

結合 VAE 與 XGBoost 之梅雨季臺灣雨量預報技術開發

李清騰^a，林昀靜^a，陳昀靖^a，羅資婷^a，潘琦^a

^a 交通部中央氣象署

摘要

中央氣象署依據具有物理概念模型(Yim et al., 2015)建立梅雨季臺灣雨量預報指引，近 10 年發現過去選取的預報因子(如 SST)與臺灣梅雨季雨量反而呈現出負相關，導致預報技術得分明顯偏低。本研究提出一種新穎的機器學習混合架構，結合變分自編碼器 (Variational Autoencoder; VAE) 與極限梯度提升樹 (XGBoost) 來提升預報技術得分，並分析梅雨季雨量的預測精準度與物理可解釋性。模型的訓練是將資料透過 VAE 方法，將氣象 4D 時空多氣象變數壓縮成低維度的「潛在特徵 (Latent Features)」，此方法不僅大幅去除了無用的雜訊，更能精準捕捉隱藏在數據背後的大尺度氣候遙相關結構（如聖嬰、季風指標）。利用此 VAE 提煉出的高濃度純淨特徵後，XGBoost 展現了處理低維度表格資料的極致效能，針對氣候科學常見的歷史樣本數較少的缺點，XGBoost 內建的雙重抽樣與嚴格的正則化 (L1/L2) 機制能有效防止模型過擬合 (Overfitting)。透過 VAE 與 XGBoost 的混合架構作成功解決了多維氣象資料維度與小樣本學習的雙重挑戰，不僅顯著提高了台灣梅雨季降雨預報的穩定度，更為探索降雨背後的氣候驅動機制提供了強而有力的新世代預測工具。

關鍵字: AI 模型、深度學習、梅雨季雨量預報