

數據與相關檔案下載

1. 數據資料下載連結

https://drive.google.com/drive/folders/1LhD7BHgP_cy06rS9vD2cSnXbKNEpM4-?usp=drive_link

2. 參考 Paper:

https://drive.google.com/file/d/1Pd8lAdjanATPT868n9UTACgqW_2yE-vg/view?usp=drive_link
https://drive.google.com/file/d/1Pd8lAdjanATPT868n9UTACgqW_2yE-vg/view?usp=drive_link

3. AI 幫我做的參考程式碼:

https://drive.google.com/file/d/17aH4j2Id63-bRWBODMwJmbMmv-Ea0aRK/view?usp=drive_link

主要分為兩個階段分析：

4. 第一階段：長期(1979-2019)日射量數據分析與特徵提取，
5. 第二階段：近期(2020-2022)日射量與太陽光電發電數據的建模與預測。

1. 長期日射量資料分析 ([Longterm_RADIATION_N].csv)

(1) 資料範圍與處理

- 使用 1979 年至 2019 年（共 41 年）的地表太陽向下輻射量（Surface Solar Radiation Downward, SSRD），作為長期日射量數據來源。
- 將每年資料依五個氣候季節分組，分別為春季（2、3、4 月）、梅雨季（5、6 月）、夏季（7、8、9 月）、秋季（10、11 月）、冬季（12、1 月）。

(2) 季節性趨勢分析與特徵提取

- 使用 STL (Seasonal-Trend Decomposition) 方法分解每一季的趨勢與季節性成分。
- 透過 Mann-Kendall 檢驗驗證長期趨勢的統計顯著性。
- 計算長期變化趨勢斜率、以及季節性特徵如(平均值、標準差)等統計特徵（每一季得到一組）。
- 每季一組固定長期特徵 {slope, mean, std, } → 作為後續模型的外生變數模板 (Template Mapping)。

2. 近期日射量與太陽光電發電數據建模與預測 ([2020-2022 mapping_N].csv)

(1) 資料前處理

- 使用 2020 年至 2022 年間的日射量 (RADIATION) 及太陽能發電量 (POWER) 觀測數據。
- 針對資料異常值、缺失值進行處理，如線性插補或季節均值填補。

(2) 機器學習與時間序列模型建構

A. LSTM 模型

- 觀測變數：2020 - 2022 年日射量 (RADIATION) 與發電量 (POWER)
- 先讓 LSTM 學習時間序列主特徵 (日射量) → 再串接靜態特徵 {slope, mean, std, } 在 LSTM 輸出後 Concatenate，最後進入全連接層進行預測。
- 使用 ASOA (Adaptive Seagull Optimization Algorithm) 優化 LSTM 超參數 (hidden units、batch size、learning rate)，調整 LSTM 模型的參數。
- 訓練集與驗證集使用 Walk Forward Optimization：逐步前推式訓練、更新。
- 使用預測指標：RMSE、MAPE、 R^2 等指標進行效能評估

B. SARIMAX 模型

- 觀測變數：2020 - 2022 年每日日射量 (RADIATION) 與發電量 (POWER)
- 依照月份對每筆資料做「季節標籤映射」，例如例如：2021-02 → 春季 → 加入「春季 slope、mean、std」等特徵。
- 將數據切分為 70% 訓練集、20% 驗證集、10% 測試集，在訓練與驗證階段，採用 Walk Forward Optimization, WFO 方法以擴增資料量。
- 保留最後 10% 資料不參與訓練及驗證，僅做為最後測試效能使用。
- 透過 ACF (Autocorrelation Function) 與 PACF (Partial Autocorrelation Function) 分析決定最適的自回歸 (AR) 與移動平均 (MA) 項數，並透過 AIC/BIC 指標選擇最優模型參數。
- 使用預測指標：RMSE、MAPE、 R^2 等指標進行效能評估

C. 季節性分組預測表現。

假設 LSTM 的預測效果比較好，以上述相同方法，進一步針對春季（2、3、4 月）、梅雨季（5、6 月）、夏季（7、8、9 月）、秋季（10、11 月）、冬季（12、1 月）等五個季節分組建立各自的 LSTM 模型，進一步以不同季節 的模型與不分季節的模型比對預測結果效能。

產出圖表：

➤ 程式需要能出以下圖形自動存檔，參考 Paper 中有範例。

Fig. 4	ASOA + LSTM Flowchart	完整流程圖：從資料處理 → 分解 → LSTM訓練 → 組合預測
Fig. 5	RSTL Decomposition	展示原始資料如何被分為Trend、Seasonal、Residual
Fig. 6	Energy generation during the study period	2020–2023年太陽能實際發電量折線圖
Fig. 7	Actual vs Prediction (各模型)	不同模型預測與實際值比對圖
Fig. 8–10	RMSE 比較圖：Train、Val、Test	不同模型的RMSE表現條狀圖
Fig. 11–13	MAPE 比較圖：Train、Val、Test	同上，改用 MAPE
Fig. 14–16	R ² 比較圖：Train、Val、Test	模型決定係數(R ²)條狀圖

- 須提供 Training and Validation Loss & Accuracy 圖， Loss & Accuracy 圖須曲線正常。
- 圖檔自動存成 png 格式，圖檔檔名和圖檔標題一樣(檔名重複時自行加入關鍵字區分)，圖檔 DPI 設定 300，圖檔大小約大概 2400 × 1800 pixels 左右就可。(參考以下 Code)

```
python

plt.figure(figsize=(8, 6))
plt.savefig("ppt_chart.png", dpi=300, bbox_inches='tight')
```

- 程式需產出以下數據。與提供表格，參考 Paper 中有範例。

Table 4	Hyperparameter Values	LSTM模型訓練參數如 learning rate, batch size
Table 5	RMSE Results	各模型在Train/Val/Test的RMSE統計
Table 6	MAPE Results	各模型MAPE 統計資料
Table 7	R ² Results	各模型 R ² 統計資料
Table 8	PACF Result on Residuals	殘差分析的統計測試結果 (如 Ljung-Box p值)

本專案之驗收須符合下列條件：

- 須提供 Python 原始程式碼(需寫繁體中文註解)，需要能在我電腦上 Windows 正常運作，須提供安裝說明文件，需遠端協助安裝程式環境與教學。
- 訓練完成後可儲存模型，並可載入進行推論。
- LSTM 與 SARIMAX 模型預測效能在測試資料中，須達成 RMSE ≤ 2.0 kWh、MAPE $\leq 10\%$ 、R² ≥ 0.85 (具體值可依 POWER 單位調整)
- 此案內容須完全保密，不可透漏給任何人。
- 需在 10 天內完成，謝謝!