

BERT 助攻！AI 賦能，大幅提升行業統計分類效能

趙明光

行政院主計總處

摘要

本研究旨在探討利用微調後之 BERT (Bidirectional Encoder Representations from Transformers) 模型，協助經濟活動描述之自動分類與編碼，以提升工作效率與精進資料品質。現行多數統計調查之行業問項，仍須倚賴人工判讀，不僅耗時費力，且易受主觀判斷影響，以 110 年工業及服務業普查為例，近 149 萬筆主次要經濟活動資料以人工判讀，須耗費 2 萬 5 千小時（每筆約 1 分鐘估算）。若改以微調後之 BERT 模型自動分類，僅需 19 小時即可完成，大幅節省作業時間，且模型預測準確率達 90% 以上。

在模型訓練部份，主要採 110 年普查資料，並導入 ChatGPT 協助資料清洗、內容修補及編碼確認，經剔除重複後，以近 17 萬筆資料進行模型之微調訓練，為提升模型效能與分類準確度，訓練過程採多次迭代方式，並根據每次測試結果持續改進，直至最佳化狀態。

目前相關成果已導入本處自行開發之「行業智能查詢系統」，並初步建置於 eBAS 平台，可協助同仁於執行相關調查（如服務業營運及投資概況調查與工業及服務業普查等）時使用，未來亦可擴展至職類別判讀等專業領域。以發揮大型語言模型在協助處理政府資料上的巨大潛力，為統計作業自動化與智慧化奠定堅實基礎。

關鍵字：BERT 模型、ChatGPT、行業統計分類、人工判讀、準確率、行業智能查詢系統