自動化爬蟲程式

本研究主要透過6個不同地方提供的資料來進行發電量的預測，這六個分別是：中央氣象局預測資料(CWB)、中央氣象局觀測站歷史資料(OBS\_CWB)、國外天氣預報(OWM)、晴空輻射資料、中興大學提供的天氣資料、彰師大提供的發電量資料

* 中央氣象局預測資料(CWB)：目前是拿中央氣象局提供各縣市未來3天的預測資料，程式檔案名稱為Crawler.CWB.3H，程式每小時會抓取一次離彰師大最近縣市的天氣資料，並將抓取下來的資料儲存在CWB.3H資料夾下的F-D0047-017資料夾中
* 中央氣象局觀測站歷史資料(OBS\_CWB)：主要是拿觀測站歷史資料，程式檔案名稱為crawler\_Observation\_CWB，程式每天中午12點會抓取一次離彰師大最近的觀測站歷史天氣資料，已獲得昨天的歷史資料，並將抓取下來的資料和以往舊有的資料進行表的合併，並儲存在Observation\_CWB資料夾下
* 國外天氣預報(OWM)：目前是拿OpenWeatherMap提供各縣市未來5天的預測資料，程式檔案名稱為Crawler.OpenWeatherMap.3H，程式每小時會抓取一次離彰師大最近縣市的天氣資料，並將抓取下來的資料儲存在OpenWeatherMap.3H資料夾下的500資料夾中
* 晴空輻射資料：目前是使用pvlib.location套件來獲得晴空輻射資料，透過該地區的經緯度獲得，經緯度資料是彰師大所提供的案場(solar\_汙水廠)位置資訊，資訊內容在Plant\_Info\_Baoshan.csv，程式每天凌晨0點會爬取一次，已獲得當天所有時段的輻射資料
* 中興大學提供的天氣資料：目前是只拿中興大學的天氣輻射資料，當下所抓取的資料會給過往4天(加今天)的輻射資料和未來4天的輻射資料，如下表

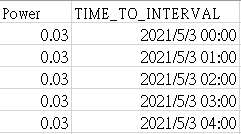
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| D-3 | D-2 | D-1 | D | D+1 | D+2 | D+3 | D+4 |

程式檔案名稱為get\_ColifeAPI\_MSM，程式每天中午12點會爬取一次，已獲得8天輻射資料，並將其儲存在MSM\_data資料夾中，程式還會將所資料夾中的所有CSV表合併成兩個大表，分別存在SAVE資料夾中：solar\_汙水廠\_dswrfobs和solar\_汙水廠\_dswrfpred，以供後續使用

solar\_汙水廠\_dswrfobs：主要抓取所有表當天的輻射資料

solar\_汙水廠\_dswrfpred：主要抓取所有表當日日期的未來四天輻射資料

* 彰師大提供的發電量資料：主要抓取他們所提供的發電量資料，程式檔案名稱為get\_power，程式每小時爬取一次，透過influxdb套件來連線彰師大資料庫，並抓取MG1\_PV表下的所有資料，並將抓來的新資料和歷史資料合併，合併後存在power\_data 資料夾下的solar\_汙水廠\_history.CSV中。資料內容如下：



資料整理和清洗

本研究有兩隻程式整理天氣預報資料，分別是read\_CWB\_3H和read\_OWM\_3H

* read\_CWB\_3H：主要是在整理抓取到的天氣預報資料，將所需要的資料拿出來，並根據時間將其分類成「24小時前」、「3小時前」和「3小時內」的資料，並將資料欄位重新命名(後續好和其他天氣預報資料合併)，將CWB.3H資料夾內的所有檔案進行資料的合併，並將其存在CWB.3H\Save資料夾中，裡面主要分成兩個資料表，CWB.3H.Merge.Multiple和CWB.3H.Merge.Multiple(new)

CWB.3H.Merge.Multiple：歷史CWB預報天氣資料(合併每日天氣資料)

CWB.3H.Merge.Multiple(new)：每日和舊資料合併的新資料

* read\_OWM\_3H：主要是在整理抓取到的天氣預報資料，將所需要的資料拿出來，並根據時間將其分類成「24小時前」和「24小時內」的資料，並將資料欄位重新命名(後續好和其他天氣預報資料合併) ，將OpenWeatherMap.3H資料夾內的所有檔案進行資料的合併，並將其存在OpenWeatherMap.3H\Save資料夾中，裡面主要分成兩個資料表，OWM.3H.Merge.Multiple(merge)和OWM.3H.Merge.Multiple(new)

OWM.3H.Merge.Multiple(merge)：歷史OWM預報天氣資料(合併每日天氣資料)

OWM.3H.Merge.Multiple(new)：每日和舊資料合併的新資料

所有資料表合併

本研究主要有兩支資料再進行合併，分別是old\_and\_new\_data\_merge和history\_merge\_new

* old\_and\_new\_data\_merge：程式主要是在每小時的02分抓取所有小表的資料，並將其合併成大表，將新的資料和舊有的歷史資料合併。
* history\_merge\_new：程式主要是在從頭開始合併大表，例如：將日期設定2021/04/01~2023/01/19，程式會抓取設定日期區間內所有小表的資料，並將其合併。

用法：當今天爬蟲程式有問題導致某天的資料都遺失時，old\_and\_new\_data\_merge所合併出的表內資料也是遺失的，當今天爬蟲的資料補齊時，可以透過該程式將大表所遺失的資料補齊。

模型預測

* **發電量提前一日預測**

發電量提前一日解釋:

使用當前的發電量去預測明天同一時間的發電量

ex.假設2023/01/19 05:00:00~05:59:59[晚上五點]發電量為20 kwh，則要透過該資料和其他天氣欄位來獲得2023/01/20 05:00:00~05:59:59[晚上五點]的發電量資料

目前的預測程式名稱為pred\_power\_d\_6-17\_2022，主要在預測2021/04/01~2022/10/31的發電量，以下是預測績效

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 發電量 | 訓練筆數 | 測試筆數 | 預測時間 | 使用欄位 | nMape | nRMSE | nMAE | RMSE | MAE |
| 隔日逐時 | 3725 | 345 | 06:00~17:00 | Radiation(SDv3)(CWB) | 7.82 | 36.2 | 24.86 | 1.1 | 0.76 |

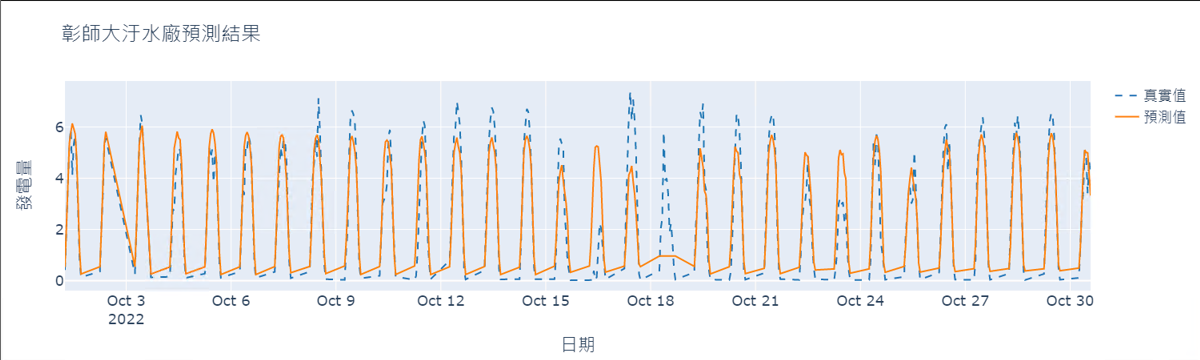
使用欄位為Radiation(SDv3)(CWB)的-1、0、+1小時資料

Ex：本研究會使用該時段發電量、前一小時Radiation(SDv3)(CWB)、該時段Radiation(SDv3)(CWB)、下一小時Radiation(SDv3)(CWB)來預測明天該時段的發電量

Radiation(SDv3)(CWB)：根據CWB預報獲得該時段的明日天氣預測，用該天氣去尋找相似的18筆歷史資料[時間由最近排到最遠]，並將18筆的輻射值做平均，以獲得明日該時段的預測平均輻射。

Radiation(SDv3)(CWB)-1：和上述一樣，只是明日天氣預測是抓取該時段的前一小時。

Radiation(SDv3)(CWB)+1: 和上述一樣，只是明日天氣預測是抓取該時段的下一小時。



* **發電量提前一小時預測**

發電量提前一小時解釋:

使用當前的發電量去預測當日下一小時的發電量

ex.假設2023/01/19 05:00:00~05:59:59[晚上五點]發電量為20 kwh，則要透過該資料和其他天氣欄位來獲得2023/01/19 06:00:00~06:59:59[晚上六點]的發電量資料

目前的預測程式名稱為pred\_power\_h\_6-17\_2022，主要在預測2021/04/01~2022/10/31的發電量，以下是預測績效

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 發電量 | 訓練筆數 | 測試筆數 | 預測時間 | 使用欄位 | nMape | nRMSE | nMAE | RMSE | MAE |
| 當日逐時 | 3952 | 365 | 06:00~17:00 | Radiation(SDv3)(CWB)  pre\_Power | 5.27 | 23.47 | 16.71 | 0.72 | 0.51 |

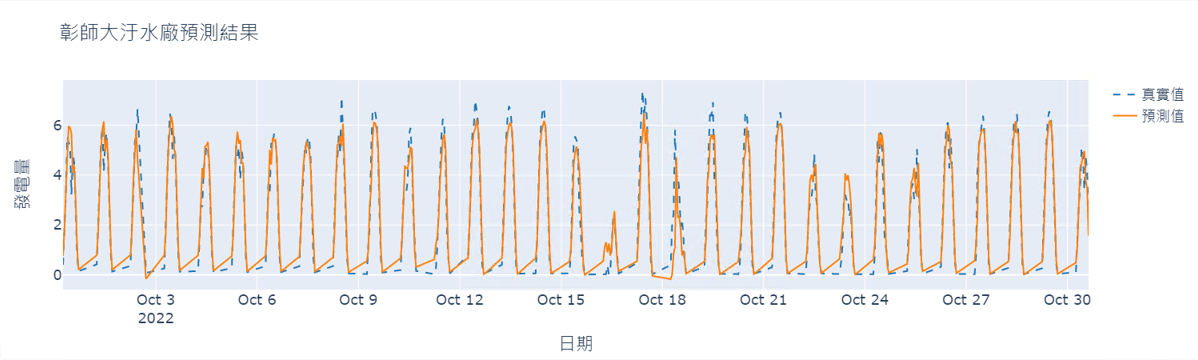
使用欄位為Radiation(SDv3)(CWB)的-1、0和pre\_Power-1資料

Ex：本研究會使用該時段發電量、前一小時Radiation(SDv3)(CWB)、該時段Radiation(SDv3)(CWB)、前一小時電量來預測明天該時段的發電量

Radiation(SDv3)(CWB)：根據CWB預報獲得該時段的當日天氣預測，用該天氣去尋找相似的18筆歷史資料[時間由最近排到最遠]，並將18筆的輻射值做平均，以獲得當日下一時段的預測平均輻射。

Radiation(SDv3)(CWB)-1：和上述一樣，只是當日天氣預測是抓取該時段的前一小時。

pre\_Power-1：使用上一時段的發電量



* **發電量提前15分鐘預測**

發電量提前15分鐘解釋:

使用當前的發電量去預測當日下15分鐘的發電量

ex.假設2023/01/19 05:00:00~05:14:59[晚上五點00分]發電量為20 kwh，則要透過該資料和其他天氣欄位來獲得2023/01/19 05:15:00~05:29:59[晚上五點15分]的發電量資料

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 發電量 | 訓練筆數 | 測試筆數 | 預測時間 | 使用欄位 | nMape | nRMSE | nMAE | RMSE | MAE |
| 逐15分 | 5320 | 1311 | 06:00~17:00 | Radiation(SDv3)(CWB)  pre\_Power | 4.87 | 24.1 | 14.78 | 0.77 | 0.47 |

使用當前的發電量去預測當日下15分鐘的發電量

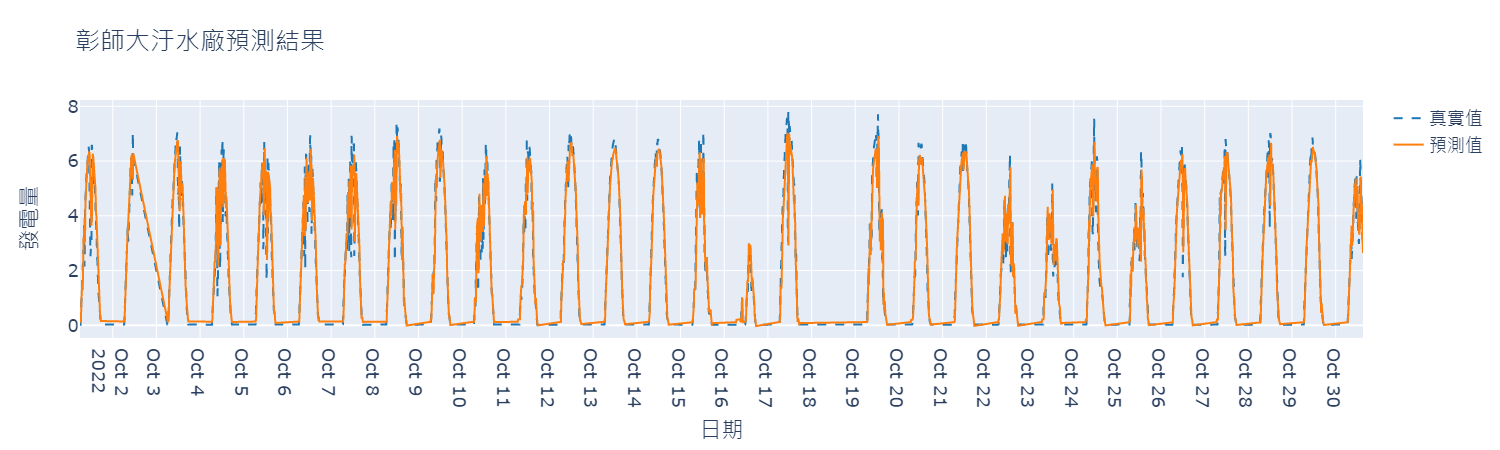
Ex：本研究會使用該時段發電量、pre\_Power\_15、next\_Radiation\_15、pre\_Power\_30來預測下15分鐘的發電量

pre\_Power\_15：前15分鐘的發電量

next\_Radiation\_15：根據CWB預報獲得該時段的當日天氣預測，用該天氣去尋找相似的18筆歷史資料[時間由最近排到最遠]，並將18筆的輻射值做平均，以獲得當日下15分鐘的輻射量的預測平均輻射。

pre\_Power\_30：前30分鐘的發電量

目前的預測程式名稱為2022\_pred\_power\_15m\_6-17，主要在預測2022/02/27~2022/10/31的發電量，以下是預測績效。



old\_and\_new\_data\_merge

目前程式內有寫註解，有問題再問我

* Powre.yml為環境檔