

全球化+公共衛生

大數據時代：誰「算計」了未來？

一步一腳印 購物皆留痕

在以前，由於記錄數據的方法較少及參考價值較低，因此被重視程度較低。以消費模式為例，以往我們到店舖購物，留下的數據可能只有購物的貨品會被記錄，但由於現時電子器材先進，連我們購物的時間都會被記錄，甚至額外的個人資料如消費者的年齡層、收入水平都可能被會員制度等方法被記錄下來。由此可見，隨着科技的進步，記錄我們行為活動為化數據的方法亦隨之增加，造成如今生產數據的速度及數量達至驚人數字。

會員吸客蒐數據

近年，其實企業都紛紛各出奇謀蒐集數據。現時不論網上電子商店或實體零售商店都紛紛提出會員制度，並提供不少優惠吸引客戶成為會員，當中一個重要原因便是為了蒐集數據。在加入成為會員後，商店便可掌握客人更多背景資料。再者，智能手機上的應用程式都是蒐集數據的主要方式，若細看一些程式的安裝說明，便可見不少程式會擷取大量儲存在手機的個人資料，例如手機的全球定位等。企業在分析數據時一般為了研究市場及發掘新商機。

分析銷情助營銷

數據比對助社群

數據對於企業而言，是創造機會的鑰匙。企業可透過大數據分析哪一款產品銷情最佳、哪一區域營業額最高和歸納客人相同的特質，從而在整個營銷策略加以配合，甚至即時掌握到市場動態並迅速作出應對，令企業更有個性。

提高效率減失救

在9月初，全球最大電腦公司微軟舉辦了有關大數據與社福界的研討會，席中帶出了大數據與社福服務的關係。以醫療為例，東非的坦桑尼亞聯合共和國便曾經有成功的例子。坦桑尼亞一直受瘧疾困擾，據統計，當地每5分鐘便有1個小孩死於瘧疾。惟瘧疾是可治療的，只是藥物的補給速度未能及時送達病患者手中，引致大量孩童因藥物不足而失救致死。直至2012年，當地政府衛生部門決定和電訊公司、科技公司及社福機構合作，並把病患的數目和身處地點輸入到資料庫中，並於已輸入的藥物庫存地點作比對，從而安排最有效率的方法送達病患手中，這方法使得因失救而死亡的比率大幅減低至30%以下。

由此可見，數據化並套用在公共政策事務上，可大大提高社會資源運用效率，使社會上弱勢社群受到更有效的幫助。再者，把社會各項資料數據化有着非常大的可塑性，例如可把長者的居住樓宇類型與健康做比對，或就着長者人數較多或未來即將老化的社區，對地區的醫療資源進行分配。另一方面，可收集新婚夫婦或新生嬰兒的居住地方，從而規劃校網及相對青年配套。人口數據的運用絕對不限於此，但若想從各方數據中找到更多組合解決社會難題，政府必須牽頭把各項資料數據化，才可進一步分析，協助制訂相關規劃及政策。

日新月異

2014年是中國接軌國際互聯網20周年，回顧過去，我們在科技及網絡上的發展改變了我們的生活。如今，我們已進入「大數據」的時代。我們的生活幾乎每分每秒都在創造新的數據並記錄在不同的平台上，例如我們每天的行蹤、不同地方的逗留時間、甚至我們每天花費在網絡上的時間等。在這些海量數據充斥的「大數據」時代下，到底可循甚麼方式改善我們的生活？本文將一一探討。

作者簡介：香港青年時事評論員協會會員。文章散見於《巴士的報》、《星島日報》等

「數據挖掘」藏商機

數據的生產方式已多得難以盡錄，但面對着無數離散的數據，往後的整合、統計及分析方法才是重點。單獨的數據本身是沒有意義的，重點是要在海量的數據中抽出相關的數據進行配對。以上述例子而言，把消費者購買的產品與其收入水平作配對並分析，便可得知產品的定位是否與消費者的收入成正比或反比的關係，這種從大量的資料中搜尋它們隱藏於其中的特殊關聯性的方法稱為「數據挖掘」。「數據挖掘」有兩個重要的功能：

- 1.要在資料中提取出隱含在過去而未知的但卻有豐富價值的潛在資訊；
- 2.它是應用是在海量資料中以科學方法中提取資訊的方法。

「數據挖掘」通常是帶有目的或前設，在挖掘的過程當中，必定會產生很多不同的配對結果，這些都會與前設的目的不同作結果驗證，直至得出最科學、客觀的結果。

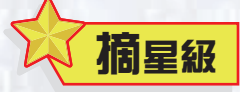


數據面前，私隱或無所遁形？

即使大數據無論在企業或公共事務都能推動社會發展，但數據往往與個人私隱扯上關係，一些有價值的數據必定是來自個人私隱的，例如企業在研究消費者時，對象的收入及資產狀況便屬個人私隱，而且現時科技發達，程式開發者能輕易編寫擷取資料程式，只要用戶下載至手機或電腦中，便可不經同意便進行收集。因此，管制收集數據方式的法例必須與時並進，防止不法分子為一己私慾而非法收集數據。

安全保護 潛在風險

數據屬於電子資料，其保護方法相對複雜，但若機構收集到數據後，安全方面未能提供足夠程度，數據亦會面臨被非法擷取的風險。數據的保護方式不外乎加密及隔離，分別可透過程式加密以及隔離整個數據庫，杜絕其他裝置接觸的可能性，惟各種保護方法都有其缺點，不能完全解決數據安全的風險。因此，數據安全的風險必須考慮在內。



大數據 (Big Data)：以海量計的資料，它的來源多元，可能直接或間接地來自我們的日常生活。大數據有兩種主要特質，分別是巨量以及雜亂無章，無法使用傳統的資料數據庫管理的方式作有系統的管理，即使如此，這些數據卻是難以想像的珍貴。數據寶貴之處是在於它代表某方面的事實或數字，若決策者，無論是個人、企業或政府能蒐集到一批數據，有效地加以分析及研究，對於事物的決策有前瞻性的影響，小至個人的投資或就業方向，大至國家政府制訂政策，都有其價值。

概念鏈接

小知識

何謂大智移雲？

大智移雲是分拆開4個組成部分。「大」是代表大數據，我們的日常生活中都會產生林林總總的數據，當這些數據聚集一起，便會形成一個巨型資料庫，稱為大數據；「智」是指智能裝置，現時最普及的智能裝置是智能電話，但同時亦有不少技術正在智能化，例如立體打印器、虛擬實景放映器等；「移」是指移動裝置，除了隨身電話外，有越來越多智能裝置變成隨身攜帶，例如智能手錶及眼鏡等；「雲」是指雲端運算，雲端運算是把軟件運作或數據儲存等運作在第三方提供的伺服器上，例如現時網絡上有不少免費儲存空間供我們使用，把資料備份在雲端上。



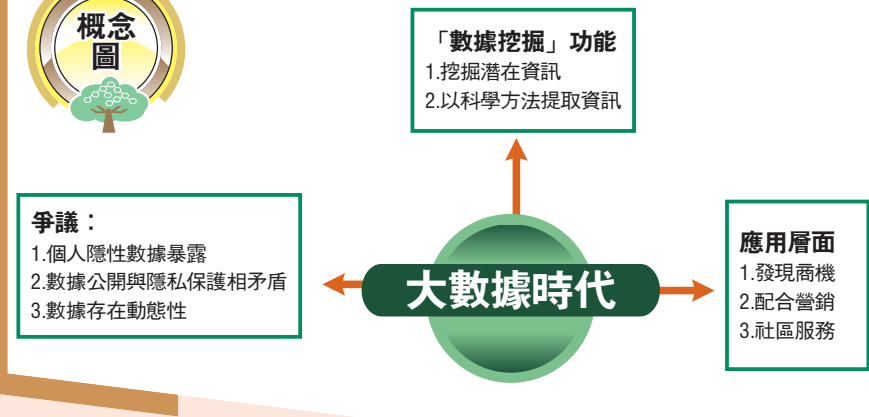
想一想

- 1.根據上文，指出何謂大數據。
- 2.參考上文並就你所知，指出「數據挖掘」的兩個重要功能。
- 3.承上題，指出大數據可應用於哪些方面？請舉例說明。
- 4.你認為「大數據」時代將面臨甚麼挑戰？解釋你的答案。
- 5.試從全球化的角度，指出「大數據」分析對全球經濟有何影響？

答題指引

1. 可指出大數據指在合理時間內達到擷取、管理、處理、並整理成為人類所能解讀的資訊。
2. 一為挖掘潛在資訊；二為以科學方法提取資訊。
3. 同學在回答時，可配合事例加以支持論點，以加強說服力，如在公共事業上，善用數據可助資源分配，這可在社區配套、建立房屋的方面體現。
4. 可從個人隱性數據暴露、數據公開與隱私保護相矛盾，及數據存在動態性等這些方面討論。
5. 可通過對數據的合理共享和利用能帶動全球產業發展，影響研發、生產、流通和社會管理這一方面入手，並舉例加以說明。

概念圖



製圖：香港文匯報記者 李慧

延伸閱讀

1. 《大數據輪轉市上下計出來》，香港文匯報，2014年11月27日，<http://paper.wenweipo.com/2014/11/27/ED1411270019.htm>
2. 《大數據穿梭網內外點出來》，香港文匯報，2014年11月24日，<http://paper.wenweipo.com/2014/11/24/NS1411240002.htm>
3. (英)維克托·邁爾·舍恩伯格著，周濤譯：《大數據時代——生活、工作與思維的大變革》，浙江人民出版社，2012。