

可解釋的 AI 系統－ 以焦爐氣儲槽液位預測為例

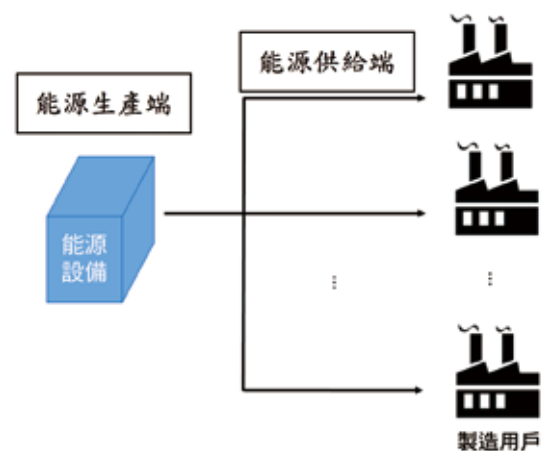
莊雅筑、葉峻賓、陳立杰、陳秉洋、夏啟峻、林順傑
工業技術研究院 資訊與通訊研究所

摘要

在石化／鋼鐵工業中，煉焦製程所產生的焦爐氣可用於發電，然而單一焦爐氣儲槽可能同時要供應給數間下游工廠或是其他能源設備，若在焦爐氣缺乏的情況下，必須使用外購電力維持下游工廠或能源設備的運作，為了避免斷電危機，如何有效的預測下游工廠與能源設備的耗能狀況，並且能夠即時調配能源產消以維持供電穩定，為製程上的一大需求。故本研究除了透過 AI 技術預測焦爐氣儲槽液位外，更進一步對儲槽液位隨時間的變化趨勢進行分析，讓算法可解釋 AI 預測決策過程，發現該變化趨勢下是由何者工廠或設備造成及其影響程度，以能即時提供現場動態產消調整的建議。本 AI 預測技術結合適應性分佈滯後模型（adaptive distributed time-lag model）與區域模型（local model），除了達成時序資料之變因鑑定，與當下最佳訓練子集合挑選之目的外，也能即時量化變因影響程度供現場人員調度決策。

一、概述

在特定的工業製程上，一套能源設備常常須同時提供能源給數個製造用戶使用，如下圖一。而在現行少樣多量的製造趨勢下，用戶的能源使用量將會存在相當程度的變異。



圖一 能源設備與製造用戶使用關係

當用戶大量使用能源時，除了能源供給上可能供不應求外，也會造成能源設備運作的不穩定，進而導致能源生產、供給的效率降低。近年來 AI 技術不斷發展，其應用層面也很廣泛，如：能源預測、影像辨識、物件偵測等，而 AI 就如同一黑盒子，大家只能相信它的結果，而大部分 AI 的運作過程以及最終的執行結果還是需要倚賴人工來理解，又或者是根本無法解讀，若將工廠完全交給 AI 運作，雖然肯定其效能，但是難免會對這無法解釋的黑盒子產生不信任感。因此本研究發展了一可解釋的 AI 系統，將以預測儲槽液位為例，讓 AI 下判斷的理由可以理解，亦可幫助確認它的正確性。本文中第二段會描述預測焦爐

氣儲槽液位的 AI 技術，第三段描述焦爐氣儲槽液位預測分析結果，最後一段為技術研究結論。

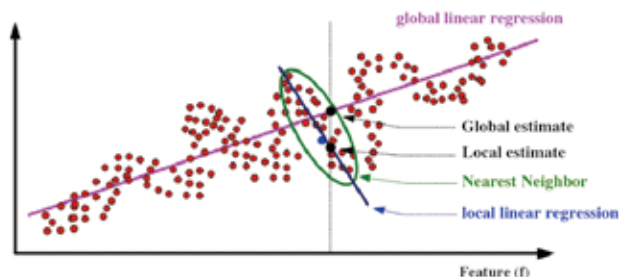
二、AI 預測技術簡介

焦爐氣提供能源給大量的下游用戶時，每個用戶對能源的使用量與使用時機不盡相同，為了避免斷電危機，此時便希望精準地預測能源設備在一段時間後是否超規，並且在當下找出導致超規的來源，以達即時調度之目的。過去對於焦爐氣儲槽液位預測方面的研究大多採以迴歸分析（regression analysis）或是時間序列分析（time series analysis）等方法來預測焦爐氣儲槽液位，但是這些方法卻有其不足或缺失之處。

本篇 AI 預測技術為了預測焦爐氣儲槽液位以及讓算法可解釋 AI 預測決策過程，分別以區域模型（local model）先強化預測準確度，再結合適應性分佈滯後模型（adaptive distributed time-lag model）提供預測影響因子，本段將個別介紹兩項技術以及焦爐氣儲槽液位的預測架構。

（一）區域模型簡介

傳統在預測議題上，習慣使用全部的資料配適模型，我們將此稱作是全域模型（global model），然而這樣的作法雖然可以捕捉到整體資料的趨勢，但是並不太適用在資料分佈太過於震盪的情況下。另一方面，區域模型則試圖在訓練資料中，找出與測試資料相似的子樣本集合進行模型訓練，以期提高預測效能。以圖二 [1] 為例，紅色樣本點為訓



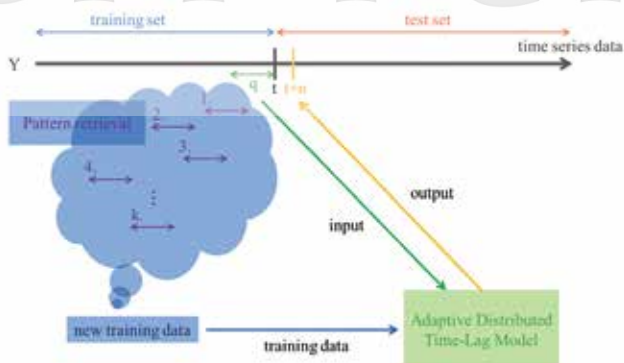
圖二 Local Model 與 Global Model 之建模與預測結果差異圖

練資料集，在其中找出與藍色測試樣本點相似的訓練樣本形成子樣本集合，以綠色橢圓框選之。兩個黑色預測點分別為使用與不使用 Local model 的結果，可以發現使用 Local model 的預測結果比較貼近實際資料的趨勢。

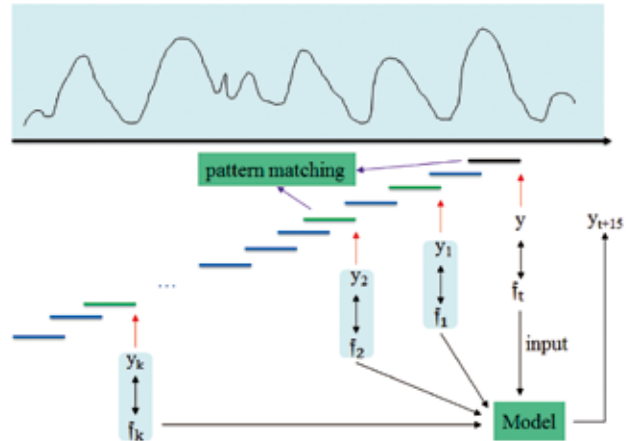
本研究設計兩種 Local Model 的參數向量：自相關預測方法，以及複迴歸預測方法。

請參考圖三，自相關預測方法在測試樣本集的時間點 t 上，由過去訓練資料集中 Y 的歷史資料，預測 $t+n$ 時間點的 Y 。特徵參數部份僅有 Y 值，而預測模型之特徵參數則是從時間點 t 往前取一段長度為 q 的序列做為代表，首先利用 k -最近鄰居法（ k -Nearest Neighbor, kNN ）找出訓練資料集中與在 t 時間上的測試樣本，較相似的 k 筆 q 維資料，並組成專於此筆測試樣本之訓練子資料集。

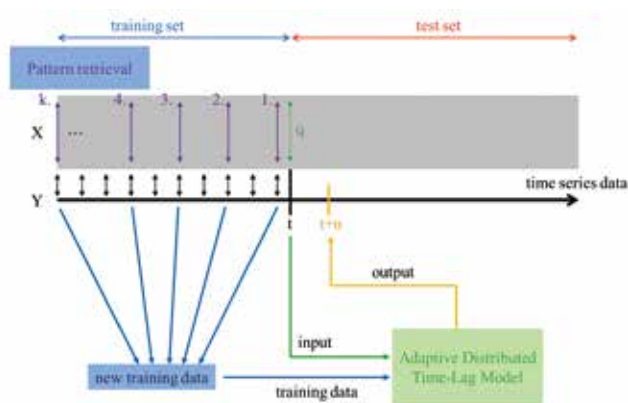
請參考圖四，複迴歸預測方法是在測試樣本集的時間點 t 上，利用過去訓練資料集裡對應 Y 的 X ，預測 $t+n$ 時間點的 Y 。一開始先使用 kNN 找出訓練資料集中與在 t 時間上的測試樣本，並組成專於此筆測試樣本之訓練子資料集，與自相關預測方法的差別在於



圖三 Local Model 搭配預測模組自相關預測方法架構圖



圖五 Pattern retrieval 示意圖



圖四 Local Model 搭配預測模組複迴歸預測方法架構圖

預測模型之特徵參數部份有 X 與 Y 值，其餘作法則相同。

1. Pattern retrieval

假設目前在 y 這個時間點想要預測下 15 個時間點 y_{t+15} ，而用來建模的訓練資料尋找方式為以下所述，首先，從 y 往前取一段 pattern（含 y 本身）做為目標 pattern， f_t ，接著尋找 y 之前的訓練資料中，與 f_t 最相似的前 k 個 pattern， $D_t = \{f_i, 1 \leq i \leq t\}$ ，視為訓練子

資料集進行建模，並對 f_t 進行 y_{t+15} 的預測。而測試樣本與各訓練 pattern 之間的相似度，依需求定義四種 pattern matching 的方式，詳述於下一節。

2. Pattern matching

此節定義四種 pattern matching 的方式，包含單變量歐式距離計算、單變量一階差分量距離計算、多變量歐式距離計算與多變量一階差分量距離計算方式，並在最末段落介紹過濾重疊率過高之相近 pattern 方式。

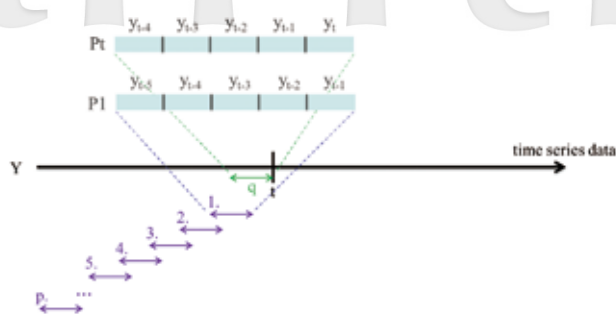
(1) 單變量歐式距離計算

計算兩兩單變量 pattern 之間的歐式距離時，可依照下述歐式距離公式來進行計算：

$$D(P_t, P_1) = \sqrt{\sum_{i=0}^{q-1} (y_{t-i} - y_{t-i-p})^2} \quad (1)$$

公式 (1) 中 P_t 與 P_1 分別代表測試與某一個訓練 pattern， q 為 pattern 的長度， p 為訓練資料 pattern 個數之 index，依圖六例 $q = 5$ ， $p = 1$ 。

(2) 單變量一階差分量距離計算



圖六 單變量 pattern 歐式距離計算例圖

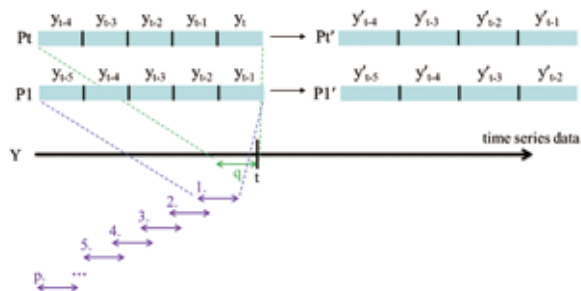
計算兩兩單變量 pattern 之間的一階差分量距離時，需先經過一階差分的計算後，再依照下述歐式距離公式來進行計算：

$$D(Pt', P1') = \sqrt{\sum_{i=1}^{q'} (y'_{t-i} - y'_{t-i-p})^2} \quad (2)$$

公式中 Pt' 與 $P1'$ 分別代表測試與某一個訓練 pattern，經過一階差分計算後的結果， q' 為 pattern 經過一階差分後的長度， p 為訓練資料 pattern 個數之 index，依圖七例，則 $q'=q-1=4$ ， $p=1$ 。

(3) 多變量歐式距離計算

計算兩兩多變量 pattern 之間的歐式距離時，可依照下述歐式距離公式來進行計算：



圖七 單變量 pattern 一階差分量距離計算例圖

$$D(Pt, P1) = \sum_{i=0}^{q-1} \sqrt{\sum_{j=1}^d (x_{t-i,j} - x_{t-i-p,j})^2} \quad (3)$$

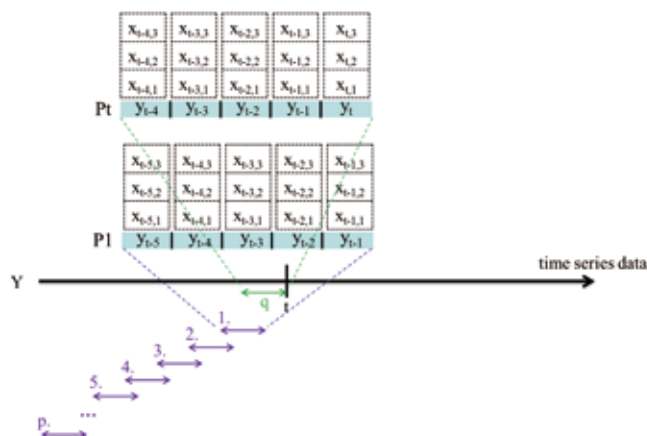
公式 (3) 中 Pt 與 $P1$ 分別代表測試與某一個訓練 pattern， q 為 pattern 的長度， d 為 pattern 上每個點之維度， p 為訓練資料 pattern 個數之 index，依圖八例，則 $q=5$ ， $d=3$ ， $p=1$ 。

(4) 多變量一階差分量計算

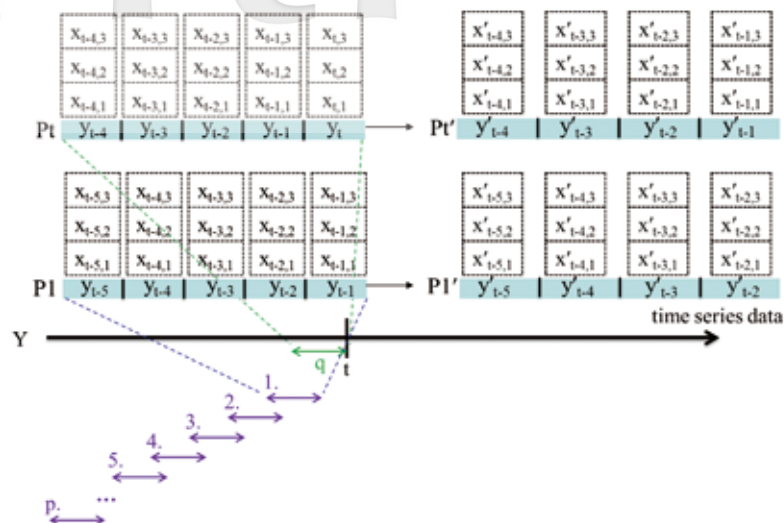
計算兩兩多變量 pattern 之間的一階差分量距離時，需先經過一階差分的計算後，再依照下述歐式距離公式來進行計算：

$$D(Pt', P1') = \sum_{i=1}^{q'} \sqrt{\sum_{j=1}^d (x'_{t-i,j} - x'_{t-i-p,j})^2} \quad (4)$$

公式 (4) 中 Pt' 與 $P1'$ 分別代表測試與某一個訓練 pattern 經過一階差分計算後的結果， q' 為 pattern 經過一階差分後的長度， d 為每個點之維度， p 為訓練資料 pattern 個數之 index，依圖九例，則 $q'=q-1=4$ ， $d=3$ ， $p=1$ 。



圖八 多變量 pattern 歐式距離計算例圖

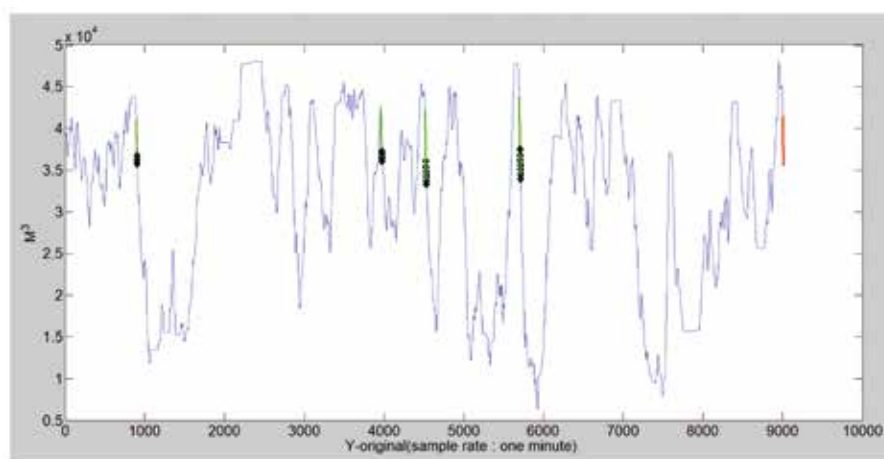


圖九 多變量 pattern 一階差分距離計算例圖

(5) 高重疊率 pattern 濾除

圖十為原始焦爐氣儲槽液位的時序圖，經由 kNN 找出與目標 pattern 相似的 k 個 pattern 後，紅色 pattern 為用來尋找相似 pattern 的目標，綠色 pattern 則為 kNN 找到的相似 pattern，黑色菱形則代表每個相似片段的最末點，將這些 pattern 畫在焦爐氣液位曲線上，觀察其分佈的範圍，發現當指定

的 k 值較小時（意指尋找的相似 pattern 數量較少），找到的相似 pattern 在原始焦爐氣液位曲線中的位置較容易集中在某幾個區段，就會發生在一段曲線中重複採樣數次，如不想讓找到的相似 pattern 集中在某些區段，就必須改變或調整尋找相似片段的方式，以下敘述如何以門檻值方式，找出相似 pattern 重疊率。

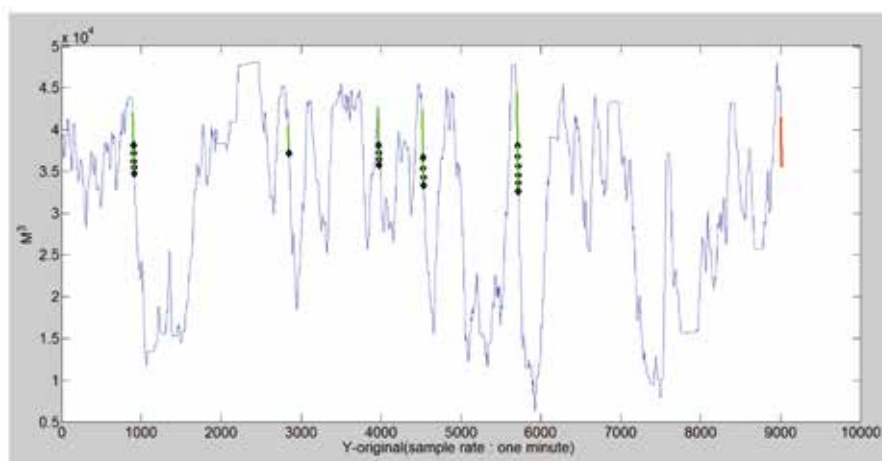


圖十 未加入重疊率門檻值之相似 pattern 在原始曲線中的位置

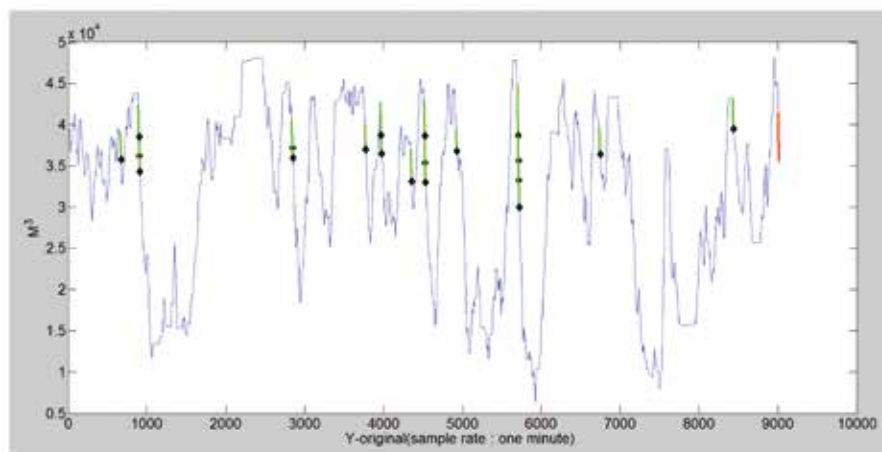
首先設定好濾除的門檻值後，利用 kNN 依序取出與目標 pattern 距離最小的 pattern，去檢查此 pattern 是否與先前已納入要做為訓練資料的 pattern 重疊程度是否超過設定的門檻值，若與先前任一個 pattern 重疊程度超過設定之門檻值比例時，則此相似 pattern 就不做為訓練資料，反之則可納入做為訓練資料。

為了避免找到的相似 pattern 會有集中在某些區間的現象，本研究加入了 pattern 重

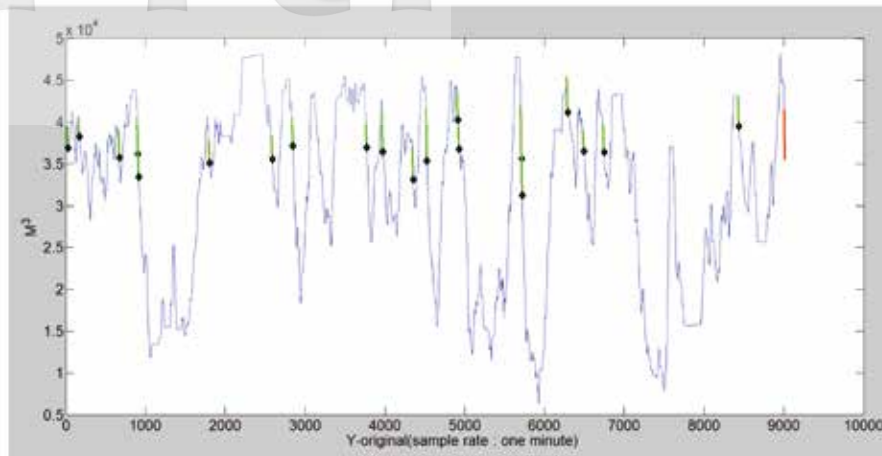
疊率門檻值的限制，當相似 pattern 彼此之間的重疊率高於所設定之門檻值時，則此相似 pattern 不做為訓練資料，圖十一、圖十二、圖十三為使用重疊率門檻值 0.8、0.5、0.2 所找到的前 20 個最相似的 pattern 在原始焦爐氣液位曲線的位置。從圖十一至十三，可以觀察到，隨著相似 pattern 重疊率門檻值的減少（代表同一線段 pattern 的可重疊率越少）時，相似 pattern 的確隨著門檻值減少而分散。



圖十一 重疊率門檻值設定為 0.8 時，相似 pattern 在原始曲線中的位置



圖十二 重疊率門檻值設定為 0.5 時，相似 pattern 在原始曲線中的位置



圖十三 重疊率門檻值設定為 0.2 時，相似 pattern 在原始曲線中的位置

(二) 適應性分佈滯後模型 (adaptive distributed time-lag model) 簡介

對於焦爐氣的能源控管，除了要預測焦爐氣的液位外，還需達成時序資料之變因鑑定，與當下最佳訓練子集合挑選之目的，即時量化變因影響程度供現場人員調度決策。為了找出影響焦爐氣在連續時間上的趨勢中，也就是當在時間點 t 所觀察到的焦爐氣液位 Y_t 可能受到當期或前期自變數 (X_t 、 X_{t-1} 、 X_{t-2} 、 \dots) 的影響時，可以使用分佈延遲 (distributed lag, DL) 模型來描述其間的關係。若資料有 p 個自變數，且已知 Y_t 可能受第 j 個自變數的影響，並最遠可追溯至延遲 q_j 期，則以 DL 模型來描述 Y_t 與各自變數及其各個期數之間的線性關係如下式：

$$Y_t = \beta_0 + \sum_{j=1}^p \sum_{l=0}^{q_j} \beta_{j,l} X_{j,t-l} + e_t, e_t \sim N(0, \sigma^2) \quad (5)$$


其中， β_0 為截距項、 $\beta_{j,l}$ 為第 j 個自變數之第 l 期延遲的迴歸係數，且 e_t 為服從常態分

配之誤差項。舉例來說，如圖十四，若從時間 $t=0$ 到 $t=8$ 收集了 9 筆反應變數 Y 與兩個自變數 X_1 與 X_2 的資料如圖十四左，考慮兩自變數各延遲 0、1、2 期對 Y 的影響，則將觀察而來的資料排列成下圖十四右之形式作為複迴歸模型 (multiple linear regression model) 的輸入資料。

在變因鑑定的方法中，Robert Tibshirani[2] 在 1996 年所提出的 Least Absolute Shrinkage and Selection Operator (LASSO) 是經常使用的手法，此方法為正規化 (regularization) 迴歸方法的一種，其目標函數 (objective function) 如下式：

$$\min_{\beta} \|Y - X\beta\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (6)$$

其中 β 代表模型參數， j 代表參數個數， λ 為一懲罰參數。此外，Hui Zou 與 Trevor Hastie[4] 為了模型預測正確性與模型解釋性，在 2005 年提出了彈力網 (elastic net)，其想

| Observed Data | | | |  | Data Used in DL Model | | | | | | |
|---------------|----------------|------------------|------------------|---|-----------------------|------------------|--------------------|--------------------|------------------|--------------------|--------------------|
| Time | Y _t | X _{1,t} | X _{2,t} | | Y _t | X _{1,t} | X _{1,t-1} | X _{1,t-2} | X _{2,t} | X _{2,t-1} | X _{2,t-2} |
| 0 | Y ₀ | X _{1,0} | X _{2,0} | | Y ₂ | X _{1,2} | X _{1,1} | X _{1,0} | X _{2,2} | X _{2,1} | X _{2,0} |
| 1 | Y ₁ | X _{1,1} | X _{2,1} | | Y ₃ | X _{1,3} | X _{1,2} | X _{1,1} | X _{2,3} | X _{2,2} | X _{2,1} |
| 2 | Y ₂ | X _{1,2} | X _{2,2} | | Y ₄ | X _{1,4} | X _{1,3} | X _{1,2} | X _{2,4} | X _{2,3} | X _{2,2} |
| 3 | Y ₃ | X _{1,3} | X _{2,3} | | Y ₅ | X _{1,5} | X _{1,4} | X _{1,3} | X _{2,5} | X _{2,4} | X _{2,3} |
| 4 | Y ₄ | X _{1,4} | X _{2,4} | | Y ₆ | X _{1,6} | X _{1,5} | X _{1,4} | X _{2,6} | X _{2,5} | X _{2,4} |
| 5 | Y ₅ | X _{1,5} | X _{2,5} | | Y ₇ | X _{1,7} | X _{1,6} | X _{1,5} | X _{2,7} | X _{2,6} | X _{2,5} |
| 6 | Y ₆ | X _{1,6} | X _{2,6} | | Y ₈ | X _{1,8} | X _{1,7} | X _{1,6} | X _{2,8} | X _{2,7} | X _{2,6} |
| 7 | Y ₇ | X _{1,7} | X _{2,7} | | | | | | | | |
| 8 | Y ₈ | X _{1,8} | X _{2,8} | | | | | | | | |

圖十四 DL 模型輸入資料方式

法是將 L1 懲罰項與 L2 懲罰項作結合，其目標函數如下式：

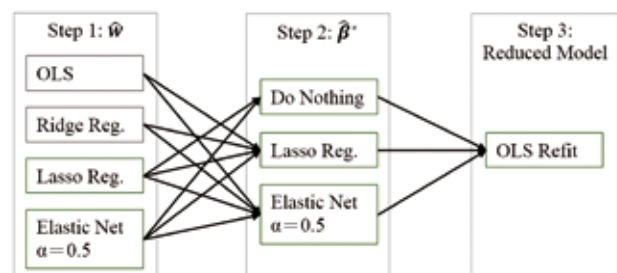
$$\min_{\beta} \|Y - X\beta\|^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p |\beta_j|^2 \quad (7)$$

Zou[3] 在 2006 年點出了傳統 LASSO 迴歸模型與傳統 Elastic Net 迴歸模型的係數可能不具有統計上一致性的性質，並提出兩階段因子篩選的 LASSO 方法，如圖十五，以得到較好的係數估計，稱為適應性 LASSO (adaptive LASSO) 方法。

此外，Zou and Zhang[5] 在 2009 進一步提出適應性 LASSO 在估計模型係數之前，事先對各變數預設了不同的重要性，實現的方法即是在其目標式內各變數的 L1 懲罰項增加了對各變數係數的預設權重 (\hat{w}_{jl})，如公式 (8) 所示。

$$\min_{\beta} \|Y - X\beta\|^2 + \lambda \sum_{j=1}^p \sum_{l=0}^{q_j} \left[(1-\alpha)^{\frac{1}{2}} \beta_{jl}^2 + \alpha \hat{w}_{jl} |\beta_{jl}| \right] \quad (8)$$

關於權重 \hat{w}_{jl} 的設定方式，文獻上提供的方法為：以傳統 LASSO 的係數估計值



圖十五 Adaptive Distributed Time-Lag Model 架構
(該模型包含變數重要性決定、參數估計，與簡化模型三個步驟，前兩個步驟分別可使用幾種不同的方法達成。)

做為適應性 LASSO 的 L1 懲罰項權重，即 $\hat{w}_{jl} = \hat{\beta}(\alpha = 1)$ ，當傳統 LASSO 的係數估計值為 0 時，便於第二階段係數估計時忽略該係數，直接指定其值為 0。同樣地，adaptive elastic net 的 L1 懲罰項權重來源便為傳統 elastic net 的係數估計值。本研究除了依循文獻上的方法之外，也認為適應性方法的權重應可使用其他合理的指定方式，如一般最小平方法的估計值 (ordinary least square, OLS)、Ridge regression 的估計值，或 LASSO 與 elastic net 方法於兩階段交叉使用。

三、焦爐氣儲槽液位分析結果

本研究採用焦爐氣儲槽液位資料集（coke oven gas, COG），此為多變數資料集，資料記錄了2013年7月以每分鐘取樣的焦爐氣數值，資料筆數約為43200筆，資料變數個數為65，本研究將預測下5、10、15分鐘的焦爐氣液位數值，依各實驗設定取得足夠的訓練資料集後，剩餘資料做為測試資料集，為了驗證區域模型對於提升預測準確度的有效性，採用了Partial Least Square（PLS）Regression、Support Vector Regression（SVR）、Classification and Regression Tree（CART）與三個方法搭配區域模型以及高重疊率 pattern 濾除，來進行預測成效之比較，模型的設定如表一所示。而表二為分析的參數設定，其中k代表訓練資料筆數，q表示每筆訓練資料使用的time lag，也就是訓練資料與過去多少筆資料有相關，t*代表要預測的時間為當前時間t的下t*分鐘，也就是時間t + t*時的儲槽液位。在分析結果中，我們記錄每個模型的預測誤差MAPE（mean absolute percentage error）來

表一 分析模型設定

| 預測模型 | 預測模型 + 區域模型（LMA）+ 重疊率門檻值（ δ ） |
|------|--------------------------------------|
| PLS | PLS + LMA + δ |
| SVR | SVR + LMA + δ |
| CART | CART + LMA + δ |

表二 分析參數設定

| 訓練資料筆數 (k) | time lag (q) | 預測下 t* 分鐘 |
|------------|--------------|-----------|
| 30,100,500 | 10 | 5, 10, 15 |

做比較，其計算式如下式(9)，其中 y_t 表時間t時的焦爐氣液位，而 \hat{y}_t 表時間t時模型所預測的焦爐氣液位。

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (9)$$

焦爐氣儲槽液位的預測結果請參考表三，從結果中可以觀察到，當預測的時間間隔t*增加，模型的預測誤差也會隨著增加。而在訓練樣本數k小的時候，加入local model不一定會對預測有幫助，但隨著訓練資料量變大，加入local model可以使預測的MAPE變

表三 焦爐氣儲槽液位資料集預測 MAPE

| Model | MAPE (%) | | | | | | | | |
|-----------------------|--------------|---------------|---------------|--------------------|--------------------|--------------------|---------------|----------------|----------------|
| | k=30 t*=5 | k=30 t*=10 | k=30 t*=15 | k=100 t*=5 | k=100 t*=10 | k=100 t*=15 | k=500 t*=5 | k=500 t*=10 | k=500 t*=15 |
| SVR | 2.84 | 5.56 | 8.16 | 3.08 | 5.67 | 8.21 | 2.74 | 5.52 | 8.11 |
| SVR + LMA + δ | 3.85 | 9.11 | 14.64 | 3.19 | 7.75 | 12.52 | 2.77 | 5.24 | 7.47 |
| PLS | 3.41 | 5.91 | 8.41 | 1.02×10^6 | 9.57×10^5 | 2.37×10^5 | 46.02 | 30.76 | 32.18 |
| PLS + LMA + δ | 4.17 | 9.91 | 16.06 | 4.28 | 14.88 | 54.27 | 2.18 | 4.72 | 7.44 |
| CART | 2.64 | 5.43 | 8.05 | 2.62 | 5.42 | 8.02 | 2.59 | 5.37 | 7.98 |
| CART + LMA + δ | 2.80 | 6.01 | 9.20 | 2.75 | 5.96 | 9.35 | 2.79 | 5.75 | 9.46 |

小，由其是對於 PLS 模型在 $k=100$ 、 $t^*=5$ 的情況，可以看到加入 local model 後，MAPE 明顯下降了。

當模型在預測焦爐氣儲槽液位異常時，必須即時分析製程中的哪個變因發生異常，在這裡我們採用適應性分佈滯後模型做即時的分析，以提供現場人員即時調整調度。以下我們分析樣本數為 600、1200、2400 的焦爐氣液位資料，並搭配不同的兩階段模型來作比較。

表四到六之中，可看出傳統方法所配適的模型，包含其縮減模型，有較高的預測能力，但其重要因子個數較多，而適應性方法能夠進一步選出較少的重要因子，但會犧牲模型的預測能力。此現象於使用前 1200 筆、2400 筆資料分析時更為明顯。值得注意的是，

若直接以分析模型的分析結果為主，適應性方法也能有較傳統方法有更好的模型預測能力，如表四中 Net 0.5+Net 0.5 使用 35 個因子所建

表五 樣本數為 1200 筆的分析方法比較

| 分析方法 | | 重要變數 個數 | 模型 預測效能 | 縮減模型 預測效能 |
|---------|---------|------------|------------|--------------|
| Step 1 | Step 2 | | MAPE% | MAPE% |
| OLS | -- | -- | 1.14 | -- |
| Ridge | -- | -- | 2.05 | -- |
| Lasso | -- | 633 | 1.32 | 1.21 |
| | Lasso | 7 | 10.75 | 10.66 |
| Net 0.5 | -- | 609 | 1.37 | 1.18 |
| | Net 0.5 | 43 | 7.68 | 6.24 |
| Net 0.3 | -- | 647 | 1.34 | 1.16 |
| | Net 0.3 | 51 | 7.86 | 5.81 |
| | Net 0.7 | 44 | 7.48 | 5.82 |
| Net 0.7 | -- | 642 | 1.32 | 1.21 |
| | Net 0.3 | 29 | 9.40 | 8.02 |
| | Net 0.7 | 23 | 9.15 | 8.01 |

表四 樣本數為 600 筆的分析方法比較

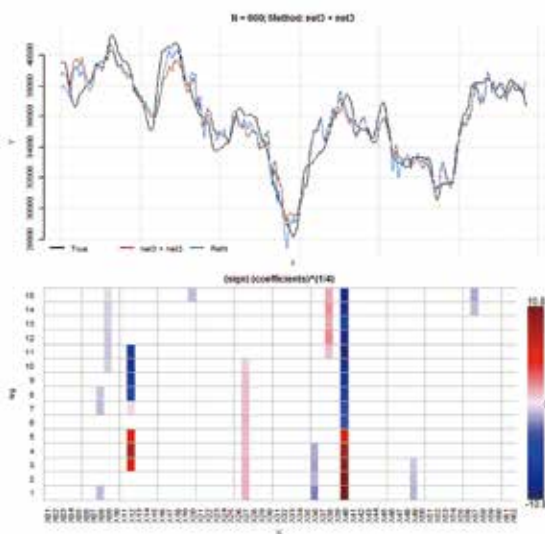
| 分析方法 | | 重要變數 個數 | 模型 預測效能 | 縮減模型 預測效能 |
|---------|---------|------------|------------|--------------|
| Step 1 | Step 2 | | MAPE% | MAPE% |
| OLS | -- | -- | 0.12 | -- |
| Ridge | -- | -- | 1.79 | -- |
| Lasso | -- | 66 | 2.48 | 1.50 |
| | Lasso | 11 | 3.19 | 2.51 |
| Net 0.5 | -- | 54 | 4.06 | 2.60 |
| | Net 0.5 | 35 | 2.98 | 2.69 |
| Net 0.3 | -- | 128 | 2.47 | 1.45 |
| | Net 0.3 | 57 | 2.11 | 1.88 |
| | Net 0.7 | 48 | 2.08 | 1.89 |
| Net 0.7 | -- | 45 | 4.04 | 2.61 |
| | Net 0.3 | 25 | 3.16 | 2.75 |
| | Net 0.7 | 26 | 3.09 | 2.80 |

表六 樣本數為 2400 筆的分析方法比較

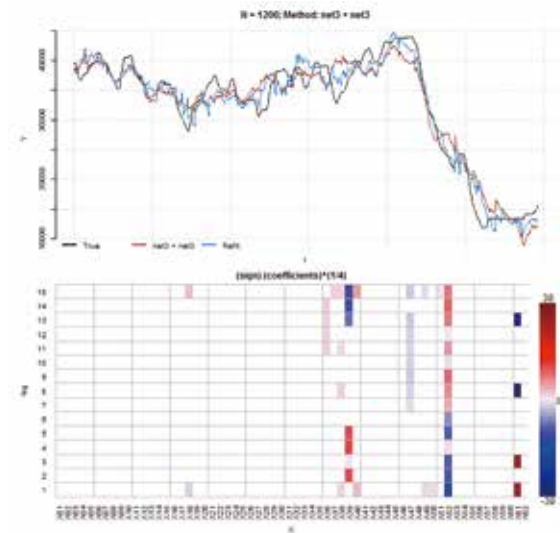
| 分析方法 | | 重要變數 個數 | 模型 預測效能 | 縮減模型 預測效能 |
|---------|---------|------------|------------|--------------|
| Step 1 | Step 2 | | MAPE% | MAPE% |
| OLS | -- | -- | 3.98 | -- |
| Ridge | -- | -- | 4.58 | -- |
| Lasso | -- | 423 | 4.32 | 4.23 |
| | Lasso | 26 | 11.53 | 11.37 |
| Net 0.5 | -- | 452 | 4.32 | 4.23 |
| | Net 0.5 | 48 | 10.47 | 8.53 |
| Net 0.3 | -- | 459 | 4.34 | 4.22 |
| | Net 0.3 | 66 | 10.09 | 7.21 |
| | Net 0.7 | 49 | 9.83 | 9.39 |
| Net 0.7 | -- | 426 | 4.35 | 4.24 |
| | Net 0.3 | 41 | 11.28 | 9.70 |
| | Net 0.7 | 38 | 10.76 | 9.80 |

立模型的預測能力比 54 個因子建立的傳統 Net 0.5 模型要好，然而，當使用縮減模型重新估計模型係數時，縮減模型的預測表現除了較分析模型出色之外，也較能忠實呈現模型的預測能力。

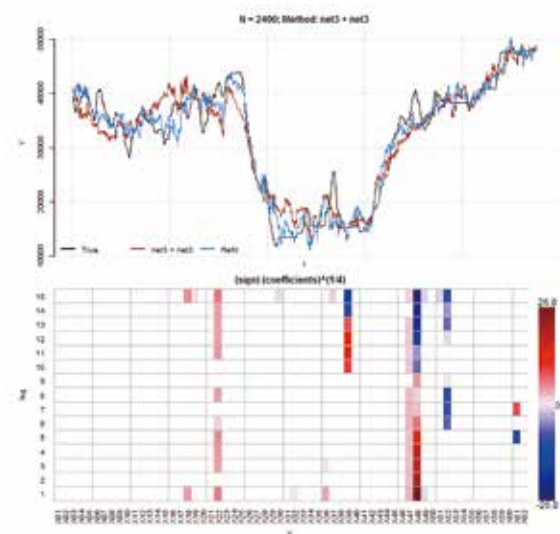
圖十六到十八分別呈現模型 Net 0.3+Net 0.3 的三組資料切割方式（資料前 600、1200 與 2400 筆數據）下的分析結果。以圖十六為例，圖片的上半部以黑色線隨著資料收集時間繪製實際液位高度，再以紅色與藍色線分別繪製分析模型與縮減模型的預測結果，可看出藍色線較貼近實際資料。圖片的下半部為模型係數估計值，其中橫軸為自變數 X1 到 X62，縱軸為從 1 期到 15 期的延遲效應，共 930 格。白色塊表示該變數之延遲期數的係數為 0，即非重要因子。分析結果中，模型係數在數值上的差異甚大，為明顯呈現模型係數值之間的差異，我們在圖形上呈現模型估計



圖十六 樣本數為 600 筆的分析結果圖



圖十七 樣本數為 1200 筆的分析結果圖



圖十八 樣本數為 2400 筆的分析結果圖

值的 4 次方根，當模型係數為正時，以紅色塊標示，係數為負值時，以藍色塊表示，且色塊顏色越深表示係數的絕對值越大。因此當現場的焦爐氣儲槽液位異常時，可藉由此可解釋的 AI 系統來解釋在每個變因對儲槽液位的影響，進而達到調配調度之目的。

表七 焦爐氣儲槽液位分析：最佳適應性模型與分析結果

| 資料筆數 | 最佳適應性法 | 重要因子 |
|------|-----------------|--|
| 600 | Net 0.3+Net 0.3 | X8, X9, X12, X20, X27, X36, X38, X40, X49, X58 |
| 1200 | Net 0.3+Net 0.3 | X18, X36, X37, X38, X39, X40, X47, X49, X50, X51, X52, X61 |
| 2400 | Net 0.3+Net 0.3 | X18, X19, X22, X30, X32, X36, X37, X39, X47, X48, X49, X51, X52, X61 |

四、結論

針對焦爐氣儲槽液位預測，考慮到能源消耗端的使用狀況與能源設備的穩定度之間關係具有時間上的落差，我們使用分布滯後模型將用戶使用狀況依時間延遲排開，試圖刻劃一段時間內的用戶使用狀況如何影響能源設備的穩定度，並且使用適應性 LASSO 與適應性 elastic net 方法，點出真正影響能源設備的用戶以及影響時機。在模擬資料測試中，當資料筆數遠低於分布滯後模型維度時，重要因子的偵測能力較差，因此我們建議在此模型之下使用適應性 LASSO 與適應性 elastic net 方法時，應盡可能地使資料筆數接近或大於模型維度。最後，透過實際資料的分析，可看出適應性 LASSO 模型輸出的重要因子

數量過低，可能有過度篩選的風險，應使用適應性 elastic net 方法以達到模型預測效能與重要因子篩選效能的平衡。為使模型挑出的重要因子為監控系統提供較佳的預測能力，根據模擬測試結果與實際資料分析中第三階段縮減模型的表現，我們建議使用適應性方法選取重要因子之後應重新建立縮減模型使用。■

【參考文獻】

- [1] He, S., Sammara, P., Burgers, J., & Schomaker, L. (2014). Towards Style-Based Dating of Historical Documents. In ICFHR (pp. 265-270).
- [2] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 58(1), 267-288.
- [3] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418-1429.
- [4] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67(2), 301-320.
- [5] Zou, H., and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733-1751.