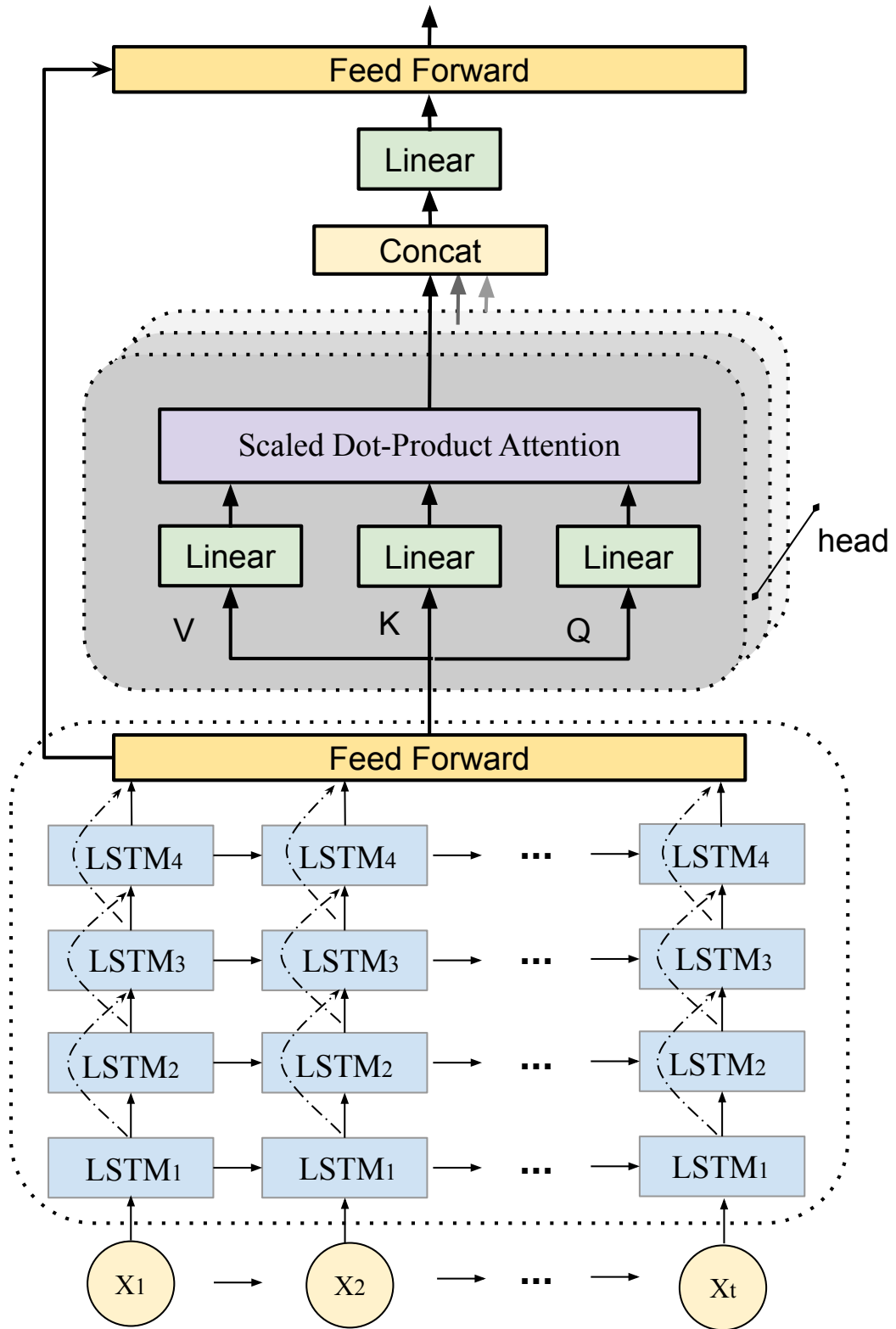


圖 3.4: Attention-Residual-LSTMs 架構圖

由訓練學習出該時間點與其他對應時間的相依程度，進而可以讓全部完整時間的資訊有效的利用，提升較難預測時段 (尖峰時段) 的結果，接下來實驗中得以證明我們模型成功提升整體預測準確度。

## 四、實驗結果與分析

此章節將評估我們的模型，進一步比較我們提出的模型及其他基準的模型。透過預測的實驗，以未來時間點的載客需求預測結果證明我們提出模型成功應用於此類型交通議題。依序說明本論文所使用的資料集、實驗環境設定、對照的模型及方法、評量指標及實驗結果。

### 4.1 資料集介紹

表 4.1: 資料集介紹

資料集	NYC Taxi	台灣大車隊
時間	2016 年 7 月-2017 年 6 月	2016 年 2 月-2017 年 1 月
時間間隔	1 小時	1 小時
資料筆數	約 1 億	約 4 百萬
網格數量	60×60	5×5
網格大小	0.5km×0.5km	1.5km×1.5km
地理投影編碼	UTM 18N	TW97

本篇論文中使用真實中計程車行車裝置所紀錄的車用資訊及 GPS 位置訊號進行處理，其原始資料包含日期時間、乘客上車經緯度、乘客數量及司機相關資訊。在實驗中我們共使用的兩組資料集以驗證模型的一般性，以下說明兩組資料集的來源及特性：

### 4.1.1 NYC Taxi

紐約市的計程車行車紀錄資料 [11]，紀錄時間從 2009 年至 2018 年，資料數量每月約有千萬筆的紀錄資料。其中資料分為分布於曼哈頓市區的黃色計程車以及分布於郊區的綠色計程車。

第一組紐約市曼哈頓地區的資料集的，我們選用從 2016 年 7 月至 2017 年 6 月的黃色計程車行車紀錄資料，每筆資料包含乘客上車地點及時間的資訊，共計一億多筆的行車紀錄資訊。我們將曼哈頓區矩形地理範圍切分為  $60 \times 60$  的地理區塊，每個地理區塊大小為  $0.5\text{km} \times 0.5\text{km}$ ，將資料以時段為區隔，根據該時段所有 GPS 座標資料分布於何者地理區塊位置處理。

### 4.1.2 台灣大車隊

台灣大車隊為台灣計程車領導品牌，此資料為台灣台北市內湖區矩型地理範圍內之計程車載客時間、地點 GPS 資料。資料涵蓋時間為 2016 年 2 月至 2017 年 1 月，共計 4,118,812 筆的行車紀錄資訊，平台無公開 2017 年 2 月份資料其部分提供預測競賽評比使用。

第二組台灣大車隊的資料集，我們選用平台公開的 2016 年 2 月至 2017 年 1 月的行車紀錄資料，平台提供的資料格式為內湖區矩型地理範圍切分為  $5 \times 5$  地理區塊，每格大小約為  $1.5\text{km} \times 1.5\text{km}$ ，每個時段為 1 小時。此資料集中我們有額外加入時間及星期的資料。

## 4.2 實驗環境

我們的原始資料透過 GIS 系統進行編碼，透過轉換至合適座標編碼可以得到較高的精準度。我們將原本的世界大地測量系統座標編碼 (WGS84) 處理各地區合適的編碼，第一組 NYC Taxi 資料集我們使用適

合美國紐約地區地理處理的 UTM 18N 編碼；第二組台灣大車隊資料集我們使用適合台灣地區地理資訊處理的 TW97 編碼。

在深度學習模型訓練過程中，兩組資料集取其中約 60% 為訓練資料集、10% 為驗證資料集及最後保留 30% 為測試資料集；在傳統模型訓練過程中，兩組資料集皆取其中約 70% 為訓練資料集及最後保留 30% 為測試資料集。

為求傳統模型、機器學習模型及深度學習模型間有公平比較，我們統一給與當前一天 (24 time steps) 資料，預測下一天 (24 time steps) 的需求分佈。以圖3.3 3.4為例：假設  $X_1, X_2, \dots, X_t$  分別代表為 2016 年 1 月 1 日 0 時、1 時、....、23 時，此時  $Y_1, Y_2, \dots, Y_t$  代表的會是 2016 年 1 月 2 日 0 時、1 時、....、23 時。

訓練模型使用 Adam 優化器 [12]，能有較好的處理稀疏梯度及噪音樣本的效果。在訓練過程中有使用 Dropout 層設定比例為 0.7 避免過度擬和的情況發生。損失函數3.4 式  $\gamma$  經反覆實驗後給定 0.01 得出較佳效果。

## 4.3 實驗對照模型與方法介紹

### 4.3.1 歷史平均

我們將過去每天同一時間的資訊的需求分佈，將同時間間隔的需求數量取平均。我們使用歷史平均的結果當作測驗的基準模型。

### 4.3.2 ARIMA

ARIMA 模型為傳統時間序列知名的模型之一。ARIMA(p, d, q) 預測方式其中使用 3 個參數來處理，分別為 p 自回歸項數，q 移動平均項數，d 為穩定序列所做的差分階層次數。過去常用於預測連續時間變化

的研究。參數  $p$ 、 $q$ 、 $d$  我們使用自動尋找參數方法 [13]，在其任務上尋找合適參數擬配出最佳的結果。

### 4.3.3 XGboost

XGboost 是一種梯度提升決策樹 (gradient boosting decision tree)，被使用於分類及回歸問題。Gradient boosting 就是通過新加入的弱學習器，更新之前所有弱學習器的殘差，最終將數個學習器相加在一起用來進行最終預測。預測有極佳的效果以及極快的運算速度，常被用於許多資料科學的競賽中。

### 4.3.4 線性回歸 (Linear Regression)

線性回歸是利用多個自變數 (independent variable) 和依變數 (dependent variable) 之間其關聯性所建立的模型。我們使用多元回歸 (multiple regression) 的方式進行預測，我們使用 Ridge 的算法得到較佳的結果。

### 4.3.5 DMVST-Net

我們比較同樣使用深度學習架構的 DMVST-Net 模型。此模型將時間點上的地理區塊資訊利用 CNN 的卷積層萃取重要的區塊資訊，經過 CNN 處理好的資訊利用全連接層至 LSTM 網路，將不同的連續時間的資訊利用 LSTM 處理，最後由全連接層預測下個時間點。

## 4.4 評量指標

我們使用 Mean Absolute Percentage Error(MAPE) 和 Root Mean Square Error(RMSE) 作為評量指標。其定義如下：

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|Y_t^i - \hat{Y}_t^i|}{Y_t^i + c} \quad (4.1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_t^i - \hat{Y}_t^i)^2} \quad (4.2)$$

其中， $Y_t^i$  和  $\hat{Y}_t^i$  分別代表實際分布於當  $t$  時間時  $i$  地理區塊的真實與預測之載客需求， $N$  代表所有資料的筆數。4.1 式中的分母加上微小的常數避免當分母為零的情況。

## 4.5 實驗結果

表 4.2: 不同模型間的效能 (第一組資料集)

模型	RMSE	MAPE	RMSE 的標準差	MAPE 的標準差
歷史平均	8.845	0.0840	7.9434	0.000413
線性回歸	10.914	0.1460	2.4451	0.000895
ARIMA	15.585	0.1660	20.8253	0.018033
XGboost	6.498	0.0806	2.0542	0.000205
LSTMs (2 層)	7.037	0.0563	3.9747	0.000056
LSTMs (4 層)	6.694	0.0595	5.1110	0.000232
DMVST-Net	7.350	0.0643	3.7034	0.000192
Residual-LSTMs (4 層)	5.187	0.0584	2.0265	0.000048
<b>AR-LSTMs (4 層)</b>	<b>4.958</b>	<b>0.0488</b>	<b>1.8909</b>	<b>0.000039</b>



表 4.3: 不同模型間的效能 (第二組資料集)

模型	RMSE	MAPE	RMSE 的標準差	MAPE 的標準差
歷史平均	11.882	0.2850	29.5423	0.001110
線性回歸	11.836	0.3436	20.8427	0.022740
ARIMA	20.754	0.815	10.7848	0.796237
XGboost	11.338	0.2938	21.9224	0.003506
LSTMs (2 層)	11.466	0.2791	22.8266	0.001683
LSTMs (4 層)	21.595	0.821	36.6752	0.903635
DMVST-Net	11.828	0.3178	21.5800	0.017413
Residual-LSTMs (4 層)	13.614	<b>0.2688</b>	42.7176	<b>0.000596</b>
<b>AR-LSTMs (4 層)</b>	<b>11.273</b>	0.2742	<b>20.6249</b>	0.003008

### 4.5.1 實驗結果分析

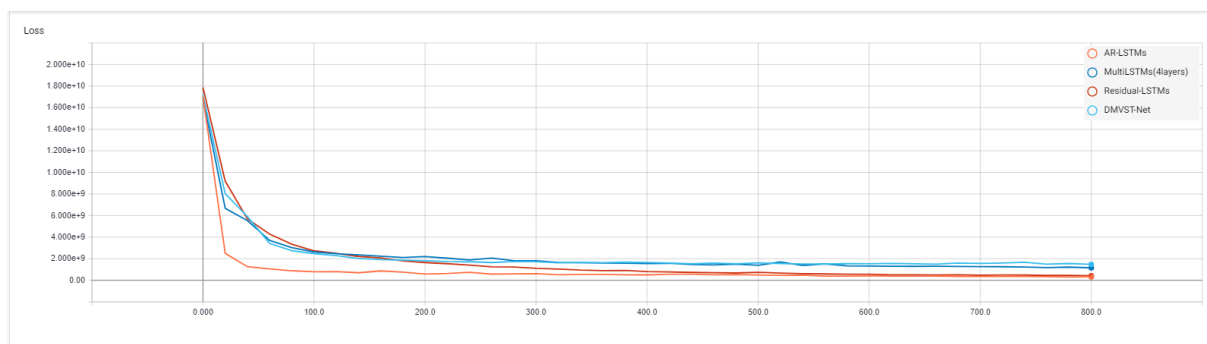


圖 4.1: Loss 變化圖

我們選用數個不同類型的模型進行比較，包含我們基準歷史平均、傳統時間序列的 ARIMA、機器學習的線性回歸模型模型、預測能力強大的決策樹 XGBoost、使用深度學習的 DMVST-Net 模型以及我們過去研究提出的 Residual-LSTMs。我們使用的評估指標為 RMSE 與 MAPE，從表4.2中說明多層式 LSTMs 在 RMSE 及 MAPE 上有很好的結果。進一步觀察我們於先前研究所提出的 Residual-LSTMs (4 層) 預測模型與相同層數的 LSTMs (4 層)，實驗結果發現當加上保留連結後提升很大的預測準確度 (表4.2 4.3)，因此我們利用此機制有效解決深層深度學習上梯度流向問題，同時於其他模型的比較中得到極佳的預測結果。

我們進一步比較本篇論文提出的 AR-LSTMs 與所有模型，不論是傳統、機器學習、決策樹及其他深度學習模型，從實驗結果中可以證明 AR-LSTMs 得到的效果優於其他模型。其中第二組資料集的地理網格數較少，各模型的結果相較於第一組資料集顯得接近。但由表4.3說明說我們的模型仍具領先

AR-LSTMs 在注意力機制上我們選用多頭式注意力，利用這樣機制可以將高維度的輸入資料切分成數個低維度資料進行運算，藉此能透過平行化方式加快運算速度，當訓練模型過程中即顯示出它的效果。我們提出的 AR-LSTMs 模型與其他三者深度學習模型在訓練過程觀察損失函數及驗證資料集的評量指標變化，包含 LSTMs (4 層)、DMVST-Net 及 Residual-LSTMs。第一組資料集相比第二組資料集維度大，我們利用第一組資料集觀察不同模型間的 Loss 變化，圖 4.1 顯示我們模型收斂速度最快，減少約 30-60% 迴圈數，時間約快 10-40%，驗證我們提出的模型在效率上的優勢。

我們觀察預測的需求量分布確認 AR-LSTMs 是否能有效的學習空間分布資訊，圖 4.2 4.3 4.4 4.5中顯示真實與預測的需求量預測，可以發現兩者不論位置及數量都很接近，證明我們提出的預測模型是可運行於現實產業上。

進一步確認 AR-LSTMs 是否能有效的學習到時間資訊與空間分布的相對應關係，觀察圖 4.2 4.3 4.4 4.5中發現預測結果和真實狀況的時間以及地理區塊上數量變化趨勢十分相近，證明我們提出的預測模型的確能學習到時間與空間上交互變化的關聯。

另外我們觀察到傳統善於處理時間序列議題的 ARIMA 模型在最後結果中得到不佳的結果。因為 ARIMA 擅於處理較長週期的資訊，此實驗屬於每小時間的變化相較於一季或月屬於高頻率的變化，我們推論在此任務的預測結果中數個區塊常無法有效呈現 24 小時週期變化。此外，ARIMA 模型僅能單一輸入及單一輸出，與多對多的模型不同，無法藉由觀察其他區塊的變化結合周邊區塊資訊進行預測。綜合上述兩點，

ARIMA 模型在此任務上沒辦法成功的預測。

透過兩組資料集測試說明我們模型不侷限於單一的資料集，說明可預測不侷限於特定的地區，而且在所有預測模型中得到最佳的結果，我們模型將可以被交通產業廣泛運用於預測地區即時的需求分布。

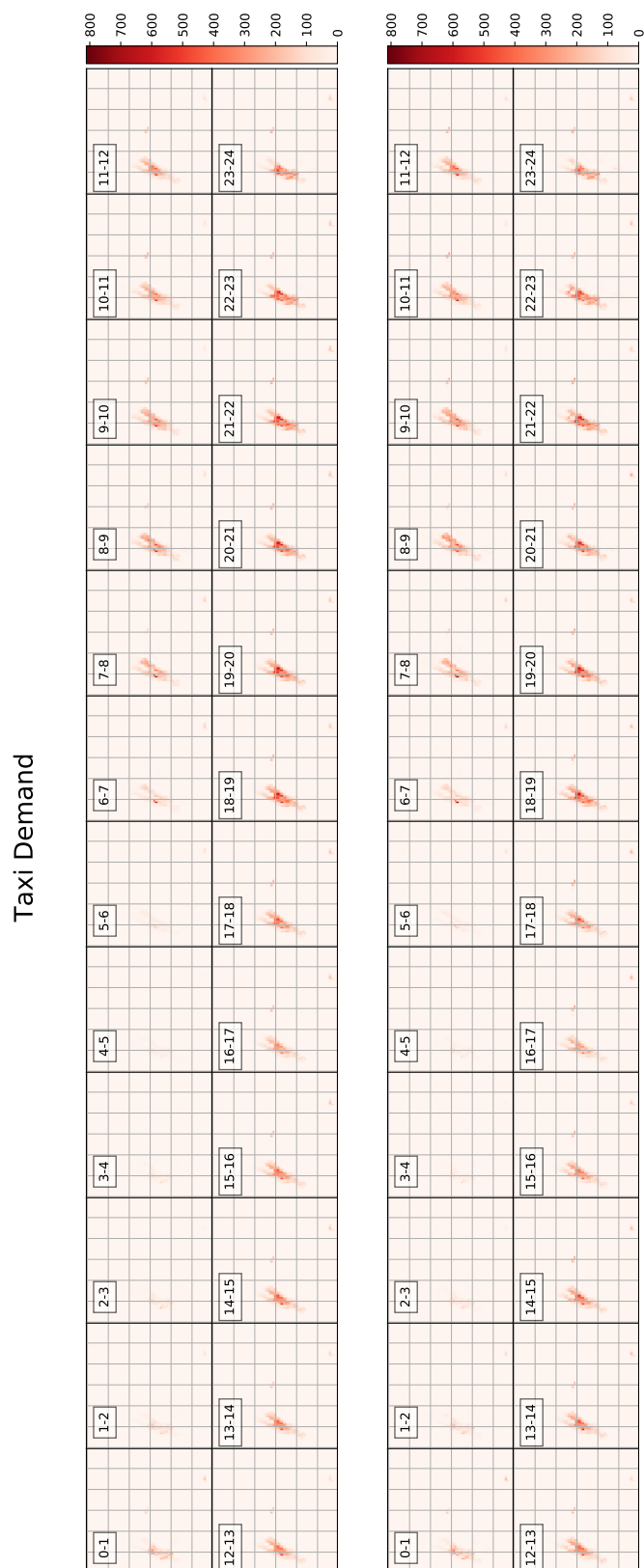


圖 4.2: 第一組資料集：2016/6/15(平日) 預測結果。上半部為真實需求；下半部為預測需求。

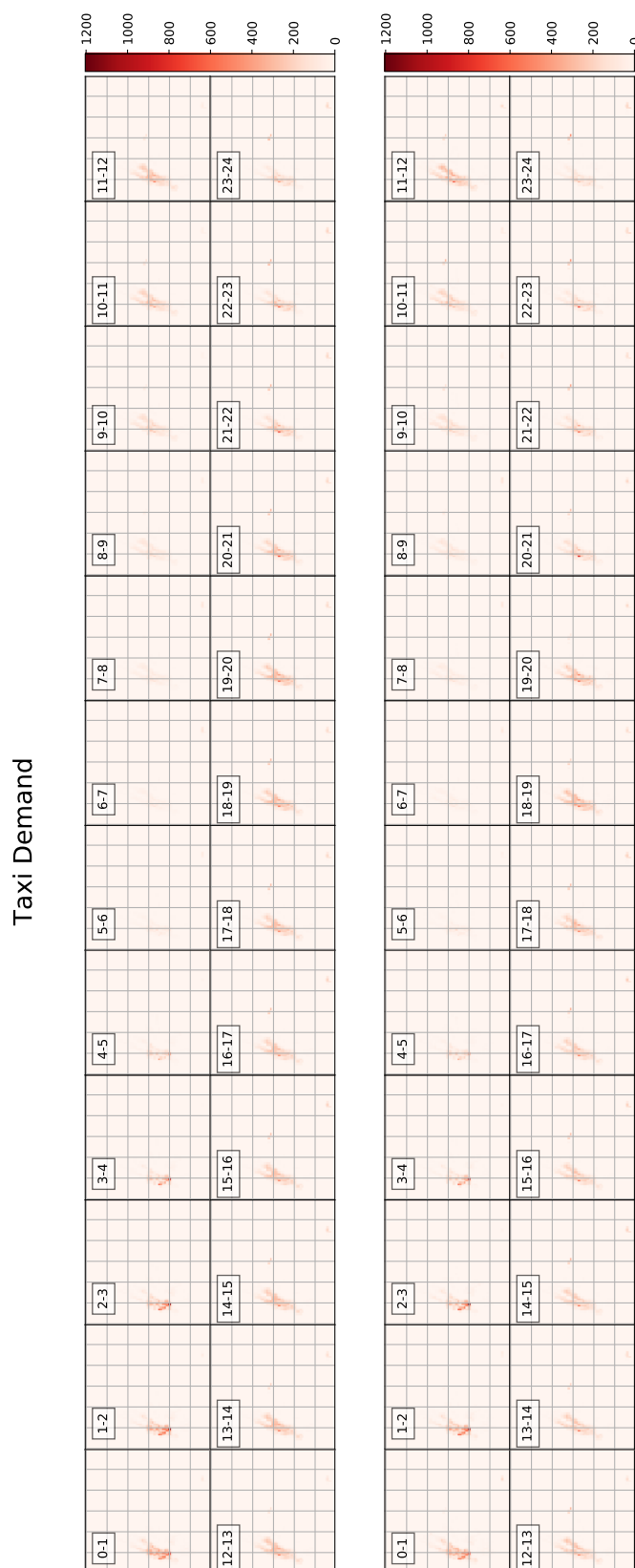


圖 4.3: 第一組資料集：2016/6/5(假日) 預測結果。上半部為真實需求；下半部為預測需求。

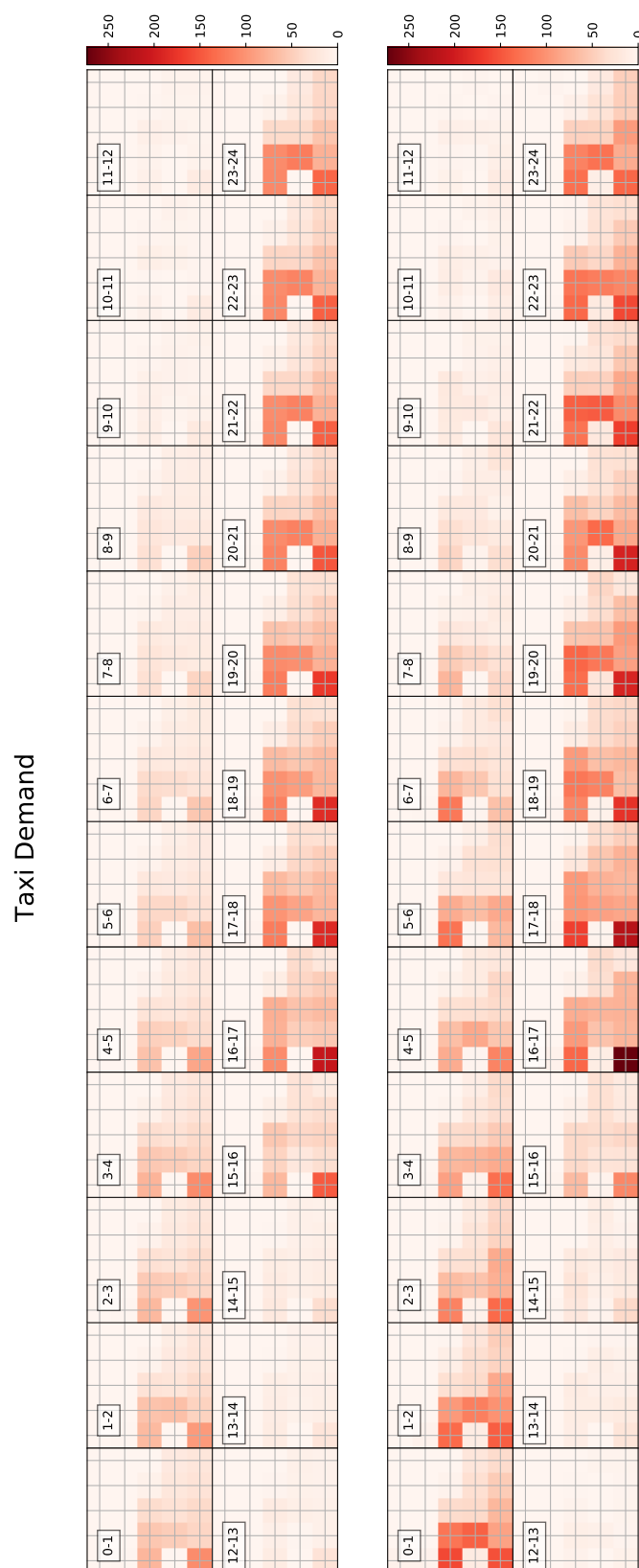


圖 4.4: 第二組資料集：2017/1/24(平日) 預測結果。上半部為真實需求；下半部為預測需求。

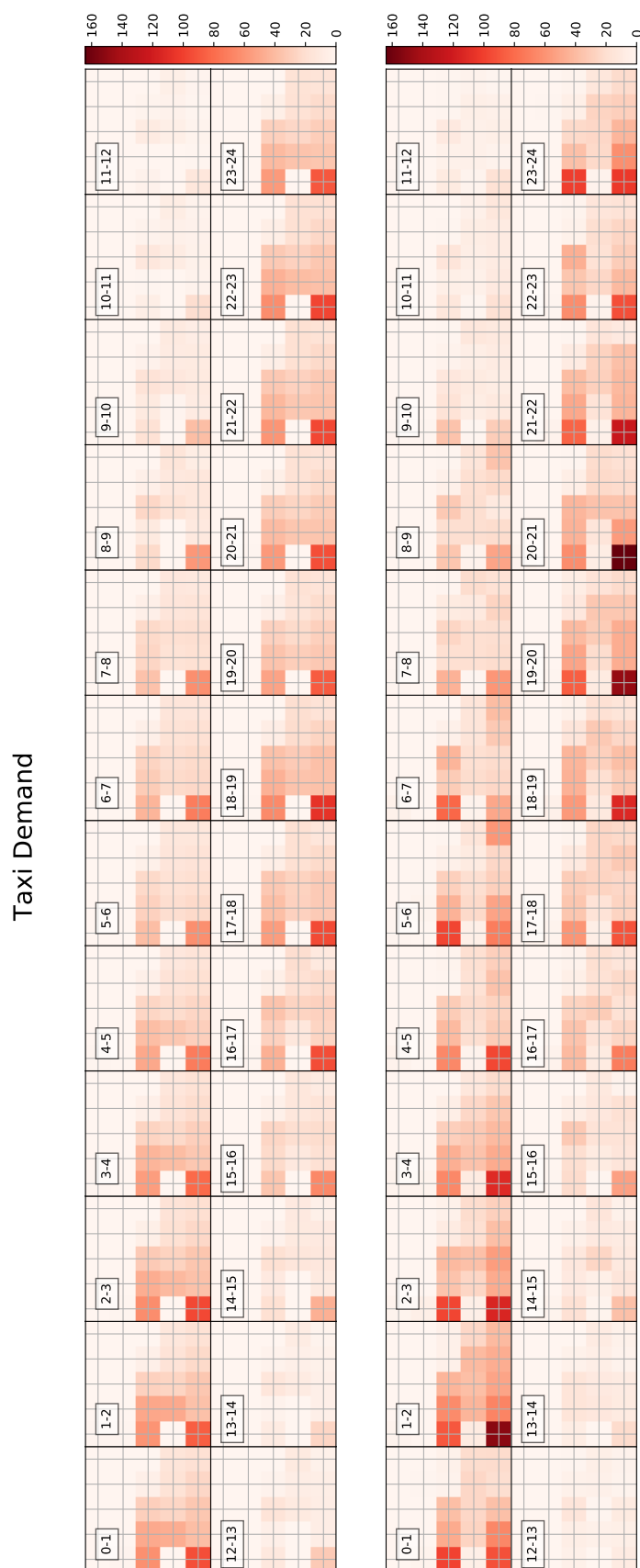


圖 4.5: 第二組資料集：2017/1/15(假日) 預測結果。上半部為真實需求量；下半部為預測需求量。

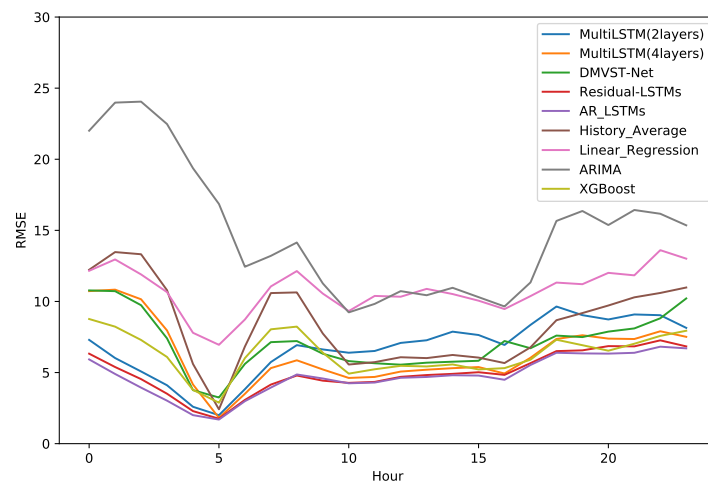
### 4.5.2 不同時間區段比較

在交通需求預測的議題上尖峰與離峰時間上需求急遽變化加劇準確預測的困難度。過去計程車需求的研究中 [8] [10] 可以發現其結果仍有明顯尖峰與離峰時間上的準確度偏差。我們運用注意力機制強化長時間前後對應關係，進而模型能獲取完整長時間資訊，用以平緩尖峰及離峰間準確度的差異。

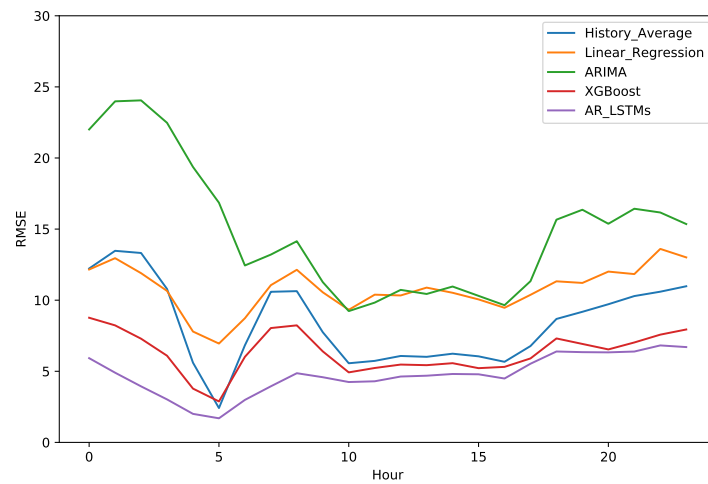
我們藉由標準差評估 RMSE 及 MAPE 以確認是否有效降低尖峰與離峰間的準確度差異，從表4.2 觀察得出我們模型確實有效地平緩 RMSE 與 MAPE 變化，進一步觀察其圖4.6視覺化結果，說明我們提出的 AR-LSTMs 模型的確相較於其他深度學習模型有效解決尖峰與離峰的預測問題。

接著我們觀察第二組資料集，從表 4.3觀察到 AR-LSTMs 模型的 RMSE 標準差比落後於 ARIMA 模型，然而從圖 4.8仔細觀察 ARIMA 模型狀況，可以了解 ARIMA 模型並沒有有效預測正確的需求量，整體而言效能仍遠不及 AR-LSTMs 模型。由於第二組資料集 (網格數量共 5\*5 格) 複雜程度遠低於第一組資料集 (網格數量共 60\*60 格)，圖中圖4.8 4.9各模型較為接近，從表 4.3 觀察在此資料集中 RMSE 及 MAPE 的標準差，我們所提出的 AR-LSTMs 模型部分評估指數相較於大多數模型平緩，在整體的預測結果中從表4.3觀察準確度我們仍具明顯優勢。

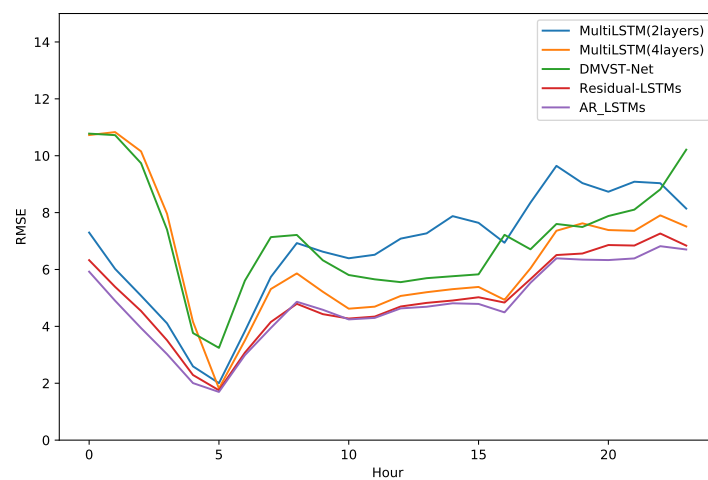




(a) 24 小時區段之 RMSE 變化比較

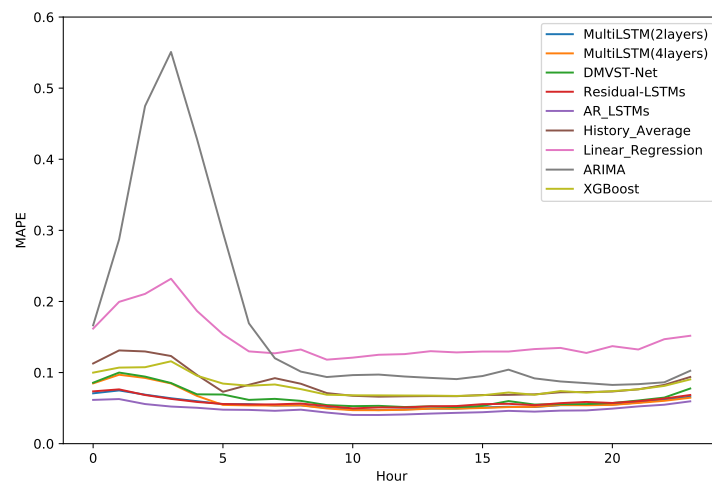


(b) 24 小時區段之 RMSE 變化比較 (傳統模型)

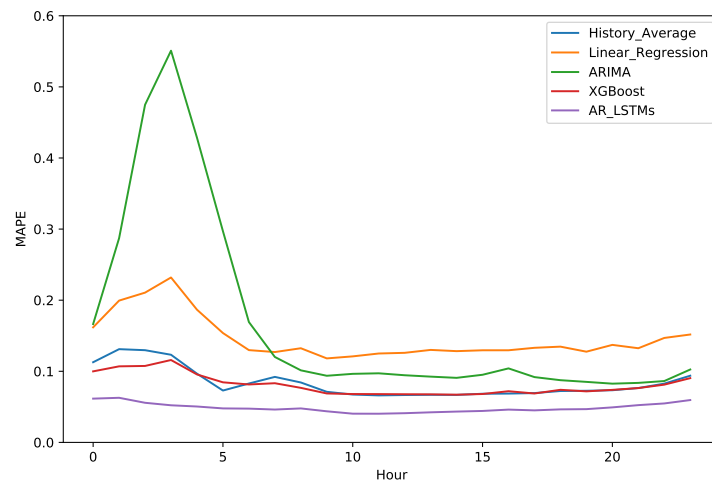


(c) 24 小時區段之 RMSE 變化比較 (深度學習模型)

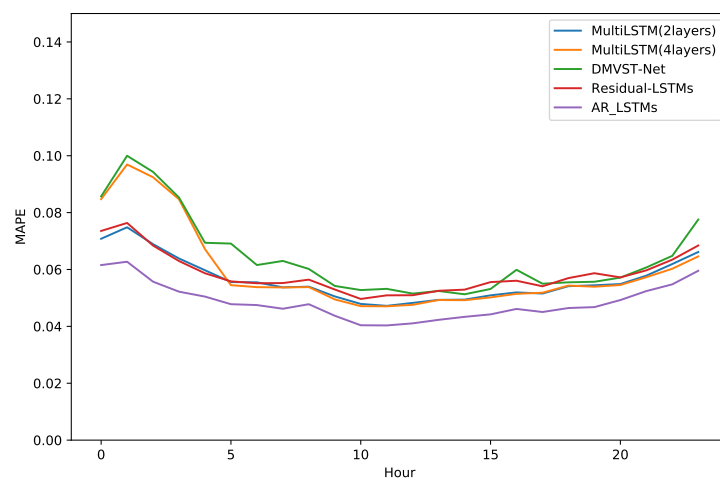
圖 4.6: 第一組資料集：24 小時區段之 RMSE 變化比較



(a) 24 小時區段之 MAPE 變化比較

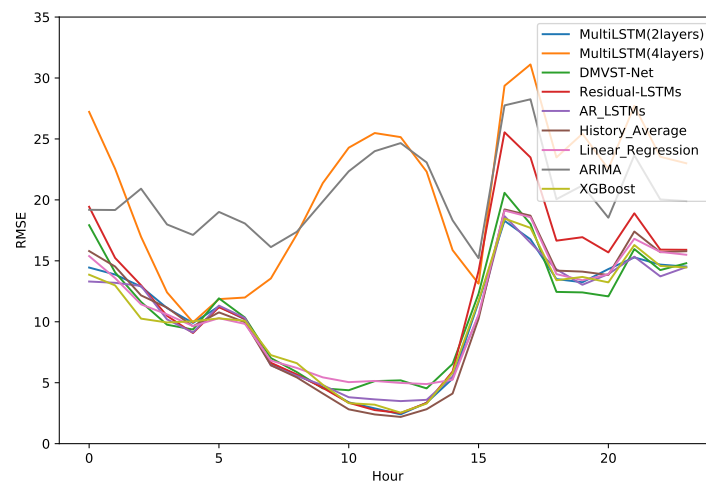


(b) 24 小時區段之 MAPE 變化比較 (傳統模型)

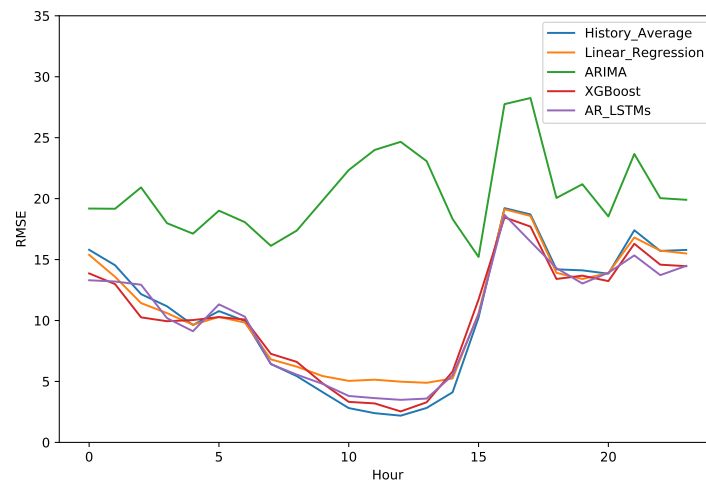


(c) 24 小時區段之 MAPE 變化比較 (深度學習模型)

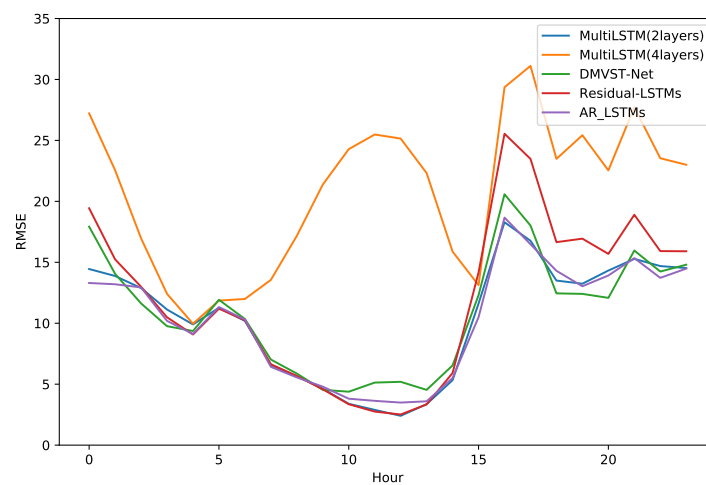
圖 4.7: 第一組資料集：24 小時區段之 MAPE 變化比較



(a) 24 小時區段之 RMSE 變化比較

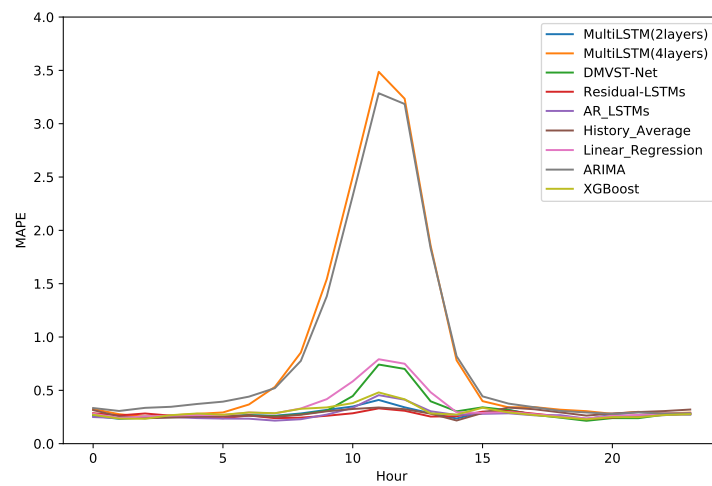


(b) 24 小時區段之 RMSE 變化比較 (傳統模型)

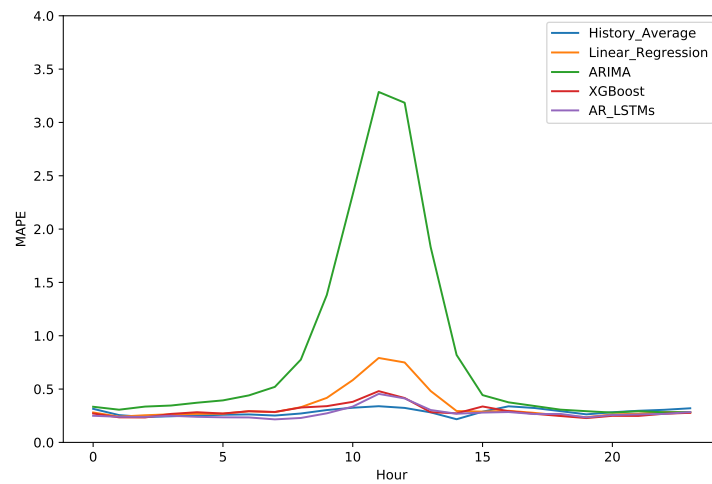


(c) 24 小時區段之 RMSE 變化比較 (深度學習模型)

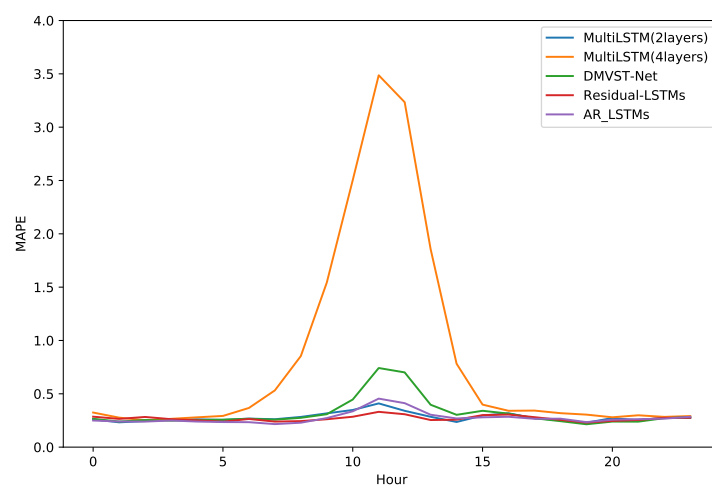
圖 4.8: 第二組資料集: 24 小時區段之 RMSE 變化



(a) 24 小時區段之 MAPE 變化比較



(b) 24 小時區段之 MAPE 變化比較 (傳統模型)



(c) 24 小時區段之 MAPE 變化比較 (深度學習模型)

圖 4.9: 第二組資料集：24 小時區段之 MAPE 變化

## 五、 相關研究

交通的需求預測一直以來對於市場上是十分有價值，人們希望得知下個時間的乘客需求位於哪裡。過去很多研究處理交通時間序列上問題，其中整合移動平均自回歸模型 (ARIMA) [14] 是最具代表性的預測時間序列問題的傳統模型之一。

近幾年來深度學習進行預測的方式日益成熟，常見模型之一的長短期記憶模型 (LSTM)，可以保留時間序列中持續長時間的資訊，其 LSTM 區塊中透過閘道的機制可以決定輸入內容重要性，適合用於處理時間序列上的問題。同樣為常見模型之一的卷積神經網路 (CNN)，由數個全卷積層、池化層及全連接層組成，能保留輸入數據結構位置上的資訊，在圖像處理上得到很好的結果 [15]。

因此近年來研究嘗試利用交通資料結合深度學習進行預測，運用車用載具所記錄的行駛間各項數據，將大數據結合深度學習網路強大預測能力，進行交通需求預測模型的相關研究。

有人提出使用 LSTM 處理空間及時間資訊的 LSTM-MDN-Conditional [10]，以及使用 LSTM 處理時間資訊及 CNN 處理空間資訊的 DMVST-Net [8]，兩者模型上皆獲得比傳統方式更好的結果，說明深度學習的框架在處理交通問題上是十分的適用。接著詳細介紹這兩種深度學習模型的架構：

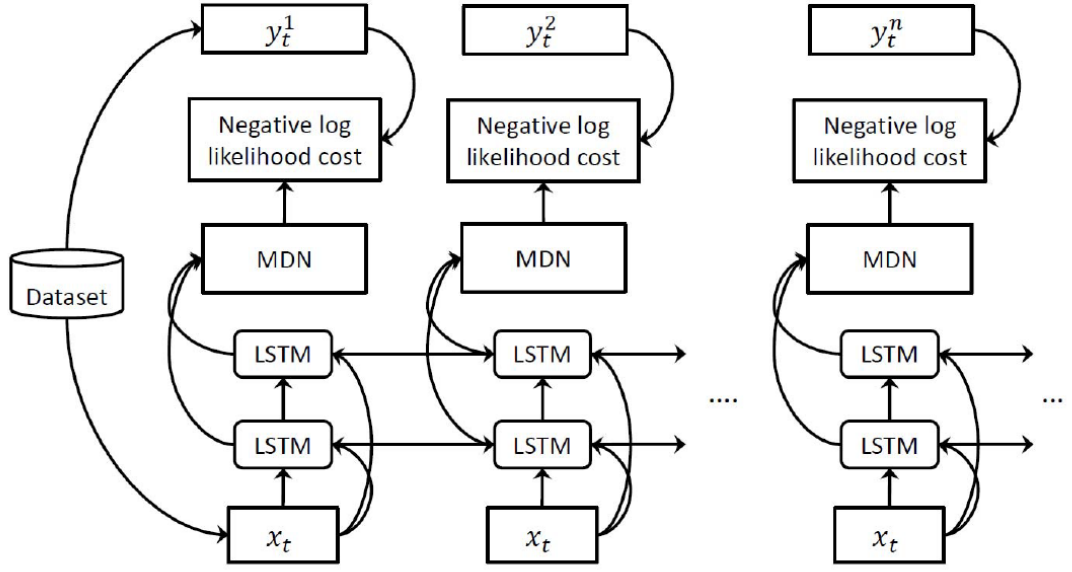


圖 5.1: LSTM-MDN-Conditional 架構圖

## 5.1 LSTM-MDN-Conditional

此篇作者在設計過程中考量計程車需求預測是連續時間的議題，牽涉的資訊不只是過去的需求，同時也包含未來的需求。

此模型以選用 LSTM 為預測模型的遞回類神經網路單元以解決上述議題，將模型的架構中考量現在時間點之前的相關資訊，使用條件式 (Conditional) 機制，該機制類似 Pixel RNNs 的方式，將預測結果依序一格一格像素生成，取代透過聯合機率分布 (joint probability) 直接生成時間點上單一全部區域的結果。

在實驗結果呈現上他們得到不錯的結果，以及成功地預測出下個時間點的乘客需求分布。

## 5.2 DMVST-Net

該篇論文除了考量計程車需求預測是連續時間的議題，同時強調空間上的議題，如何萃取空間上資訊以達到有效預測為此篇特色之處。

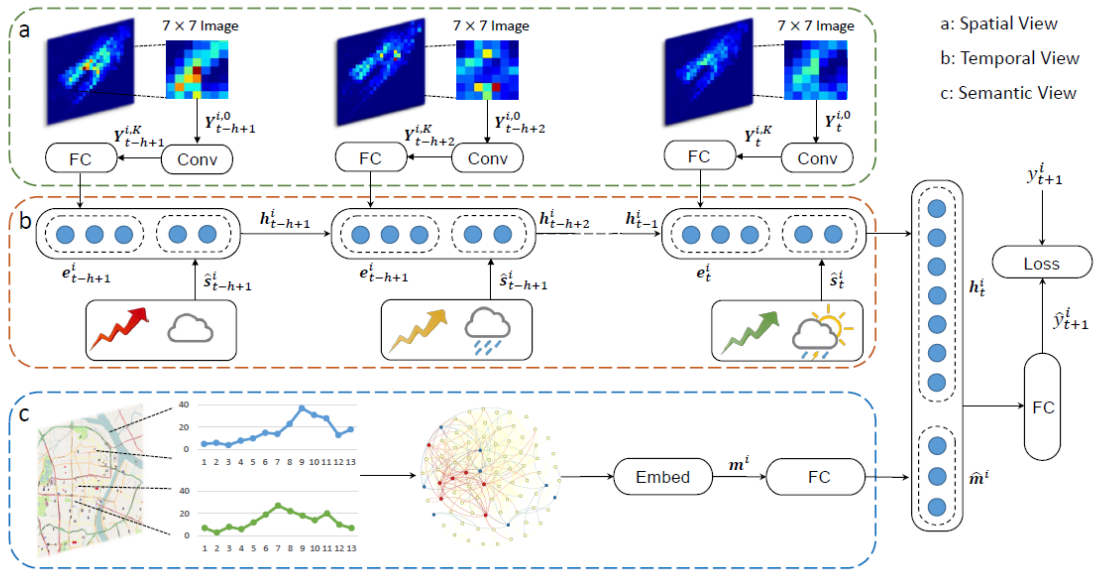


圖 5.2: DMVST-Net 架構圖

此模型將不同的連續時間的資訊利用善於處理時間序列的 LSTM 處理，時間點上的地理區塊資訊利用善於處理空間上資訊的 CNN 的卷積層萃取重要的區塊資訊。利用這樣的架構可以有效地將高維度的地理區塊資料降至較低維度，並且在訓練過程中依然保留部分地理對應位置的資訊。

在實驗結果呈現上他們在空間上的表現得到很好的效果，他們結合預測該區塊其他資訊，包含地區的交通變化路徑圖、天氣及乘客需求趨勢等，模型處理複數混合資訊使得結果優於其他模型。

## 六、 結論及未來展望

### 6.1 結論

我們提出的模型已經能在訓練過程中學習真實載客需求的分布，能有效的預測下個時間點的載客需求情況，同時我們也解決過去研究所無法有效解決的尖峰區段準確度不高的問題。

過去相關研究的方法無法有效解決尖峰時間及離峰時間的變化差距，在其他相關研究中可以發現當早上 4 點至 8 點的通勤時間評估指標 RMSE 及 MAPE 會顯示最高峰會落在此區間。我們使用注意力機制成功地去降低此點的 RMSE 及 MAPE，且在預測平滑度更勝其他模型，有效解決尖峰時段預測提高準確率的議題，尖峰時間相較於離峰時間預測準確的需求量對於交通領域方面是極為重要。

在實驗結果中也證明在深層的 LSTMs 通常會得到比單層的 LSTM 更加的準確度的論述，以及利用保留連結解決梯度爆炸或消散的問題，因此我們提出的 AR-LSTMs 預測模型基於多層的 LSTM 其結果相較於其他模型得到更佳分數。

### 6.2 未來展望

不過在實驗結果中也發現，圖層中少量的叫車需求容易在模型學習的過程中消失，導致預測的結果會偏向熱門的需求地點，是否要保留少數的載客位置，現實市場中是值得探討的議題，未來研究的方向可以朝



向如何去保留一些少量載客的地點資訊，以達到預測的位置更為精確目標。

我們的研究僅使用時間因子及乘客上車的地點進行預測已達到很好的效果，相信在未來如果整合其他額外的資料包含天氣、交通路網及乘客司機資訊將能達到更好的效果。

## 參考文獻

- [1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” pp. 1097–1105, 2012.
- [2] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [3] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” pp. 770–778, 2016.
- [4] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [5] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, *et al.*, “Google’s neural machine translation system: Bridging the gap between human and machine translation,” *arXiv preprint arXiv:1609.08144*, 2016.
- [6] R. Pascanu, T. Mikolov, and Y. Bengio, “Understanding the exploding gradient problem,” *CoRR*, *abs/1211.5063*, vol. 2, 2012.
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” pp. 5998–6008, 2017.
- [8] H. Yao, F. Wu, J. Ke, X. Tang, Y. Jia, S. Lu, P. Gong, J. Ye, and Z. Li, “Deep multi-view spatial-temporal network for taxi demand prediction,” 2018.
- [9] 徐志榮 and 陳弘軒, “多層式短中長期記憶模型之即時計程車需求預測,” *Conference on Technologies and Applications of Artificial Intelligence*, 2018.
- [10] J. Xu, R. Rahmatizadeh, L. Bölöni, and D. Turgut, “Real-time prediction of taxi demand using recurrent neural networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2572–2581, 2017.
- [11] NYC Taxi Limousine Commission and others, “Taxi and limousine commission(tlc) trip record data,”
- [12] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

- [13] R. J. Hyndman, Y. Khandakar, *et al.*, “Automatic time series for forecasting: the forecast package for r,” no. 6/07, 2007.
- [14] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, “Predicting taxi-passenger demand using streaming data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, pp. 1393–1402, 2013.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” pp. 1097–1105, 2012.

# 附錄一

實驗程式碼：<https://github.com/ncu-dart/AR-LSTMs>

## 1. main.py

此程式為建構並執行完整的類神網路。

## 2. modules.py

此程式提供各種深度學習運算單元元件，包含卷積層、注意力層及長短期記憶層。

## 3. process\_data.py

此程式將原始資料處理成深度學習可以接受的資料格式。

## 4. visualization.py

此程式視覺化曲線圖。

## 5. visualization\_grid.py

此程式視覺化地理網格圖。