

大數據

穿梭網內外 點出來

蒐集瀏覽紀錄 深諳人類喜惡

編者按：古人認為，萬物一體，任何事物都息息相關。來到今時今日的數碼電子年代，透過大數據 (Big Data) 的挖掘與發現，人類可以把情緒與股市、啤酒與尿布、天氣與保險、點擊與喜惡，透過每一則訊息紀錄連繫起來，並產生有價值的導向，真正地從現實意義中，證明萬物如何息息相關。本報從今日起會透過一系列報道，由各範疇的數據專家，向讀者展示如何透過挖掘大數據，在教育、交通、金融和醫療方面提升人類的生活質素，甚至改變社會。

香港文匯報訊 (記者 歐陽文倩) 我們身處大數據時代，在網絡世界按一個點擊，經統計分析，就可以成為精準地投放廣告資源的依據；在現實世界留下的每項紀錄，如到超市購物都成為商戶營銷參考；甚至連內心情緒也成為股市走勢指標。箇中奧秘何在？多位大數據專家將為大家解說。

到底大數據從何而來？香港科技大學霍英東研究院院長倪明選表示，以網絡上收集的大數據而言，就是從每個滑鼠點擊而來，無論收發電郵、瀏覽網頁、上傳文件或圖片，都成為一項有潛在價值的數據。能收集得有多細緻呢？倪明選坦言：「比你想像的更細，不止是電郵的內容，還有你在頁面停留的時間。」

他進一步解釋：「Google 就是用廣告去賺錢。可是大家都做廣告，為何要跟 Google 做？因為它的針對性特別好，有很多服務給你，收集你很多數據，就用大量數據去分析你的行為和喜好，針對性地做廣告，效益和準確度都特別高。」這亦解釋為甚麼網站上「你可能感興趣」的建議，都能有極高中率。

尿布啤酒放在一起 兩者銷量增

單純的數據分析，可精準反映人的愛惡，但倪明選指，有關分析是「知其然而不知其所以然」，「業界就有個經典故事：沃爾瑪在十幾年前做了分析，發現買尿布的人很有可能買啤酒，他們不知道為甚麼，但簡單地做了決策，把啤酒和尿布擺在一起，這就增加兩者的銷量。」

心情好股票漲 看透金融市場

另一個實例就是利用大數據去預測連金融從業員都看不透的股票市場。科大霍英東研究院研究助理教授譚浩宇表示，美國有一間基金公司將推特 (Twitter) 用戶發表的內容加以收集，並分析當中情緒，了解當中的「情緒總和」是傾向正面或負面，「結果他們發現，如果是正面的，也許20分鐘後股價會上漲。於是他們做了計算模型，開了基金公司，用這指標進行買賣，只開了一個月，收益就超越大多數基金公司、跑贏標普500指數，賺了2.2%。為何會這樣？其實大家不在意為甚麼，只在意是否能起作用，是否有關聯。」

而大數據的影響力更在於數據挖掘所產生的連鎖效應，譚浩宇繼續講解上述預測股市的例子，「那基金公司之後就關閉掉了，因為他們發現更好的賺錢辦法，就是不用自己交易，把數據產品賣給日間交易者 (day trader)，這樣自己不用承擔風險，還一定能賺錢。」

如標清變高清 發展一日千里

香港理工大學互聯網及移動計算實驗室主任曹建農就大數據帶來的改變打了個比喻：「大數據的好處是，可幫助人觀察得細一點。簡單而言，就像由普通電視變成高清電視，大家看到更多東西，也就多了東西可以講。」

他坦言，在現今技術配合下，大數據發展可謂一日千里：「現在從各個方面收集到的數據愈來愈多，可能今天一天的數據，就可能是比前一年的數據更多。比如說5年前很少人自拍，因為沒應用，也沒技術，現時大家都拍完馬上放上社交網站。」

他進一步指，未來的數據會愈來愈多，「網上的流量多數都是應用的流量，現在手機應用只有3%至5%，到2015年至2020年會增加20%。」他更說，大數據將是以後每所公司都感興趣的議題，「過去是有問題了，再看數據，現在則是從數據裡找機會。」

倪明選(左)及譚浩宇(右)認為，大數據分析能幫助人們作出決策。
香港文匯報記者 彭子文攝

應用大小通吃 輔證上帝粒子

大有所為

大數據的應用領域，雖然現在更多時用於網上營銷，或超市用來研究如何吸引顧客，但其實其應用範疇絕不限於此，它能把更多看似風馬牛不相及的東西聯繫起來，並改善交通、提升空氣質素，規模小至學校的學術研究、大至國家的國防都有作用。

曹建農表示，大數據在其他方面的應用，如智能交通、環保、污染監控、天氣等都

大有所為，並能從數據分析得出更客觀的結論，「例如過去做交通，看的是車流量、坐甚麼車，但如果把天氣加起看，又會有不同的現象。就好像說下雨天，地鐵多很多人，其他交通工具少一點。又比如說出租車的用車情況，這都要靠數據去做可靠分析。例如雷暴雨時，大家都覺得很難打車，以為這是因為大家都在打車，所以一輛空的士變得難求，但其實可能是因為危險，少司機出車也說不定。」

擴科研領域 促藥物開發

倪明選亦從不同規模上講解大數據的應用。他表示，從院校研究的層面，學校做了很多大數據相關的應用，如深度學習；去年證明上帝粒子存在，也是透過大數據做的，「所以大數據能擴大科研領域。」

譚浩宇亦表示，現時生物系、物理系都要學用大數據，「例如做不同藥物和蛋白質的互動關係，科學家想去看某些藥跟其他蛋白質是否有互動，有的話就可以用來治病，可是當中的檢測是十分昂貴的，如果有1,000種藥，10,000種蛋白質，前後可能要1,000萬次才可測出來。然而，我們可透過大數據，根據現有資料去找出更有可能互動的藥物與蛋白質拿去做實驗，如果是的話，就可開發藥物。」

追蹤恐怖分子 找一揪餘

在城市層面上，大數據所衍生出的「數據產品」，能讓中小企更掌握市場資訊，提升其競爭力；在國家安全層面，大數據更可用來反恐，靠的就是分析社交網絡上的數據。倪明選表示，在這個萬物互聯的「互聯網」上，只要能找出一個恐怖分子，就可找出其他，「這都是現時國家需要的技術。」

香港文匯報記者 歐陽文倩

公諸於世

大數據影響力既廣且大，倪明選和譚浩宇均認為政府應建立自己的「數據中心」，公開所擁有的數據，讓整個社會一起從中「挖掘財富」。他們認為，數據中心是未來城市的基建，就像高速公路一樣是城市的「必需品」。譚浩宇更一針見血地表示，「數據不公開，就是成本，因為你要維護；數據公開了，就是資產，有可能產生價值。」

對於Google放棄來港設立數據中心，倪明選稱羨，因為這樣重要的設施，不應假他人之手來代勞，「香港政府應該做數據中心，對市民來講，知道不同的數據資料，就像知道各區PM2.5懸浮粒子的濃度一樣。」像天氣、交通、房地產等資料，倪明選認為都可公開，他以交通數據為例：「像公車數據、有無誤點、幾點幾分到哪個站、地鐵有多少人、每個站多少人出入等，有了這些資料，就可以做深入分析，可是現在都沒有。」

倪明選最想做「停車」App

倪明選笑言，自己最希望看見的大數據應用之一是「停車資料庫」，「現在停車很麻煩，我繞來繞去，一直找不到車位。如果有這個數據，可以分析到每天多少車在停，賺多少錢，這就能分析出來如何分配停車位。」再加上手機應用程式的配合，就可看到哪裡有停車位分配給自己，並用手機付賬。

存儲平台是基建 必不可少

除資料公開外，倪明選認為提供數據存儲平台亦十分重要，「這是基建之一，就像高速公路一樣。」譚浩宇進一步補充說：「當你提供了這樣的基建，未來大家可利用它產生更多價值。美國有個城市，就把它的城市規劃、下水道、電網鋪設等資料公開，很多人可以看這個數據，發揮他們的創造力，去讓城市更好。」就像美國做全球定位系統 (GPS)、中國的「北斗」衛星導航一樣，所產生的價值遠超原本所想。他們更指，中小企亦可靠此發財，「如果能給人家機會，香港那麼多博士生，他們各自去開公司，用數據分析去做判斷，這更是在幫香港賺錢。」

香港文匯報記者 歐陽文倩



曹建農表示，大數據在智能交通、環保、污染監控、天氣等範疇上都有大有所為。
香港文匯報記者 莫雲芝攝

學者倡設政府數據庫「挖富」