台灣自來水股份有限公司

108 年度

LSTM 深度學習模型之應用並與 ARIMA 比較-以板新廠基本電價最適契約容量之訂立為例

研究單位：第十二區管理處會計室

研究人員：高宇正

研究期間：108 年 1 月至 108 年 7 月

內容

壹、 前言

第十二區管理處(以下簡稱本區處)係本公司為因應大台北都會區，工商迅速發展，人口急劇增加，供水需求量大幅成長及加強用戶服務，應地方政府及民眾之要求，於民國77年10 月24日由第二區管理處劃出成立，轄區包括新北市板橋、土城、三峽、鶯歌、樹林、新莊、泰山、五股、蘆洲、八里等十個區及三重、中和部份地區，並透過尖山加壓站支援二區處之供水。截至108年上半年底用戶數計840,508戶，轄區供水人口數為2,094,643人，供水普及率為99.14%。

近年來隨著板二計畫完成，向臺北市自來水事業處(以下簡稱北水處)購買清水量日益增加，從107年平均每日43.7萬噸，到本年度截至6月平均每日50.2萬噸，清水成本大幅提高，造成本年度預估至年底盈餘僅2億5千萬元至3億元，較107年盈餘5億2千萬元相比之下大幅下降，故如何提高收入及降低成本將日益重要。

本人認為要如何有效降低成本，可以分為兩個大原則，一是在數量無法調低之情形下想辦法拉低單價(如必要之機修費，透過多家廠商詢價後找出最低價修理)；二是評估是否

能降低委外數量來減少成本(如一些外包費用，評估是否有多餘人力可以自行負擔而不需委外)。

而先前 106 年應用統計報告「板新廠基本電價最適契約容量之訂立」中，為找出一項費用中之可控因素，該因素可透過人為因素去改變，且其難度並不高，只需在年初時改變其合約的定價即可之費用，以期透過人為管控達最佳節省支用狀態。板新廠的基本電價應屬於數量無法調低而想辦法拉低單價的項目，故如果能成功預測下一個月之最高需量，可能就能訂出最適契約容量，而該契約容量不會超約且避免因容量訂太高而浪費基本電價。

在 106 年應用統計報告「板新廠基本電價最適契約容量之訂立」中，以 ARIMA 時間序列預測模型利用歷史當月最高需量資料來預測下一個時間點之當月最高需量，而透過評估模型指標 MAPE(平均絕對值誤差率)獲得 20.49%，依 Lewis 訂定之相關評估標準位於 20%~50%之範圍，為合理的預測模型，雖然如此 20.49%預測力仍略顯不足，探其原因 ARIMA 模型在對於非線性時間序列預測力較差，而對線性時間序列預測力較好。經查板新廠當月最高需量除受制於當月水情外，亦常受到板二計畫水源調度政策而調整其出水量，故並沒有

明顯季節性因素而無法使用預測力更強的SARIMA。

綜上所述，是否能透過其他方法達到更好的 MAPE 值來做更好的預測呢?在目前大數據時代的來臨，大家都想透過人工智慧 AI 來創造更高的利潤，以機器學習、深度學習的方式去找出潛藏在數據中的有利訊息來幫助決策，尤其 2015 年 Goodgle 團隊 Tensflow 開源軟體庫、Keras 以 Python 編寫的開源神經網路庫出現後，許多人可以用簡單短短的程式碼就能輕易開發深度學習模型，在各式的免費軟體中，Python 因好學、程式碼簡潔有力而成為新一代資料科學家所使用的程式設計語言(R 語言一開始學習曲線較高，學習後可以用更短的程式碼來實現建模，但其語法的邏輯性並不符合一般程式語言的邏輯，而 Python 則因著重可讀性和易用性，使得它的學習曲線相對比較低且平穩)。故本次應用統計報告以 Python 之 Keras 深度學習庫中 LSTM 長短期記憶模型，用來預測板新廠下一期之當月用電最高需量，並且與 ARIMA 模型做比較，評估哪個模型預測力較佳，係為本報告研究之目的。

貳、 分析方法

運用 2007 年 8 月至 2019 年 6 月之板新廠電號 05-91-8820-11-7 之每月最高需量資料來預測 2019 年 7 月最高需量，先以 ARIMA 模型做預測並得出測試集的 MAPE，再利用 LSTM 長短期記憶模型做預測也得出測試集的 MAPE 及 RMSE，並比較兩者 MAPE 及 RMSE 的高低，以判斷哪個模型預測力較佳。

參、 分析內容

一、使用 ARIMA 差分自回歸移動平均模型進行時間序列預測

ARIMA 模型的整體名稱為差分整合移動平均自迴歸模型，又稱整合移動平均自回歸模型(移動也可稱作滑動)，時間序列預測分析方法之一，ARIMA(p，d，q)中，AR 是"自回歸"，p 為自回歸項數；MA 為"滑動平均"，q 為滑動平均項數，d 為使之成為平穩序列所做的差分次數（階數)。其實本模型係 AR(自迴歸模型)及 MA(滑動平均模型)的合體，再加上使非平穩的時間序列數據進行平穩化處理，合稱 ARIMA 模型，如經過平穩化處理後再進行之流程即為 ARMA 模型流程處理。

(一) AR(p)模型簡介(自迴歸模型 Autoregressive model)

要先了解 ARIMA 模型，就先了解最重要的 AR(p)(自迴歸模型)模型，其公式為：$y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t$，$a_0$ 為一常數之截距項，p 代表落後期數，y_t 代表第 t 期之用電最高需量，受到 $\sum_{i=1}^{p} a_i y_{t-i}$ (如 p 為 1 則為受到前 1 期 $a_1 y_{t-1}$ 影響，如為 2 則為前 1 期 $a_1 y_{t-1}$ 及前 2 期 $a_2 y_{t-2}$ 影響)及當期的 ε_t 誤差影響，意即自回歸模型描述的是當前值與歷史值、誤差之間的關係。

(二) MA(q)模型簡介(移動平均模型 Moving-Average model)

公式為:$y_t = a_0 + \sum_{i=1}^{q} b_i \varepsilon_{t-i} + \varepsilon_t$ ，a_0為一常數之截距項，q 代表落後期數，y_t代表第 t 期之用電最高需量，受到$\sum_{i=1}^{q} b_i \varepsilon_{t-i}$（如 q 為 1 則為受到前 1 期$b_1 \varepsilon_{t-1}$影響，如為 2 則為前 1 期$b_1 \varepsilon_{t-1}$及前 2 期$b_2 \varepsilon_{t-2}$影響)及當期的$\varepsilon_t$誤差影響，意即滑動平均模型描述的是自回歸部分的誤差累計。

(三) ARMA(p,q)模型簡介(自回歸滑動平均模型 Autoregressive moving average model)

該模型為 AR(p)模型及 MA(q)模型之合體，其公式為兩模型之合體: $y_t = a_0 + \sum_{i=1}^{p} a_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} b_i \varepsilon_{t-i}$ ，意即當期的數值受到過去歷史值、誤差值、當期誤差值…等等影響的一種函數關係。

(四) 自相關函數(ACF)及偏自相關函數(PACF)之介紹

1. 自相關函數(ACF)

描述時間序列觀測值與其過去觀測值之間的線性相關性。其函數式為:$R(s) = \frac{E[(y_i - u_i)(y_{i+s} - u_{i+s})]}{\sigma^2}$ ，其中u_i為 第 i 期之預測值，s 為落後期數。該函數圖形可以辨別我們所要觀察之時間序列數值是否為平穩,及決定 MA(q)中 q 的階數為何，產出之圖形如在第 s 期截尾，而 s 期後之 ACF 均於顯著水準中(95%信賴區間內)，且 PACF 圖為拖尾(數值漸漸消失)，則

決定 MA(q)模型，其 q=s。

2. 偏自相關函數(PACF)

描述在給定中間觀測值的條件下時間序列觀測值與其過去的觀測值之間的線性相關性，其函數式為：

$$\phi(s)=\frac{R_s-\sum_{j=1}^{s-1}(\phi_{s-2,j}-\phi_{ss}\phi_{s-2,s-j})R_{s-j}}{1-\sum_{j=1}^{s-1}(\phi_{s-2,j}-\phi_{ss}\phi_{s-2,s-j})R_j}$$ ，其中 R_s 就是第 s 期之自相關函數，s 同樣為落後期數，j 同樣與 i 都是代表 t 期。該函數圖形可以決定 AR(p)中 p 的階數為何，產出之圖形如在第 s 期截尾，而 s 期後之數值均於顯著水準中，且 ACF 圖為拖尾(數值漸漸消失)、指數衰減或振盪，則決定 AR(p)模型，其 p=s。

106 年應用統計報告以 2007 年 8 月至 2016 年 9 月資料為樣本內預測，而以 2016 年 10 月至 2017 年 9 月為樣本外預測，透過評估模型指標 MAPE(平均絕對值誤差率)獲得 20.49%，而本次因資料已更新至 2019 年 6 月，為了與 LSTM 深度學習模型做比較，本次再利用 2007 年 8 月至 2018 年 6 月為樣本內預測(訓練集)，而以 2018 年 7 月至 2019 年 6 月為樣本外預測(測試集)。另外因上次報告總處建議用更多 p、d、q 參數來調整 ARIMA 模型，以比較不同參數下模型之 MAPE，故本次將利用多種 ARIMA 模型做出預測，再選出最佳 MAPE 之

模型與 LSTM 深度學習模型做比較。

（五） 序列是否平穩-利用 ADF 檢驗

一段時間序列是否能預測，是希望這段時間序列的平均數及變異數盡量變化不要太大，如果時間序列的平均數及變異數常常變來變去，則對於下一時間點的預測將很難估測，故我們希望時間序列是能定態穩定的。簡單來說，檢驗序列的平穩性是時間序列分析的關鍵步驟。時間序列中很多估計量的統計特性都依賴於資料是否平穩。一般意義上，一個（弱） 平穩過程的期望、變異數和自相關係數應不隨時間變化。而對於時間序列是否有穩定，除透過折線圖來看之外，亦可以利用 ADF 單位根檢驗來檢驗出該序列是否平穩，其結果如下：(為樣本內預測與樣本外預測做比較，本檢驗針對 2007 年 8 月至 2018 年 6 月資料檢驗，以下 ARIMA 建模亦針對 2007 年 8 月至 2018 年 6 月資料做建模)

表一、ADF 單位根檢驗	
統計量 Test-statistic	-3.426231526676526
P-value 值	0.010096
1%臨界值	-3.4816817173418295
5%臨界值	-2.8840418343195267
10%臨界值	-2.578770059171598

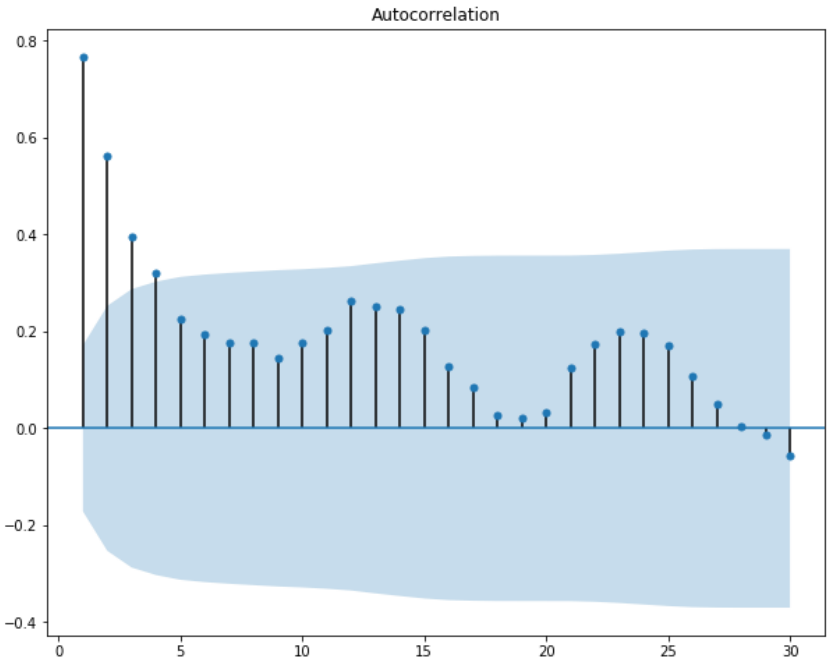
表一所要表達的意思是，統計量如果小於相對應信心程

度的臨界值的話，則可以該信心程度拒絕時間序列存在單位根的虛無假設，而本例統計量約為-3.426，小於 5%臨界值而大於 1%臨界值，表示我們有大於 95%~98%的信心(亦即 2007 年 8 月至 2018 年 6 月之板新廠電號 05-91-8820-11-7 每月最高需量之時間序列可以拒絕有單位根之存在)，表示該時間序列是定態恆定的(弱) 平穩過程，故在做 ARIMA 模型是可以不需要做差分的。

(六) 利用 ACF 圖及 PACF 圖判斷 p、q 之階數建模

決定 p、q 階數，慣例上是使用 ACF 圖及 PACF 圖來分別判斷 p 及 q 使用哪個階數，本次 ACF 圖及 PACF 圖如下:

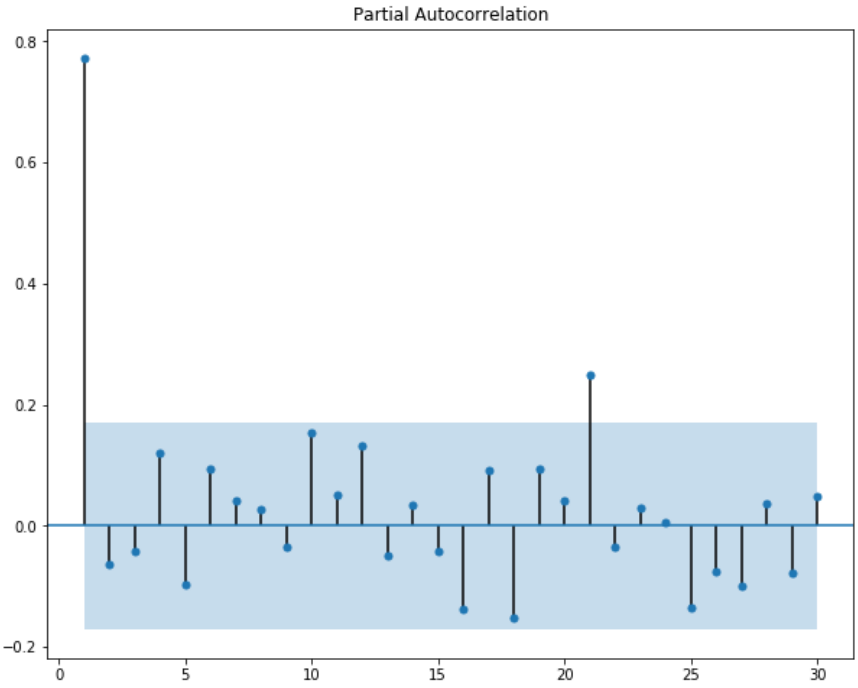
圖一、ACF 圖



圖表說明:X 軸表示與當期落後數，Y 軸表示相關係數，藍色色塊表示 95%信賴界線

由上圖看出,相關係數由第 5 階開始就已落入信賴區間,表示每一期的Y_t與 5 期(包含第 5 期)以後的落後數有 95%的信心是互相獨立的，而與前 4 期是有線性相關的，而 MA(q)是取決於 ACF 圖，故 q 可以取 0 至 4 試試看(取 0 之原因為只有用 AR(p)建模表現不一定比較差，也許更好)。

圖二、PACF 圖



本圖相關係數由第 2 階開始就已落入信賴區間，表示從每一期的 Y_t 與第 2 期(包含第 2 期)以後的落後有 95%的信心是互相獨立的，而與前 1 期是有線性相關的，而 AR(p)是取決於 PACF 圖，故 p 只取 0 至 1 應該就可以(取 0 之原因為只有用 MA(q)建模表現不一定比較差，也許更好)。

綜合以上條件,本次使用 ARIMA(1,0,0)、ARIMA(1,0,1)、

ARIMA(1,0,2)、ARIMA(1,0,3)、ARIMA(1,0,4)、ARIMA(0,0,1)、ARIMA(0,0,2)、ARIMA(0,0,3)、ARIMA(0,0,4)等來做建模，再利用 AIC(赤池信息量)、BIC(貝葉斯信息量)來選出最佳模型。

(七) 利用 AIC(赤池信息量)、BIC(貝葉斯信息量)選出最佳模型

以 ARIMA(1,0,0)、ARIMA(1,0,1)、ARIMA(1,0,2)、ARIMA(1,0,3)、ARIMA(1,0,4)、ARIMA(0,0,1)、ARIMA(0,0,2)、ARIMA(0,0,3)、ARIMA(0,0,4)等來做建模，再利用 AIC(赤池信息量)、BIC(貝葉斯信息量)來選出最佳模型，其結果如下：

表二、AIC、BIC 比較表

p	q	AIC	BIC
0	1	2529	2535
0	2	2436	2445
1	0	2059	2064
1	1	2059	2068
1	2	2059	2070
1	3	2051	2066
1	4	2053	2071

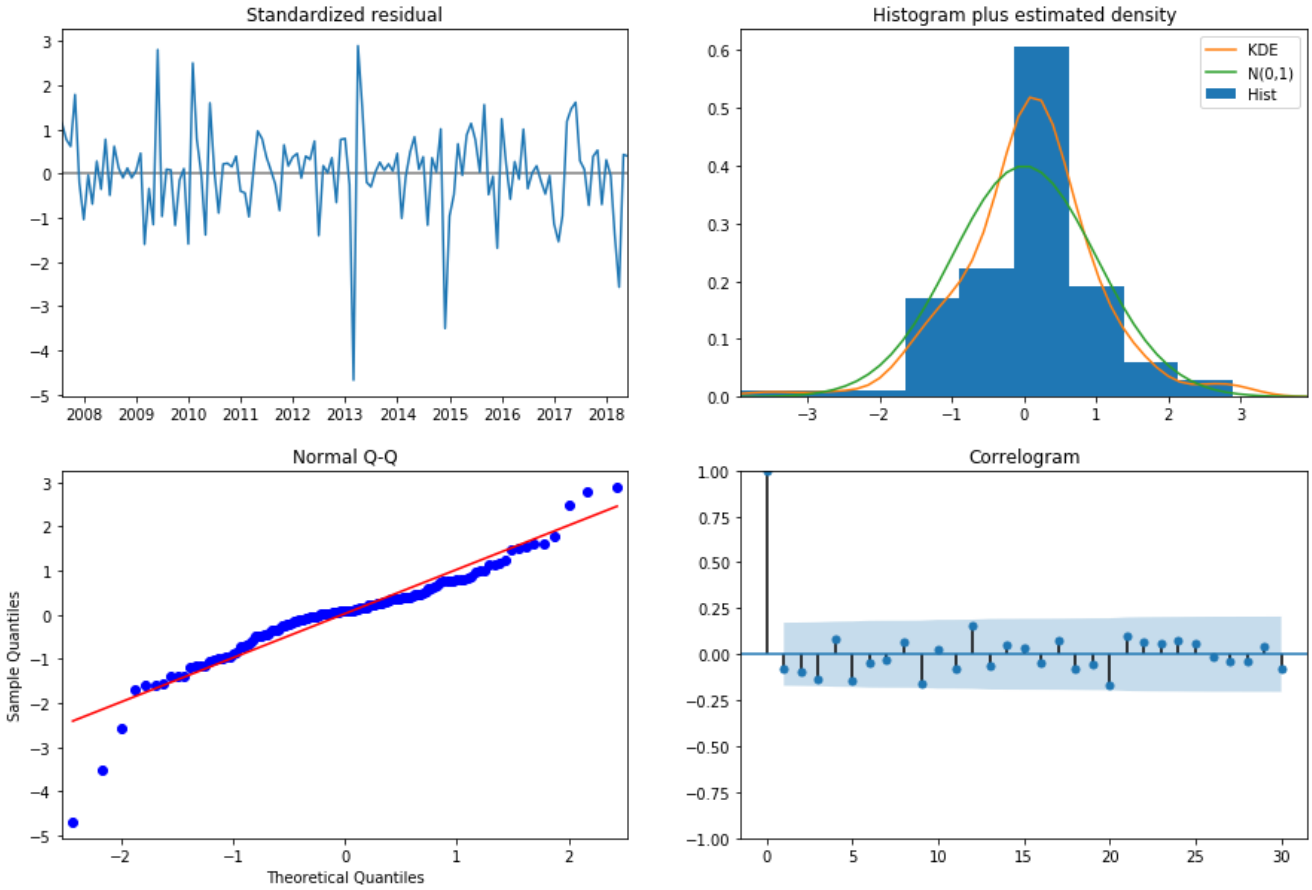
由表二可知，AIC 最低的為 ARIMA(1,0,3)，而 BIC 最低的為 ARIMA(1,0,0)，而究竟以哪一個標準來評估模型好壞是沒有絕對的標準，此兩者均除有評估模型的預測能力外，亦有懲罰項，懲罰參數多的模型而可避免「過度擬合」(Overfitting)，而 AIC 及 BIC 公式的差異主要是針對懲罰

項懲罰的程度不同，本例樣本數為 131 個，當樣本數大於等於 8 時，BIC 對參數數量多的模型懲罰較大，故 BIC 比較偏向參數數量較少之模型 ARIMA(1,0,0)，而 AIC 在預測能力與過度擬合中找出一個平衡而選出 ARIMA(1,0,3)模型，為保險起見，本次將兩個模型均列入以下 Ljung-Box Q 統計量檢定並且做後續預測。

（八） 殘差圖檢定

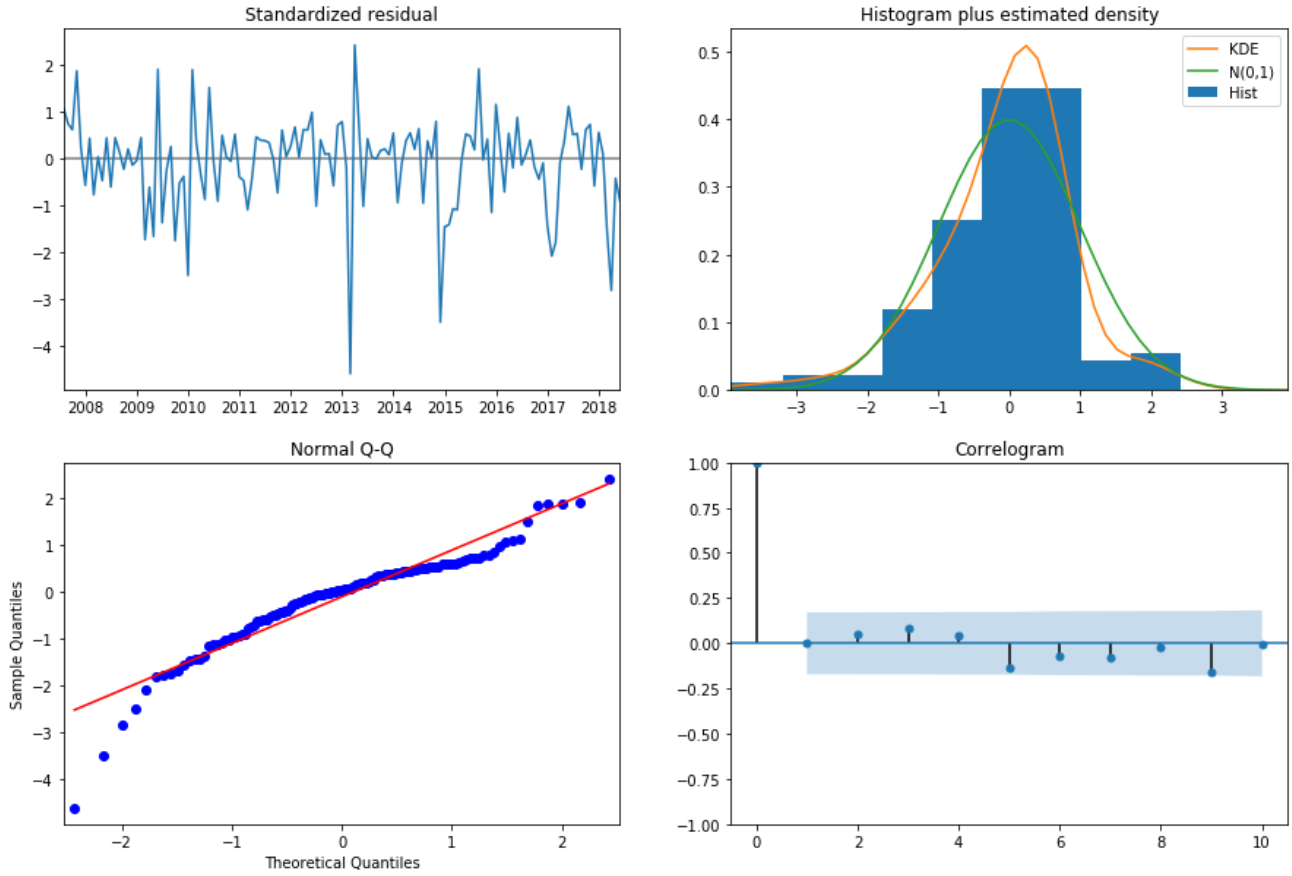
我們製作多元線性回歸分析，要假設誤差項為互相獨立且服從常態分配，換言之我們不希望線性 OLS 公式下誤差項間有「自我相關」，故本次透過觀察殘差圖，來檢測每個殘差間的自我相關係數是否能在信賴區間 95%內:

圖三、ARIMA(1,0,0)模型殘差圖：



由 ARIMA(1,0,0)的殘差自我相關圖(如圖三右下圖)，顯示在落後 1 期後就直接進入 95%信賴區間內(lag=1 至 lag=30 均於信賴區間內)，而直方圖(如右上圖)大致為常態分配，雖 Q-Q 圖(左下)看起來雖然有點歪歪的看仍可接受範圍(數據越靠近直線則越接近常態分配)，故殘差應可視為白噪音(長期下平均數為 0，誤差項為互相獨立且服從常態分配)。

圖四、ARIMA(1,0,3)模型殘差圖：



如上圖，除 Q-Q 圖稍為歪之外，其餘差不多跟 ARIMA(1,0,0)相同，故殘差應可視為白噪音(長期下平均數為 0，誤差間互相獨立且服從常態分配)。

綜上所述兩模型均通過殘差檢測(其實還有 Ljungbox 等檢定，惟本次主題並非 ARIMA 故省略，且透過殘差自我相關圖即可粗略檢驗)，故以下利用預測值與測試集做比較找出兩模型的 MAPE。

(九) ARIMA 模型 MAPE 及 RMSE 計算

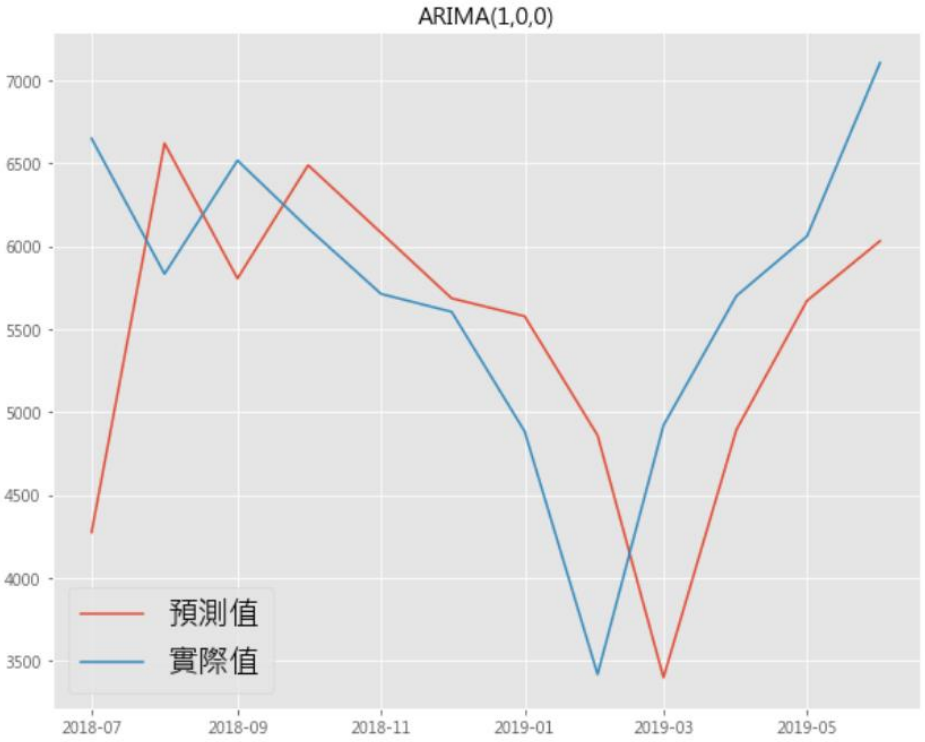
為了要評估模型的績效，除在訓練集(樣本內)上的績效

之外，亦要用測試集(樣本外)來評估，以避免有模型過度適應訓練集而在新樣本的表現很差，這就是要避免所謂的「過度擬合」(Overfitting)，故每次做預測模型都要將樣本切成「訓練集」及「測試集」，以訓練集訓練出來的模型去評估在測試集上的表現。而這次的 ARIMA(1,0,0)及 ARIMA(1,0,3)是以 2007 年 8 月至 2018 年 6 月資料訓練出來的，故本次 MAPE 是以該訓練出來之模型，評估 2018 年 7 月至 2019 年 6 月測試集。

另外因 ARIMA 模型對於下一個時間點的預測力是最準確，故本次測試集將以滾動式滾入模型做擬合，較能體現出模型表現。

另外本次再加入 RMSE(均方根誤差)來評估，公式為 $\sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}}$，RMSE 是一種常用的測量數值之間差異的量度，其數值常為模型預測的量或是被觀察到的估計量，因 RMSE 可以大概估測平均誤差值，以樣本之單位(瓩)來評估預測值與真實數據間相差之單位數，是目前有效的衡量工具。

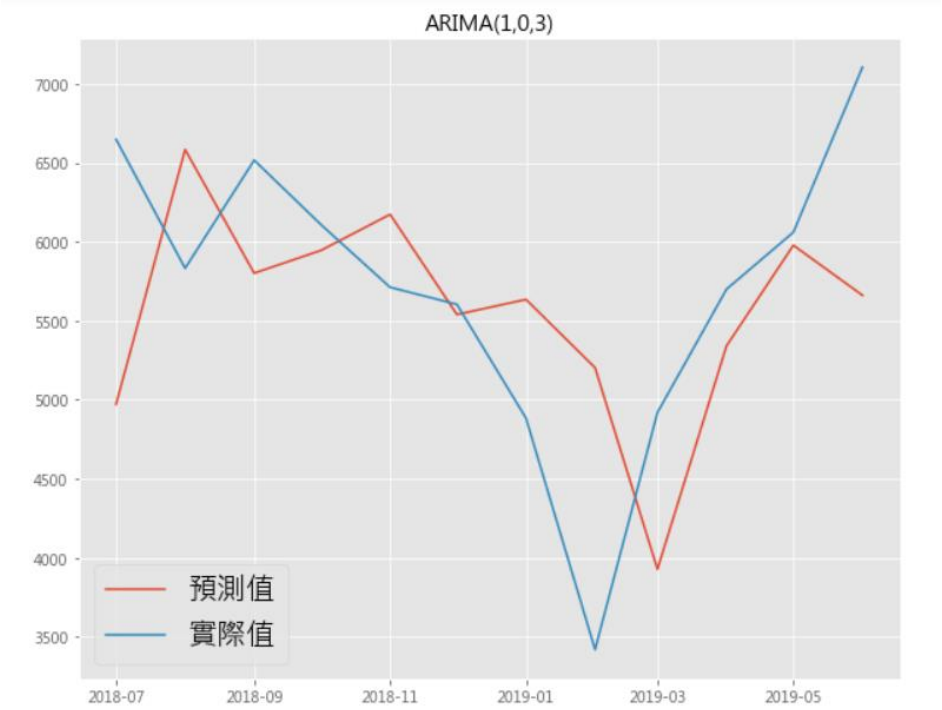
圖五、ARIMA(1,0,0)測試集及預測比較圖及 MAPE、RMSE：



表三、ARIMA(1,0,0)差異表												
年月 項目	2018/7/1	2018/8/1	2018/9/1	2018/10/1	2018/11/1	2018/12/1	2019/1/1	2019/2/1	2019/3/1	2019/4/1	2019/5/1	2019/6/1
實際值	6,648	5,832	6,516	6,108	5,712	5,604	4,884	3,420	4,920	5,700	6,060	7,104
預測值	4,276	6,619	5,805	6,487	6,080	5,685	5,578	4,860	3,401	4,894	5,671	6,031
差異值	-2,372	787	-711	379	368	81	694	1,440	-1,519	-806	-389	-1,073
差異值絕對值百分比	35.68%	13.50%	10.92%	6.21%	6.45%	1.45%	14.20%	42.10%	30.88%	14.14%	6.41%	15.11%
差異值平方	5,627,644	619,523	506,069	143,755	135,701	6,566	481,139	2,072,696	2,307,763	649,585	151,066	1,151,884
MAPE(差異值百分比相加再除以12)	16.42%											
MSE(差異值平方相加再除以12)	1,154,449											
RMSE(MSE開根號)	1,074											

ARIMA(1,0,0)在測試集表現 **MAPE 為 16.42%，RMSE 為 1,074 呎誤差**(計算結果如上表)，表現尚可，圖形線條大部分看起來與預測值都晚實際值一個月，以 2018 年 12 月預測為最準，僅高出 81 需量。

圖六、ARIMA(1,0,3)測試集及預測比較圖及 MAPE、RMSE：



表四、ARIMA(1,0,3)差異表												
年月 項目	2018/7/1	2018/8/1	2018/9/1	2018/10/1	2018/11/1	2018/12/1	2019/1/1	2019/2/1	2019/3/1	2019/4/1	2019/5/1	2019/6/1
實際值	6,648	5,832	6,516	6,108	5,712	5,604	4,884	3,420	4,920	5,700	6,060	7,104
預測值	4,971	6,583	5,801	5,946	6,173	5,540	5,635	5,203	3,928	5,342	5,978	5,661
差異值	-1,677	751	-715	-162	461	-64	751	1,783	-992	-358	-82	-1,443
差異值絕對值百分比	25.22%	12.89%	10.97%	2.66%	8.07%	1.15%	15.38%	52.14%	20.15%	6.29%	1.36%	20.31%
差異值平方	2,811,779	564,693	510,922	26,396	212,382	4,139	563,925	3,179,751	983,114	128,472	6,793	2,081,421
MAPE(差異值百分比相加再除以12)	14.71%											
MSE(差異值平方相加再除以12)	922,816											
RMSE(MSE開根號)	961											

ARIMA(1,0,3)在測試集表現**MAPE為14.71%，RMSE為961**

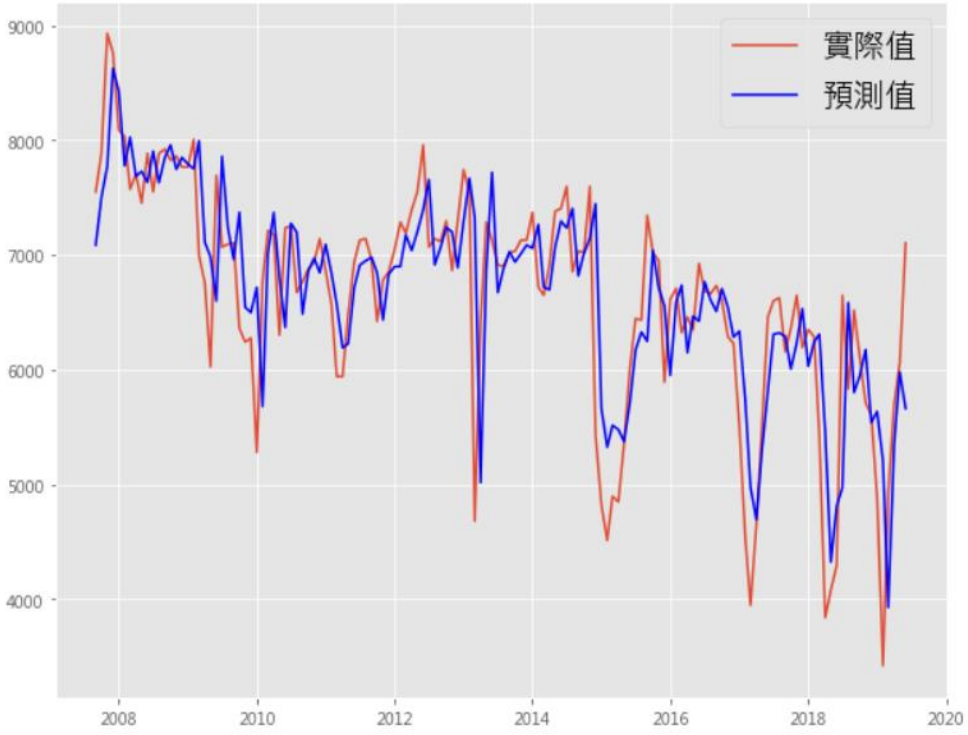
瓩誤差(計算結果如上表)，表現較ARIMA(1,0,0)佳，以2018年12月、2019年5月預測為最準，分別低64瓩、82瓩。

（十） ARIMA預測較佳之模型為ARIMA(1,0,3)

綜上所述，利用2007年7月至2018年6月當訓練集，

透過 AIC、BIC 找出表現相對好的模型，選出 ARIMA(1,0,0)、ARIMA(1,0,3)，再將測試集丟入該兩模型後，分別觀察兩者之 MAPE、RMSE 後，選出 ARIMA(1,0,3)為較佳之預測模型，將訓練集及測試集丟入同一張擬合圖如下：

圖七、訓練集及測試集放入同一個擬合圖



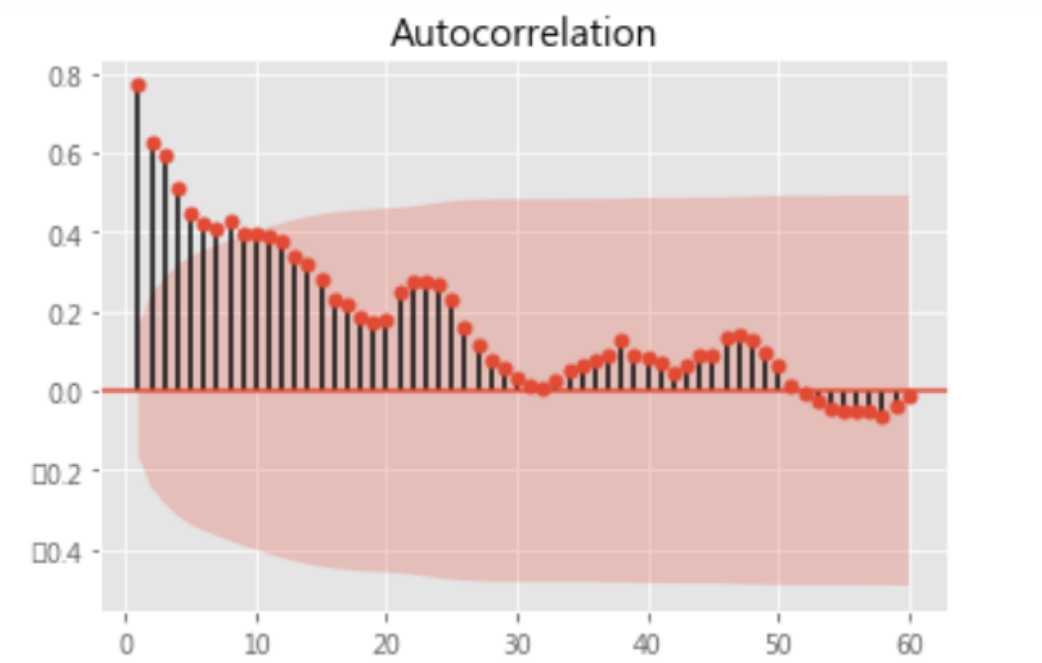
我們可發現紅線與藍線之高低起伏相當接近，顯示其預測值之趨勢線與實際值之趨勢線擬合結果相當不錯。

（十一） 是否能用 SARIMA 之說明

另外為了可以預測的更精準，如有季節性因素可以用 SARIMA(含季節性因子之 ARIMA 模型，將季節性趨勢帶入模型中做出更精準之預測值)，故本次透過觀察 acf 圖是否有

季節性因子：

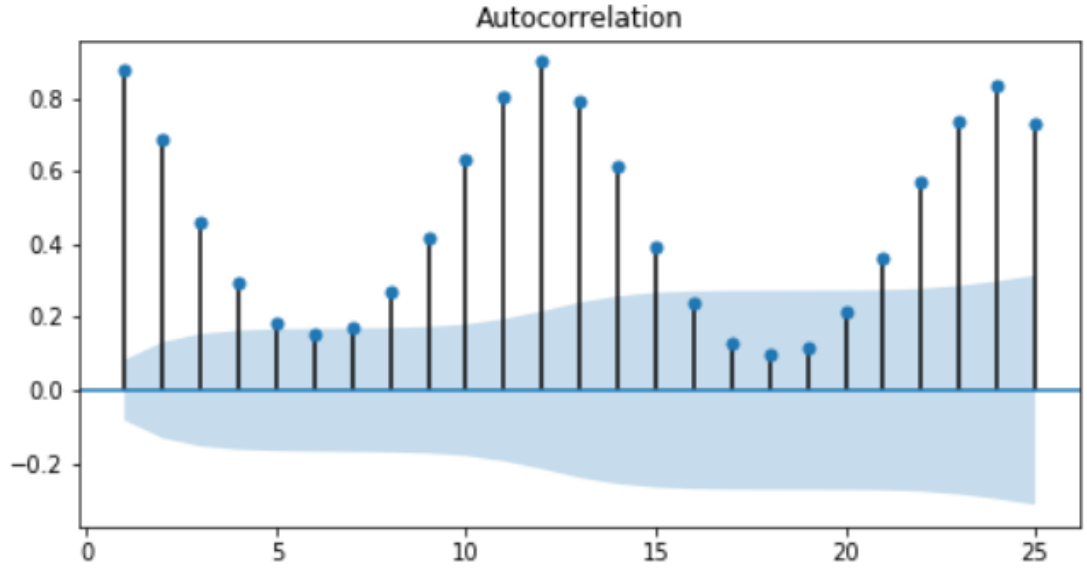
圖八、判斷是否有季節性因子之 acf 圖



由上圖判斷，如果有季節性因子，在相對應之月份理應有高相關係數，惟本圖從 lags=1 以後就慢慢往下掉，並無在某些時點有規律性的高點，故本例應不適用 SARIMA。

（另外，補充附上有明顯季節性因子之例子及 acf 圖，如圖九）:

圖九、有明顯季節性因子的 acf 圖例



在該圖中，在 lag=12、24 中有明顯高度相關性，表示該時間序列每 12 個月都有一個循環周期，當月的數據跟上年同月的數據是相似的，故該例可用 S=12 去做 SARIMA 模型，而本例則不適用 SARIMA 模型。

二、 利用長短期記憶模型(LSTM)進行時間序列預測

(一) 機器學習介紹

大數據時代的來臨，許多人將具有可以對各種數據做出預測之能力，以幫助公司或政府等做出較佳之決策，目前最流行的屬「**機器學習**」及「**深度學習**」兩種方式。機器學習最基礎的用法，是透過演算法來分析數據、從中學習，以及判斷或預測現實世界裡的某些事，並非手動編寫帶有特定指令的軟體程序來完成某個特殊任務，而是使用大量的數據和演算法來「訓練」機器，讓它學習如何執行任務。我們常常用的方法，在回歸方面有單變量線性回歸、多元線性回歸、多項式線性回歸、具有懲罰多係數的 Lasso 回歸、Ridge 回歸及 ElasticNet、廣義線性回歸、XgboostRegression…等等；而在分類方面有邏輯斯回歸、決策樹、隨機森林、SVM 向量機、XgboostClassicifier…等等。

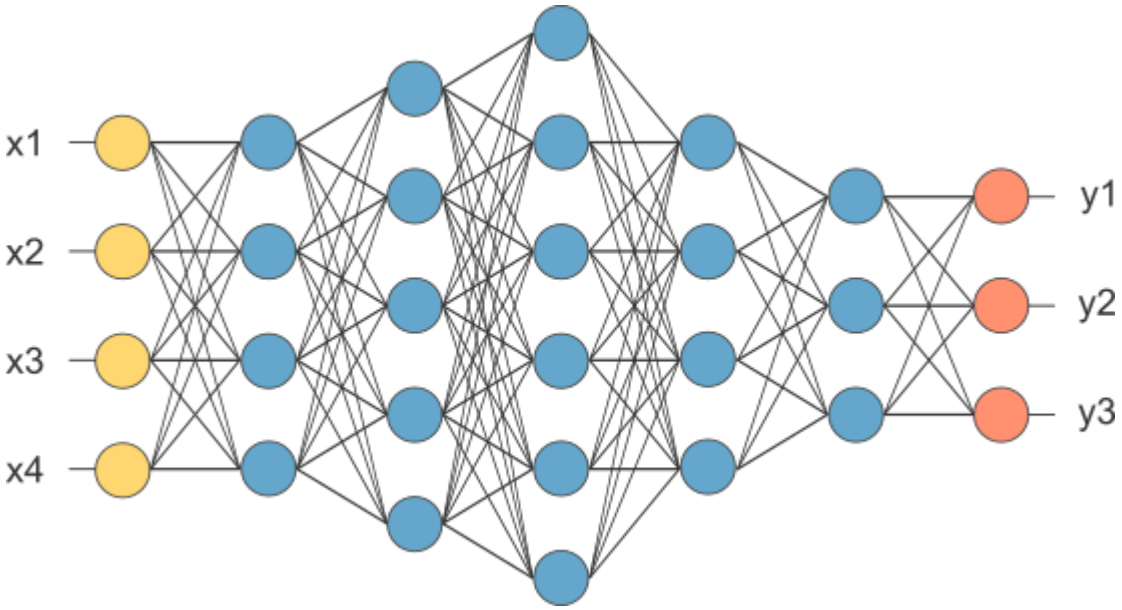
在 Python 上以 **Sklearn-Learn** 函式庫為首，僅透過輸入短短程式碼，即可跑出許多預測模型，雖然模型建置方便且預測快速，但在多維度的資料上，大部分仍需要進行「**降維工程**」**才得**以提高模型的表現，而降維工程需要有 **Domain Knowledge(領域知識)**，舉簡單例子，如果要透過身高及體重來判斷一個人是男生還是女生(二元分類問題)，將身高、體重分別設為X_1及X_2，將男生及女生設為 Y，透過簡單的二元邏輯斯回歸分類器做出預測判斷，兩個因變數較一個因變數還容易有過度擬合的現象，但我們知道如果將身高、體重換算成 BMI，兩變數將變為一個變數，且更能判斷出性別，對於模型的預測能力上會有更好的表現，惟 BMI 的換算是需要有 Domain Knowledge(領域知識)，如果在 BMI 公式發明以前，除非是專業人士，我們很難了解為什麼是這樣子計算，故機器學習對於預測大型多維資料集的好壞很需要透過良好的降維才能有比較好的表現，或是直接丟入深度學習訓練，當然對於維度較少的資料集，其實用機器學習就綽綽有餘且運行快速。

其實機器學習的學習方式有分**監督式學習、半監督式學習、非監督式學習、增強式學習**…等等眾多學習方式，惟篇

幅稍大在此不做介紹，而本次報告係**使用監督式學習來研究**〔係每一筆樣本均有相對應的標籤(X→Y，如每一筆 BMI 值均對應一個性別)〕；而半監督式學習、非監督式學習等方式都是資料有缺陷或根本沒有相對應的標籤所另外學習的配套措施(PCA 主成分分析、t-SNE 降維視覺化…等，日後或可再利用不同方式進行研究。

(二) 深度學習

深度學習是類似模擬人類的神經元建模，藉由建置輸入神經元(輸入層)，中間的隱藏層，及最後的輸出層來做出預測，如下圖：



在訓練過程中每個輸入層輸入X_1、X_2、X_3…X_n，對應輸出層為Y_1、Y_2 …Y_n（Label，標籤），讓機器自行學習，透過最關鍵的 Gradient Descent(**梯度下降法**)來更新每次訓練的權重(係數)w，讓 Loss Function(損失函數)達到最低而做出良好的預測能力，深度學習對於感知類的預測比機器學習有較好的效果，例如目前較流行的手寫辨識系統、圖片辨識系統、人臉辨識系統、語音辨識系統、自然語言翻譯、輿情分析…等，均可透過深度學習建模分辨，如圖片辨識系統為輸入一張黑白圖片，來判斷是狗還是貓等，其輸入為將圖片上每個 pixel 值作為 input，如一張圖為 19x19 的大小，則將有 361 維的資料輸入，在機器學習訓練上是非常緩慢且無效率，且在降維工程方面，感知方面的資料是比較抽象而難以將維度下降，你可能很難將聲音輸入的資料降低其維度，

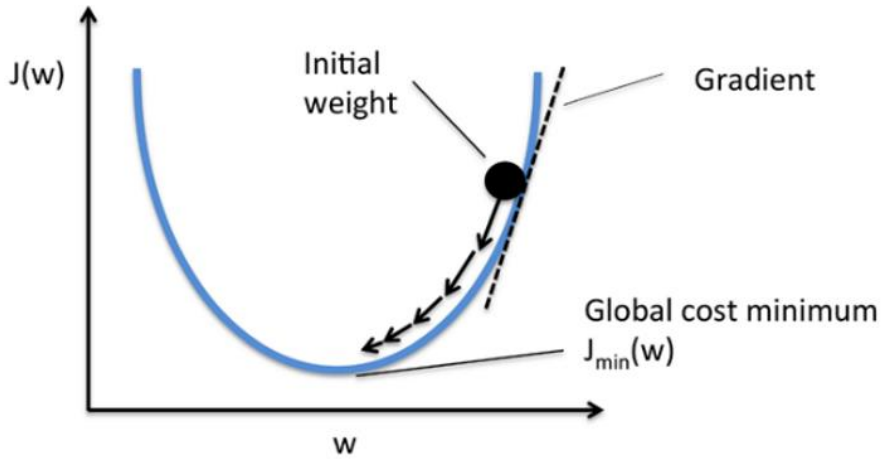
故直接透過建置深度學習模型，建置多個隱藏層，且每個隱藏層有多個神經元，直接將資料丟到深度學習模型裡訓練學習是相對簡單而暴力的，也因此不太需要有領域知識。但對於簡單的資料還是建議用機器學習即可。另外多層的人工神經網路也是非線性模型的一種。

1. 深度學習的第一關鍵-梯度下降法

*補充說明:簡單說深度學習的「深度」意思就是有多層的隱藏層，而機器學習一般只有一層，在簡單的資料上還是透過一層隱藏層學習較為快速準確。

前面介紹機器學習，能讓機器自行學習獲得最佳解的運算方式，最重要的非「梯度下降法」莫屬了，為什麼我們給很多樣本，讓他知道 BMI 跟相對應的性別後，他能夠慢慢學習調整權重呢?完全就是透過梯度下降法，藉由計算某個權重(係數)w 對損失函式(就是預測出來的數值與實際值的差異，越少表示預測力越好)作偏微分後(意思就是某個權重(係數)w 的變動影響損失函式的變動)，了解目前權重 w 在整個損失函式所在的斜率是正的還是負的，而做相對應的調整，如斜率為正，表示 w 應該要少一點，如斜率為負，表示 w 應該要多一點，讓他能慢慢接近損失函數最低的地方(如下圖

表示)，而要調整多大呢?主要是靠學習率來決定，如果學習率太多可能永遠跳不進最佳解，太少可能要很慢很慢才收斂，而決定學習率的方法很多，如 SGD、Momentum、AdaGrad、Adam(本報告使用 Adam，是目前深度學習最常用的優化器，Adam 算法和傳統的隨機梯度下降(SGD)不同。隨機梯度下降保持單一的學習率（即 alpha）更新所有的權重，學習率在訓練過程中並不會改變。而 Adam 學習率會隨著斜率的降低而下降)



圖中 Global cost minimum 表示該損失函數最低之地方(最佳解)，intial weight 表示權重(係數)，Gradient 表示斜率(梯度)，此圖為係數(w)與損失函數 J(w))的一種散佈圖，藉由 w 的變動來影響損失函數的高低，J(w)對 w 偏微分如呈現正數表示 w 的增加會增加損失函數的數值(簡稱損失值)，必須要降低 w 才會降低損失值，反之呈現負數表示 w 的增加

會降低損失值，必須要增加 w 才會降低損失值，故透過梯度下降法調整 w 來降低損失值，讓模型的預測能力越來越好。

2. 深度學習的第二關鍵-反向傳播(Backpropagation)

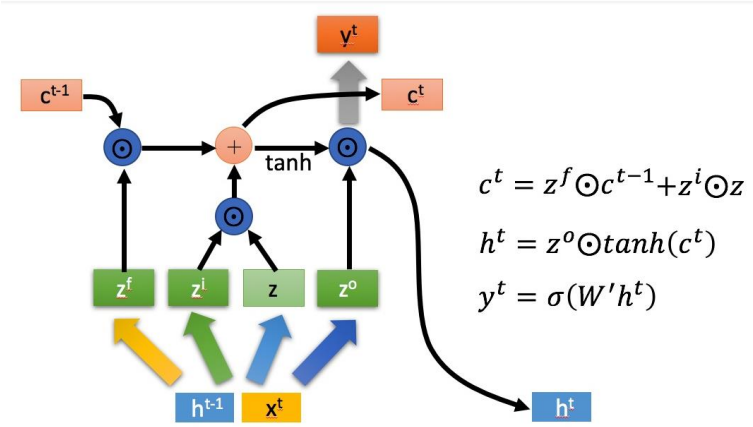
深度學習能快速的更新權重還有一個重要的關鍵，就是反向傳播，簡單來說正常的更新權重為由輸入神經元透過中間許多隱藏層而到達輸出層，而反向傳播則是反過來透過微分連鎖率將整個模型反過來，原來的輸出層變成輸入層，而輸入層而轉為輸出層，進而提高更新權重的效率。雖然反向傳播不是本模型的必要條件，但是本研究有用到這個方法來加速更新模型的速度(因程式碼已寫入此方法，故使用時就有直接用到)。

3. 深度學習的第三關鍵-激勵函數 (Activation Function)

每一個神經元除了權重(係數)w 外，尚存在決定是否要把答案送到下一層的**激勵函數**，在深度學型模型中使用激勵函數中，主要是利用非線性方程式，解決非線性問題，若不使用激勵函數，類神經網路即是以線性的方式組合運算，因為隱藏層以及輸出層皆是將上層之結果輸入，並以線性組合

計算，作為這一層的輸出，使得輸出與輸入只存在著線性關係，而現實中，所有問題皆屬於非線性問題，因此，若無使用非線性之激勵函數，則深度學型模型訓練出之模型便失去意義。所以為了預測非線性時間序列，本研究使用含有激勵函數之深度學習來預測板新廠每月契約最高需量之非線性時間序列(基本上深度學習模型每個神經元幾乎都有使用激勵函數)。

(三)　長短期記憶模型(LSTM)簡介



本次報告使用深度學習之長短期記憶模型(LSTM)，它是屬於時間遞歸神經網路(RNN)的一種，該模型的一個神經元如上圖，簡單說除一般神經元有 input 之外，還多了 input gate(輸入門)、forget gate(遺忘門，實際上它的功用應屬於記憶功能，但英文上翻譯仍稱遺忘門)、output gate(輸出門)來決定本顆神經元是否要留住本次樣本的一些相關訊息，以延續到下一個樣本，這模型讓每個樣本丟進去跑的時候，神經元具有記憶功能，例如簡單的一個樣本規則如下，禮拜一到禮拜日的晚餐分別吃義大利麵(禮拜一)、滷肉飯(禮拜二)、披薩(禮拜三)、小籠包(禮拜四)、日本料理(禮拜五)、泰式料理(禮拜六)、火鍋(禮拜日)，而這些規則假如不變，如果把星期幾吃什麼丟到一般人工神經網路訓練他可能要算半天還無法得到很好的解答，但其實我們一眼看就知道義大利麵的後一天一定是滷肉飯，後兩天一定是披薩等簡單的

順序規則，以記憶前一天吃什麼的方式來預估隔一天是吃什麼，就是遞歸神經網路(RNN)與人工神經網路(DNN)主要不同的地方，而長短期記憶模型(LSTM)又可以回顧更久之前的事情來推斷目前的事情，此模型在文字翻譯上及時間序列預測上都有很強大的表現。本報告利用 LSTM 模型，透過板新廠每月最高需量歷史資料來預測下一個月甚至未來更多個月之最高需量，並與 ARIMA 進行 MAPE、RMSE 的比較。

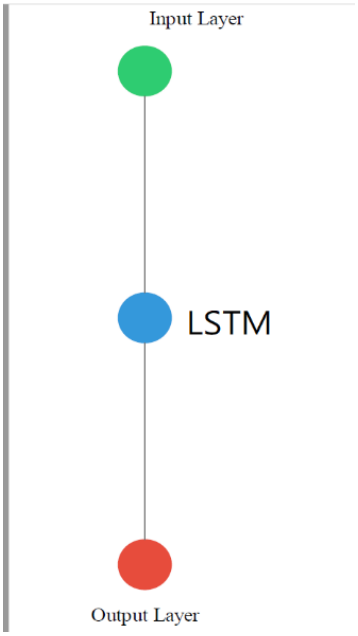
(四) 以 1 個 LSTM 建立模型

要讓一個深度學習模型運作，我們需要自行製作人工神經網路的輸入層、隱藏層、輸出層等，並決定隱藏層的神經元顆數，本報告以實際操作面重點進行，我們先假設板新廠當前值僅受前一期值的影響，故以下以層數、神經元數量、受到前 n 期的影響來做預測分析(同樣以 2007 年 8 月至 2018 年 6 月資料為訓練集，以 2018 年 7 月至 2019 年 6 月資料為測試集，以同樣標準跟 ARIMA 模型做比較，以下模型優化器均為 adam，模型損失函數均為 mse，epoch 次數均為 100 次，訓練批次為每次 1 次)：

1. 模型層配置：隱藏層→LSTM(數量:1，激活函數: ReLU)，時間步為 1 步(表示當前值受前 1 期影響)。

附註：ReLU 稱為線性整流函數（Rectified Linear Unit），又稱修正線性單元，是一種人工神經網絡中常用的激勵函數（activation function），函數為 $f(x) = \max(0, x)$ ，在神經網絡中，線性整流作為神經元的激活函數，定義了該神經元在線性變換 $\mathbf{w}^T\mathbf{x} + b$ 之後的非線性輸出結果。換言之，對於進入神經元的來自上一層神經網絡的輸入向量 x ，使用線性整流激活函數的神經元 $\max(0, \mathbf{w}^T\mathbf{x} + b)$ 的結果會輸出至下一層神經元或作為整個神經網絡的輸出。

2. 模型示意圖：



3. 預測擬合圖：

4. 差異表(測試集):

表五、1個Lstm、時間步1步之各月份差異表

年月 / 項目	2018/7/1	2018/8/1	2018/9/1	2018/10/1	2018/11/1	2018/12/1	2019/1/1	2019/2/1	2019/3/1	2019/4/1	2019/5/1	2019/6/1
實際值	做為參數	5,832	6,516	6,108	5,712	5,604	4,884	3,420	4,920	5,700	6,060	7,104
預測值		6,722	5,832	6,583	6,125	5,717	5,621	5,183	4,899	5,198	5,706	6,072
差異值		890	-684	475	413	113	737	1,763	-21	-502	-354	-1,032

5. 模型評估：

透過 LSTM 模型訓練後，訓練集的 RSME、MAPE 分別為 604.16、6.86%，測試集的 RSME、MAPE 分別為 784.22、12.64%，表示當只有一顆 LSTM 神經元，MAPE 竟高達 6.86%，表示在訓練集的擬合程度相當高，惟在測試集表現較差，MAPE 由 6.86%上升至 12.64%，其中以 2019 年 2 月差距最大，

可能原因為除本(108)年度上半年除乾旱外，石管局將核放水庫水條件修改為每日需向北水處購買平均每日 65 萬噸清水，才核放 2CMS(每秒 2 噸)水給予本區處，而此核放條件在往年均不曾有過，故以歷史資料推斷尚未發生的事情，是會有很大的預測偏差的，但 LSTM 模型有記憶功能，它可能學習到前一期的資料往下掉後，他後續的值往下掉的機會比較大。例如表 5,在本年 2 月份差異值非常大(1,763 瓩)之後的 3 月份，預測偏差僅 21 瓩，此可看出該模型的強大功能，會透過歷史條件而做調整。另外模型尚可再做修正，當時間步不只僅有 1 步，則當前值可能受更多前期數值影響，故以下再調整時間步進行修正。

6. 時間步 1 步至 6 步之 RMSE、MAPE 表：

表六、1個Lstm、時間步1~6步之RMSE、MAPE比較表

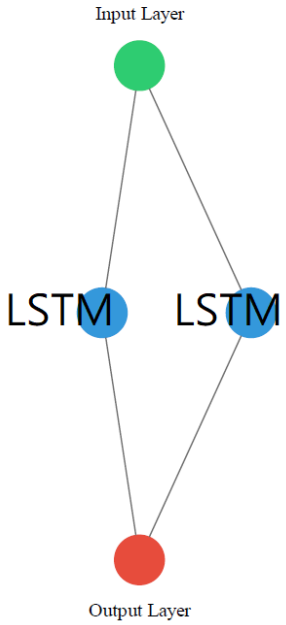
時間步 / 項目	1	2	3	4	5	6
訓練集RMSE(需量)	604.16	604.39	598.13	585.62	564.81	566.32
訓練集MAPE(%)	6.86	6.74	6.83	6.87	6.58	6.68
測試集RMSE(需量)	784.22	757.15	770.15	780.74	834.54	899.99
測試集MAPE(%)	12.64	11.32	11.93	12.96	12.93	14.67

由上表可見，時間步越多步，在訓練集的擬合程度越高(以 RMSE 為比較基準)，但在測試集表現卻越來越差，表示在時間步較高時(如時間步為 5~6 步)，會有較嚴重之過度擬合情況(過度擬合意即此模型的參數夠多，與訓練集擬合的程

度很高，但對於測試集不一定適用)，也表示時間步越高在模型預測能力上不一定越高。在本表中，1 個 LSTM 的模型下 RMSE 最低的是時間步為 2 步。

(五) 以 2 個 LSTM 建立模型(LSTM 神經元數量越多，函數參數增加，模型複雜度亦相對提高)

1. 模型層配置：隱藏層→LSTM(數量:2，激活函數:默認 ReLU)，時間步為以 1 至 6 步
2. 模型示意圖：



3. 時間步 1 步至 6 步之 RMSE、MAPE 表：

表七、2個Lstm、時間步1~6步之RMSE、MAPE比較表						
時間步 項目	1	2	3	4	5	6
訓練集RMSE(需量)	588.29	567.62	562.97	545.76	518.15	526.23
訓練集MAPE(%)	6.61	6.27	6.16	6.18	6.02	6.16
測試集RMSE(需量)	779.91	741.13	682.83	721.35	655.22	809.61
測試集MAPE(%)	12.99	12.66	11.78	12.23	11.70	14.90

由上表可見，LSTM 數量為 2 個的情況下，在訓練集的 RMSE 及 MAPE 均比 1 個 LSTM 模型表現上還要好，但在測試集表現卻不一定比較好，表示在 LSTM 數量提高後亦有過度擬合，也表示 LSTM 數量增加在模型預測能力上不一定越高。在本表中，2 個 LSTM 的模型下 RMSE 最低的是時間步為 5 步。

（六） 分別以 1~4 個 LSTM 模型建置測試集 RMSE 差異表

表八、1~4個Lstm、時間步1~6步之測試集RMSE比較表						
時間步 lstm數量	1	2	3	4	5	6
1	784.22	757.15	770.15	780.74	834.54	899.99
2	779.91	741.13	682.83	721.35	655.22	809.61
3	777.30	711.47	685.76	639.44	699.99	735.00
4	778.99	726.88	691.64	654.33	589.75	768.13

由表八可知，若以 RMSE 為比較基準，則 4 個 LSTM 及本期與 5 個時間步(表示本期與前 1~5 期均有關係)在測試集上 RMSE 可達到最低，為 **589.75 瓩**；而時間步為 6 時，除 LSTM 為 3 個(RMSE 為 735.00)外，其餘之 RMSE 表現均較差。

（七） 分別以 1~4 個 LSTM 模型建置測試集 MAPE 差異表

表九、1~4個Lstm、時間步1~6步之測試集MAPE比較表						
時間步 / lstm數量	1	2	3	4	5	6
1	12.64	11.32	11.93	12.96	12.93	14.67
2	12.99	12.66	11.78	12.23	11.70	14.90
3	13.23	12.06	12.01	10.68	12.17	13.91
4	13.86	12.34	12.12	11.51	11.14	14.92

由表九可知，若以 MAPE 為比較基準，則 3 個 LSTM 及 4 個時間步(表示本期與前 1~4 期均有關係)在測試集上 MAPE 可達到最低，為 10.68%，我們也發現到時間步在第 6 步時表現都不太好，可能是因為訓練集本身為 12 個月，如果以當期對前 1~6 期均有關係，則其樣本數將僅剩 6 個，而有樣本數過少而致標準差較大之問題。

但是不管怎麼樣，LSTM 模型表現不論是 RMSE 或 MAPE，在測試集表現均較 ARIMA 表現較佳，如 ARIMA(1,0,3) RMSE 為 961， MAPE 為 14.71%；而 4 個 LSTM 配上 5 個時間步其 RMSE 為 589.75，MAPE 為 11.14%；3 個 LSTM 配上 4 個時間步其 RMSE 為 639.44，MAPE 為 10.68%。為什麼會有這樣的差距?原因在於 ARIMA 畢竟為線性模型，可能一個時間序列真正的函數是很多曲線而形成，而在 LSTM 深度學習模型則為非線性模型，它有記憶前一個樣本之功能，並透過多個閥

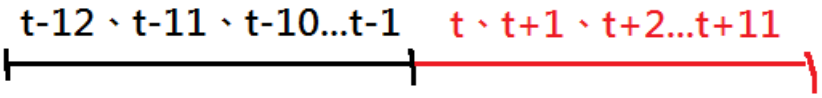
門開關(為分析方便，本次主要都是用 ReLU 線性整流函數來當激勵函數)來決定是否要留住記憶、留多少、是否要更新記憶、這記憶在下一個樣本是否要使用…等等。

三、 使用 LSTM 深度學習模型預測未來 12 個月之用電最高需量

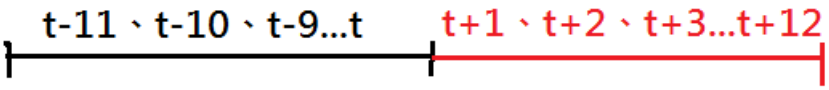
前面做 ARIMA 及 LSTM 之模型比較，是為了凸顯 LSTM 深度學習的強大預測能力，藉由分別比較在測試集上的 RMSE、MAPE 的表現，來看出模型之好壞。而在實務上 ARIMA 對於預測下一個時間點(月)的預測能力是最強的(除非使用 SARIMA 可對較長遠的值做預測，但適用條件較嚴苛)，惟契約容量之訂立期間為一年，本例仍嘗試預測未來 12 個月之用電最高需量以達最大化降低成本，透過 LSTM 深度學習模型預測未來 12 個月之非線性趨勢表現應較 ARIMA 好，故本次將利用 LSTM 深度學習模型方式來製作以前 12 個時間步，以預測未來 12 個時間點的模型。

(一) 以前 12 個時間步來預測未來 12 個時間點示意圖(多步預測)

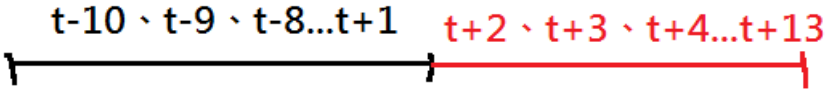
第一個樣本:(黑色為輸入值、紅色為輸出值)



第二個樣本:



第三個樣本:

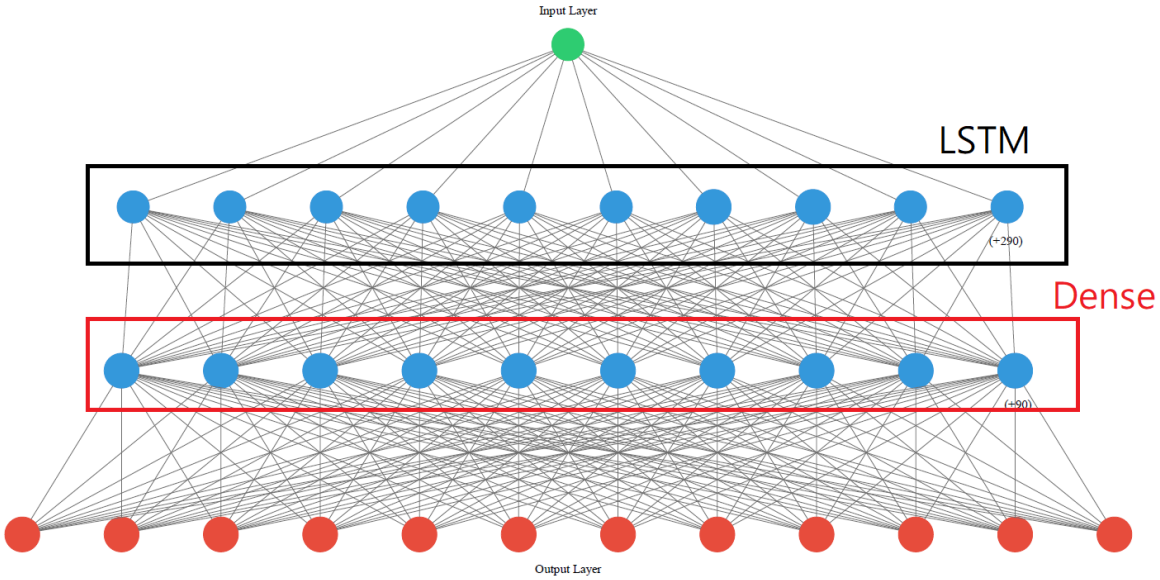


以此類推…，如我們最後一個月之資料時間為 2019 年 6 月，那我們將以 2018 年 7 月至 2019 年 6 月的值為輸入值，預測期間為 2019 年 7 月至 2020 年 6 月。

而本次訓練集為了可以讓 12 個月整除，故將 131 個訓練月(2007 年 8 月~2018 年 6 月)降至 120 個訓練月，以 2008 年 7 月至 2018 年 6 月之用電需量來預測 2018 年 7 月至 2019 年 6 月之用電需量，來推估 RMSE 及 MAPE，並透過此模型來預測 2019 年 7 月至 2020 年 6 月之用電需量。而模型的參數及隱藏層層數、神經元數量等，經過多次嘗試後，將以 LSTM 300 個、Dense 100 個(Dense 就是人工神經網路一般的神經元，函數 $\mathbf{w}^T\mathbf{x}+b$)、輸出層有 12 個(因為要輸出 12 個月的數字)、訓練次數 200 次、每次訓練樣本 1 個等參數建模(以上

參數之決定為嘗試多次後，表現較好之參數所構建出來之模型）。

（二） 以 300 個 LSTM→100 個 Dense 建立模型



經過訓練後，其模型結果 RMSE 為 939，MAPE 為 13.82%，其實際值與預測值比較如下：

實際年月	實際值	預測值	實際值-預測值
2018-07-01	6648	5713.401855	935.0
2018-08-01	5832	5690.663086	141.0
2018-09-01	6516	5739.218262	777.0
2018-10-01	6108	5633.604004	474.0
2018-11-01	5712	5821.663574	-110.0
2018-12-01	5604	5601.058105	3.0
2019-01-01	4884	5480.458496	-596.0
2019-02-01	3420	5480.439453	-2060.0
2019-03-01	4920	5549.394043	-629.0
2019-04-01	5700	5415.862793	284.0
2019-05-01	6060	5319.479980	741.0
2019-06-01	7104	5307.391113	1797.0

經發現預測值的波動相當小，可能因為以前一期每 12 個月數據去推估當期 12 個月的數字而有類似移動平均的因素，使波動較小，而預測差較多的月份為 2019 年 2 月及 2019 年 6 月，與前述原因類似：本年度上半年除乾旱外，石管局將核放水庫水條件修改為需向北水處購買平均每日 65 萬噸清水，才核放 2CMS(每秒 2 噸)水給予本區處，而此核放條件在往年均不曾有過，故以歷史資料推斷尚未發生的事情，是會有很大的預測偏差的。而本年 6 月份石門水庫核放水量條件因本區處處長之努力使本處向北水處購買量下降至平均每日 30 萬噸清水致出水量增加，致預測偏差在

經過 3~5 月較低後又回升提高至 1,797 瓩。

（三） 推估 2019 年 7 月至 2020 年 6 月每月最高需量

我們以前述模型推估 2019 年 7 月至 2020 年 6 月每月最高需量，以訂立較佳的契約容量，期可節省基本電價費開支，其結果如下：

實際年月	預測值
2019-07-01	6497.0
2019-08-01	6308.0
2019-09-01	5873.0
2019-10-01	5666.0
2019-11-01	5375.0
2019-12-01	5247.0
2020-01-01	5190.0
2020-02-01	5335.0
2020-03-01	5356.0
2020-04-01	5700.0
2020-05-01	5811.0
2020-06-01	5743.0

	預測值
count	12.000000
mean	5675.083496
std	411.778076
min	5190.000000
25%	5350.750000
50%	5683.000000
75%	5826.500000
max	6497.000000

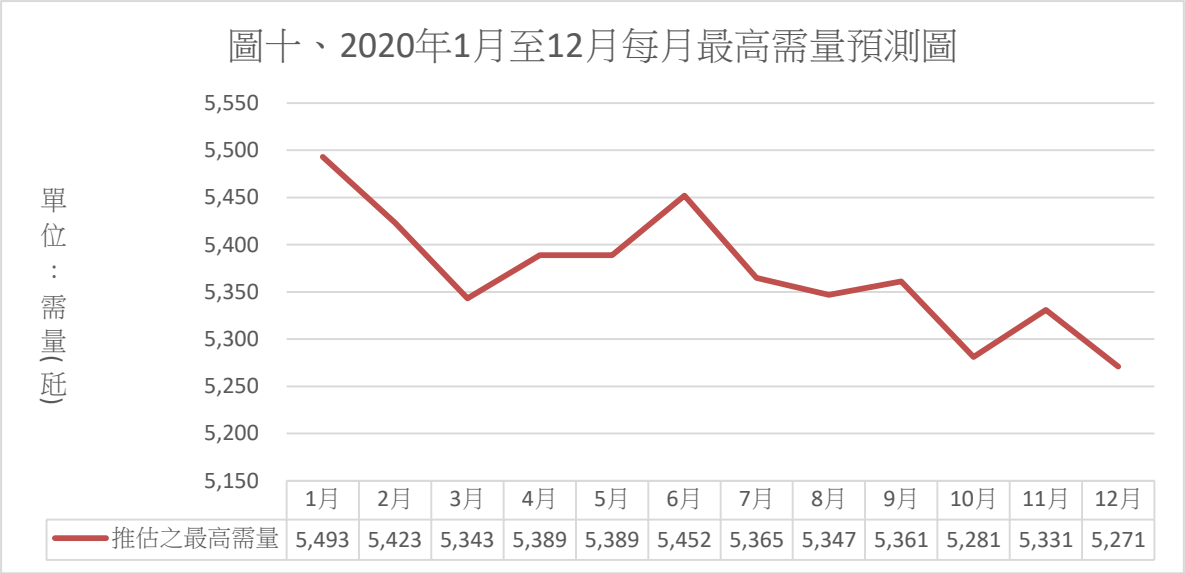
由上表顯示，預測值在乾旱機率較大之月份確實較低(11 月、12 月、1 月、2 月、3 月等)，而在豐水機率較大之月份較高(4~10 月)，故在常理上看本預測實屬合理。

（四） 推估 2020 年 1 月至 2020 年 12 月每月最高需量

為推估 2020 年 1 月至 2020 年 12 月每月最高需量，因本次預測模型需要前面 12 期的數據，故缺少的 2019 年 7 月至 12 月的最高需量，以前面推估 2019 年 7 月至 2020 年 6 月每月最高需量時所取得之預測值來填補，其模型運作後跑出之結果如下：

表十、以2019年1月至12月每月最高需量推估2020年1月至12月每月最高需量

項目 / 年月	變數X	輸出Y
2019年1月	4,884	
2019年2月	3,420	
2019年3月	4,920	
2019年4月	5,700	
2019年5月	6,060	
2019年6月	7,104	
2019年7月(推估)	6,497	
2019年8月(推估)	6,308	
2019年9月(推估)	5,873	
2019年10月(推估)	5,666	
2019年11月(推估)	5,375	
2019年12月(推估)	5,247	
2020年1月		5,493
2020年2月		5,423
2020年3月		5,343
2020年4月		5,389
2020年5月		5,389
2020年6月		5,452
2020年7月		5,365
2020年8月		5,347
2020年9月		5,361
2020年10月		5,281
2020年11月		5,331
2020年12月		5,271



我們再以規劃求解方式解出 2020 年應訂立之契約容量：

表十一、預測 109 年 1 月至 12 月最適契約容量

月份	當月最高需量(瓩)	契約容量(瓩)	109 年合計基本電費(元)	基本電費單價(元)		超約附加費	基本電費
1 月	5,493	5,361	11,669,181	夏月單價	217.3	42,301	861,025
2 月	5,423	5,361		非夏月單價	160.6	19,817	861,025
3 月	5,343	5,361				0	861,025
4 月	5,389	5,361				8,896	861,025
5 月	5,389	5,361	0			8,896	861,025
6 月	5,452	5,361				39,416	1,165,011
7 月	5,365	5,361				1,606	1,165,011
8 月	5,347	5,361				0	1,165,011
9 月	5,361	5,361				0	1,165,011
10 月	5,281	5,361				0	861,025
11 月	5,331	5,361				0	861,025
12 月	5,271	5,361				0	861,025

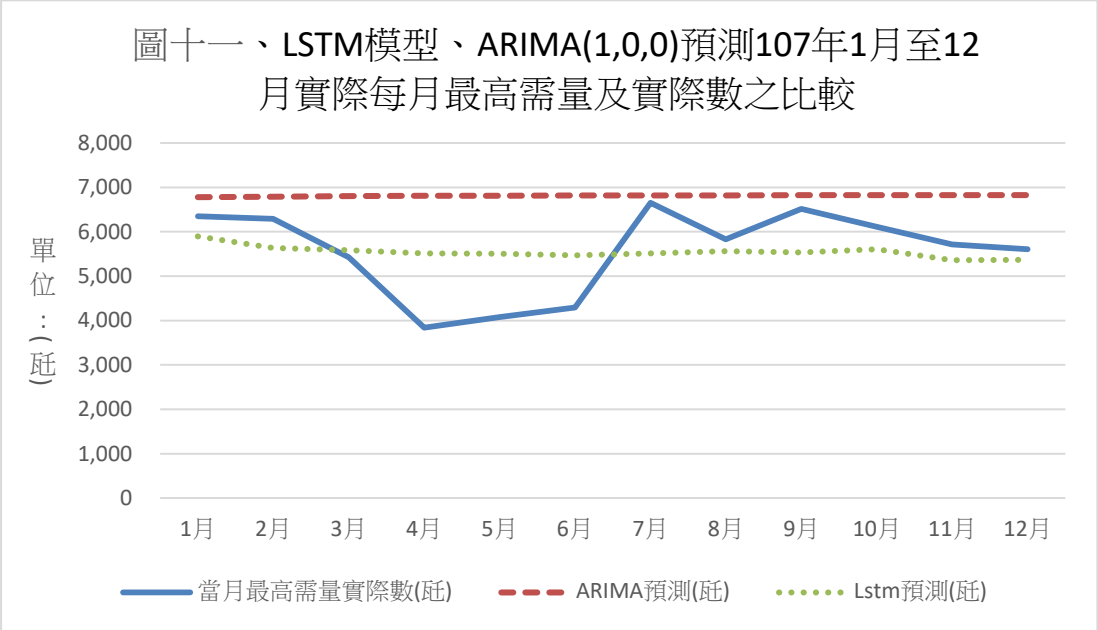
由上表可知，建議 2020 年板新廠電號 05-91-8820-11-7 契約容量可訂立 5361 瓩，基本電費推估為 11,669,181 元，如與目前契約容量 6800 瓩來計算，則需負擔基本電費 14,647,200 元來比較，預計可節省 2,978,019 元，該節省之金額近 3 百萬元實不可小覷。

（五） 回測106年研究報告:LSTM模型預測107年1月至12月每月最高需量與106年ARIMA預測及實際數之比較

106年應用統計研究報告係使用ARIMA(1,0,0)模型，透過2007年9月至2017年9月資料對2018年1月至12月資料進行預測，並訂立最適契約容量為6,818瓩。

本次則利用LSTM深度學習模型(模型參數與前面相同)，透過2007年10月至2017年9月資料(湊集12的倍數，故捨棄2007年9月)，先對2017年10月至12月做出預測；再利用2008年1月至2017年12月(捨棄2007年10月至12月以湊集12的倍數，並加入2017年10月至12月預測值作為建模的數據)之數據建模，並對2018年1月至12月資料做預測，再與2018年1月至12月實際發生之每月最高需量做比較，其比較表如下：

表十二、LSTM模型、ARIMA(1,0,0)預測107年1月至12月實際每月最高需量及實際數之比較			
實際用電月份	當月最高需量實際數(瓩)	ARIMA預測(瓩)	Lstm預測(瓩)
1月	6,348	6,778	5,895
2月	6,288	6,790	5,632
3月	5,424	6,799	5,582
4月	3,840	6,806	5,511
5月	4,080	6,811	5,502
6月	4,296	6,815	5,466
7月	6,648	6,818	5,512
8月	5,832	6,820	5,559
9月	6,516	6,822	5,535
10月	6,108	6,823	5,605
11月	5,712	6,824	5,358
12月	5,604	6,824	5,368
平均值	5,558	6,811	5,544
MAPE(%)		26.67%	15.13%
RMSE(瓩)		1,562	895



由表十二、圖十一可知，LSTM 所預測之每月最高需量是比 ARIMA(1,0,0)預測值還接近實際每月最高需量，ARIMA(1,0,0)所預測之值從 1 月開始一直在 6,778 瓩以上，而 LSTM 竟能落在 5,358 瓩至 5,895 瓩之間變動，LSTM 之預

測值相對較低；平均值方面 ARIMA(1,0,0)為 6,811 瓩，而 LSTM 為 5,544 瓩(接近實際值平均數 5,558 瓩），LSTM 之預測值亦相對較低。另外在 RMSE 方面，ARIMA(1,0,0)為 1,562 瓩，LSTM 為 895 瓩，而在 MAPE 方面，ARIMA(1,0,0)為 26.67%，LSTM 為 15.13%，兩個評估績效亦均以 LSTM 較優。

我們從兩個模型預測跑出規劃求解後之最佳契約容量，再與實際每月最高需量做比較，看能省下多少錢：

(1) 以 LSTM 預測 107 年 1 月至 12 月最適契約容量：

表十三、LSTM 預測 107 年 1 月至 12 月最適契約容量

實際用電月份	當月最高需量(瓩)	契約容量(瓩)	107 年合計基本電價(元)	基本電費單價(元)		超約附加費	基本電費
1 月	6,348	5,518	14,048,888	夏月單價	217.3	311,275	886,191
2 月	6,288	5,518		非夏月單價	160.6	282,367	886,191
3 月	5,424	5,518				0	886,191
4 月	3,840	5,518				0	886,191
5 月	4,080	5,518				0	886,191
6 月	4,296	5,518				0	1,199,061
7 月	6,648	5,518				616,741	1,199,061
8 月	5,832	5,518				136,464	1,199,061
9 月	6,516	5,518				530,690	1,199,061
10 月	6,108	5,518				195,643	886,191
11 月	5,712	5,518				62,313	886,191
12 月	5,604	5,518				27,623	886,191

以 LSTM 預測之每月最高需量再以規劃求解獲得最適契約容量 5,518 瓩，107 年基本電價需負擔 14,048,888 元。

（2）以 ARIMA 預測 107 年 1 月至 12 月最適契約容量：

表十四、ARIMA(1,0,0)預測 107 年 1 月至 12 月最適契約容量							
實際用電月份	當月最高需量(瓩)	契約容量(瓩)	107 年合計基本電價(元)	基本電費單價(元)		超約附加費	基本電費
1 月	6,348	6,818	14,685,972	夏月單價	217.3	0	1,094,971
2 月	6,288	6,818		非夏月單價	160.6	0	1,094,971
3 月	5,424	6,818				0	1,094,971
4 月	3,840	6,818				0	1,094,971
5 月	4,080	6,818				0	1,094,971
6 月	4,296	6,818				0	1,481,551
7 月	6,648	6,818				0	1,481,551
8 月	5,832	6,818				0	1,481,551
9 月	6,516	6,818				0	1,481,551
10 月	6,108	6,818				0	1,094,971
11 月	5,712	6,818				0	1,094,971
12 月	5,604	6,818				0	1,094,971

以 ARIMA(1,0,0)預測之每月最高需量再以規劃求解後，獲得最適契約容量 6,818 瓩(106 年研究報告結果)，而以 107 年基本電價需負擔 14,685,972 元。

兩個模型比較，因 LSTM 預測的數值比較準確，107 年基本電價 LSTM 較 ARIMA(1,0,0) 可少支付 637,084 元，其金額雖然不大，但在十二區處因折舊、原料費等致經營情況日益變差之情況下仍不無小補，且僅須僅年初時重新訂立契約容量即可達成。

肆、 結論與建議

一、 LSTM 深度學習模型預估能有效降低 109 年板新廠基本電費將近 3 百萬元

在經過 ARIMA 與 LSTM 的綜合比較後，我們可發現在預測板新廠電號 05-91-8820-11-7 每月最高需量資料時，LSTM 模型表現不論是 RMSE 或 MAPE，在測試集表現均較 ARIMA 表現較佳，例如： ARIMA(1,0,3) RMSE 為 961， MAPE 為 14.71%，而 4 個 LSTM 配上 5 個時間步其 RMSE 為 589.75，MAPE 為 11.14%，3 個 LSTM 配上 4 個時間步其 RMSE 為 639.44，MAPE 為 10.68%。可能原因為板新廠電號 05-91-8820-11-7 每月最高需量容易受到水情好壞、政策因素、設備跳電…等等綜多因素影響，因 ARIMA 為線性模型，故在預測此種變化性大的時間序列下預測是有受限的，而 LSTM 為非線性模型透過歷史資料學習並記憶，且每個神經元有各種閥門控制數值的進出，故其資料擬合度較佳，ARIMA 用來對線性關係的資料應更好地建模，而 LSTM 則更好地建立了具有非線性關係的模型資料。

而透過 LSTM 預測模型預測 2020 年 1 月至 12 月每月

最高需量後，透過規劃求解方式建議契約容量可訂立5361瓩，基本電費推估為11,669,181元，如以目前契約容量6800瓩來計算，則需負擔基本電費 14,647,200 元，預計可節省2,978,019元，該節省之金額近3百萬元實不可小覷。

透過LSTM深度學習模型回測107年每月最高需量，107年基本電價LSTM比ARIMA(1,0,0)實際上少支付637,084元，其金額在十二區經營情況因折舊、原料費等而日益變差之情況下實屬甘霖，且僅須僅年初時重新訂立契約容量即可達成。

二、 ARIMA、LSTM與前三年實際平均數三者之簡單比較：

1. 如以 RMSE 為比較基準，**以前三年實際平均數(104 年~106年)6,063瓩回測107年RMSE為3,671瓩，或前三年加權平均(20%、30%、50%)3,606瓩**，均比ARIMA 1,562瓩及LSTM 895瓩差，表示ARIMA及LSTM表現均比以前三年實際平均數比較要好，因此我們是值得花時間透過預測模型去建模，以取得比實際平均數更好的績效。
2. ARIMA對短期預測有較好的預測效果，而LSTM對長期模型有較好的預測效果。
3. 傳統的時間序列預測方法(ARIMA)側重於具有線性關係和固

定人工診斷時間依賴的單變數資料。

4. 對大量資料集的機器學習問題研究發現,與 ARIMA 相比,LSTM 獲得的平均錯誤率在 84-87%之間，代表 LSTM 優於 ARIMA。
5. 傳統的方法，如 ETS 和 ARIMA，適用於一元資料集的單步預測。
6. 像 ARIMA 這樣的經典方法側重於固定的時間依賴性：不同時間觀測值之間的相互關係，這就需要分析和說明作為輸入的滯後觀測值的數量。

三、 日後可再改進的地方：

1. 將影響當月出水量的石門水庫供應政策因素轉為虛擬變量套入模型，以增加模型表現，例如有發生則為 1，未發生則為 0，增加一個變數。
2. 因板新廠電號 05-91-8820-11-7 供應許多原水、淨水、供水等設備，可找出此類設備當月最高出水量套入模型，來判斷當月最高出水量是否會影響當月最高需量，以增加模型表現。
3. 深度學習模型中隱藏層間的建構，可再多嘗試不同的組合(如 CNN 卷積神經網路、Adaboost 集成等方法)，以增加模型表現。
4. 深度學習模型尚有許多參數可調整，如梯度下降機制的選用、

是否需要特徵縮放(如將數值標準化)…等，以增加模型表現。

5. 目前有相當多的演算法尚待我們發掘及應用。

伍、 資料來源

1. 台電網站電子帳單資料

2. 參考書本:時間序列分析三版-經濟與財務上之應用 楊奕農著

3. 李宏毅機器學習線上教學(免費)

4. 林軒田機器學習線上教學(免費)

5. 吳恩達機器學習線上教學(免費)

6. 各大網站 python 線上教學

7. Deep learning 深度學習必讀：Keras 大神帶你用 Python 實作

8. Python 機器學習(第二版)

9. DataCamp 網站上各種教學資源(年費)