**研究計畫書**

**研究題目：**

**暫時想不到**

**申請人：劉宗妮**

目錄

[**壹、** **摘要** 3](#_Toc145164472)

[**貳、** **研究動機與目的** 3](#_Toc145164473)

[**參、** **文獻回顧** 4](#_Toc145164474)

[一、空氣品質指標與發展 4](#_Toc145164475)

[二、非監督式學習與集群分析 4](#_Toc145164476)

[三、自組織映射神經網路 5](#_Toc145164477)

[四、類別型資料的空間相依性 6](#_Toc145164478)

[五、灰階共生矩陣 6](#_Toc145164479)

[**肆、** **研究架構與方法** 6](#_Toc145164480)

[一、研究資料與區域 6](#_Toc145164481)

[二、資料前處理 7](#_Toc145164482)

[三、自組織映射神經網路演算法 7](#_Toc145164483)

[四、類別空間自相關 8](#_Toc145164484)

[五、灰階共生矩陣 8](#_Toc145164485)

[**伍、** **預期成果** 8](#_Toc145164486)

[**陸、** **參考文獻** 8](#_Toc145164487)

**暫時想不到標題**

1. **摘要**

關鍵字：

1. **研究動機與目的**

隨著科技日新月異，人類消耗資源、發明新產物的速度越來越快，以追求更便捷的生活。但與之同時面對的是地球環境破壞造成負面影響，大自然的反撲帶來極端氣侯、環境污染等問題，是對人們生活的品質和其餘生物的生存空間而言都是一大考驗，有鑑於資源消耗所帶來的環境變遷愈發顯著，用於環境品質變遷監測的量化方法已成當代研究矚目的議題之一。

空氣品質是其中一項環境品質的重要指標，許多醫學研究證實空氣汙染對於人體的危害極大。聯合國世界衛生組織（WHO）在2005年提出適用於全球的空氣品質指南（Air Quality Guideline），將幾項已被認知的有害物質立定健康標準以供參考，由此可見，空氣品質的監測與改善是全球性的重要議題。

隨著通訊技術的提升，物聯網於各領域的應用蔚為盛行，環保署結合該科技建構「臺灣空氣品質感測物聯網」，截至2023年中，微型感測器的布建範圍已擴及全台17縣市，累計數量達10,506台，為當今全球空氣品質感測網絡密度最高的國家。相較國家標準精密儀器，微型感測器成本低廉，因此在大量布建上極具優勢，且資料時間解析度可達分鐘等級，在分析汙染物的時間序列樣態與空間上熱區等具備優良的應用價值。

近年來使用機器學習與深度學習等技術進行空品預測儼然已成為趨勢，相較於傳統統計模型，這些方法能更好的建立空品的非線性映射，但較少有研究關注於量化的環境品質指標開發，以及時空現象變遷模型的建立上，且研究多以單一測項為主，特別是以PM2.5佔多數。然而環境品質的優劣並不是僅由PM2.5就能判斷，需要一個能結合各項數值，在空間上將各指標做綜合分析的模型，而這正是過去研究較少擁有的。

因此本研究欲組建一個混合模型，透過環境因子的變化，找出環境品質於空間與時間尺度上的差異，建立一個客觀明確的環境品質分區，並期望能達成以下目標：

1. 使用機器學習中的自組織映射神經網路進行資料降維與像素分群，並透過複迴歸分析集群內變數，判斷分群各自的特性。
2. 將群的像素重新映射回地理空間後，使用LICD與灰階共生矩陣對地理空間的類別資料進行分析，以此找出其空間及時間相依性。
3. **文獻回顧**

一、環境品質指標與發展

根據行政院主計總處的報告，我國環境品質指標的建立是基於經濟合作暨發展組織(OECD)所訂定的美好生活指數之環境品質，選取空氣品質及水質滿意度兩項指標，加上自訂的接近綠地所組成。空氣品質透過感測器監測而得，水質滿意度是透過問卷調查獲得，接近綠地則為計算都市計畫區內人口平均享有已闢綠地面積。

在空品監測上，最傳統的方式是使用固定式空氣品質監測站，其具有準確且可靠的優勢。但龐大的體積與重量加上造價高昂，導致佈建上較為稀疏，而粗略的時空解析度無法真實反映出空氣汙染於個人健康所造成的風險(Yiet al. 2015)。為了提升觀測的時空解析度，體積小且成本低廉的無線感測器網路(Wireless Sensor Network, WSN)，被廣泛應用在空氣品質監測中，例如都市空氣品質監測(Liu et al, 2011; Liu et al., 2012)、交通工具污染源感測（Hu et al., 2011）、室內空氣品質監測（Abraham and Li, 2014; Bhattacharya et al., 2012）以及空品地圖繪製（Boubrima et al., 2019）等。

過去環境品質資料的建立是專家法，若以監督式學習進行建模，由於專家的答案不等同於毫無誤差的正確答案，透過專家的標準作為答案進行學習會造成誤差傳遞，影響機器學習的最終成果，因此本研究欲使用非監督式學習讓資料透過個別的特性自行群聚，以避免人為誤差的產生。

二、非監督式學習與集群分析

非監督式學習（Unsupervised Learning）為機器學習的分支，常被運用於資料的分群、降維，與監督式學習相比，非監督式學習最大的特色在於不需要資料標籤（Label），適用於事先沒有正確答案的資料進行機器學習，該方法透過輸入資料的相似性與距離進行運算，透過演算法找出數據間的關聯性。

透過分群找出資料的同質性（homogeneity）與異質性（heterogeneity），讓同分群內的資料相似、相異分群的資料盡可能相異，又被稱為集群分析（Cluster Analysis），K-means是傳統上經常被使用的方法。隨著資訊科技的蓬勃發展，出現了以機器學習分群的技術，例如自組織映射神經網路（Self-Organizing Map，SOM）是基於類神經網路為基礎的分群方法，過去分群的應用範圍十分廣泛，從股價與公司類型的分群(

三、自組織映射神經網路

Kohonen在1982年提出自組織映射神經網路（Self-Organizing Map，SOM），基於類神經網路，透過模擬大腦神經元的演算法，將高維度的資料映射至二維的拓樸空間。根據楊東昌（2004）的整理，自組織映射神經網路在資料分析的應用上可用於大型資料集合的擷取與組織化、視覺化，以達成分群。

與其他分群法相比，自組織映射神經網路最大的優勢在於不需要事先給予分群數目，舉例而言，K-means的分群結果受到起始分群數目影響，然而由研究者決定良好的分群數目是不易處理的問題之一，涉及到經驗與足夠的專業知識，而自組織映射神經網路能透過模型訓練自行組成最終集群。

自組織映射神經網路的模型（圖一）可分為輸入層與輸出層，變數透過權重連結至輸出層的神經元，在訓練模型的過程中不斷調整權重，映射回地理空間可獲得神經元間的拓樸關係。

|  |
| --- |
| *w* |
| 圖一 SOM模型示意圖 |

在進行權重調整前，需要根據競爭學習演算法，決定出優勝神經元(winner)後，只有優勝神經元及其鄰近神經元進行權重調整。透過歐基里德距離公式比較輸入向量X與各神經元的連結權重()的距離，最小值即為優勝神經元，以下為其計算公式：

鄰近神經元的範圍由鄰近函數**η**所決定，隨著模型訓練次數的增加、鄰近半徑*R*的縮小而遞減。

四、類別型資料的空間相依性

根據地理學第一定律：所有事物都與其他事物相關，但是近處的事物比遠處的事物更相關(Tobler, 1970)，可見空間相依性自古以來就是地理學研究中所關注的重點之一。常見的方法有全域型分析的空間自相關（Global Moran’s I）以及區域型的空間自相關（Local Indicators of Spatial Association, LISA）。但前述兩種方法僅適用於連續資料的計算，在分析類別資料的空間自相關時，類別空間自相關（Local Indicators for Categorical Data, LICD），舉例而言，在考古學上能用於

|  |
| --- |
|  |
| 表 |

五、灰階共生矩陣

過去的研究大多僅專注於單一時間的資料於空間上的分析與視覺化，然而地理資料往往是大量且長時間性的，近年來將時間維度納入研究、找尋時間序列上的趨勢是地理研究的重要議題之一。灰階共生矩陣（GrayLevel Co-occurrence Matrix, GLCM）多被用於判視鄰近紋理特徵，該方法能整理出相鄰像素的空間分布與相鄰次數，過去研究用於

透過統計八方位相鄰像素點的連接關係，獲得計算結果的矩陣(圖)，*i*代表起始點值，*j*代表觀測相鄰值。

|  |
| --- |
|  |
| 圖 灰階共生矩陣 |

改GLCM丟空間自相關

1. **研究架構與方法**

一、研究資料與區域

根據前人的研究，台中的地形不利於氣流的擴散、加上境外、工業與交通等汙染因素，造成了台中嚴重的空汙問題。本研究設定台中沿海地區的行政區：大甲區、大安區、清水區、梧棲區、龍井區（不包含大肚台地範圍）為研究區（圖），使用行政院環境部架設之感測器資料，以PM2.5、PM10、VOC、溫度、濕度、風向、風速作為自變數，研究區間為2018的冬季，以每周的資料平均。

|  |
| --- |
|  |
| 圖 研究區範圍 |

二、資料前處理

由於獲取之資料型態無法直接使用，因此需要進行資料前處理。以下為主要處理方法：

(1)資料聚合：若以單日資料進行分析可能會造成特例影響結果，因此本研究將資料進行整合，將….年冬季每周之自變數資料數值進行平均。

(2)空間內插：由於自組織映射神經網路的運算需要網格資料，而各測站獲得的是單點向量資料。本研究使用克利金法（Kriging）獲取無數值的研究區估計資料，透過已知點位之數值來估算未知區域的值，獲得網格資料。

三、自組織映射神經網路演算法

以

som跑出來之後用原始資料再跑一個複回歸解釋 or PCA

四、類別空間自相關

五、灰階共生矩陣

1. **預期成果**

本研究藉由實作出SOM模型，將其應用於環境品質分群，後續進行質性分析，LICD、GLCM分析相依性，預期成果如下：

1. 透過SOM建立明確客觀的環境品質集群，並能比較各區域時空上的差異
2. **參考文獻**

吳治達, & 曾于庭. （2023）. 氣候變遷下的空氣污染分布: 地理人工智慧技術之應用. 土木水利, 50（1）, 16-23.

楊東昌（2004）。自組織映射圖神經網路改善模式與分群應用之回顧研究。﹝碩士論文。華梵大學﹞臺灣博碩士論文知識加值系統。 https://hdl.handle.net/11296/c45qqq。

Carrer, F., Kossowski, T. M., Wilk, J., Pietrzak, M. B., & Bivand, R. S. （2021）. The application of Local Indicators for Categorical Data （LICD） to explore spatial dependence in archaeological spaces. *Journal of Archaeological Science*, *126*, 105306.

Cormack, R. M. （1971）. A review of classification. Journal of the Royal Statistical Society, Series A, 134, 321-367.

Kohonen, T. （1990）. The self-organizing map. Proceedings of the IEEE, 78（9）, 1464-1480

Liu, J. H., Chen, Y. F., Lin, T. S., Lai, D. W., Wen, T. H., Sun, C. H., ... & Jiang, J. A. (2011, November). Developed urban air quality monitoring system based on wireless sensor networks. In 2011 Fifth International Conference on Sensing Technology (pp. 549-554). IEEE

Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. International journal of scientific and research publications, 3(5), 1-5.