

摘要

近年來台灣的空氣污染議題逐漸成為社會所關注的焦點，這一議題引起了民眾、政府以及環保團體的重視，其中 PM2.5 近年來常被作為空汙指數的代表，因此如何監測與預測 PM2.5 濃度是一項重要的課題。

本研究將會以「行政院環保部-環境資料開放平台」所提供的資料做 PM2.5 濃度預測，利用線性插值法和平均插值法進行缺失值補值，並利用三倍標準差進行異常值處理，接著透過 K-means 將 72 個站點進行分群後，利用兩種特徵選擇方法(SelectKBest、MI)找出各群間影響 PM2.5 濃度的特徵，最後輸入時間序列模型(LSTM、GRU)進行 PM2.5 濃度預測，並利用三種評估指標(MSE、MAE、RMSE)進行模型績效評估。

本研究旨在透過 K-means 分群後，降低建立模型的數量及建立成本，並比較不同補值方法、特徵選擇、模型選擇共 8 種組合後，研究結果表明，透過此種不同方法之組合，可以針對各群找出最適合的建模組合，更好地滿足各群間不同的特性，在降低模型數量的前提下，維持模型的精確度。

關鍵詞:PM2.5、長短期記憶模型、門控循環單元、相互資訊、SelectKBest、K-means