



主計大數據之應用

大數據應用在各領域不斷的衝擊現況，行政院主計總處各項業務產出的龐大數據，不斷地發出數據存在的價值信號，利用主計相關資料分析政府決策並即時彙整預測情勢，對於政府單位未來發展是一大突破。105 年臺北醫學大學大數據研究中心與主計總處合作進行研究，探討主計資料大數據，也探討業務單位對於大數據的需求及創新發想，以數據做為主計的基礎與核心，用數據顛覆傳統業務。

謝邦昌、丁台怡、江志民、吳格非、張葦憶（臺北醫學大學管理學院院長兼大數據研究中心主任、臺北醫學大學管理學院大數據研究中心顧問、經國管理暨健康學院餐旅管理系助理教授、輔仁大學商學研究所（博士班）博士生、臺北醫學大學管理學院大數據研究中心秘書）

壹、大數據介紹

大數據的處理分析正成為新一代資訊技術融合應用的結點：互聯網、社交網路、智慧家庭等是新一代資訊技術的應用形態，不斷的產生大數據。大數據是資訊產業持續高速增長的新引擎，也是提高核心競爭力的關鍵因素，大數據將對晶元、存儲產業產生重要影響，各行各業的決策正在從“業務驅動”轉變“數據驅動”。例如

經濟部資訊中心利用台電太陽能資料建立「太陽能發電效能分析平台」了解太陽能實際效益；Uber 利用乘客數據分析精準到站時間及金額等，完美運用大數據造成顛覆以往計程車新時代；台灣集中保管結算所利用數據分析投資戶使用電子投票的習慣，調整並改善行銷方式；ZARA 藉由全球的網絡，對每一件銷售出去的商品數據精準記錄。將資料用自動化程式分析，總結出顧客的消費喜

好，作為未來推出產品的決策及依據；行政院主計總處（以下簡稱主計總處）建立「薪情平臺」利用數據分析各產業與勞動薪資的狀況。

2015 年機器學習的周邊市場規模約 3.6 億美元，至 2020 年預估將突破 29 億美元，並在 AI 整體市場的 50 億美元中佔了約 6 成比重，可以說機器學習的技術突破，就是 AI 市場發展的原動力。大數據深度學習讓 AI 大躍進，自從 Google 的

人工智慧 AlphaGO 成爲圍棋界的百勝將軍開始，AI (Artificial Intelligence, 人工智慧) 這兩個英文字，瞬間成爲科技業最熱門的關鍵字之一。Big Data to AI, Create a perfect life。

貳、開放數據應用實例

臺灣政府防範登革熱，社群推群聚感染地圖。利用雲端大數據的彙算，每天自動更新，由於統計範圍較村里小，可以呈現疫情重點區域，隨時檢視街道分布、衛星空照或擴大村里檢視，讓民衆清楚登革熱集中區域，強化社區動員。結合內政部統計區分類及運用雲端大數據的彙算，每日自動更新一級發布區內近兩週病人數達 2 例以上之地理分布。一級發布區是指約 450 人的生活圈，面積較村里小，能更貼近在地民衆日常活動區域 (圖 1)。

台灣企業利用經濟部水利署的公開資料，讓民衆了解全台灣缺水的情況，隨時掌握停水資訊與水庫蓄水量，並用預測公式算出剩餘供水天數，

讓民衆對於節能用水更加有感 (圖 2)。

我國政府在統計大數據也

有亮眼成績，目前主要應用在五個部會，包含：ETC 資料應用於制訂交通政策、財稅資



資料來源：政府資料開放平台。



資料來源：政府資料開放平台。



料應用於稅制政策，改善貧富差距、電子發票用於分析商業情勢、健保資料用於特殊或亞太疾病研究，開發科技新藥與健保制度改進、圖資災防資料應用於國土規劃及救災、聯合徵信中心資料用於分析產業發展趨勢及國家發展委員會建置「政府物價資訊看板平台」，整合財政部、行政院消保處及主計總處等單位的資訊，可以查到賣場、便利商店近期的食品、飲料、日常用品售價。經常引起各界關注的貧富差距問題，財政部將五百多萬戶家庭的稅檔資訊彙整成十等分位、二十等分位的所得分配統計，成為外界研判台灣貧富差距的重要參考。不論是電子發票、勞保統計、入出境資料、綜所稅檔皆屬公務統計，無需抽樣推估即可取得，有助於我們了解台灣經濟社會，其對政府抽樣調查所取得的相關統計也有相輔相成之效。

參、大數據時代的會計、審核六大發展趨勢

面對大數據所帶來的新思維、新技術和方法的變革，政府會計、審核人員需要應時而變來適應思維模式及數據處理模式的變化。大數據對會計、審核發展的影響，主要表現在以下：

一、財務報告從事後往即時發展

傳統會計人員只是在企業生產經營業務發生後才編制財務報告，且編制過程漫長，一般約三至四個月才能完成編制，影響了會計信息的即時性和利用效率，大數據技術使即時財務報告成為可能。

二、從反映過去轉為預測未來

通過各種技術收集、儲存和傳遞的大數據改變工作的重心，從數據分析和挖掘過程中向企業主管提出預測性的重要趨勢，並為股東和利益方創造新的財富。要做到預測未來需要先完成以下三件工作：

1. 制定數據評估方法和服務。
2. 利用大數據提供更具針對性

的決策。

3. 評估生產經營活動中所面臨的短期和長期風險和規避。

三、向綜合管理理念發展

財務管理將不再局限於傳統的財務領域，而向銷售、研發、人力資源等多個領域延伸和滲透，成為財務管理的主要定位和主導任務。

四、從抽樣朝總體會計審核模式發展

由於抽取樣本的有限性，忽視了大量的業務活動，無法完全發現和揭示被審計單位的重大舞弊行為。不用隨機抽樣方法，而採用搜集和分析被審計單位所有數據的總體審計模式。

五、向綜合會計審核成果應用發展

審計人員的審計成果主要是提供給被審計單位的審計報告，其包含的信息較少。大數據技術在審計中廣泛應用，審計人員的審計成果除了審計報告外，過程中採集、挖掘、

分析和處理的大量的資料和數據，可以提供給被審計單位用於改進經營管理，促進審計成果的綜合應用，提高綜合審計成果的應用效果。

六、朝高效的數據會計審核發展

相比依賴於小數據和精確性的時代，大數據因為更強調數據的完整性和混雜性，在實施審計時，審計人員應使用分布式拓樸結構、雲數據庫、聯網審計、數據挖掘等新型的技術手段和工具，以提高審計的效率。

肆、統計發展趨勢

大數據與互聯網的發展相輔相成。互聯網資料是大數據中重要的資訊與資源，另一方面，大數據為互聯網的發展提供了更多支撐、服務與應用。大數據是互聯網發展到現今階段的一種表象或特徵，在以雲計算為代表的技術創新襯托下，這些原本很難收集和使用的資料開始變得容易利用。對於政府統計而言，互聯網資料

主要有社群網站資料、媒體資料和搜尋引擎資料三種類型。互聯網大數據在政府統計諸多專業中都具有廣闊的應用前景。如在宏觀層面，互聯網搜索資料能夠為官方統計提供分析、預測與決策支援。

一、經濟發展

傳統官方統計按月度、季度或年度統計各項經濟指標，以GDP、社會消費品零售總額、固定資產投資額、採購經理指數等各項資料來分析經濟發展趨勢；而互聯網企業可以利用大數據來探索和完善各項經濟指標，及時有效地反映國民經濟運行狀況，提高總體經濟監測的全面性和及時性，為總體經濟部門把握經濟發展趨勢、監控企業景氣狀態提供分析、預測與決策支持。

二、價格統計

在CPI統計方面，電子商務交易資料、互聯網企業資料都是價格統計的新資料來源，這些資料量大、更新快，充分利用這些資料有助於減少調查

成本，提高指標發布的頻率。應用大數據進行價格統計的途徑有三種：一是採用搜索方式收集網上交易價格資料；二是與電子商務企業進行合作，獲取交易價格資料；三是建立商場、超市、醫院等實行電子計價的採價點向統計部門報送交易記錄的制度。

三、批發零售業統計

由於網上電商交易資料量非常大、更新速度快，而且在全社會商品零售貿易中所占比重越來越大。因此，充分利用這些資訊可以為改善傳統的批發零售貿易業統計。

四、人口統計

傳統官方統計投入大量人力物力財力，進行人口普查，可獲得資料包括全國和地區人口數量、城市和農村人口數量、人口性別比例、人口地域分布、年齡結構、出生率/死亡率等；而利用互聯網，可以快速及時地統計PC端和移動端使用者，統計維度包括地域、年齡、性別、學歷等，將來還可以根據

Text Mining (文字探勘)、Text Mining_Auto (自動分析) 以及 Facebook (臉書) 四大功能。經由爬網收集到的文檔資料，文章詞彙分析機會自動分析所有文章內有哪些字詞，並且統計有哪些字詞出現的頻率高，頻率高的字詞在文字雲中的字體就會較大，反之頻率低的字詞就會較小(上頁圖3)。可以分別對 FACEBOOK 粉絲團小編的貼文以及每篇文章下粉絲的流言進行分析探討小編們及一般粉絲的想法是否有差異，更深入的瞭解民衆的想法(圖4)。透過此平台可以快速獲取客觀的新聞稿文章以及主觀的民衆發言貼文，整合兩者不同面向的分析更能清楚反應媒體將消息傳給決策者時是否存在偏差，而決策者獲取資訊後有哪些不同的解讀，如此一來在民衆輿情的部份能夠掌握更多的徵兆。

二、跨域資料整合平台

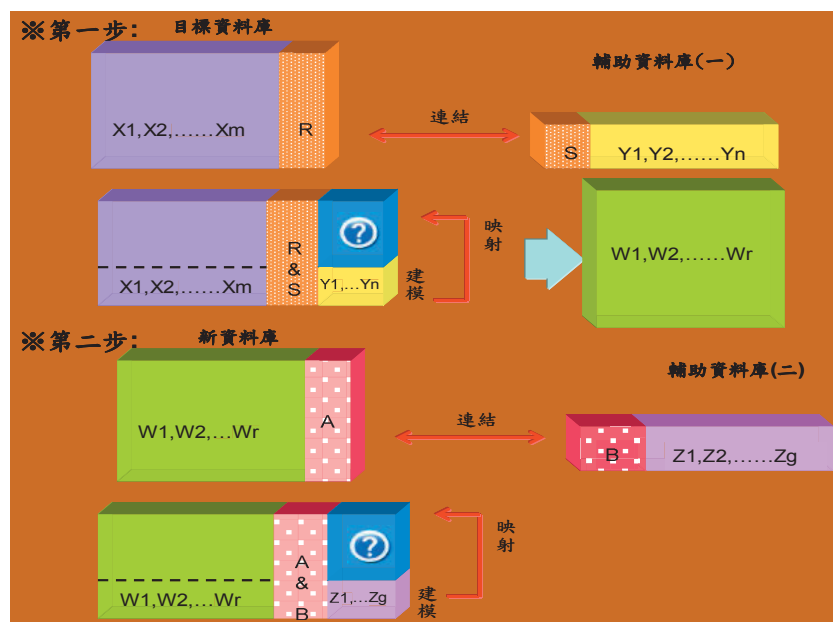
國內有很多調查係由不同部會執行及調查組織進行，想利用這些資料進行研究的使

圖 4 客製化 Text Mining 平台 -Sentiment 分析頁



資料來源：「主計資料大數據之研究」研究計畫分析繪製。

圖 5 跨域整合資料模型



資料來源：「主計資料大數據之研究」研究計畫分析繪製。

論述》專論 · 評述



用者，存在最根本且難以解決的問題：調查結果各自表述與使用，無法進行完整的對應關連，故欠缺整合性的分析功能。為解決上述問題，105 年度研究計畫依循政府進行開放資料（open data）的策略，從大數據思維與角度出發，以資料採礦 CRISP-DM 分析技術為本，嘗試將過往執行的企業面的科技議題相關的調查，如工商服務業普查及產業創新調查，以資料映射（data mapping）方式進行資料庫的串聯進行科技

跨域整合資料模型建立（上頁圖 5），完成創新活動及研發經費之映射與推估研究（圖 6）。科技跨域整合資料庫繼續收集、儲存與管理國內主計總處與科技部各項調查所收集之企業研發與創新資料，未來提供多元資料整合供應與加值應用服務；另外，透過整合模型建立，可提供資料使用者整合數個大小不同資料集，將可創造更多資料價值效益。

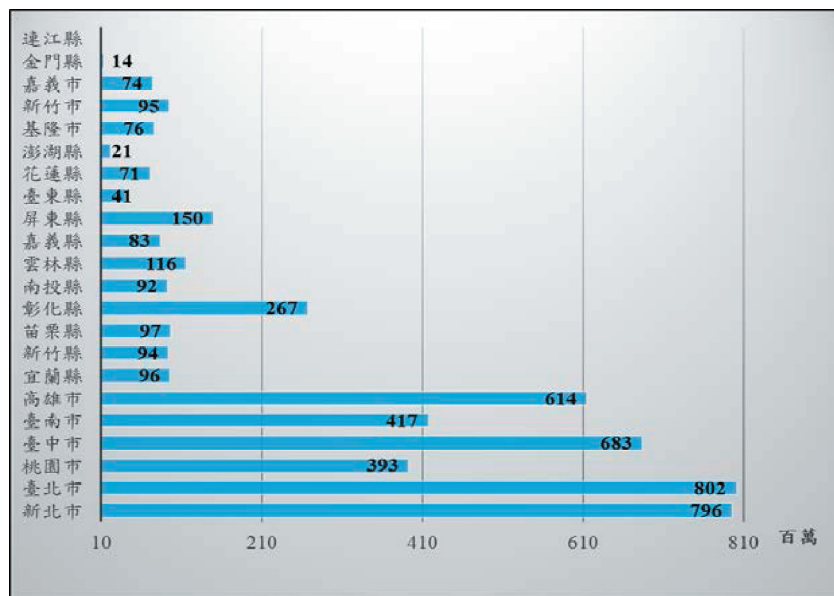
三、歲計估算模型

本次歲計估算模型建立以主計總處政府歲計會計資訊管理系統（GBA）資料為主，並以法務部人事費資料為主要分析項目，因法務部資料種類較多，以法務部作為預測建模資料較具代表性。透過建立歲計預算模型可以即時並且充分反映政府整體的財政狀況，以及減少不必要的資源浪費之情形，除了預測之外，也可以用來監測政府預算，通過建立基本建模的資料庫來分析預算偏離的情況。分析主軸為以下兩點：一是找出預算—決算（借—貸）差異最大的用途別，並且找出其原因；二是結合決算資料及指數平滑法建立預測模型，並進行未來兩年的預算推估，以作為明後年預算編列時之參考依據。

陸、結論

為促進整合歲計、會計、統計資料，運用大數據分析技術，經過主計總處「主計資料大數據分析」研究計畫已實地與主計人員訪談收集意見，也透過相關資料發展 4 大應用

圖 6 全國各縣市之研發經費



資料來源：「主計資料大數據之研究」研究計畫分析繪製。

範例：在歲計估算模型上，由於人事費預算金額與實際員額變動、考績晉級具有高度相關性，為使歲計估算模型更加精準預測，未來建議以主計總處為例，整合主計總處 GBA、人事、銓敘及薪資系統資料，分析瞭解其關係變化，並找出人事費預算經費樣態（pattern），強化模型預測之精準度。

此外，台灣青壯年世代常利用社群媒體工具，建議未來可透過輿情分析運用於預算、統計資料（經濟成長率、失業率、物價等）發布前後輿情比較，觀測搜尋大部分社群平台之輿論動向，掌握新世代民意。另外，大數據分析技術提供收集多樣資料的潛力，遂也帶動證據基礎制定決策理論 (evidence-based decision making) 被廣泛運用至政府部門領域，如運用公務登記與調查資料整合成決策所需資訊，提供政府進行決策。然隨著民眾逐漸重視隱私權保護議題，以及機關職掌本位主義，多會面臨缺乏跨組織單位統籌整合

與隱私權考量困難。未來可建置去識別化且不影響欄位原有特性之智能化及去識別化資料整合模型，將可俾利資料連結分析，提高跨機關分享公務登記與調查資料之意願，進而引用連結外部數據，解決歷年調查瓶頸。

由於「大數據」提高深度學習精準度，演算法及硬體條件躍進提供了機器學習發展的優良條件，再加上數位化聯網的蓬勃下帶來的「大數據」，引爆了科技大廠爭相投入深度學習技術的浪潮。目前不管是 NVIDIA 這類的晶片商，或擅長演算法的 Google、Facebook 等軟體商，最常提到從事的機器學習的主流技術，就是深度學習。要讓機器人照顧老人、汽車自動駕駛，面對陌生的環境路線、隨時有危險狀況，機器怎麼判斷？透過深度學習，需要提供機器海量的數據資訊，包含生活習慣、路標、號誌等，讓它學習並順利執行任務。數據越豐富完整，機器就越能提高辨識的精準度，利用數據堆疊出 AI 人工智慧，將

會讓未來人類生活大躍進，未來主計數據分析應用也可朝此方向邁進。

參考文獻

1. 李欣宜，2015，一次搞懂大數據（上），數位時代。
2. Emily Badger, Dec 19, 2012, The Best Open Data Releases of 2012, From The Atlantic CityLab.
3. Robert Lee Hotz, Feb. 5, 2015, Big Data and Bacteria: Mapping the New York Subway's DNA, THE WALL STREET JOURNAL.
4. <https://wechatinchina.com/thread-151870-1-1.html>
5. 謝邦昌，江志民 (2016)，「主計資料大數據之研究」研究計畫。

