

作業報導

● 巨量資料應用發展趨勢

國家發展委員會分析師 楊耿瑜

工業技術研究院產經中心研究員 葉乃嘉

一、萬物聯網環境驅動巨量資料崛起與價值

巨量資料(Big Data)從何而來？關鍵在於”網際網路”的出現，網路科技的興起帶動了產業變革，1995年起，Amazon、Google等大型電子商務網站與網路搜尋服務陸續出現，Twitter、Facebook等社群網站逐漸盛行，智慧行動裝置廣泛使用，以及近2年來穿戴裝置和工業網路(Industrial Internet)等物聯網(Internet of Things)應用再度成為關注議題。新興網路科技發展帶來大量聯網的人與物，因而產生一般傳統資料庫軟體工具難以儲存處理的大量資料，此外，有別於傳統結構化的交易資料，新興科技所帶來的資料更多屬於非結構化的資料(例如：網站記錄Log、社群網頁、影音圖檔等)，而且資料是源源不斷迅速的產生，每秒鐘達到數千次以上的交易(例如：Google平均每秒鐘有近5千次搜尋、風力渦輪機的感測器每40毫秒要接收分析100筆以上資料)，也因為這些特徵，一般稱「巨量資料」具有4V特性：Volume(大量性)、Variety(多樣性)、Velocity(迅速性)、Veracity(真實性)。

巨量資料分析的目的在於創新應用，不論政府或企業，運用巨量資料的核心思維在於為提供客戶更好、更具差異化的服務，包括做好客戶關係管理、優化營運服務、研發創新產品，最終目的是要創造商業價值，以增加公司收益或提升人民福祉，達成企業經營目標。也因為這些需求，巨量資料除了具備4V特性之外，巨量資料分析也有別於傳統資料環境，不僅是分析樣本資料，而是分析全部的資料，甚至整合其他領域的各種資料源，透過多種演算法分析以發掘未知的洞見，下表列出巨量資料分析與傳統資料分析的差異。

表 傳統資料分析與巨量資料分析之差異比較

特性	傳統資料分析	巨量資料分析
資料範圍	分析樣本資料，為特定目的蒐集取得的資料	分析全部資料，包括跨領域從各種來源、其他目的取得的資料，以及自動產生的資料
資料型態	結構化資料，已被關聯定義	結構化與非結構化的資料
資料運用	查詢資料以驗證一個特定的假設	採用多種演算法以產生新的洞見或發現之前未知的資料關係
資料分析	延遲分析，資料先被結構化、整理儲存之後，才進行分析	即時分析，資料在擷取同時就進行分析

二、新興資料源帶動巨量資料創新應用模式

巨量資料應用模式可以用一個簡單的公式來表示：「資料」加上「處理演算」發掘出「商業智慧」。「資料」包括大量、多樣、即時的資料；「處理演算」在整合儲存各種來源的各式資料，並且符合成本效益(金錢與時間效益)；「商業智慧」則是要分析了解客戶的需求或者提高經營效率等。

巨量資料在各個領域有各式各樣的應用，伴隨著網路科技發展所產生的新興資料源，也成為創新巨量資料應用的重要趨勢，茲說明如下：

(一)分析用戶網站互動紀錄，提供個人化推薦機制

推薦機制是許多網站服務公司會使用的一種巨量資料應用模式，目的在於能夠推薦符合客戶需求的產品或服務，進而提高銷售。以線上影音服務業者 Netflix 為例，其蒐集分析大量用戶線上互動紀錄(如：影片觀看行為)，並且同時考量社群等多種資訊，根據用戶最新狀態即時回應需求，Netflix 提供多面向個人化推薦機制服務，其結果顯示用戶選擇之影片有 75%來自系統推薦，因而提高用戶的黏著度。

(二)分析社群媒體輿情資料，幫助市場決策與即時事件警示

社群媒體分析的目的，在於透過社群網站的發文資料，以掌握群眾的想法，作為協助企業市場策略規劃之依據，此作法有別於傳統問卷調查方式，社群網站分析可以更即時而全面的了解群眾的情感需求與特定事件消息，在金融、新聞、公部門、消費市場等領域都有實際應用案例。例如日本 Cool Japan 計畫中，他們蒐集社群網站 Twitter 和 Wikipedia 有關日本數位媒體(漫畫、動畫、遊戲)的資料，透過預測分析繪製成類似氣象圖的「Asia Trend Map」，了解不同數位媒體在亞洲各國的熱門趨勢，不僅降低海外市場調查研究的成本，也能精準的針對不同國家，推出當地最受歡迎的商品。

(三)分析行動位置資料，提供適地性服務與群眾行為洞察

分析手機行動位置資料，可以根據民眾所在位置提供即時服務，也可以了解群眾行動路徑及聚集的地方，不僅可以作為商業決策也可以作為城市規劃之參考，目前在行銷廣告、交通運輸、旅遊觀光、零售業等領域都有實際應用案例。例如：日本觀光廳與電信公司合作，分析用戶手機位置資料，了解觀光客的出發地、停留時間、所利用的交通工具、是否有住宿等資訊，目的在了解觀光活動的實施成效，並能開發新的觀光路線和景點，促進觀光產業發展，例如：透過此應用可以發現被政府忽略的熱門景點，可作為觀光推廣的新景點，或者由旅遊公司推出新的旅遊行程。

(四)分析物聯網資料，提升機器設備營運效益

物聯網結合巨量資料分析在工業上有很重要的應用發展，包括德國政府提出的工業 4.0(Industry 4.0)策略，以及 GE 公司提倡的工業網路概念，都朝著工業

製造智慧化應用邁進。從 GE 所推出在能源發電、鐵道航空等領域的解決方案來看，工業應用主要模式，係在機器設備上安裝大量感測器，收集即時資料並整合歷史資料進行分析，一方面可以提供預測性維護功能(Predictive Maintenance)，主動診斷設備問題進行事先維修，以降低非預期的停機時間，另一方面可以自動最佳化機器運作，以提高產能(例如：自動調整風力發電渦輪機的葉片間距，提高電力輸出量)或者減少能源使用，進而提升營運效益。

三、發展資料驅動型組織發揮巨量資料最大價值

巨量資料對產業帶來的影響不只是「量」的改變，更重要的是「質」的變化，且各個領域都有相關，也就是都有創造價值的機會。根據調查，企業與政府發展巨量資料時，最大的挑戰即在於“不知道如何獲取巨量資料的價值”，這包括不知道要解決什麼問題、不知道要分析什麼資料源、不知道怎麼發展應用案例，因此，建立資料驅動(Data Driven)的文化認知是組織成功發展巨量資料創新應用的基礎，也就是要依據資料證據進行決策，例如：有哪些是根據人工“經驗法則”或者“統一原則”判斷的地方，都可以進一步運用巨量資料分析做自動化、個別化的智慧決策，應用發展過程中也要不斷試驗洞察，持續提高實施成效。除此之外，組織需要培養自己的資料科學家團隊(Data Scientist)，深入了解自身的業務需求，為公司持續思考有哪些資料可以搜集，以及如何運用於決策，資料使用上，也不能只是分析現有內部記錄，更需要思考彙整分析大量非現有的資料，如log、外部資料(社群媒體、Open Data等)、或透過感測器收集用戶/機器/環境的資訊，可以帶來更多不一樣的洞見。在文化認知的基礎之上，結合人才、資料與技術，為發揮巨量資料最大價值的關鍵成功要素。