



# 以 LINE 建構 AI 行職業統計編碼機器人

行職業統計編碼正確性影響統計結果及施政成效甚鉅，為非常重要之統計基礎工作，本文就如何運用 AI 演算法，創建自然語言處理辨識模型，並以國人常用通訊軟體 LINE 為查詢介面，建置 AI 行職業統計編碼機器人，有效提升統計編碼工作效能及確度。

高雄市政府主計處（黃專員河川）

## 壹、現況及問題

統計編碼工作係將資料依特性分類編碼，以作為計算運用基礎，如人力資源調查、受僱員工薪資調查、工業及服務業普查、人口及住宅普查等，須將受訪對象之行職業依行政院主計總處（以下簡稱主計總處）所訂行業統計分類及職業標準分類（已移勞動部編修）判定編碼，其正確性影響統計結果及施政成效甚鉅，為非常重要之統計基礎工作。

陳述行職業內容之文字因係自然語言之呈現，易因人而異，且囿於統計工作人員勤前教育訓練時間有限，統計編碼能力未臻純熟，故須投入可觀人力時間於解讀詞義及正確編碼。

## 貳、執行方式

### 一、文獻探討

研析國內外有關自然語言辨識之文獻，比較各種理論架構及模型後，選擇以人工

智慧深度學習循環神經網路（RNN）中之長短期記憶模型（LSTM）建立 AI 編碼模型。

### 二、建置 AI 行職業統計編碼模型

AI 編碼模型之建置須經過一系列程序處理，本研究主要建構行職業詞義解析編碼（下頁圖 1），相關程序及內涵說明如下：

#### （一）資料前置處理

將行職業內容文字資料先行處理，成可整理為模型運

用之數字序列，包含切字分詞、將字詞轉換成數字序列、切割訓練及驗證資料集等。

(二) 以詞嵌入技術自行學習字詞語義

將文字數字序列之數字序號轉為張量 (Tensor, 係為 N 維向量), 俾以詞嵌入技術解析詞義, 神經網路會自動調整每個字詞之張量 (N 維向量) 中各維度數值,

找到潛藏於大量詞彙中之語義關係, 經學習提升預估準確度。

(三) 字詞資料輸入 LSTM 層訓練學習

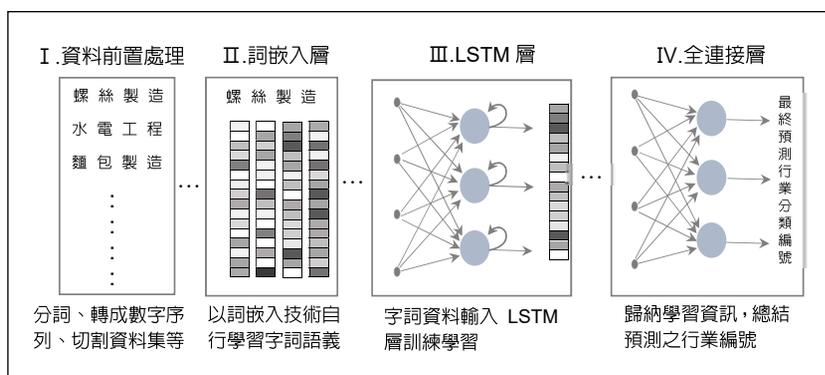
文本數字序列之張量資料讀入長短期記憶模型 LSTM (Long Short-Term Memory) 層, 經調整記憶及遺忘等參數以達預估最佳化。

(四) 歸納學習資訊, 總結預

測之行職業編號

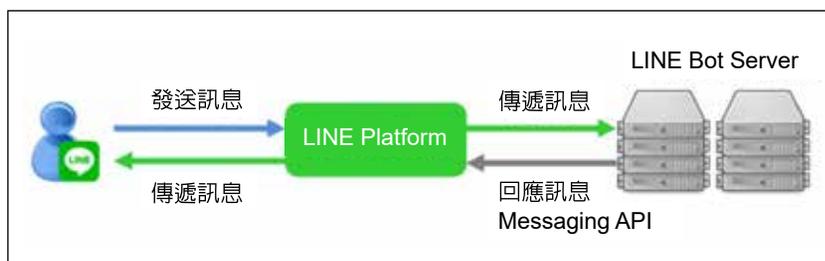
依據所有神經元資訊, 歸納各行職業內容文字相對各行職業分類編號之機率。全連接層使用 Softmax 函數將本層所有神經元裡數值正規化 (Normalization), 使其全壓縮為 0 至 1 之間的範圍, 並讓它們的和等於 1, 運算結果可將各神經元數值解釋為模型預估對應各行職業分類編號之機率, 數值最高者即為模型對行職業內容文字之預估行職業分類編號。

圖 1 行職業詞義解析編碼神經網路架構



資料來源：作者自行繪製。

圖 2 LINE 聊天機器人架構



資料來源：作者自行繪製。

### 三、AI 編碼模型介接 LINE 聊天機器人

將 AI 編碼模型介接 LINE 聊天機器人 (架構如圖 2), 以 LINE 聊天畫面作為 AI 編碼模型的查詢介面, 不僅提供親民便利的查詢介面, 後端 AI 編碼模型及資料庫之更新維護, 亦因採伺服器模式, 大幅提升使用者及管理者之更新維護作業效能。

### 參、具體成效

茲就效益性、應用性及革

# 創新變革精進獎勵項目

新性逐一說明 AI 編碼模型之成效評估。

## 一、效益性

高雄市 107 至 111 年計完成約 78 萬筆行職業資料統計編碼，估計費時約 6,500 小時，如以本案 AI 模型進行編碼，預估花費時間為約 6.6 小時，較人工快近千倍，大幅提升工作效率。（圖 3）

若以高雄市 111 年人力資源行職業編碼約 3.6 萬筆進行測試，行職業編碼模型預估準確率可達 99%，有效提升編碼確度。

本案 AI 模型訓練資料彙整、AI 建模 python 程式撰寫、LINE API 平台介接等工作均自行採用免費軟體程式開發完成，節省委外開發經費估逾 130 萬元。

## 二、應用性

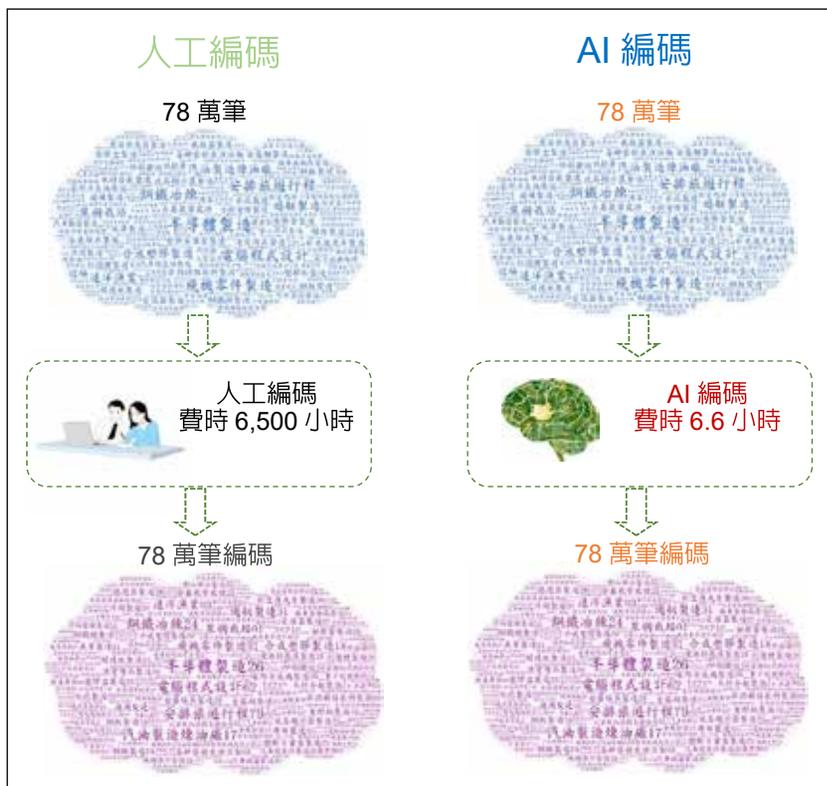
彙整主計總處行職業查詢網站資料、高雄市工業及服務業普查、高雄市人力資源調查等行職業相關內容文字及編碼，建置逾 4.7 萬筆不重複內容詞彙資料庫，與 Line 機器人介接應用，使用者查找資料一站完成。

經由彙集機器人接受查詢資料，建立歷史查詢資料庫，依此定期檢視調校 AI 模型參數及訓練資料，經此資料回饋機制，讓 AI 模型滾動學習精進，機器人持續優質進化。（下頁圖 4）

高雄市平均每年約須完成行職業統計編碼逾 15 萬筆，本案編碼機器人可供上述編碼工作運用，應用效益每年可惠及相關統計工作人力逾千人次，亦可推廣至有編碼作業需求之其他縣市。

除統計工作之外，本案 AI 技術可應用於市政客服文案分類、會計帳務摘要文字與對應科目檢核等市政業務，作為跨部門應用 AI 專案合作基礎，

圖 3 AI 模型提升行業編碼工作效率示意圖



資料來源：作者自行繪製。

以加速推動智慧市政。

### 三、革新性

本案以 python 程式語言，運用自然語言文字處理技術成熟穩定之循環神經網路 (RNN) 長短期記憶模型

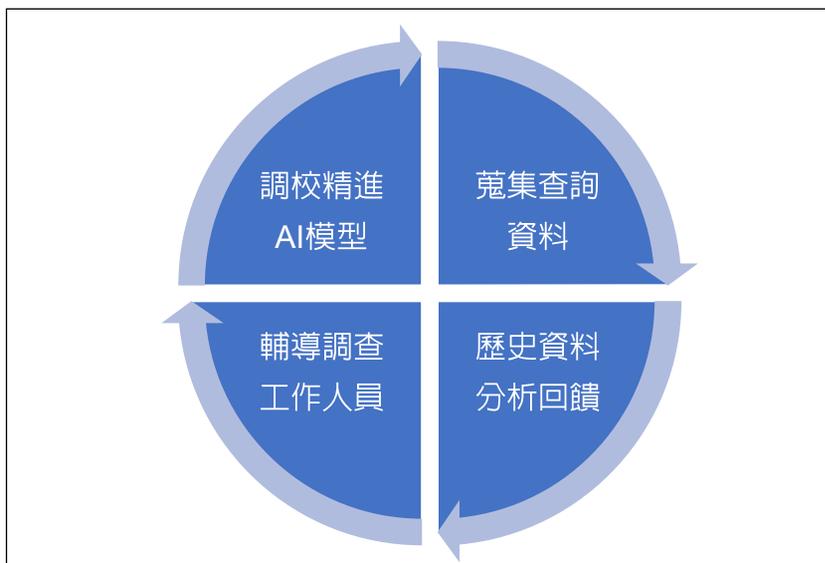
(LSTM) 建立 AI 編碼模型，模型參數逾 483 萬個，並經由介接 LINE Messaging API，以聊天機器人形式分別創建行業及職業 AI 編碼機器人，係全國首創運用人工智慧深度學習，以及運用 LINE 聊天機器

人作為統計編碼與查詢工作。

(圖 5)

本案機器人即時蒐集使用者利用各式通訊設備傳送之查詢資料，經物聯網數據分析，可了解使用者行職業內容文字陳述方式慣性，如發現使用者敘述行職業內容文字完整性不足或有明顯偏誤時，可據以個別輔導調查工作人員，有效提升統計工作效能及確度。

圖 4 AI 模型資料回饋持續調校精進



資料來源：作者自行繪製。

圖 5 Line 聊天畫面查詢行職業編碼



資料來源：作者自行繪製。

### 肆、結語

經本案實證，運用人工智慧演算法確可協助統計人員更快更準確地完成行職業編碼工作，進而提升統計資料正確性及決策施政品質。

中文詞彙變化範疇大，人工智慧語意解析難度相對較高，本案在此領域初窺門徑，幸有所獲，惟 AI 語意解析相關技術日新月異，又應用層面無所不及，為知識共享及擴大創新效益，期能以本案累積知能為基石，深入各領域業務開拓應用視野，與各機關協力開啓 AI 運用之窗，看見嶄新的統計面貌。❖