

工研院巨量資訊科技中心主任 余孝先

資料分析品質，決定企業成敗

不論是大企業還是中小企業，都可以用資料分析來做決策。但最重要的是，要先想好這個分析對企業的重要性，以及資料的正確性、完整性及可用性；同時應擁有軟體、數學及應用領域的專業人才，才能將資料分析的價值極大化。

近年，大數據（Big Data）已成為各產業的新顯學，更被視為與對手競爭的殺手級應用。資料分析的目的，就是用資料分析的結果來做決策，不管是個人、公司、團體或政府單位，每天都在做決定，有些時候會用資料分析的結果，有時會用學理或經驗判斷，而多數時候是這些的綜合；這一大數據分析之所以會受到重視，主要是因為隨著科技的發展，使得運用資料分析做出決策的效果跟準度都有很大的提升。

以傳統柑仔店與便利商店為例，柑仔店老闆可能經營了幾十年，他最大的本事就是會隨著季節變換以及店家附近有沒有舉辦活動，來決定在有限的店面空間陳列哪些商品，靠的是經驗累積；但現在很多連鎖店，許多店長都是剛退伍的年輕人，沒什麼社會經驗，他們怎麼決定在有限的貨架上擺放哪些商品才能賣得最好，靠的就是後端的資訊系統，例如在全省有 5000 多家門市的 7-Eleven，透過銷售時點情報系統（Point of Sale, POS）掌握最新銷售狀況，並決定未來幾周要陳列哪些商品、以及哪些商品要下架。這就是資料分析跟以往的經驗判斷兩者的差異，善用資料分析，就可以容許沒有經驗的人做的決策甚至不比有經驗的人差。

人為參與程度越少，分析價值越高

資料分析不是一個新生事物，過去就有所謂的資料探勘（Data Mining）或機器學習（Machine Learning），只是巨量 Big Data 資料本質有很大的變異，使得這項新科技的出現，可以解決很多以前不能解決的問題，甚至也解決了



很多新生的問題。因此，如果還是用舊的思維、方法，（資料探礦）挖的就淺，用新的工具就能挖的深、挖的快。

一般來說，資料分析可以分為幾個層次，最淺的是描述性的（Descriptive），例如統計過去幾年藍綠兩黨的得票率，只要有計算器就可以做；其次是診斷性的（Diagnostic），也就是把資料統計出來，然後由人去解讀，但每個人解讀的可能不一樣；再往上一層是預測性的（Predictive），是否能利用過去的資料，預測下一次的得票率，這就很難。

以美國線上影音公司 Netflix 來說，想做的就是預測用戶打開電視想要看什麼，同樣是 A 客戶，周一跟周三、周末，想看的可能不一樣，是早上打開電視還是晚上打開電視，可能也不一樣，所以 Netflix 想要預測用戶的行為，就要直接自動化到預測這一層，光是做到統計（最淺的那一層）是沒有用的。同樣是資料分析，做的越深、人的參與程度越少的（自動化程度越高），價值越高；人的參與程度越多的，價值越低。

再往上是處方性（Prescriptive），就更困難了，比如做完預測，發生方案 A 的機率是 70%，方案 B 是 20%，到底要挑哪個方案？因為每個方案的價值、損失、風險都不一樣，很多時候是停在預測，然後交給人去判斷，但如果能做到由機器直接幫你判斷，自動化程度就更高。

大數據的核心：資料品質

不論是大企業還是中小企業，都可以用資料分析來做決策，但最重要的是，企業要先想好這個分析對你是不是很重要，比如可以幫公司增加 1000 萬的收入，投資上限就是 1000 萬。其次，

你想發現的事情，最好不是你的經驗值或學理就可以做到的，如果已經是這樣，那按照原來的做法也可以，資料分析的價值就沒有那麼高；或是當一個企業的資料量大到無法用人工可以處理，這時 Big Data 是最適合的。

舉例來說，Netflix 在 2013 年時有 3000 萬家庭用戶，員工有 1500 人，如果要做到個人化推薦，平均每位員工要處理兩萬個用戶過去的收視行為、居住的地區、社經地位、年齡、性別等資料，然後給予適當的推薦，光是靠人力幾乎不可能完成，但在技術上是可行的。資料分析厲害之處就在於整件事透過機器就可以完成，只要程式設計好、分析工具做好，資料倒進系統後，

3000 萬戶的分析結果就出來了，而且每天都可以做。

但企業要應用 Big Data，一定要先把資料

整理好，E 化是第一步；E 化之後，還要檢查資料的正確性及完整性，然後對資料有足夠的瞭解，以及盤點目前的資料足不足以支撐你想要做的事情，如果不夠，還要重新累積，甚至有些資料會過期，資料價值已經不存在，比如道路的交通流量變化，會隨著新增道路或紅綠燈號的改變而不同。而資料是結構化、非結構化，資料形態的問題；有沒有法令的障礙，有沒有隱私權、個資法的問題，也都必須予以考量。

最後是系統跟人才，一個理想的 Big Data 團隊需要 3 種人才，一類是軟體人才，學電腦科學的；另外一類是數學，尤其是統計學的；第三類就是產業的應用領域的專業人才，比如電子商務的，就需要有些人懂電子商務，如果是半導體工廠，就需要有人懂半導體。前兩類人才可運用外部資源協助，唯有這三類人才整合在一起，Big Data 才可能成功。

資料分析價值

資料分析做的越深、人的參與程度越少（自動化程度越高），價值就越高；而人的參與程度越多，價值就越低。