

經濟部110年度
《人工智慧產業關鍵技術拔尖計畫(4/4)》
合作研究計畫

《組合式分層建模之效能優化方案研究》
建議書徵求文件

財團法人資訊工業策進會

中華民國 110 年 1 月 11 日

110 年度合作研究計畫建議書徵求文件

一、簡介

多樣化的產品造成製造業生產系統日益複雜，產線中各種產品與設備的組合之間存在複雜的交互作用關係，對於加工時間與生產效率有不同的影響，導致預測不易，須經過多次試誤，才能建立適合的預測模型，過程不僅耗時且耗費大量資源，難以精準規劃人、機、料的投入安排及交期訂定。

因應主計畫「人工智慧產業關鍵技術拔尖計畫」研發「產線資源智慧調配系統」之目標，針對具有複雜交互作用的資料集，擬研發利用隱藏群的階層式展開機器學習方法，期開發混合型資料集的階層式展開與選模流程，並運用類別屬性的隱藏群，找出切分資料集最佳的方法以及最適合該預測目標的模型子集，在大量且即時的資料流中，達到建模過程可追溯、預測結果更精準同時兼具運算效能。

二、計畫目標

本計畫預計以機器學習方法如XGBoost為基礎，利用分群、特徵選擇與相關選模指標，逐步研究資料集中的隱藏群，進而建立各階層子資料集的預測模型且評比其預測能力，進行模型解釋與結果分析，達到預測結果可對應原始數據，利於瞭解資料的影響範疇且更能深度剖析資料，同時提升運算效能、預測準確度與可解釋度。目標為建立工業用之可解釋性的機器學習方法，貼近場域使用需求，加速AI平民化與產業化之進程。

三、計畫範圍

1. 研發混合資料集選模演算法：
將原始資料集進行階層式的展開，以屬性之間的相似性研究隱藏群，進而針對階層中不同屬性分群訓練預測模型，利用不同的測試資料交叉驗證，增進模型運算效能與預測準確率。
2. 使用Gitlab共享程式碼編修：
使用Gitlab建立遠端數據庫，利於多人共同開發，記錄編修歷程，有效管控程式碼版本，縮短開發與測試時程。

四、預期成果

預期使用者可追溯機器學習的推論結果，即使沒有資料科學家從旁協助，也能建構出良好的預測模型，以解釋實際場域面臨的各種可能情境，減少決策誤差，維持最佳產效能。

預期成果產出包括：

1. 期中、期末報告各一份
2. 專題演講一場
3. 論文投稿一件

4. 程式碼一份
5. 教育訓練一場

※前述成果如有專利構想或專利申請產出時，需注意專利申請之新穎性 (novelty)。因凡經公開發表之研發成果，如擬申請專利，須於公開發表後 6 個月內完成，前述成果如是以論文方式公開發表，將無法取得大陸與歐盟等國之專利。

五、執行方式

本計畫擬研發利用隱藏群的階層式展開機器學習方法，期發展混合資料集之階層式展開流程，同時觀察類別屬性對於目標值影響的相似性，研究資料集中的隱藏群，進而動態學習生產組合的數據特徵。演算法開發工作如下：

1. 混合型變數資料前處理
2. 建立資料集階層式展開方法
3. 開發類別屬性隱藏群偵測方法
4. 訓練各階層預測模型
5. 驗證各種模型預測準確度與解釋度

六、計畫期程及預估計畫總經費

計畫期程：自 110 年 03 月 01 日起 至 110 年 12 月 15 日

總經費：新台幣 600,000 元

七、驗收標準(含教育訓練)

1. 期中報告：110 年 07 月 30 日完成混合資料集選模演算法開發說明報告
2. 期末報告：110 年 11 月 30 日完成混合資料集選模演算法測試成效報告
3. 提供教育訓練：針對計畫成果轉移、相關程式建構說明等提供專題演講
4. 進度討論會議：每月召開一次會議

八、技術能力需求

1. 人工智慧專業知識應用能力：參與人員需具備動態最佳化、機器學習建模能力與人工智慧相關專長。
2. 計算機程式與資訊系統開發能力：參與人員需具備程式撰寫理論知識與實作能力，以C#、Python、R或其他程式語言開發。
3. 結果分析與實驗設計能力：參與人員需具備作業研究、大型系統優化、分散式決策分析能力，及實驗設計學理與巨量資料分析背景。