

國立臺北大學
專題研究

應用維度縮減法於企業電力需求之預測

系所：經濟學系
學生：劉冠麟 410273091
李大維 410373059
康益豪 410473014
王姿文 410473074

報告日期：2018-06-01

國立臺北大學 專題研究

應用維度縮減法於企業電力需求之預測

劉冠麟、李大維、康益豪、王姿文¹

摘要

電力對於民生安全以及經濟產業十分重要，因此本研究決定藉由預測企業電力需求，探討主成分分析法、切片逆迴歸法，以及等軸距特徵映射法在支持向量迴歸模型的表現。我們認為企業電力需求與產業發展、景氣循環有關，因此決定以多個經濟指標來預測企業電力需求。首先透過探索性資料分析發現可以利用相關係數及階層式分群來初步篩選變數的類別，再以不同維度縮減方法降低資料維度，研究結果發現可能因上述維度縮減方法不適用於時間序列資料，以致預測結果皆不及無維度縮減之模型。

¹ 臺北大學經濟學系

目錄

壹、 研究動機.....	3
貳、 文獻回顧.....	4
參、 分析方法.....	5
一、 階層式分群 (Hierarchical Clustering).....	5
二、 動態時間扭曲法 (DTW)、相關係數.....	4
(一) 動態時間扭曲法 (Dynamic Time Warping)	4
(二) 相關係數	6
三、 維度縮減法.....	5
(一) 主成分分析法 (PCA).....	6
(二) 切片逆迴歸法 (SIR).....	6
(三) 等軸距特徵映射 (ISOMAP)	7
四、 維度縮減衡量指標.....	6
五、 支持向量迴歸 (SVR)	7
肆、 實證分析	8
一、 資料蒐集與整理.....	8
(一) 資料蒐集	8
(二) 資料整理	9
二、 探索性資料分析.....	10
三、 模型設定.....	13
四、 維度縮減法.....	14
五、 支持向量迴歸模型.....	14
六、 預測結果.....	14
七、 結論.....	14
八、 致謝.....	17
參考文獻.....	18

壹、研究動機

臺灣於2017年8月15日發生了大規模停電事件，起因於天然氣管線意外停止運作，造成備轉容量不足，供電系統啟動保護措施而分區停電。在這約六小時的停電時間內，不僅交通停擺、受困於電梯的危險事件發生，經濟損失也不容小覷，經濟部評估當晚全臺產業損失高達8800萬臺幣。因此，若能事先預測電力需求，則可以避免電力供應不足而影響到臺灣人民安全與生計。

蔡英文總統上任後，於2016年提出新能源政策措施，欲透過停運核電、降低火力發電，並提升天然氣及綠能發電比例，盼臺灣於2025年成為非核家園，達到燃氣占總發電量的比例50%、再生能源占總發電量比例20%的綠能低碳環境。然而臺灣2017年的總發電量中，燃氣比例占34.6%且再生能源僅占4.6%，未來不再仰賴核能發電的狀況下，或許可能面臨電力供不應求的狀況，因此本研究旨為因應未來非核家園的環境，先進行電力需求之探討。

在文獻探討的過程中，發現國內外文獻較常使用主成分分析法。本研究在維度縮減方面，除了使用傳統的主成分分析法，另外運用切片逆迴歸法與等軸距特徵映射法於電力需求預測，欲比較此三種維度縮減方法代入支持向量迴歸模型之後的預測優劣。傳統的支持向量迴歸模型的維度縮減過程，並無考慮反應變數的資訊。然而，本研究額外採用考慮反應變數的監督式方法—切片逆迴歸法。其中，主成分分析法與切片逆迴歸法均為線性維度縮減方法，因此我們決定再採用非線性的維度縮減方法—等軸距特徵映射法，來與線性維度縮減方法做比較。

本研究採用支持向量迴歸模型來探討臺灣企業用電之電力需求。在閱讀文獻的過程中，發現企業對於電力需求之價格彈性相比住宅對於電力需求之彈性來的較小，表示倘若電力價格上升幅度固定，那麼企業之電力需求量減少程度會相比住宅之電力需求量減少程度低，代表假定電價上漲，企業用電之成本也會上升，電價會反應在企業用電成本上。基於企業成本會受企業用電需求影響，本研究鎖定以企業用電為主要研究及探討範疇。

貳、文獻回顧

表 1、文獻回顧

名稱	作者 (年份)	方法	目的	結論	資料來源
我國住宅與服務業部門之電力需求預測模型準確度比較	吳易樺、黃朝熙、陳庚轅、劉子衙 (2015)	1.VAR 模型 2.因素模型 3.PCA	使用因素模型(factor model)預測我國住宅部門與服務業部門之電力需求，並與傳統向量自我迴歸模型比較何者具有預測優勢。	發現除了 FAVAR 因素模型，IMS, UN, 以及 DMS 等因素模型的樣本外預測誤差皆小於傳統的 VAR 模型。	http://km.twenergy.org.tw/Publication/thesis_down?id=60
Modeling and prediction of Turkey's electricity consumption using Support Vector Regression	Kadir Kavaklioglu (2010)	SVR	用 SVR 來預測土耳其的電力消費	此模型預測土耳其的電力消耗未來 20 年將達到 284.9 TWh，也就是將近 2006 年的兩倍。這意味著土耳其政府應該計劃增加電力的能力未來的發電廠。	https://www.sciencedirect.com/science/article/pii/S0306261910002904
Seasonal variations in residential and commercial sector electricity consumption in Hong Kong	Joseph C. Lam, H.L. Tang, Danny H.W. Li (2007)	1. PCA 2. 複迴歸	季節性變動的電力消費和氣候之間的關聯性，以及迴歸模型在商業和住宅電力消費間的差異。	迴歸模型配適商業比住宅來得好，且配適程度隨著年增加而有輕微上升，顯現氣候逐年影響電力消費程度也跟著增加。	https://www.sciencedirect.com/science/article/pii/S030605442007001831
台灣電力需求預測模型之探討－ARIMA 模型及迴歸模型	賈繼德 (2009)	ARIMA 及迴歸模型	對電力需求探究其計量的預測模型方法，期望能對用電建立具簡單且配適度良好的 ARIMA 模型及迴歸模型。	實證結果顯示，建立迴歸模型可使變數間產生共整合關係；經線性組合成為平穩的序列。迴歸模型預測能力優於 ARIMA，只是兩者差異不大。	https://ndltd.ncl.edu.tw/cgi-bin/gsweb.cgi?o=dnclcdr&s=id=%22097SCU05389036%22.&searchmode=basic#XXX

參、分析方法

一、階層式分群法 (Hierarchical Clustering)

本研究採用聚合式階層分群法，是由樹狀結構的底層開始，將相似的資料點層層向上群聚而成。群聚的過程，以平均連結法 (Average) 來定義群間的距離，並使用上述動態時間扭曲法 (DTW) 作為序列間的距離衡量。對於分群法來說，如何決定群的數目是相當重要的問題，R語言的套件「ComplexHeatmap」能夠幫助我們客製化熱圖，透過視覺化，可以更清楚地判斷群間的關係。平均連結法的公式如下：

$$d(C_i, C_j) = \sum_{a \in C_i, b \in C_j} \frac{d(a, b)}{|C_i| |C_j|} \quad \text{式 1}$$

式 1 中， C_i 、 C_j 個別代表兩群， $|C_i|$ 、 $|C_j|$ 為群內資料點的數量。

二、相似度測量方法

(一) 動態時間扭曲法 (Dynamic Time Warping, DTW)

相較於直接在兩序列的點和點之間計算歐氏距離，DTW 考量序列的走勢，更適合用來判斷兩序列間的相似程度 (Hui Ding等人，2008)。如圖 1，基於單調遞增的條件，能到達點 (i, j) 的前一點只有三種可能： $(i-1, j)$ 、 $(i, j-1)$ 、 $(i-1, j-1)$ ，藉由動態規劃的方法，計算序列中元素的累積距離，DTW能夠尋求一最短路徑。公式如下：

$$A(i, j) = D(i, j) + \min \begin{cases} A(i-1, j) \\ A(i-1, j-1) \\ A(i, j-1) \end{cases} \quad \text{式 2}$$

式 2 中， $A(i, j)$ 為原點 $(1, 1)$ 到 (i, j) 的最小累積距離。

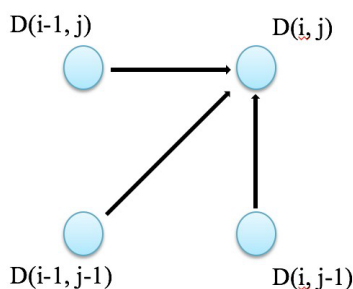


圖 1

(二) 相關係數 (Correlation Coefficient)

相關係數是衡量兩變數間線性關係強弱的指標，若大於零，代表變數間呈正相關，即一變數的值會隨另一變數的值增加而增加，反之亦然；若小於零，代表變數間呈負相關，即一變數的值會隨另一變數的值增加而降低，反之亦然。公式如下：

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}} \quad \text{式 3}$$

式 3 中， \bar{x} 、 \bar{y} 為樣本平均。

三、 維度縮減法

(一) 主成分分析法 (Principal Component Analysis, PCA)

PCA 是一個非監督式的維度縮減方法，藉由預測變數的線性組合，也就是主成分，解釋資料矩陣的變異結構，最後選擇可以解釋最大變異的主成分，達到維度縮減的目的。主成分分析法的公式如下：

$$PC_i = W_i^T X = w_{i1}X_1 + w_{i2}X_2 + \cdots + w_{ip}X_p, i = 1, 2, \cdots k \quad \text{式 4}$$

$$Var(PC_i) = W_i^T \Sigma_X W_i = \lambda_i, i = 1, 2, \cdots k \quad \text{式 5}$$

式 5 中 PC_i 為第 i 個主成分， X_1 、 X_2 、 \cdots 、 X_p 為預測變數， Σ_X 為預測變數的共變異矩陣，而 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$ 為其特徵值， W_1 、 W_2 、 \cdots 、 W_k 為特徵向量，而預測變數 X_i 一般都會予以標準化，避免單位不同影響主成分分析的結果。

(二) 切片逆迴歸法 (Sliced Inversed Regression, SIR)

SIR 在 1991 年由 Ker-Chau Li 提出，是一個監督式的維度縮減方法，藉由排序反應變數後，將預測變數投影至有效維度縮減 (Effective Dimension Reduction, EDR) 的方向上，最後以卡方檢定決定投影的維度，達到維度縮減和保有預測變數對反應變數的迴歸訊息的目的。以下為 SIR 假設的迴歸模型：

$$Y = f(\beta_1^T X, \beta_2^T X, \cdots, \beta_k^T X, \varepsilon) \quad \text{式 6}$$

式 6 中 Y 是單變量的隨機變數， X 是維度為 $p \times 1$ 的隨機向量，且 $k \leq p$ ， f 為一未知函數， ε 是一與 X 獨立的隨機誤差， β_i 是維度為 $p \times 1$ 的有效維度縮減方向。以下為有效維度縮減方向的推導公式：

$$\hat{\Sigma}_X^{-1} \hat{\Sigma}_W \hat{\beta}_i = \hat{\lambda}_i \hat{\beta}_i, i = 1, 2, \cdots, k, k \leq p \quad \text{式 7}$$

$$\left\{ \begin{array}{l} \bar{x}^{(h)} = n_h^{-1} \sum_{i \in slice_h}^{n_h} x_i \\ \hat{\Sigma}_w = \sum_{h=1}^H \frac{n_h}{n} (\bar{x}^{(h)} - \bar{x})(\bar{x}^{(h)} - \bar{x})^T \end{array} \right. \quad \text{式 8}$$

式 8 中 $\bar{x}^{(h)}$ 為切片的樣本平均數， n_h 為第 h 切片的樣本數， $\hat{\Sigma}_w$ 為加權共變異矩陣，特徵值符合 $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_k$ 。

(三) 等軸距特徵映射法 (Isometric Feature Mapping)

ISOMAP是J. B. Tenenbaum等人於2000年提出的一種非線性流形學習方法，試圖在高維度空間發現非線性結構。採用多維標度法(MDS)的概念，ISOMAP 將歐氏距離矩陣換成測地線距離，例如：地球上兩點間的距離衡量不適合用歐氏空間的距離將兩點穿越地球相連，藉由測地線尋找最短路徑可能是更好的測量方法。其方法為：透過最近鄰居法(KNN)尋找 K 個最近鄰居，與鄰居之間的距離設定為歐氏距離，鄰居以外的資料點則設成無限大，並使用Dijkstra演算法找出最短路徑，便形成一測地線距離矩陣。最終將此距離矩陣輸入MDS，達到非線性的維度縮減。

四、 維度縮減衡量指標

Local Continuity Meta Criterion(LCMC)是一個維度縮減的衡量指標，藉由比較維度縮減前後與鄰近資料點間相對距離的差異，進而判斷維度縮減的好壞。例如，若在維度較高的空間中，A點的鄰近點由近至遠分別是B、C、D點，而E點不是A的鄰近點，經過維度縮減後，A點的鄰近點由近至遠變為B、E、C點，而D點不是A的鄰近點，則在LCMC的衡量標準下，此維度縮減的效果不佳。公式如下：

$$LCMC(K) = Q_{NX}(K) - \frac{K}{N-1}, \quad \text{式 9}$$

$$\text{其中，} Q_{NX}(K) = \frac{1}{KN} \sum_{i=1}^K \sum_{j=1}^K Q_{ij}$$

式 9 中， Q 為 co-ranking matrix， K 為鄰居數， N 為樣本數。

五、 支持向量迴歸 (Support Vector Regression, SVR)

支持向量機(SVM)是一種監督式學習的方法，用來處理二元分類問題，其概念是將資料點從低維空間映射到高維空間，並找到一超平面(Hyperplane)，利用此超平面在高維空間，處理原空間線性不可分的問題。相較於SVM，SVR可看成其延伸型態，透過建立一超平面，處理連續變數的迴歸及預測問題。如下：

$$y = f(x) = w\phi(x) + b \quad \text{式 10}$$

式 10 中， $\phi(x)$ 為向量 x 的映射函數； w 為超平面的法向量； b 為誤差項。類似 SVM 的邊界 (margin)，SVR 在迴歸線旁建立一管狀空間 (tube)，忽略管狀空間內的誤差，並對空間外的資料點施加懲罰。為求 w 和 b 最小化，得到一目標函數：

$$\min_w \quad C \frac{1}{n} \sum_{i=1}^n L_\epsilon(d_i, y_i) + \frac{1}{2} \|w\| \quad \text{式 11}$$

$$\text{其中，} \quad L_\epsilon(d_i, y_i) = \begin{cases} |d_i - y_i| - \epsilon, & |d_i - y_i| \geq \epsilon \\ 0 & \text{otherwise} \end{cases} \quad \text{式 12}$$

式 11 中， $C \frac{1}{n} \sum_{i=1}^n L_\epsilon(d_i, y_i)$ 為經驗誤差，由(3)式的 ϵ 不敏感損失函數計算而得，2 倍的 ϵ 即是管狀的寬度。在訓練模型時， ϵ 和 C 為自行選擇的參數，增加 C 值可能造成過度配適 (overfitting) 的問題，因對噪音 (noise) 的容忍度越小；反之，則可能造成配適不足 (underfitting) 的問題。式(2)的最小化問題，使用拉格朗日函數可將其轉換為對偶問題，再透過解對偶問題可將式(1)表示為：

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) K(x, x_i) + b \quad \text{式 13}$$

式 13 中， α_i 、 α_i^* ($i=1,2,\dots,n$) 為拉格朗日乘數； $K(x, x_i)$ 為核函數。透過核函數，可以在低維空間做高維空間的複雜運算。本研究採用徑向基核函數，為一種常見的非線性核函數，如下：

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad \text{式 14}$$

式 14 中， γ 為寬度參數。

肆、實證分析

一、資料蒐集與整理

(一) 資料蒐集

本文採用國家發展委員會提供的電力(企業)總用電量(以下簡稱：企業用電量)作為反應變數，以及數個經濟相關指標作為解釋變數，如：景氣領先指標、物價指數(吳易樺等人，2015)、經常性薪資、就業人口(吳易樺等人，2015)、勞動力生產指數、工業生產指數(吳易樺等人，2015)、進出口總值(Kadir Kavaklioglu，2010；Jing 等人，2012)、原油價格(吳易樺等人，2015)。

表 2、資料來源

資料名稱	資料來源	資料型態	資料期間
企業用電總量	景氣指標查詢系統 - 國家發展委員會 https://index.ndc.gov.tw//n/zh_tw/data/eco/index	月資料	1989 至 2017 年
消費者物價指數	總體資料庫 http://statdb.dgbas.gov.tw/pxweb/dialog/statfile9L.asp	月資料	1989 至 2017 年
就業人數	薪資及生產力資料查詢系統 - 行政院主計總處 http://win.dgbas.gov.tw/dgbas04/bc5/EarningAndProductivity/QueryPages/More.aspx	月資料	1989 至 2017 年
經常性薪資	薪資及生產力資料查詢系統 - 行政院主計總處 http://win.dgbas.gov.tw/dgbas04/bc5/EarningAndProductivity/QueryPages/More.aspx	月資料	1989 至 2017 年
勞動生產力指數	薪資及生產力資料查詢系統 - 行政院主計總處 http://win.dgbas.gov.tw/dgbas04/bc5/EarningAndProductivity/QueryPages/More.aspx	月資料	1989 至 2017 年
原油價格	Market Index https://www.marketindex.com.au/crude-oil	月資料	1989 至 2017 年
工業生產指數	統計指標簡易查詢 - 經濟部統計處 http://dmz9.moea.gov.tw/GMWeb/common/CommonQuery.aspx	月資料	1989 至 2017 年
景氣領先指標	景氣指標查詢系統 - 國家發展委員會 https://index.ndc.gov.tw//n/zh_tw/data/eco/index	月資料	1989 至 2017 年
進口貿易統計值	進出口貿易統計 - 經濟部國際貿易局 https://cus93.trade.gov.tw/FSC3010F/FSC3010F	月資料	1989 至 2017 年
出口貿易統計值	進出口貿易統計 - 經濟部國際貿易局 https://cus93.trade.gov.tw/FSC3010F/FSC3010F	月資料	1989 至 2017 年

(二) 資料整理

本研究中我們蒐集到政府提供的變數資料，各變數資料集會因為附檔不同且格式不同而產生亂碼，或是出現混亂資料排序的狀態，但一個給定的數據資料集應該容易判斷何為觀察值以及變量。當我們即將開始處理數據資料進行分析時，發現本研究的變數資料集確實屬於混亂的資料集(Messy Data)，在本研究的資料集內，出現部分欄標題為變數標籤而非變數名稱，這使得我們無法精確的定義變量以及觀察值，因此我們決定來整理變數資料使其成為整齊的資料集(Tidy Data)。

Hadley Wickham(2011)提及整齊的資料集(tidy data)可以較為方便地被操作、建模和視覺化處理，而這符合我們想對本研究之變數資料集做的處理，故我們決定以tidy data的原則來整理變數資料，Tidy Data的基本原則有三：

1. Each variable forms a column.
2. Each observation forms a row.
3. Each type of observational unit forms a table.

故我們分別將每一個變數形成一欄，共整理出14欄，包含年、月、1個反應變數、8個預測變數以及3個類別變數。整理前的資料集原本共有9個預測變數，然而在整理過程中發現勞動生產力指數的遺失值比例高達81.2%，因此決定刪除該變數，最終為8個預測變數。接著將每一個觀察值形成一行，使得本研究的資料集成為一乾淨整齊的數據資料集。整理成Tidy Data後，我們的資料集擁有特定的結構，在處理數據資料時將更為順利且不易出錯。表格如下：

表 3、Tidy Data

年	月	企業用電量	物價	石油價格	物價類別	石油類別
1998	1	3.74	45.80	18.01	食物類	布蘭特
1998	1	3.74	45.80	16.96	食物類	西德州
1998	1	3.74	94.44	18.01	居住類	布蘭特
1998	1	3.74	94.44	16.96	居住類	西德州
1998	1	3.74	70.06	18.01	衣著類	布蘭特
1998	1	3.74	70.06	16.96	衣著類	西德州
⋮	⋮	⋮	⋮	⋮	⋮	⋮

註：因版面配置，此表僅節錄部分變數內容。

二、探索性資料分析

由圖 2(a) 顯示，企業用電量呈現每年為一週期循環和上升趨勢，但是從平滑曲線可以看出，近幾年的趨勢漸漸平緩，且每年的變異程度隨著年份增長而上

升。此外，可以觀察到大約在2009年企業用電量異常的下迭許多，我們推測其原因為2008年的全球金融海嘯，造成台灣景氣巨大地波動，在一個不安定地社會環境下，消費者的購買力下跌，生產者可能隨之減少生產，因而造成用電量暴跌。

圖 2(b) 為企業用電量在1989年到2017年間，每月的分佈情形。從盒鬚圖可以更清楚地看出企業用電量的夏峰冬谷的現象。其中，用電量中位數的高峰主要位於七月至九月，而低點則位於一、二月，隨著夏天來臨，不僅是機械、設備需要避免過熱，公司內部的冷氣耗電量也將隨之上升，我們認為季節因素可能與溫度變化有關。然而，公開資料庫並無提供長期氣溫和其他氣候因素的月資料，因此我們選擇將研究內容著重於經濟指標對用電量的影響。

圖 3(a) 為景氣領先指標與企業用電量的時序圖，臺灣景氣領先指標（Leading Indicator）是由外銷訂單指數、實質貨幣總計數、股價指數、製造業存貨量指數、工業及服務業加班工時、核發建照面積（住宅、商辦、工業倉儲）及SEMI接單出貨比七項構成項目組成，此指標可以提前反應臺灣未來的景氣方向，指標上升象徵著建商準備開始興建大樓、企業的訂單數量增加、著手於興建廠房與增設生產線，造成企業總用電量上升，最後才帶動經濟的成長。

圖 3(b) 為工業生產指數與企業用電量的時序圖，工業生產指數是相對指標，衡量製造業、礦業與公共事業的實質產出。該指數反映的是某一時期工業經濟的景氣狀況和發展趨勢，文獻回顧中，劉美琦在臺灣地區電力尖峰負載預測之研究，發現工業生產指數對尖峰負載影響同時來自短期動態的調整及長期均衡，我們也由此時序圖，發現到工業生產指數與企業總用電量有著相同的成長趨勢。

圖 3(c) 為進出口產值與企業總用電量的時序圖，進出口是指與外國當事人通過締結契約進行買賣商品，包括勞務、技術等。國內生產總值(GDP)能反應臺灣的經濟狀況，而 $GDP = \text{總消費} + \text{總投資} + \text{政府購買} + \text{出口總值} - \text{進口總值}$ ，因此進出口總值會影響臺灣的生產經濟，故進出口總值之趨勢與企業總用電量相符。

圖 3(d) 為原油價格與企業用電的時序圖，黑線為企業用電量，紅線為布蘭特原油價格，淺藍線為西德州原油價格，兩種原油價格非常相近，其中西德州原油價格的變異程度較小，所以我們選擇將其作為我們的解釋變數。原油價格與天然氣具連帶關係故會影響燃氣發電，並且原油為火力發電的原料，因此原油價格與發電成本相關，當原油價格上升時發電成本會上升，導致企業總用電成本上升，影響企業總用電量，反之亦然。

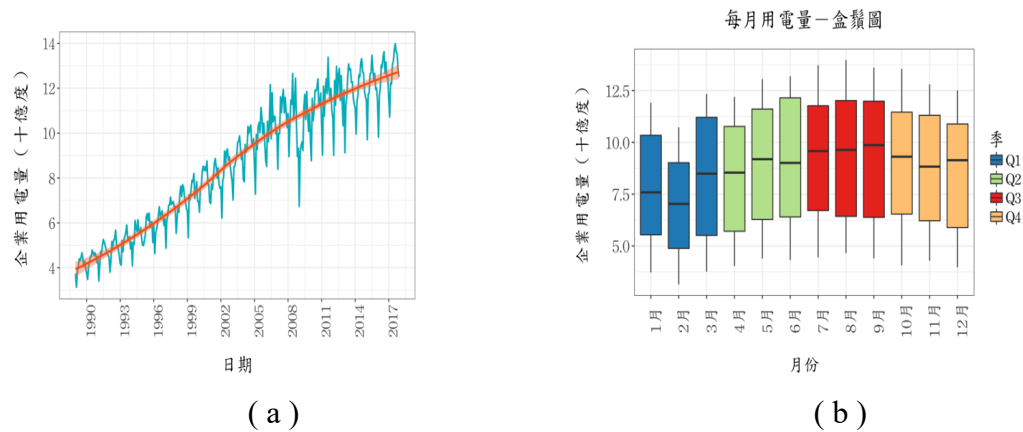


圖 2

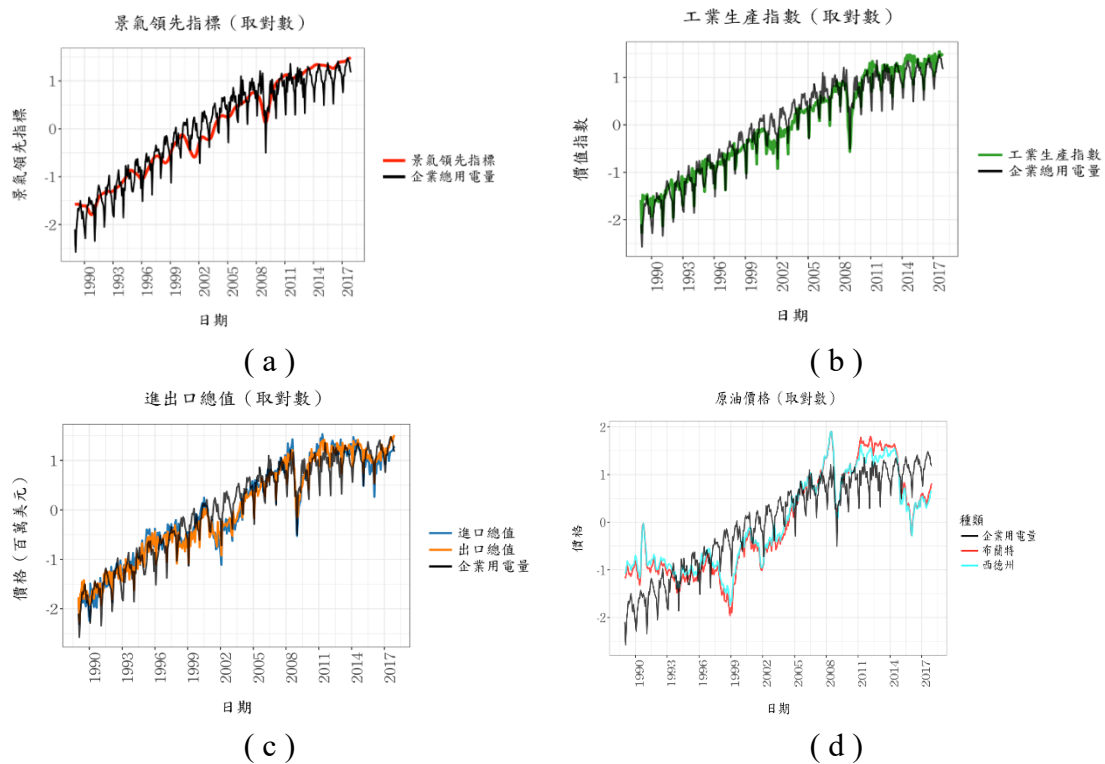


圖 3

考量ISOMAP及LCMC計算的複雜度，我們決定藉由計算相關係數和以DTW計算距離的階層式分群兩種方法，篩選變數的類別達到加快運算速度的效果；再從各變數類別的折線圖輔助判斷何種方法的效果較好。最後我們選擇使用相關係數作為相似度指標，因其選取的類別與企業用電量的趨勢較為相似。一連串的保留與淘汰過程後，我們發現距離矩陣的運算仍超過記憶體容量限制，於是便將年份縮減為1998年至2017年，資料集的維度變為：5760個觀測值；14個變數。

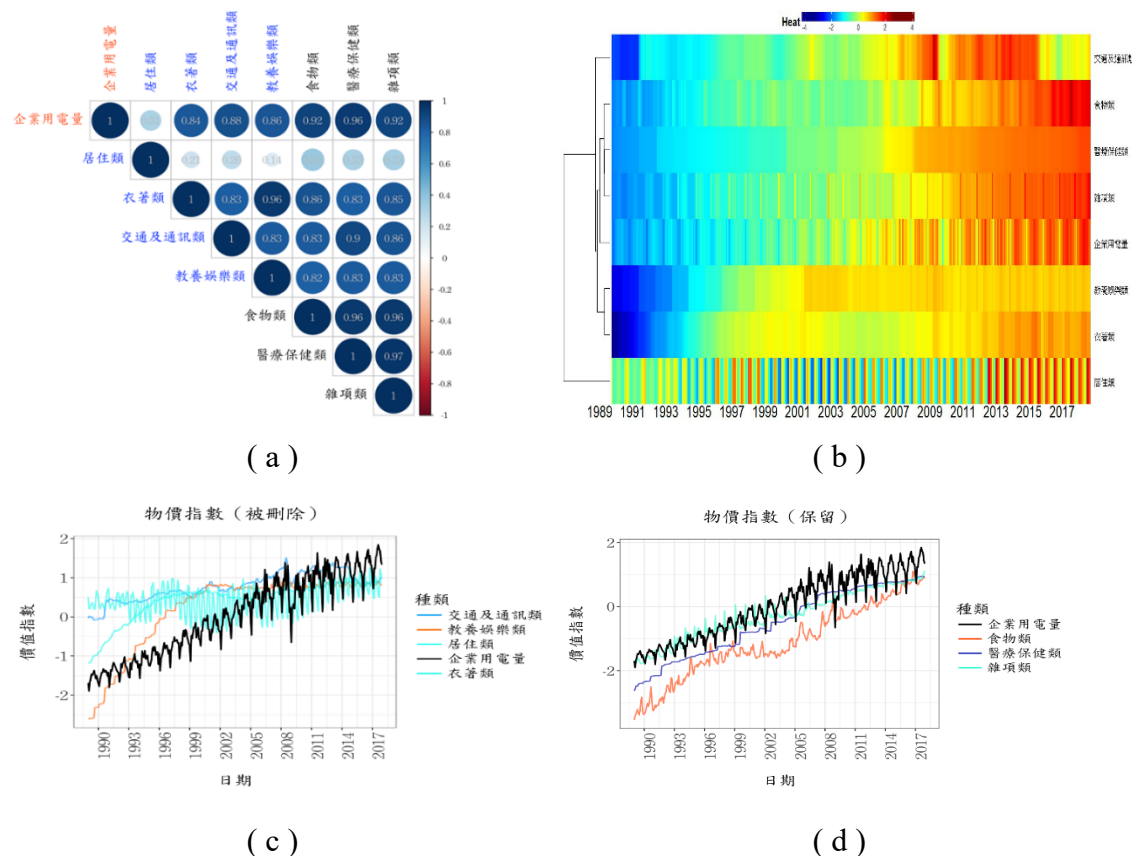


圖 4

圖 4，(a)為各類別物價指數的相關係數矩陣圖，可看出企業用電與食物類、醫療保健類和雜量類間的相關係數皆大於 0.8；(b)為各類別物價指數的階層式分群熱圖，若分四群則企業用電與食物類、醫療保健類和雜量類為同一群，群間距離的計算方法為DTW，其忽略負相關性大的類別，是本研究放棄此方法作為相似度判斷基準的原因；(c)、(d)為物價指數透過相關係數篩選類別後，保留、捨棄類別的折線圖。

三、 模型設定

考量企業用電量與經濟指數的靜態關係、未來用電量的動態關係，我們參考了兩篇文獻：Kadir Kavaklioglu (2010) 使用SVR模型預測土耳其電力消費；Xuchan Ju 等人(2014) 建立預測用水量需求的靜態模型，且透過ARMA模型事先預測影響因素(模型的解釋變數)，達到預測 (Forecasting) 的效果。然而，本研究的類別變數經過相關係數篩選後，再一一建立ARMA模型，可能仍不是個有效率的方法。綜合以上兩篇文獻的模型建立思維，我們決定嘗試在預測變數中，加入企業用電量過去的資訊。模型中，除經濟指數外，我們額外放入年、月作為預測變數，以避免虛假迴歸的問題，並希望能捕捉長期上升的趨勢和季節性變化。本研究共建立未來一期至三期預測模型，如下：

$$Y_{t+1} = Y_t + DR(X_t^1 + X_t^2 + \dots + X_t^{10}) \quad \text{式 15}$$

$$Y_{t+2} = Y_{t+1} + Y_t + DR(X_t^1 + X_t^2 + \dots + X_t^{10}) \quad \text{式 16}$$

$$Y_{t+3} = Y_{t+2} + Y_{t+1} + Y_t + DR(X_t^1 + X_t^2 + \dots + X_t^{10}) \quad \text{式 17}$$

其中， Y_{t+3} 、 Y_{t+2} 、 Y_{t+1} 分別為未來一期至三期的企業用電量； DR 為維度縮減法的函數； $X_t^1 + X_t^2 + \dots + X_t^{10}$ 為整理成 Tidy Data 形式後的預測變數。

本研究將訓練集設定為1998年至2015年的所有觀測值，維度縮減後，再將投影至低維度的新向量輸入模型，進行訓練；測試集為2016年至2017的所有資料，用以衡量維度縮減法的預測能力。此外，本研究並將類別變數納入模型加以探討。

四、維度縮減法

(一) 主成分分析法 (Principal Component Analysis, PCA)

圖 5(a) 的長條圖代表每個主成份對於原資料的解釋力大小，可以看到第一個主成份已達約百分之六十四的解釋力，雖然三個主成分的累積解釋力已達百分之八十四，但我們仍決定保留第四個主成份，因其解釋能力與前兩者相差不遠。

圖 5(b) 能夠呈現出每個預測變數對於前四個主成份的貢獻度，其中，原油價格 (OilPrice) 和消費者物價指數 (CPI) 的貢獻度相較其他變數而言較低，這可能導致在建模的過程中，他們對於模型的影響較小。

然而，PCA在做維度縮減的過程，僅考慮預測變數自身的變異，而忽略反應變數的資訊。因此，我們更進一步嘗試切片逆迴歸法 (SIR)，希望該方法能夠達到更好的維度縮減效果。

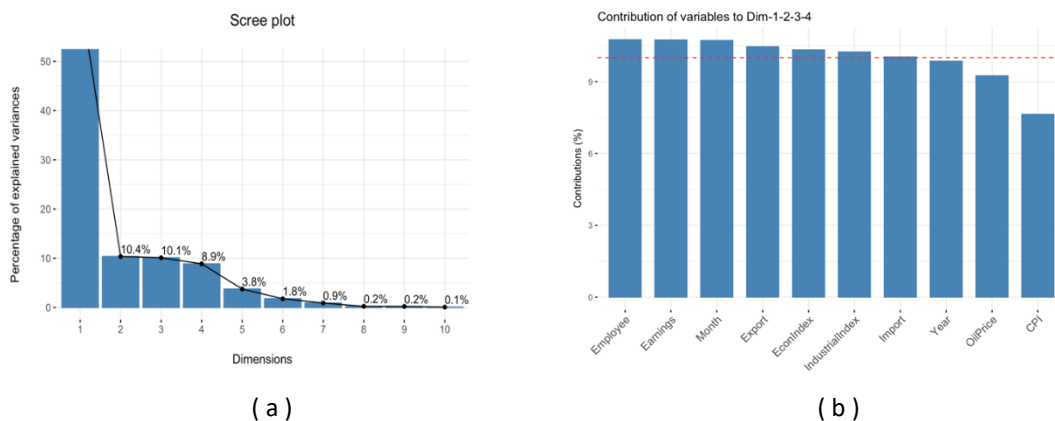


圖 5

(二) 切片逆迴歸法 (Sliced Inverse Regression, SIR)

Li (1991年)假設預測變數X的邊際分配為常態，所以 SIR 中邊際維度的檢定漸進服從卡方分配，則可以進行卡方檢定來決定維度縮減後的變數個數，從表1可看出本次實證透過 SIR，切片數設定為18，可將10個維度縮減至7個維度，其中切片數的選擇為觀察企業用電量將數值排序後的分布的結果。

表 4、Large-sample Marginal Dimension Tests

	Stat	df	p-value
0D vs \geq 1D	8361	170	0.000e+00
1D vs \geq 2D	3689	144	0.000e+00
2D vs \geq 3D	2472	120	0.000e+00
3D vs \geq 4D	1499	98	0.000e+00
4D vs \geq 5D	824.5	78	0.000e+00
5D vs \geq 6D	292.6	60	0.000e+00
6D vs \geq 7D	113.4	44	4.924e-08
7D vs \geq 8D	31.67	30	0.3829

(三) 等軸距特徵映射法 (Isometric Mapping, ISOMAP)

ISOMAP 縮減維度效果的優劣，在分類問題中，能夠以散佈圖視覺化辨別類別之間的分散程度，但是時序型的連續變量在圖形視覺化上較不易解釋、呈現，若使用交叉驗證的方法，檢驗 SVR 的 RMSE，可能會受限於建模效率，所以我們決定以 LCMC 法作為鄰居數及維度選取的衡量標準。從圖 6可看出縮減至5維的 LCMC 最大，但與4維的 LCMC 差異不大。為了嘗試使用更精簡的模型，達到同樣的預測能力，我們決定將 ISOMAP 的鄰居參數設定為8個並縮減至4維。

五、支持向量迴歸模型 (Support Vector Regression, SVR)

我們使用臺灣大學林智仁教授開發的 R 語言套件「e1071」建立了未來一期至三期的 SVR 模型。該模型內有三個參數必須設定：Cost (C), Gamma (γ), Epsilon (ϵ)。參照 Kadir Kavaklioglu (2010)的參數選取範圍，Cost 選擇 10 到 100，每次增加 30 單位；Gamma 選取 1 至 3，每次增加 0.5 單位；Epsilon 為 0.001 到 0.01 每次增加 0.003 單位，總共 80 個組合。本研究採用 R 語言「caret」套件的 Time Slice 交叉驗證方法，初始期間設定為五期，每次向後加入一期作為驗證集，總共施行 13 次交叉驗證，以平均結果代表每組參數的 RMSE，最終選取平均 RMSE 最低的參數組合作為最佳的模型設定。因此，包含 80 個參數組合、13 次交叉驗證，在不同維度縮減法下，未來每期的預測模型必須先建立了 1040 個模型，得以獲得最佳參數組合。

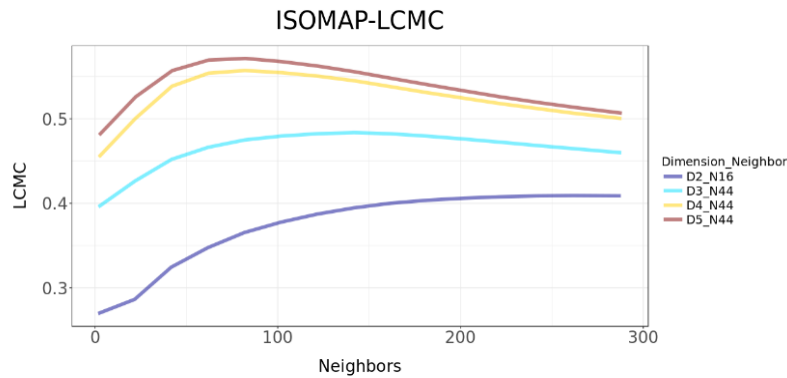


圖 6

六、 預測結果

從RMSE的結果來看(表4、5、6)，三種維度縮減方法在未來二期的模型都有較嚴重的過度配適情形；維度縮減法不僅沒有達到預期的減緩維度災難問題，預測能力也沒有比無維度縮減的模型還好，我們認為可能是原變數數量就不多。圖7為運用三種不同維度縮減方法及不做維度縮減代入模型進行未來一期、二期及三期的預測結果，並與實際值比較。總的來說，未來一期預測的結果皆比二、三期來的好，其中又以無維度縮減的模型預測最佳，其次是運用SIR的模型。不論是未來一期、二期或三期預測，預測值皆與實際值相差一個垂直截距項，我們認為可能是因為我們運用的維度縮減方法皆沒有考慮序列關係，而運用SIR得到較好的預測結果可能是因為此方法有考慮企業用電量的變異程度。

表 5、未來一期預測模型結果

一期	PCA	SIR	ISOMAP	RAW
Train	0.1086	0.1712	0.1202	0.1773
Test	0.2338	0.1802	0.2766	0.1506

註：RAW 代表無維度縮減

表 6、未來二期預測模型結果

二期	PCA	SIR	ISOMAP	RAW
Train	0.1263	0.1636	0.1270	0.1767
Test	0.3287	0.3445	0.3485	0.1592

註：RAW 代表無維度縮減

表 7、未來三期預測模型結果

三期	PCA	SIR	ISOMAP	RAW
Train	0.1066	0.1493	0.1222	0.1667
Test	0.2452	0.2723	0.2709	0.1889

註：RAW 代表無維度縮減

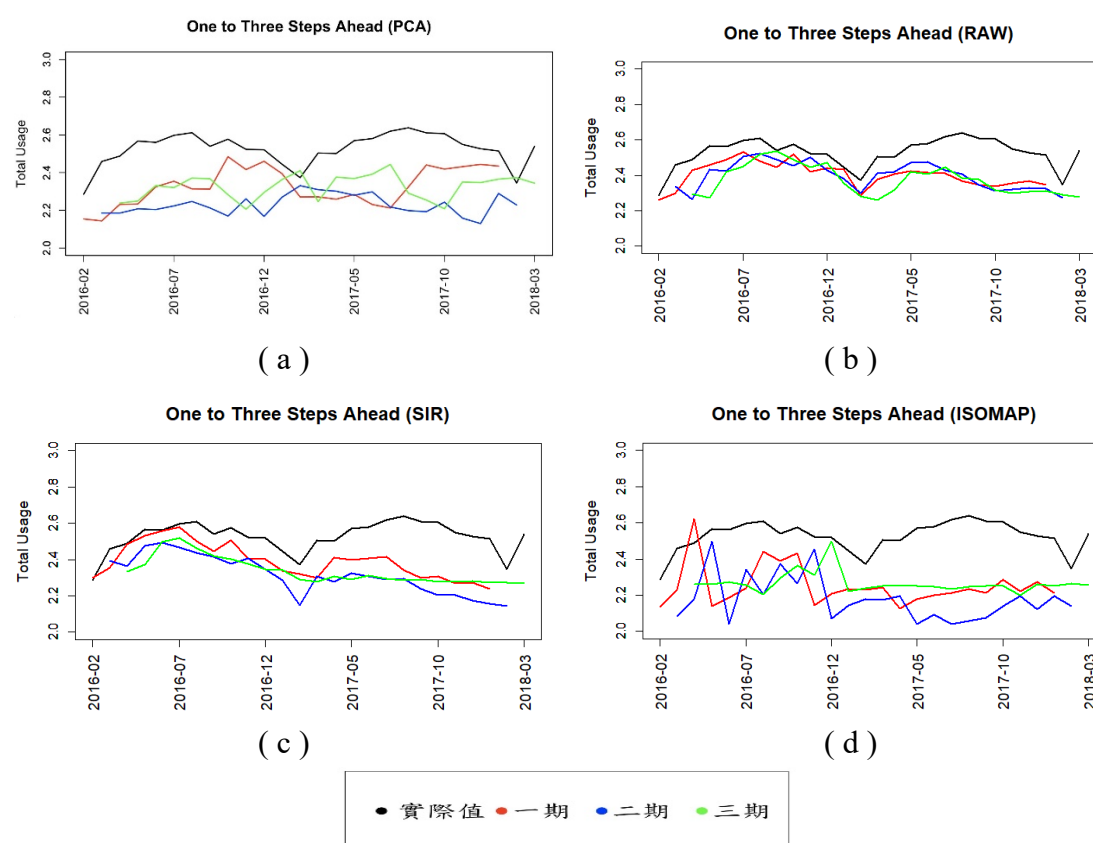


圖 7

七、 結論

本研究發現三種維度縮減方法：PCA、SIR、ISOMAP，並無達到降低RMSE的效果，且從2016年至2017年的企業用電量折線圖來看，以上方法在未來一到三期的預測值與實際值皆向下平移，並無捕捉到企業用電量上升的趨勢。然而，SIR的未來一期模型及無維度縮減的一至三期模型，其預測值與實際值的變異較為相似。我們認為日後可以嘗試加入電價、氣溫等變數，及採用考慮序列關係的維度縮減方法，例如：Time Dependent Principal Component Analysis。或許能夠獲得更好的預測結果。再者，若不使用傳統的多變量時間序列模型，要如何運用SVR建立更長期的用電量預測模型，也是值得我們更進一步探討的問題。

八、 致謝

感謝吳漢銘教授半年來的傾囊相授，即使工作屢屢忙碌到深夜仍不忘回答我們的問題，當我們一再忘記過去的提醒，教授始終耐心地重述應注意的事項直至我們學會為止，每一次的建議有如醍醐灌頂，點出我們學習的盲點並幫助我們快速了解，教授是我們得以完成本次的專題研究最大的助力。

參考文獻

壹、中文部分

- 尹邦嚴、趙景暉，2016，〈以 Cyber Swarm Algorithm 演算法進行台灣電力需求預測 (p1-p7) 〉。「第十五屆離島資訊技術與應用研討會論文」
- 吳易樺、黃朝熙、陳庚轅、劉子衙，2015，〈我國住宅與服務業部門之電力需求預測模型準確度比較〉。《台灣能源期刊》2(1)：1-14。
- 涂宗裕、戴吟芳、吳昀珊、吳貞霖、陳碧利、劉亞函，2008，〈台灣電力需求的研究以及預測〉。
- 賈繼德，2009，《台灣電力需求預測模型之探討－ARIMA 模型及迴歸模型》。東吳大學經濟學系碩士論文。
- 劉美琦，2001，〈台灣地區電力尖峰負載預測之研究〉。淡江大學機構典藏。

貳、英文部分

- Kavaklioglu, Kadir. 2010. "prediction of Turkey's electricity consumption using Support Vector Regression".
- Lam, Joseph C. , H.L. Tang, and Danny H.W. Li. 2007. "Seasonal variations in residential and commercial sector electricity consumption in Hong Kong".
- Mohamed, Zaid and Pat Bodger. 2003. "Forecasting electricity consumption in New Zealand using economic and demographic variables".
- Zhang Jing , Xin-yao Yang, Fei Shen, Yuan-wei Li, Hong Xiao, Hui Qi, Hong Peng, and Shi-huai Deng. 2012. "Principal Component Analysis of Electricity Consumption Factors in China".
- Kumar Vasimalla, Narasimham Challa and Manohar Naik S. 2016. "Efficient Dynamic Time Warping for Time Series Classification"
- Lisha CHEN, Andreas BUJA. 2006. "Local Multidimensional Scaling for Nonlinear Dimension Reduction, Graph Drawing, and Proximity Analysis"
- Ker-Chau Li. 1991. "Sliced Inverse Regression for Dimension Reduction"
- XuchanJu, Manjin, Cheng, Yuhong Xia, Fuqiang Quo, Yingjie Tian. 2014. "Support Vector Regression and Time Series Analysis for the Forecasting of Bayannur's Total Water Requirement"