聯合學習發展趨勢與對台灣產業的影響

台灣亞太產業分析專業協進會 107 年認證產業分析師 葉逸萱

一、全球聯合學習解決方案市場與垂直產業應用

聯合學習於 2016 年由 Google 提出，是一個分散式機器學習的概念。在過程中，擁有數據的參與者可獨自進行 AI 模型的訓練，之後藉由貢獻自己的 AI 訓練模型參數，共同來優化所有的模型，以不同於過往獨自的模型訓練方式，突破 AI 廣適性的限制與資料不足的問題。也因聯合學習為訓練 AI 模型中解決數據安全與隱私的問題，近年來受到廣泛的關注。

根據 ReportLinker 的研究報告指出，2020 年全球聯合學習解決方案市場規模約 9,283 萬美元，預計至 2025 年將增長至 1.59 億美元。主要潛力產業為醫療保健、零售、電子商務、能源與公共事業、以及製造業等垂直產業，其中，又以醫療領域占最大市場規模。而年複合成長率增長最高者為製造業，主因為工業物聯網(IIoT)應用與競爭日益劇烈，製造業將優先分析從各方收集來的數據，以獲取競爭優勢。

二、數據應用需求與聯合學習發展趨勢

（一）產業數據量與分析需求劇增

現在是數據當道的世代，數據應用已是全球貫徹決策的重要核心，加上物聯網相關技術成熟，不同領域爭相採用 IoT 裝置，所蒐集到的數據量十分龐大，根據 IDC 報告指出，2019 年全球數據量為 45 ZB，預計至 2025 年將成長到 175 ZB。此外，從許多國際智庫的研析中發現，僅有不到 1%的非結構化數據被妥善運用、且不到 50%的結構化數據用於企業決策中，顯示大部分的數據尚未充分發揮其潛力。

隨著 2020 年疫情危機，全產業面臨停工、人力不足、消費者需求、以及市場供給的轉變，為了在嚴峻的競爭環境中勝出，各產業開始投入智慧轉型，調整營運模式，更加強跨單位與多源數據整併，減少現有組織的數據孤島，以獲得最佳管理生產數據或淬煉客戶洞見的能力。然而，隱私權與個人資料保護等議題一直受到大眾的重視，如何妥善應用數據又兼具合規與安全性，將影響未來新興科技導入的重點。

經濟部技術處科專計畫成果，未經同意不得轉載

（二）聯合學習分析的商機與挑戰

聯合學習分析可以改善單一企業因數據孤島無法建構準確的 AI 模型、或是單一企業多源資料源彙整不易問題，透過各點/各企業分散式機器學習，互利又不需交換數據的機制，來達到 AI 模型優化的結果。由於聯合學習分析於醫療的成效(特別是 COVID-19 應用)，讓更多的企業了解其帶來的好處，以及釋放數據價值的潛力，因此也提高企業採用聯合學習的意願。

聯合學習可為數據隱私帶來解決方案，但仍有一些挑戰需要克服。例如(1)溝通效率問題：由於聯合學習由多方共同參與，因此須找出共通性關鍵議題，並針對問題進行定義、認知共識、產出項目、至初始 AI 模型的產生，將耗費許多溝通成本。若問題改變，則須重新定義和討論，導致效率不佳；(2)異質性問題：例如每個參與者所具備的系統、運算能力、存儲等方面不一致；參與者所擁有的數據豐富性不同，可能使得擁有豐富數據者，因模型參數平均效果下，導致自身精準度下降；(3)間接資訊洩漏：雖然聯合學習不須將資料上傳，但在傳送模型參數的過程中，仍可能受到模型逆推攻擊(Model Inversion Attack)，透過參數回推特定使用者的資料。

目前常見的解決方法，包括利用個性化聯邦學習(Personalized Federated Learning)降低數據、模型和系統異質性問題；透過新分散式架構(如 Client-Edge-Cloud Hierarchical Federated Learning)或 Asynchronous Federated Learning 改善系統異質性問題；利用同態加密、差分隱私(Differential Privacy)，在傳送的參數中添加雜訊(Noise)、或是結合區塊鏈技術，來強化數據的隱私和安全性等。

（三）國際大廠佈局與應用案例

國際主要兩大聯合學習框架為 Google 發布的 TensorFlow Federated (TFF)和 Facebook 與 OpenMined 開發的 PySyft。此外，還有 Nvidia Clara、美國南加大聯合 MIT、Stanford 等大學共同發布的 FedML、以及 Awesome MLOps 等。雖然聯合學習技術處於早期發展階段，但已有國際科技大廠(如 Google、NVIDIA、Intel、IBM…)、新創業者(如 Giant Oak、Consilient、Data Republic、DataFleets、Xayn…)、金融企業(如新加坡聯合海外銀行、澳洲 ANZ 集團、西太平洋銀行…)等投入探索與試驗。

例如 2019 年 NVIDIA、Owkin 和倫敦國王學院(King's College London)合作，在 NVIDIA Clara 上運行 Owkin Connect 來建立一個醫療保健服務的聯合學習平台。此平台透過區塊鏈技

經濟部技術處科專計畫成果，未經同意不得轉載

術讓參與的每家醫院都可以取得和追蹤訓練模型的各項數據，所訓練出的 AI 模型可用於癌症、心臟衰竭及神經退化性疾病等研究預測。而後，亦與英國藥廠聯盟 Melloddy 合作，試圖在保護患者隱私下，運用藥物化合物資料集來訓練 AI 模型，以提高藥物發現能力；2020 年 NVIDIA 與麻省布萊根綜合醫院合作的 EXAM (EMR CXR AI Model)計畫，研發可判斷新冠肺炎患者在初步檢查後的數小時或是數天內是否需要補充氧氣的 AI 模型，此研發成果有效輔助醫師判斷應給予新冠肺炎患者何種程度的醫療照顧。此計畫有世界各地共 20 間醫院參加，包含台灣大學醫學影像與數據人工智慧(MeDA)實驗室、全幅健康照護子中心(MAHC)、三軍總醫院及健保署等。

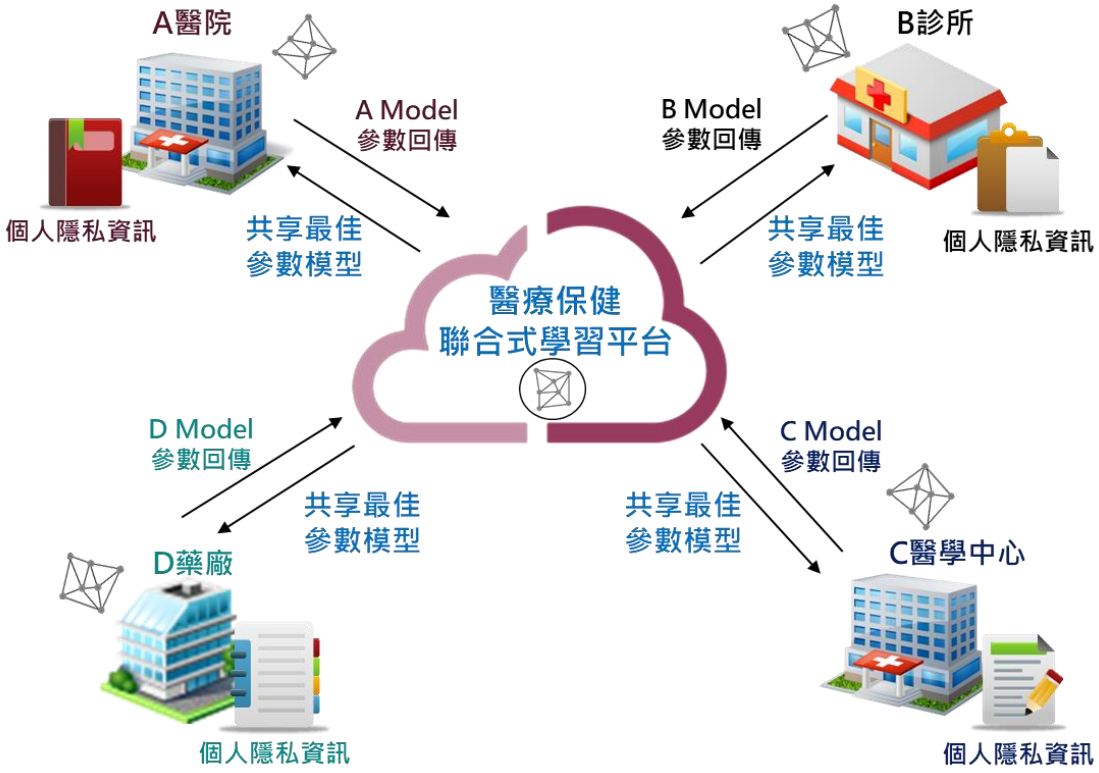

圖 1　醫療保健聯合學習平台

資料來源：工研院產科國際所 ITIS 研究團隊(2021/07)

金融領域方面，Intel 與澳洲 Data Republic 合作共同推出 Senate Platform，協助大型金融企業管理各數據資料庫的數據共享與分析，具數據治理機制包含審核日誌、用戶權限控管，進而降低各子集團間數據交換的風險。IBM 則運用聯合學習打造信用卡防欺詐偵測模型。在跨金融企業間的數據應用，則以反洗錢、反恐怖主義(AML/CFT)為居多。過去受到監管、國際數據

經濟部技術處科專計畫成果，未經同意不得轉載

政策、隱私、技術、同業競爭等因素，金融機構對金融犯罪事件的預防措施或是通報準確率皆不足，透過共享演算法的合作模式，讓參與使用該系統的金融業者可從彼此的交易數據中受益，進而建構功能更完善與更強大的模型。因此，此類型的聯合式學習系統對各國監管機構(如英國 FCA、美國 FinCEN)非常具吸引力，它可提升大小規模不同的銀行、相關金融業者至相當的能力共同打擊金融犯罪。其他領域應用如 IBM、Cloudera、Edge Delta、Snowflake 等提供客戶安全且快速的數據分析與 AI 模型訓練，有效挖掘多源數據的潛在價值。

三、我國聯合學習發展現況

我國近兩年也陸續有研發單位、企業、醫療機構投入相關應用，例如 2020 年 11 月 3 日 Taiwan AI Labs 宣布成立「台灣聯合學習醫療聯盟」，將採用其研發之開源框架 Harmonia，以分享模型取代資料共享，解決醫療數據機敏性問題。其中，台北榮總透過 Harmonia 進行腦轉移瘤 AI 模型訓練及優化、台大醫院建立 COVID-19 胸部 X 光自動檢測系統、以及心臟電腦斷層冠狀動脈最佳相位 Al 智慧選取等研究。2021 年 2 月 19 日國發會與 Taiwan AI Labs 號召，政府各部會響應下，共同組成「台灣聯合學習產業大聯盟」，期望透過各產業共同合作，推動臺灣優質的聯合學習服務與發展環境，建立國際可信任的 AI 解決方案。目前此聯盟研發應用的領域有醫療、金融、交通、製造、城鄉等。儘管全球聯合學習發展才剛起步，要達到大數據解密尚待一段時間，但台灣亦不落人後，政府正攜手產學研共同推動中，相信未來將有機會從安全數據共享機制中獲得更多的利益。

四、結論

由國際技術發展趨勢與產業應用得知，數據孤島和數據隱私是 AI 發展的重要挑戰，聯合學習提供一個較安全的解決方案，也因此吸引國際科技公司、大型企業、新創公司積極投入。在台灣方面，我國醫療業(具健保數據、影像標註資料庫、專業人力)、製造業(如晶片、網通及終端設備等)為高度優勢產業，面對疫情衝擊，各國積極研究醫藥和打造自己的半導體產業，為了維持我國產業的競爭優勢，更應善用大數據。然台灣產業環境以中小企業為主，具備完善 AI 分析環境、工具與人才並不容易，因此，本研究建議除了政府策略推動外，也建議我國雲端服務提供者，加強邊緣機器學習與聯合學習的投入，以平台即服務的概念，整合與簡化由本地訓練、邊緣推論到雲端的過程，提供產業跨組織、多源數據庫的安全協同機制，亦是提升客戶黏著度的重要應用。

經濟部技術處科專計畫成果，未經同意不得轉載

(本文作者為工研院產科國際所執行產業技術基磐研究與知識服務計畫產業分析師)

原文出處：ITIS 智網 http://www.itis.org.tw/

經濟部技術處科專計畫成果，未經同意不得轉載