

以人工智慧方法進行系集溫度預報場推估至測站溫度研究

簡報者：張鶴齡 harry_chang@leadtek.com.tw

麗臺科技股份有限公司 張鶴齡 羅世軒 薛宏宇 劉家豪

中央氣象局氣象科技研究中心 張庭槐

大綱

- 前言
- 資料格式
- 人工智慧模型建置
- 研究結果
- 結論與分析

前言

前言

- 進行溫度推估時常使用統計學概念的線性回歸模型。
- 資料複雜時使用非線性模型可能會有較好的推估結果。
- 本文透過一種以人工智慧方法進行系集溫度預報場推估至測站溫度研究，使用非線性深度學習模型進行溫度推估。
- 深度學習模型具有遷移式學習的優勢

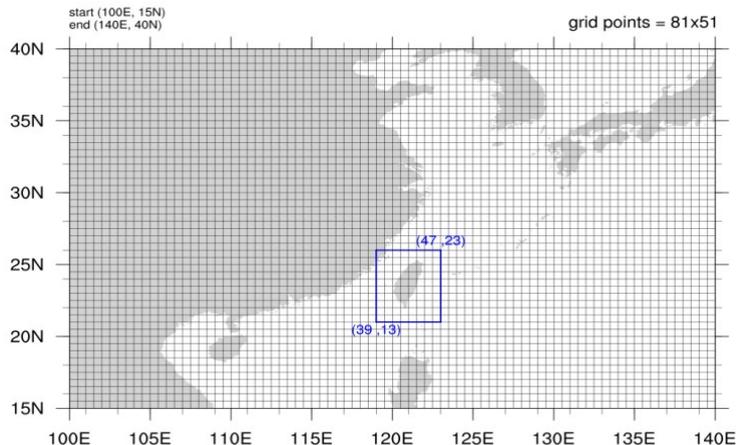
資料格式

系集資料與觀測站資料

系集資料

- 使用中央氣象局提供的全球數值預報模式系集資料
- 時間範圍: 2019年10月1日至2020年1月31日
- 預報時: UTC 00至24(逐1時)
- 系集資料格式:

名稱	內容
氣象變數	距地面2公尺高的溫度(T2M)
經度	東經119至123度
緯度	北緯21至26度
網格解析度	0.5°
格點數	11 x 9
系集成員數	21



- 以人工智慧方法生成高解析度預報延時資料

使用人工智慧方法進行系集溫度預報增加預報延時預報場研究模型由低時間解析度資料生成高時間解析度的系集資料

測站資料

- 時間範圍: 2019年10月1日至2020年1月31日
- 預報時: UTC 00至24(逐1時)
- 測站資料格式:

名稱	內容
氣象變數	測站溫度
資料格式	NetCDF
測站	32站人工觀測氣象站

-扣除:金門、馬祖、五分山、墾丁

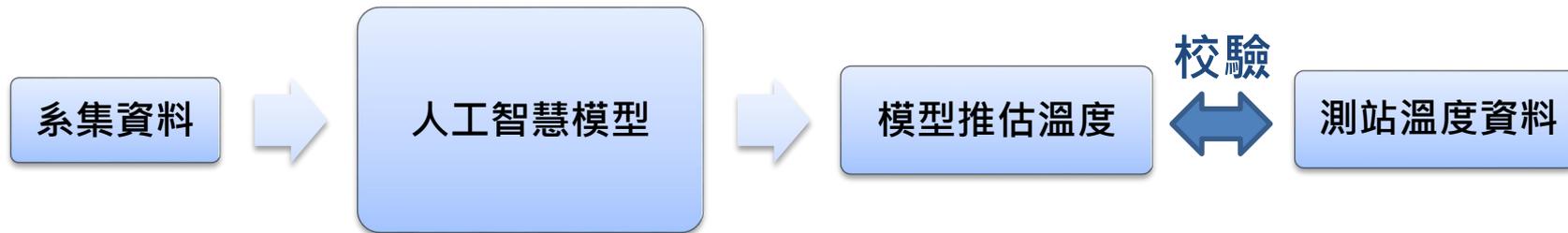
實際建模觀測站代號與名稱表

#	代號	名稱	#	代號	名稱	#	代號	名稱
1	466880	板橋	11	467080	宜蘭	21	467550	玉山
2	466900	淡水	12	467300	東吉島	22	467571	新竹
3	466910	鞍部	13	467350	澎湖	23	467590	恆春
4	466920	臺北	14	467410	臺南	24	467610	成功
5	466930	竹子湖	15	467420	永康	25	467620	蘭嶼
6	466940	基隆	16	467440	高雄	26	467650	日月潭
7	466950	彭佳嶼	17	467480	嘉義	27	467660	臺東
8	466990	花蓮	18	467490	臺中	28	467770	梧棲
9	467050	新屋	19	467530	阿里山			
10	467060	蘇澳	20	467540	大武			

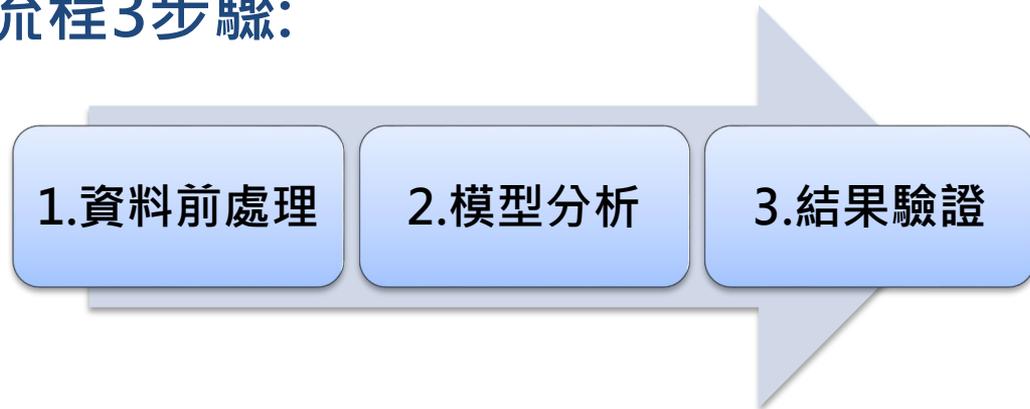
人工智慧模型建置

人工智慧模型建置

- 建置人工智慧模型流程:



- 模型建置流程3步驟:



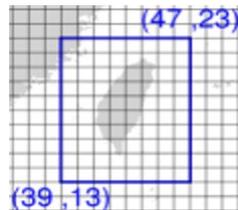
資料前處理

- 資料異常值、缺值處理:
 - 當日某一時間點少一筆資料，則當日的資料全部不用
- 將資料以10月至11月為秋季、12月至1月為冬季分批訓練
- 以一UTC時為一筆資料
 1. 一天UTC00、UTC01至UTC24時 (25個時間點)
 2. 資料格式: (25*天,21,11,9) = (資料筆數,系集成員,11,9)
- 將一季資料隨機以8:2比例分為訓練數據集與校驗數據集

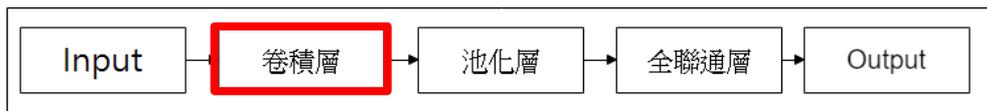
模型分析

- 使用三種不同結構的**CNN(Convolutional Neural Network)**模型:
1D,2D,3DCNN

- Q:為甚麼是CNN模型?A:系集資料的型態



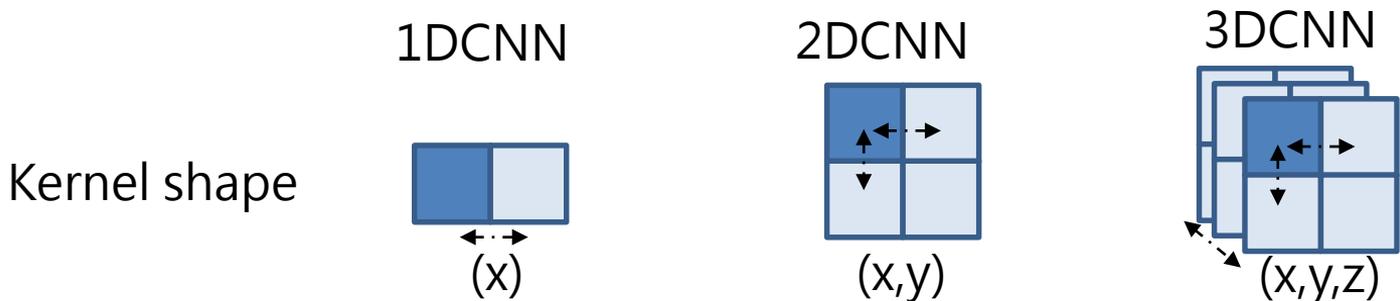
- CNN結構:



- 1D,2D,3DCNN結構**差異**:
模型卷積層的**卷積核(kernel)維度不同**

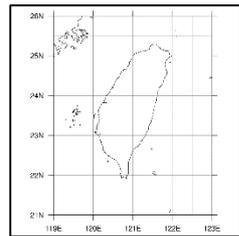
kernel維度與資料格式

- CNN使用Kernel以移動窗格的方式來提取特徵



- 將資料格式轉成模型的輸入格式

Data shape: $(1,21,11,9)$ # (UTC,系集成員數,台灣區域網格點)



Data reshape: $(1,21,11*9)$ $(1,21,11,9)$ $(1,1,11,9,21)$
 (UTC, C, X) (UTC, C, X, Y) (UTC, C, X, Y, Z)

結果驗證

- 使用均方根誤差(RMSE, Root-Mean-Square Error)作為模型推估溫度與實觀測溫度的差異值。

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}}$$

- 使用皮爾遜相關係數(PCCS, Pearson product-moment correlation coefficient)表示模型推估和實際觀測的相近程度。

$$\text{PCCS} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

研究結果

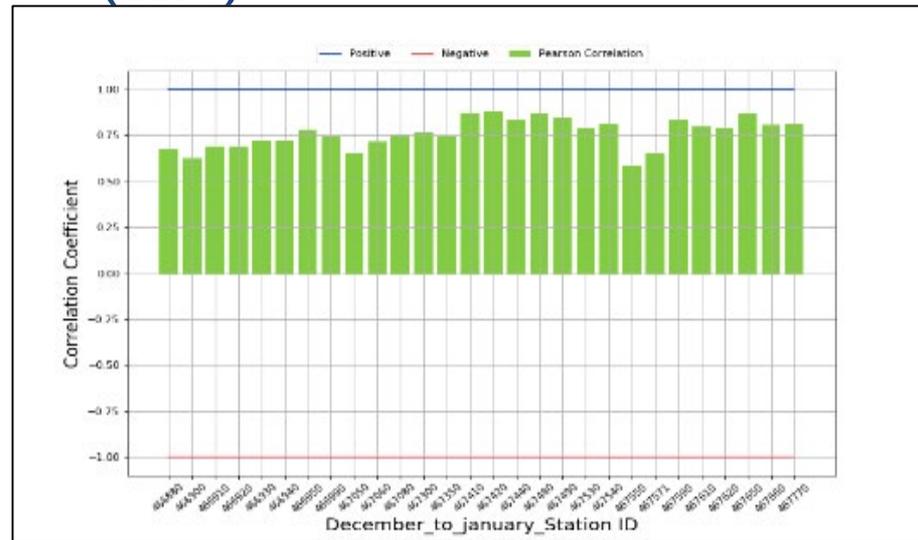
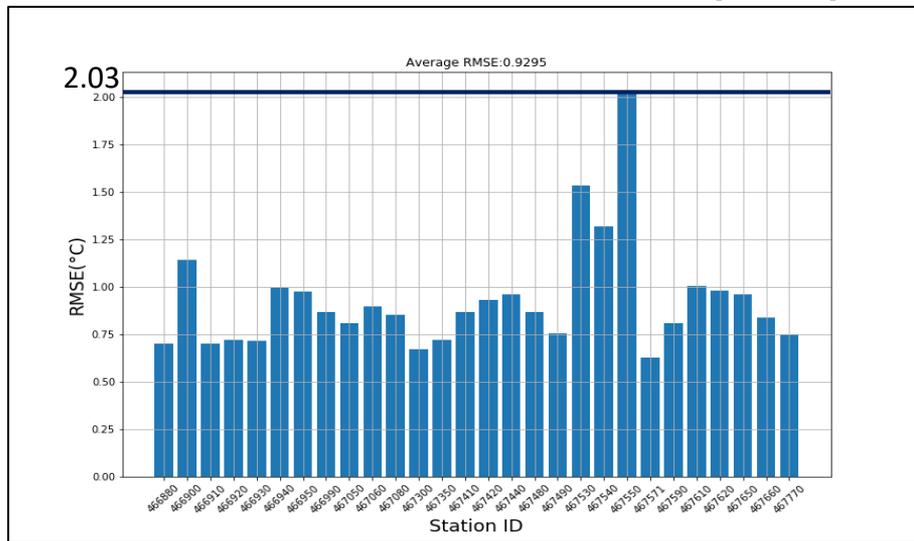
研究結果

- 使用訓練數據集訓練模型，用校驗數據集做推估，保存5次中推估結果最好的模型權重。

預報時	Count 1	Count 2	Count 3	Count 4	Count 5	Average
AVG	0.915	0.933	0.934	0.922	0.944	0.929

(AVG:25個UTC時平均RMSE)

- 5次的推估的平均RMSE(左圖)、PCCS(右圖)

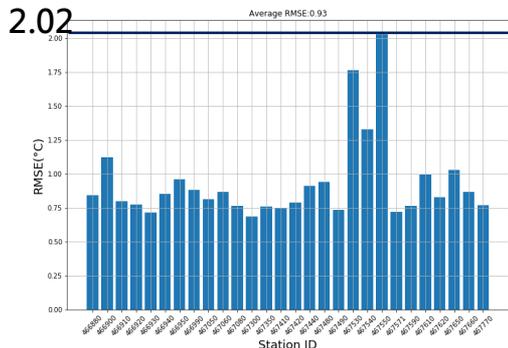


秋季研究結果

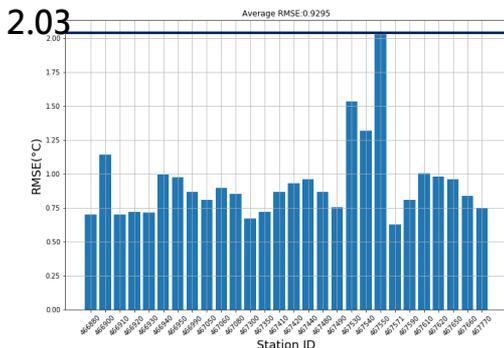
1DCNN

2DCNN

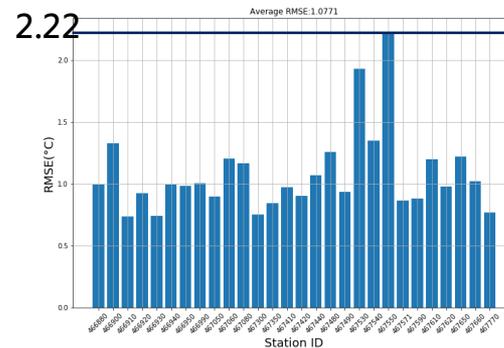
3DCNN



AVG RMSE = 0.93



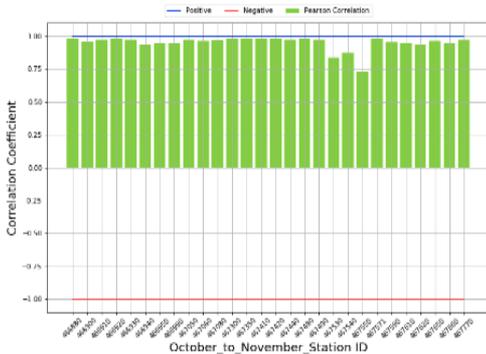
AVG RMSE = 0.9295



AVG RMSE = 1.0771



AVG PCCS \approx 0.9



AVG PCCS \approx 0.9



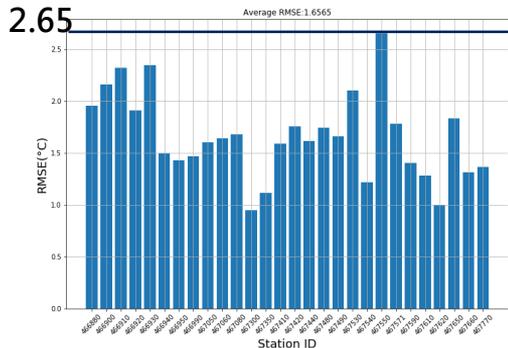
AVG PCCS \approx 0.9

冬季研究結果

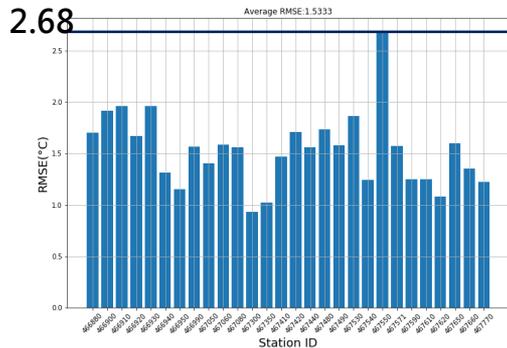
1DCNN

2DCNN

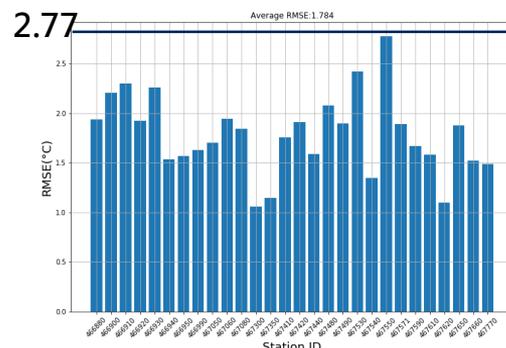
3DCNN



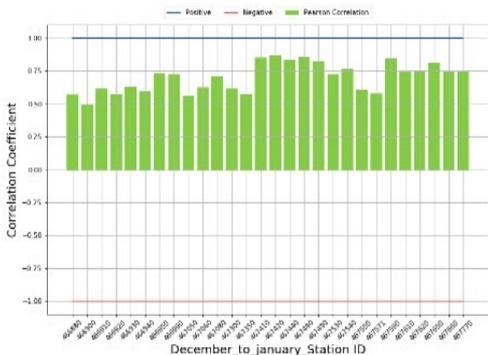
AVG RMSE = 1.6565



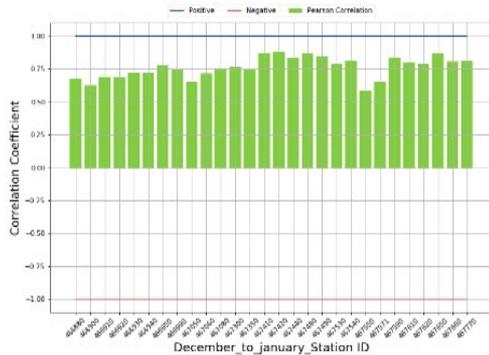
AVG RMSE = 1.5333



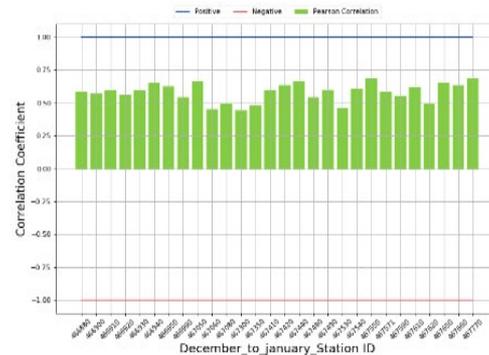
AVG RMSE = 1.784



AVG PCCS \approx 0.7



AVG PCCS \approx 0.75



AVG PCCS \approx 0.6

綜合分析結果

平均RMSE(°C)	1DCNN	2DCNN	3DCNN
秋季	0.93(°C)	0.929(°C)	1.077(°C)
冬季	1.656(°C)	1.533(°C)	1.784(°C)

綜合分析結果顯示

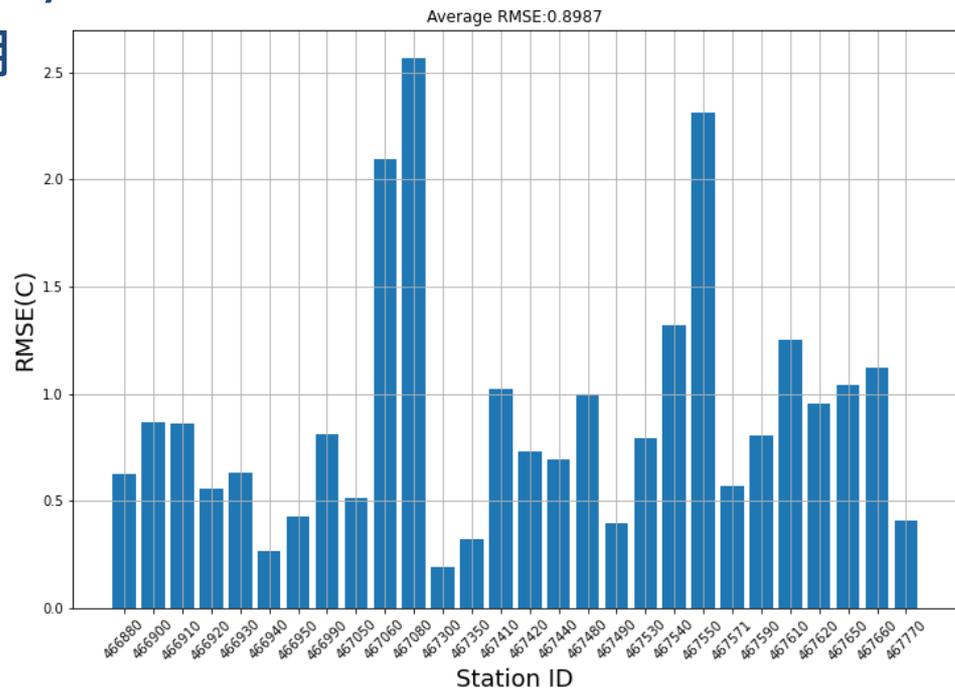
- **2DCNN**有最低RMSE及最穩定表現
- 整體PCCS平均在7成以上(秋季近乎9成)
- 各模型普遍最高RMSE值落於高山區(玉山與阿里山)

多變數3DCNN模型建置

- 資料型態:系集全球模式
- 變數層場:u 850, v 850、H 500、H 850、SLP、T2m
- 預報時間解析度/長度:逐6小時/72小時
- 年分:2018年1月~2020年3月

初步分析結果:

- AVG RMSE = 0.8987**
- Top1: 宜蘭 2.5682(°C)
- Top2: 玉山 2.3136(°C)
- Top3: 蘇澳 2.0946(°C)



結論分析

結論分析

- **2DCNN**有最低RMSE值及最穩定表現
- 各模型普遍最高RMSE值落於山區觀測站
 - 使用**高空間網格解析度**資料有機會降低台灣山區的溫度推估誤差值
- **多變數3DCNN**模型分析結果
 - 使用六種變數**增加資料複雜性**的多變數3DCNN有較低的RMSE值
- **提升網格解析度與增加資料複雜性能有效的提高精準度**
- **相同時效完成模型建置與推估需提升運算速度**
 - 本研究使用**GPU**加速運算，達成模型再增加資料量的情況下**訓練時間與原先相近**

YOU NEED AI

YOU NEED AI

WE Are HERE

harry_chang@leadtek.com.tw

