



攝影：陳志偉

《專家觀點》陳昇瑋的12年自我修鍊之路

撰文·吳沛璇

台灣資料科學教父： 好奇心比程式能力重要

如果要找功力最高強的大數據人才，恐怕不在業界，而是在學術界。中研院資訊科學研究所研究員陳昇瑋，是國內玩過資料種類最多的人。目前他手上超過十個專案與業界、政府合作。「我想與一家麵包店合作，只要兩個多月就可以告訴麵包店老闆，麵包如何配料，如何取名會賣得更好。」陳昇瑋腦中就像魔術師的那頂神奇高帽，隨時變出花樣來。

**求廠商「施捨」資料
從中發現「沒人看出來」的祕密**

戴著細框眼鏡以及一臉自信的笑容，陳昇瑋一談到「資料」就滔滔不絕。在今年夏末舉辦的台灣資料科學愛好者年會上，他可以從早上九點聊到下午四點，讓台下聽眾無不凝神聆聽，信手拈來好幾個案例，靠的全是十多年來扎下的基本功。

二〇〇三年，就讀台大電機系博士

仔細推敲，一、二月是領年終獎金的時候，大家希望過年能多做點善事，剛好也有點閒錢，所以出手特別闊；而七、八月是旅遊旺季，旅行花費已讓手頭吃緊，自然無暇再顧及他人。

更出乎意料的是，就連一星期內的不同日子，也會影響捐款的金額。整體來說，周二、周三收到的捐款最多，周末收到的捐款則最少。陳昇璋分析，周二、周三捐款額最高，可能是因為大家上班很辛苦，所以也對弱勢族群有更多的同理心；但是到了周末，大家忙著去玩，可能對這些議題沒那麼關注。

此外，只要《蘋果日報》的標題含「老弱」、「單身」、「婦女」等字眼，收到的捐款就比較多；但是，若將體型分成七分，受款人（身材）每「胖」一分，平均捐款人數就少二十個。這樣的研究結果，擁有十幾年採訪經驗的記者或許都不曾發現，透過資料科學，只要短短一、兩個星期，就能挖掘出經驗也無法比擬的洞見。

台灣輕忽資料科學 態度過於保守 落後美國二十年

然而，台灣廠商至今仍不知道資料的力量。「十幾年來我接觸了至少五十家公司，但大部分的人都不理我，因為台灣從來就不把資料放進 Big Data（優先順位）。」陳昇璋感嘆，知名電商亞馬遜

好的資料科學家 必備的四種能力

陳昇璋認為，一位好的資料科學家具備四種能力，但四種能力都有的人就是獨角獸（指不存在）。想進入這個熱門行業，要在程式語言、數學統計、領域知識及溝通能力，起碼要具備其中兩項，之後努力統整出三種，就可以成為企業間炙手可熱的人才。

就算具有領域知識，也懂得使用統計工具，若沒有好的溝通能力，將難以說服老闆，好的計畫也可能因此胎死腹中。因此，陳昇璋認為，好的溝通能力是資料科學家亟需的個人特質。

寫程式的人如何具備領域知識？陳昇璋引述阿里巴巴數據委員長車品覺的說法，「就是『混』出商業敏感度。」車品覺認為，每周起碼要與第一線做業務的人吃飯，還有周工作報告內容，不能遺漏對接商業部門的主要專案。

另外擁有金融、零售領域知識的人，要如何接近資料分析領域？陳昇璋建議，至少要熟悉基本的統計工具，譬如R語言（主要用於統計分析、繪圖、資料探勘的程式語言）。

從一九九五年就開始使用推薦系統，台灣電商卻只注重物流速度，「真的是落後別人二十年！」句句道出台灣企業的保守心態。

創辦相關論壇 金融、零售業高階主管都來參加

為了不落人後，去年起，陳昇璋開始策畫「台灣資料科學愛好者年會」，希望能提升業界對資料科學的重視。第一屆年會，吸引了八百多人參加，其中更不乏金融業、零售業的高階主管，都來一睹資料的龐大威力。

除了舉辦活動，陳昇璋還收了不

少弟子，甚至有知名企業送員工來「練功」。跟了他四年的博士後研究員許懷中憶起，當時他會應徵，就是看到陳昇璋網頁上，一張鮮奶油蛋糕與貓的照片，底下寫著短短一句：你也想一起來體會貓得到奶油的感覺嗎？許懷中說：「老師對我的啟發不只是 Believe in Data（相信資料），還有培養你對探索世界的胃口、對精進自己的一種渴望。」

不僅僅只是一位出色的資料科學家，陳昇璋還要再培養一批批優秀的資料科學人才。因為，他視資料科學為一種信仰，要將它發揚光大。



他判讀資料解決難題 準確度逾9成

——陳昇璋近年研究過的案例

狀況	從資料發現……	從資料判讀……
這支電話是詐騙或行銷電話嗎？	詐騙、行銷電話都是單方打出，且每天通數較多、通話時間短。另，只在周間上班時間撥出。	只要5個通聯紀錄，就能抓出詐騙/行銷電話， 準確率超過95%
有沒有人偷登入我的臉書？	陌生人偷用你臉書時，會花較多時間觀看你的照片跟個人資料，但幾乎不會留言或按讚。	只要被偷看超過10分鐘，就會被抓到， 準確率超過90%
這個玩家是不是開外掛？	外掛每次上線時間遠超過真人，對話頻率很規律，且與真人玩家行走路徑不同。	只要觀察400秒，就能抓出外掛， 準確率高達95%

整理：吳沛璇

班的陳昇璋，在思考論文題目時發現，相較於寫出一套完美的理論模型，他更喜歡研究貼近現實的資料。他認為：「模型雖然漂亮，但你一定沒辦法完整描述這個世界；相較之下，data driven（沒有先入為主的直覺，而是純用數據佐證）比較 nice（不好看），但是它是從現實資料來的。」

但是，要取得業界的資料談何容易？在台灣，除了業界、學界鮮少合作外，業界都對自己的資料保密到家。更何況，當時「資料」一詞，不像現在有大數據的加持，許多廠商雖空有資料，卻不清楚重要性。「所以，那時候很辛苦，都要求人『施捨』資料給我，就像窮人借錢，只能一個一個朋友問。」陳昇璋無奈地表示。

把所有朋友都問過一輪後，終於靠朋友牽線，認識了宇峻奧汀遊戲公司（旗下知名遊戲包含：〈絕代雙驕〉、〈神舟 online〉）的員



Profile 陳昇璋

出生：1976年
現職：中研院資訊科學研究所研究員、悠遊卡公司獨立董事
經歷：Gogolook顧問
學歷：台大電機所博士、清大資工所碩士

工，願意提供資料給他做博士論文。經過分析，陳昇璋發現這款線上遊戲，在伺服器設定上出了問題，導致原先應該輪流輸送的流量，變成同時輸送，反而讓總流量變少。在那之前，遊戲公司雖發現網路不穩，卻一直認為是品管出了問題。

「直到那時，我才知道資料的威力，那個遊戲我只玩過一小時，卻比日夜泡在遊戲裡的開發人員還懂。」陳昇璋補充：「透過資料，可以發現『大家都在看，卻沒有人看到』的東西。」

在不疑慮有疑 分析資料 找出捐款金額落差原因

但可別認為，要成為一位資料科學家，就要做一堆阿宅才會想做的事。其實，比起程式能力，對周遭事物產生好奇心，才是資料科學家最須具備的能力。自〇三年起，《蘋果日報》成立慈善基金會，替台灣貧困弱勢族群募款，至今已十二個年頭，卻從沒人分析過，為什麼每個個案的捐款金額差異這麼大？

直到有一天，陳昇璋無意間找到了這個公開的資料庫，一經分析後，發現不少有趣的現象，更清楚描繪了台灣人民捐款的動機。在一整年裡，一、二月募到的金額特別高，而後慢慢遞減，直到七、八月跌至谷底後，又漸漸回升。