

機器學習於政府統計之應用

資通訊技術不斷快速發展，加上疫情洗禮，各國更加正視政府統計現代化課題，並開啓探索運用機器學習演算法，提高作業效率與拓展新資料來源的風潮，本文透過國際的機器學習實證經驗，探討對政府統計導入機器學習應有的認知。

趙明光（行政院主計總處綜合統計處研究員）

壹、前言

大數據時代，科技帶領公共政策論述漸由民意取向轉為循證（Evidence Based）基礎，另一方面受疫情襲擊，各界對於經社資訊取得簡便性與即時性的期盼更甚以往，在這兩股力量交互衝擊下，政府統計該如何與時俱進導入新技術，並拓展新的資料來源，成為各國政府統計現代化過程所須面對的共同課題。

1980 年代，因人工智慧（AI）的發展而孕育出「機器學習¹」（以下簡稱 ML）新

學門，惟受限於當時的硬體技術無法因應大量訓練資料和有效處理龐大的運算能力，致研究熱潮漸趨沉寂。近年來，受軟硬體技術大幅躍進（如圖形處理器 GPU 開創全新的運算架構）影響，加上 AlphaGo 及 ChatGPT 等成功典範的鼓舞，重新燃起 AI 及 ML 的研究與應用熱潮，國際上並廣泛認為導入 ML 是政府統計現代化可行途徑之一。

聯合國歐洲經濟委員會（以下簡稱 UNECE）於 2019 年成立 ML 專案小組，系統化執行多項前導研究（Pilot

Study）計畫，嘗試將 ML 技術應用於精進政府統計作業，並於 2022 年協同英國國家統計局將研究成果共同發布在「機器學習於公務統計應用」（Machine Learning for Official Statistics）報告中。本文將透過 UNECE 的實證經驗，探討對政府統計導入 ML 應有的認知。

貳、公務統計導入機器學習技術之實例

UNECE 主要針對「資料分類與編碼」、「資料檢核與插補」及「影像資料探勘」三

類主題規劃前導研究：

樣的輔助資源與政府統計資料，建置自動分類及編碼系統（表 1）。

一、資料分類與編碼

目的在於有效整合多態

二、資料檢核與插補

利用屬性間邏輯關係或以 ML 探索屬性間的內在結構，

表 1 UNECE 資料分類與編碼研究案例

機構	美國勞工統計局	加拿大統計局	國際貨幣基金組織	墨西哥國家地理統計局	比利時法蘭德斯統計局	塞爾維亞共和國統計局	挪威統計局	波蘭統計局
目標	工作場所傷害和疾病編碼	行、職業編碼	IMF 時間數列目錄之自動編碼	應用自然語言進行經濟活動及職業別之歸類與編碼	對推特 (Twitter) 資料進行情感分析	經濟活動編碼	行業歸類與編碼	依歐洲個人消費目的分類 (ECOICOP) 將產品描述進行歸類及編碼
應用資料	職業傷害和疾病調查	社區家庭調查；勞動力調查；健康措施調查	會員國的時間數列資料	家庭收支資料	推特資料	勞動力調查	經濟活動的描述相關歷史資料及官方關鍵詞 (編碼)	網路爬梳所得之商品名稱
ML 演算法	羅吉斯迴歸；支援向量機；單純貝氏分類；隨機森林；多層感知器；卷積神經網路；循環神經網路	fastText 或極限梯度提升	羅吉斯迴歸；K-近鄰	支援向量機；羅吉斯迴歸；隨機森林；類神經網路；極限梯度提升；K-近鄰；單純貝氏分類；決策樹	懲罰羅吉斯迴歸；隨機森林；梯度提升決策樹；多層感知器	隨機森林；支援向量機；羅吉斯迴歸	fastText；羅吉斯迴歸；隨機森林；單純貝氏分類；支援向量機；卷積神經網路	單純貝氏分類；羅吉斯迴歸；隨機森林；支援向量機；類神經網路
應用軟體	Python; scikit-learn	G-Code	Python	Python; scikit-learn; Keras	Python	Python; Pandas; scikit-learn; Pyzo IDE	Python	Python; scikit-learn
結果	準確度：69.8%~91.9%	準確度：>95.0% 精確度：100.0% 召回率：100.0%	準確度：約 80.0%	準確度：83.1%~87.7% 精確度：57.8%~66.0% 召回率：57.3%~64.5%	精確度：80.0% 召回率：81.0%	準確度：69.0%~75.0%		準確度：90.5%~92.2% 精確度：90.0%~92.0% 召回率：90.0%~92.0%

說明：準確度 (Accuracy)、精確度 (Precision) 及召回率 (Recall) 為分群演算法常用的評核指標，以將產品描述歸入食品、衣物和電子產品三類的模型為例，

1. 準確度是指模型正確分類的占比，例如在 100 個產品描述，模型將 90 個描述正確分到上述三個類別中，那麼模型的準確度為 90%；
2. 精確度是指模型正確分入某 (食品) 類，真正屬於該類別的占比，如模型將 30 個描述分為食品類，但其中只有 20 個是真正屬於此類，則模型在食物類之精確度為 $20/30 \times 100\% = 66.7\%$ ；
3. 召回率則是指所有真正屬於某 (食品) 類中，模型成功分類出來的占比，假設有 50 個描述真正屬於食品類，而模型只成功分類出其中的 40 個，那麼召回率為 $40/50 \times 100\% = 80.0\%$ 。

資料來源：UNECE (2022), Machine Learning for Official Statistics.

論述》統計 · 調查

據以進行特異值檢核與修正，或遺漏值插補（表 2、表 3）。

結合影像處理技術及時空（Spatio-Temporal）屬性資料協助位址辨識或統計資訊探勘等（下頁表 4）。

三、影像資料探勘

UNECE 的研究結果證實，透過 ML 技術與專業人力協同合作，確能提升機構對新資料源（如航空影像、網路爬梳資料等）的處理技能與應用層面，對具重複性且勞力密集作業（如行職業編碼等）明顯有效提高品質及效能，並使部分業務程序自動化（如資料檢核等）。此外，這些研究歷程亦給後續仿效者提供對 ML 更深層的认识。

表 2 UNECE 資料檢核研究案例

機構	英國國家統計局	義大利國家統計局
目標	生活費用和食物調查 (LCF) 之收入資料分類及修正	編修公共行政登錄資料
應用資料	2018 年第 2 季及第 3 季生活費用和食物調查	公共行政資料庫 (BDAP) 及廣義政府交易信息系統 (SIOPE)
處理方式	透過資料變動趨勢建模	辨識不同來源資料間資料不一致情形，並透過資料單位及資料背景知識進行深入分析，以探究資料檢核與修正之規則
ML 演算法	決策樹；隨機森林；類神經網路	決策樹；隨機森林
應用軟體	Python	R

資料來源：UNECE (2022), Machine Learning for Official Statistics.

參、由 UNECE 實證經驗看 ML

一、政府統計導入 ML 技術作業宜有長遠規劃

UNECE 的前導研究計畫大部分皆屬對某些單一特定目的之小規模議題（如分類與編碼）實現，以證明其可行性的「概念驗證」（Proof of Concept），尚未發揮 ML 能由實際案例的驗測過程中，累積經驗（Learning）不斷優化模型的特質（此亦是 ML 與傳統統計方法明顯的差異點之一），顯示導入 ML 仍處啟蒙階段，

表 3 UNECE 資料插補研究案例

機構	德國聯邦統計局	義大利國家統計局	比利時法蘭德斯統計局	波蘭統計局
目標	資料插補	個人教育程度資料插補	能源平衡與統計之預估	旅遊調查資料插補
應用資料	製造業、礦業及土石採取業營運調查資料	教育部、大學及研究所公務登記資料；2011 年普查資料；相關統計調查	2000 年第 1 季至 2019 年第 1 季能源相關資料	2016 年至 2018 年旅遊調查及其他大數據
ML 演算法	K-近鄰；貝氏網路；隨機森林；支援向量機	多層感知器；隨機森林；對數線性迴歸	各式迴歸；隨機森林；類神經網路；集成預測	各種線性模型；迴歸樹；隨機森林；K-近鄰；各式支援向量機
應用軟體	R	Python	Python	R

資料來源：UNECE (2022), Machine Learning for Official Statistics.

表 4 UNECE 影像資料探勘案例

機構	澳洲統計局	荷蘭統計局	瑞士聯邦統計局	墨西哥國家地理統計局
目標	位址自動辨識	從圖像中擷取統計資訊	土地應用分類及其變化偵測方式改進	應用衛星數據觀察非普查年的城市擴張
應用資料	航空影像	航空及衛星影像	航空及衛星影像	衛星影像
ML 演算法	卷積神經網路	卷積神經網路；隨機森林；支援向量機	卷積神經網路；隨機森林	卷積神經網路；極限隨機樹
應用軟體	TensorFlow	TensorFlow；scikit-learn	TensorFlow；scikit-learn	TensorFlow；scikit-learn

資料來源：UNECE (2022), Machine Learning for Official Statistics.

也非一蹴可及，應有長遠願景的規劃及面對挫折的心理準備。

二、導入 ML 技術作業應強化跨領域技術整合

ML 善於處理大量且複雜型態資料 (Complex Data)，如網頁、圖像影片、聲音或衛星數據等，惟 ML 係為針對特定目的建構具自動學習機制的系統泛稱，並無所謂萬用的 ML 機制存在，ML 系統的開發建置須先面對問題本質確立標的，並依資料與問題形態選取合適演算法進行模型訓練。

ML 對複雜資料處理能力來自各相關領域知識的整合應用，如自然語言技術解析語

音資料，多媒體技術建置影片索引結構，文字探勘前置程序對網頁內容進行斷詞、同義詞對應、詞庫更新、字詞文章索引等，模型的訓練過程離不開演算法及統計學的知識支援，而整個學習過程也高度依賴執行者的專業判斷，以將偏離學習軌跡的路徑予立即修正或刪除。

面對單純型態資料的小議題，單打獨鬥或許能立竿見影，惟若要建置跨公務機關間的循證決策基礎，須能拓展新的資料來源、架構協力分享平臺、建立跨領域專業團隊，及不斷機動強化成員專業素養等，才能讓 ML 技術於機關發揮較大功效。

三、ML 典範移轉不易

UNECE 的研究結果顯示，針對「分類與編碼」議題，各組織所應用的演算法相當雷同，普遍應用羅吉斯迴歸 (Logistic Regression)、隨機森林 (Random Forest)、支援向量機 (SVM) 及單純貝氏分類 (Naive Bayes) 等常見的分類 (Classification) 統計模型或演算法，惟若以準確度 (Accuracy) 來看，不同國家各種應用的實證結果差異卻相當大，有不及 7 成者，也有高達 9 成 5 (第 73 頁表 1)。

輸入資料集的清理 (Data Cleaning) 品質、對資料型態的認知與資料表達結構、演算法評核機制、訓練資料與驗證資料間的背景相似度、多個演算集成 (Boosting) 為強演算法的權衡機制，以及執行者於建模過程各階段的專業判斷等，都會對最終結果產生影響，以致他人的成功經驗往往無法僅透過資料處理機制及演算法的仿效而得。也就是說，應用 ML 演算法並不代表



已將 ML 技術導入組織，因為整個機器學習建置歷程，是執行者對資料與技術背景的深入探索過程，必須充分瞭解相關背景知識，並不斷訓練測試，才可以產製出投入與結果間能清楚解釋，及具重複驗證性的 ML 模型，提高模型的穩定性與可信度，這也是 UNECE 於報告中強調選取合宜的 ML 演算法不能忽略可解釋性（Explainability）與可重現性（Reproducibility）的原因。

四、摒棄為 ML 而 ML 的心態

ML 是資料探索的程序，與一般套裝軟體的應用模式不同，導入 ML 須評估資訊資源、對技術的掌握能力，以及對機關文化的影響，最重要的是就要先盤點現階段所擁有的資料特質及機關的最終需求，以評估任務是否具引入 ML 的成熟度，此外，如手邊所蒐集的資料案例未能充分涵蓋各態樣情形，或用於訓練之資料與驗證資料分屬不同情境案例，這些情況所訓練出來的 ML 模型並

沒任何意義。

肆、ML 演算法選取的正確思維

資料清理並確認任務目標是 ML 重要的基本工作，之後的 ML 演算法選取則為影響 ML 成敗的另一關鍵課題，UNECE 提出 QF4SA（Quality Framework for Statistical Algorithms）的評核概念，來協助判斷所選取演算法的合適性。QF4SA 以準確度、可解釋性、可重現性、時效性（Timeliness）及成本效益（Cost Effectiveness）五面向為評定演算法優劣的參據，強調除了關注 ML 模型對設定目標推算的確度外，演算法的邏輯是否易於理解與可重複驗證，以及 ML 運作的時效與成本效益亦是重要評核項目。雖五個評核面向孰輕孰重，並無明確準則，須視每一個 ML 個案之目標而定，惟深入瞭解資料與演算法內涵，及不拘泥於準確度的追求，均有助於適當演算法的選取，進一步說明如下：

一、資料及技術的背景知識同等重要

大家常以「garbage in, garbage out」來提醒對情境的認知是否正確，這句話亦顯示背景知識的重要性，資料的背景知識是資料分析的根本基礎，ML 所得結果若與資料本身背景特質明顯相矛盾，就須認真重新審視整個建模過程。

同樣地，應用 ML 演算法須瞭解各種演算法的特質，如我們常認為應用愈多維度屬性，所涵蓋資訊量愈大，結果會更精確，惟部分常用的 ML 演算法有「多維魔咒」（Curse of dimensionality）問題（如 K-近鄰演算法），當資料屬性維度過大（10 維度以上），其結果會變得相當不穩定（無意義）²；又如神經網路於訓練初始神經元權數設定不當，其訓練過程常會發生無法收斂情形。此外，資料及技術的背景知識亦是 QF4SA 讓演算法擺脫黑箱，具「可解釋性」的重要基礎，深入瞭解資料面或技術面的內涵都很重要。

二、準確度的追求迷思

準確度是 ML 模型評核的常用指標，惟其比較前提應以某特定的相同「模型空間」下，以準確度較高者較優，所謂的「模型空間」係指我們建模所列入考量的所有模型所成的集合（如二次多項式迴歸），倘若我們未將「模型空間」設限，我們可輕易透過增加模型的複雜度來提高準確度（例如將 N 次多項式迴歸之 N 值不斷提高），甚至讓訓練與測試資料集的錯誤率同時逼近零，但此結果將對新發生案例幾乎無實質預測效果，這即是 ML 理論的「一般化」（Generalization）議題。

三、簡約法則

由於模型複雜度與錯誤率具相互消長關係，以較高複雜度模型提高準確度，除衍生前面所討論的「一般化」外，亦會降低計算效能及模型可解釋性，UNECE 之 QF4SA 準則雖未對模型複雜度有所討論，但可由簡約法則（Occam's

Razor）所提出具相同解釋度的模型，以簡明為優的觀點，補足這項缺口³。

伍、結語

ML 最終產品常以自動化、高效率及不斷自我優化的優雅形態呈現，然整個建置過程是透過不斷的人機互動，交互激勵成長，一步步匍匐前進（Increment），逐步揭開資料背後所隱藏的神秘面紗。

UNECE 的成功經驗對加速政府統計服務創新步調，提升組織的應變韌性，強化循證決策基礎，及拓展新統計資料來源等政府統計現代化議題帶來相當正向的鼓舞，惟導入 ML 前我們仍須仔細自我評估資料面、技術面、軟硬體資源及機關文化等現況，才能讓這些成功經驗激勵更多 ML 應用於政府統計的新思維。

註釋

1. 人工智慧係指「透過電腦程式或相關科技硬體來呈現人類智慧的一種技術」，機器學習則是人工智慧的一種，可從大量資料中理

出頭緒並學會某些事情。

2. Beyer K., Goldstein J., Ramakrishnan R. and ShaftU (1999), When is Nearest Neighbors Meaningful? ICDT Conference Proceedings.
3. Ian H Witten, Eibe Frank and Mark A. Hall (2011), Data mining: practical machine learning tools and techniques, 3rd Edition. Morgan Kaufmann, 183 頁；ML 研究者依據簡約法則後續推演出一套「最小描述長度」（Minimum Description Length）理論，為 ML 建模之模型複雜度、計算複雜度與準確度間關係及模型選取原則等建構相關理論基礎，請參閱 Peter D. Grunwald (2007), The Minimum Description Length Principle, the MIT press.

參考文獻

1. Alison Baily & InKyung Choi (2022), ONS-UNECE Machine Learning Group 2022 Final Report, UN.
2. Barbara Ubaldi, Charlotte Van Ooijen, & Benjamin Welby (2019), A Data-Driven Public Sector: Enabling the Strategic Use of Data For Productive, Inclusive and Trustworthy Governance, OECD Working Papers on Public Governance.
3. UNECE (2022), Machine Learning for Official Statistics. ❖