

摘要

在工業生產環境中，機器設備屬於高成本的固定資產，因此故障診斷在製造業的設備維護中非常的重要。隨著物聯網與大數據等新興技術的發展，預測性維護透過機器學習對資料做分析，以進行設備的狀態監測與故障診斷等預測，而故障診斷當中的故障分類可以確定故障的發生、故障的類型以及故障的原因，因此在工業製造中能有效作為設備維護管理的參考依據。由於在工業製造過程中所收集到的數據普遍具有遺漏值、維度與數據不平衡等問題，加上以往的分類器皆建構在數據平衡的情況下，使分類錯誤率相當高且無法真正觀察到我們主要想關注的對象，這將使故障分類面臨了挑戰。

本研究以一組來自 UCI 資料庫具有高維且資料不平衡的資料集以及一組某工廠機器中實際收集的資料不平衡資料集作為研究對象，以減少多數法與 SMOTE 兩種採樣方法來處理資料不平衡的問題，搭配卡方特徵選取方法找出得分較高的重要特徵後，以 10 交叉驗證進行參數的調整，最後以決策樹(DT)、隨機森林(RF)、XGBoost 與 K 近鄰(KNN)四種機器學習演算法進行分類模型的建立，經測試組資料的測試後，以正確率與 AUC 值作為績效評估的指標，最後結果顯示 XGBoost 的績效較其他三個模型好，在兩組資料集中皆獲得 0.9 以上的正確率與 AUC 值，進而探討機器設備故障的原因，作為設備維護的重要資訊。

關鍵字：機器學習、預測性維護、故障診斷、故障分類、資料不平衡、XGBoost