

數據、謊言與真相讀後感

一、前言

「大數據」這三個字，是目前網路最夯的字，各大網路入口網及收尋引擎的關鍵字，都遍及於我們生活中的許多角落，大數據是一種科技演進下的產物，近年來因為資訊科技越來越複雜，人們已經無法負荷如此龐大的資料量，所以有人問「大數據時代」真的來臨了嗎？然而，什麼叫做「大數據」，「大數據」又有什麼作用呢？「大數據」在今天這個資訊十倍數高數成長的時代，真的已深深改變了我們的世界，大至國家安全、科學研究、總統大選的選前選情預測，小至個人的日常生活消費習慣，都無所不包無所不在，尤其現代人不分男女老少、幾乎人手一機，每天從早到晚都有大量的訊息在網路上在行動電話中大量流傳，現在人的溝通方式已從以前必須面對面從對方口中獲得的訊息，這種溝通方式因為可以看到對方的臉跟動作，所以我們很容易判斷出他所提供出來的訊息的真偽度，而今天溝通方式藉著強勢大量的數位方式傳播，許多資訊可能都是經過加工的不實資料，而我們都很難判別它的真假，所以生活在資訊發達時代的我們，對大數據對我們所造成的影響應該要特別關注，才不會生活在資訊的迷宮中迷失了自己。

這本書的作者為賽斯·史蒂芬斯-大衛德維茲，他重新解讀「大數據」對人們帶來的影響，以有趣的方式來看待目前網路的現況及實際人們最真實的想法，透露人們內心真正的想法與期望，為何會這麼說呢？因為人們可能會對朋友、家人甚至是醫生說謊，但人們對自己上網收尋打的關鍵字” keyword” 以及提出疑問？渴望尋求解答，常常有許多難以啟齒的問題因為害羞不敢直接透露，甚至是面對親密愛人，有時候也是會有語帶保留的時候，不會全部一五一十地說出自己真正要表達的想法。2008年9月14日，雷曼兄弟在美國聯準會拒絕提供資金支持援助後提出破產申請，同一天美林證券宣布被美國銀行收購。在這兩件事爆發後，引發投資人信心恐慌，接下來這一星期 2008年9月全球股市大崩盤的序幕..由於對大環境不確定因素還有投資人恐慌指數上升，全球股市跳水式崩盤。儼然成為窺探內心的新窗口，在數據「爆炸」的現在，大數據常常被寄予厚望。大數據時代，每個人都會「自發地」在網路上提供數據。我們的各種行為，如查詢瀏覽網頁、刷卡消費、觀看網路電視、搭乘大眾運輸工具、駕駛汽車、機車，都會生成數據及軌跡，人們的行為會一筆一筆老老實實地被記錄下來，近些年來宅經濟商機越發成長，很多電商在搶購眼球商機，消費者從網路點選商品、購物等信息，都會被從大數據中挖掘出來，以利分析商機。

二、本書內容與概要分析：

這本書共分為八個章節，分別是美國總統川普的勝選是讓你跌破眼鏡還是實至名歸？-歐巴馬的勝選是不是代表著所有的種族歧視已經消失？管他大數據還是小數據-別讓直覺扯你的後腿。大數據背後藏著龐大驚人的力量，是永遠跟你所想像的不一樣。

書中還提到要怎樣的頭條標題才能真正吸引住人們的眼球，讓閱聽眾能產生繼續往下閱讀的興趣與動力？作者在本書最後的第七、第八章提出：「大數據，大垃圾！愈多數據，愈多問題？」，「數據會說話？」，但我們也不得不承認，數字也會胡說八道，提醒我們不要過分依賴數據，因為數據本身是一個非常籠統的概念，也就意味著不可能精確。作家馬可·吐溫曾經說過這樣的話：「謊言有三種：謊言、該死的謊言與統計數字」。我打個比方，大數據看起來可能是很萬能，但看走眼的機率也是相當的高。資料的發掘與探勘正在流行，但大部分的人們挖到的是金礦，還是垃圾？

分享到這裡又讓我想到另外一本也是有關描述大數據的經典好書-《機率思考—大數據時代下，不犯錯的決斷武器》，這本書是由英國人羅伯·麥斯修（Robert Matthews）所寫的，本書內容取材自日常生活的人事物，以生動活潑的寫法，揭開數據機率千變萬化的樣貌。不管是個人或群體，從大數據時代到之後真相世界，矛盾恐慌層出不

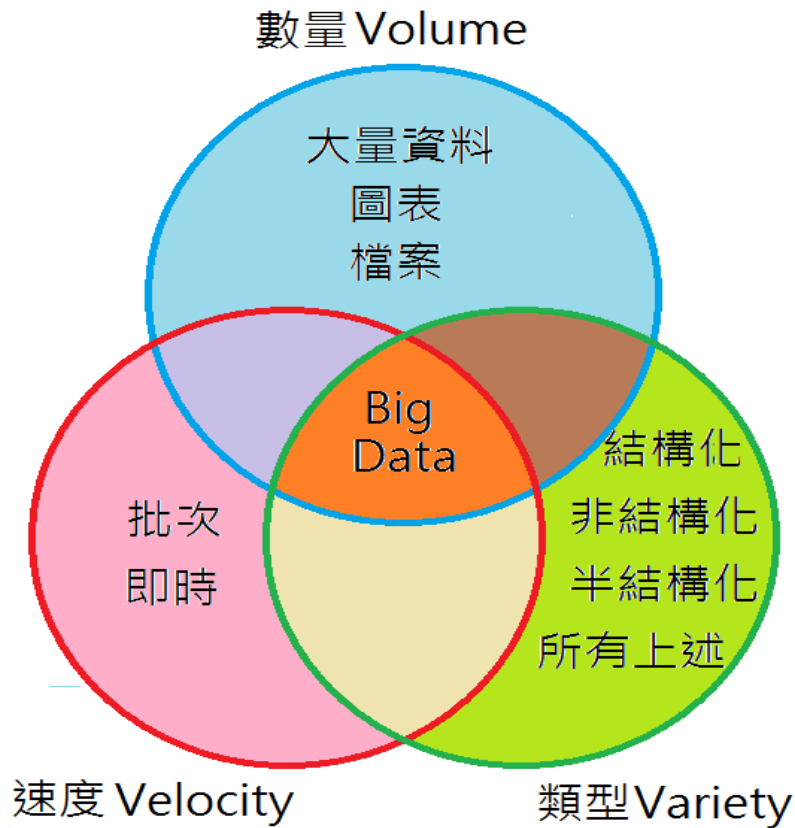
窮，決策風險日益高漲。惟有對機率、風險和不確定性有更深入而正確的理解，才能培養足夠的思辨能力，擺脫理盲目的困境。藉由機率思考，在你我無法捉摸之際（to be or not to be），期待能幫助讀者，擁有決斷的勇氣與勝出的能力。此剛好可以跟賽斯·史蒂芬斯-大衛德維茲寫的這本書內容相互呼應。

「大數據」的定義是甚麼？在資訊爆炸時代，人類一天可以產生巨量的訊息，到底要多「大」才夠「大」？但是，作者也說擁有大量的數據並不會自動產生精闢結論。有時候擁有的可能只是一些垃圾無用的訊息。大數據是應用的能力，在眾多或多樣化資訊中，迅速獲得資訊，且有助於決策、行動或行銷。

此外大數據它是具有預測能力的，當資料量大到接近全部的母體，也就是所被研究對象的總體的總數或信心得以代替母體時，可取代傳統透過抽樣預測母體的方式。

另外也可以從以下「3個V」來探討大數據：- Volume（大量）、Velocity（速度）和 Variety（多樣性）。Volume（大量）：以過去的技术無法管理的資料量，資料量的單位可從 TB（terabyte，一兆位元組）到 PB（petabyte，千兆位元組）；Variety（多樣性）：企業的銷售、庫存資料；網站的使用者動態、客服中心的通話紀錄；社交媒體上的文字影像等企業資料庫難以儲存的「非結構化資料」。

Velocity(速度): 資料每分每秒都在更新, 技術也能做到即時儲存、處理。



圖片來源：維基百科

雖然如此，大數據並非萬能，作者賽斯·史蒂芬斯-大衛德維茲點出了大數據做不到的事情，那就是股市開盤走向，投資客想要從股市獲利，最快的當然就是股市的內部消息，但如果你想要從網路、或是新聞事件來摸索出線索，是無法得到即時且迅速的答案的。另外大數據對於突如其來的意外也無法事先預測，所以看似萬能的大數據，仍有辦不到的事情。

在這一場數位革命中將會如何為每個人打開找尋隱藏真理的大門，一窺人們的內心世界真實的想法，數字會說話，讓提出大數據時代人們真正看到的是什麼。我們的大數據足跡已然將全世界打造成一個世界級地球村的實驗室，這是讓我們在進入網路新時代的這幾十年來在政治、健康、消費能力、醫療、商業、兩性，以及種族等各類議題上有著更深入的了解與認知。

三、心得分享：

在今年眾多的專書中，這本書的封面相較於其他書本，讓我第一眼就覺得，就是它了。書名簡潔活潑，且書中本文字據的描述也是簡單明瞭。其實大家都對「大數據」這詞耳熟能詳，卻不知道是誰先發想起這個詞，其實「大數據」這詞首先是由 IBM 公司發明出來的新名詞，一開始只是個商業手段，當時會發明出這個詞，是因為 IBM 公司為了要賣它的機器，但想不出聳動的詞來吸引消費者，然後就發明出這個新名詞，以下是 IBM 公司的廣告詞~~

大數據時代，面對數百 TB 的資料需求，眾多企業面臨身型走樣的危機：暴吞資料量，卻缺乏完善部署。企業轉型之際，核心資料無法有效管理。預算有限，但不知如何分配資源... 本公司在最合適的地方，提供所有正確的資料整合分析、預測分析並結合物聯網。

以上如此聳動的標題，馬上吸引許多企業的注意，但講「數據」，會讓人感覺好像局限於試算表之類的，所以有些人喜歡用「巨量資料」來形容大數據。

以前我認為數據就只是一種統計結果，可是在看完這本書後，我發現數據事實上可以和許多領域結合，說明了人們內心是可用數據去深入發掘的。假設能利用這點瞭解人們的慾望和需求，那企業或許就能分析出現在市場的趨勢，占得市場的先機，但有時，數據並非都是真實的，也並非所有資訊都是有價值的。

我們在日常生活中常常會聽到：數據會說話，的確正確的數據會提供我們做正確的判斷與決策，大到一個國家的領導者如何治國，如何把國家帶向未來的發展方向，讓人民未來是過幸福的生活還是痛苦的日子至關重要，小至一個家庭或個人不管在事業上，在子女教育上或在消費購物上……現在都會先上網 google 一下，其實這也是大數據的運用，只要輸入你要查詢資料的關鍵字，就有一大堆相關資料等你去選取去閱覽，你一定會想先聽聽看別人的意見後才能做自己的決定，完全失去自主判斷的能力，或許多參考一些資料可免去自己摸索思考的時間，增進不小效率，但如果只是依樣畫葫蘆照單全收，可能也會付出不少代價，因為別人的模式不一定適合你，你可能會花一大堆錢買進你根本不需要的商品，然後懊惱不已。我們經常在網路上查

詢資料或購物，查詢只是個普通的行為，但在不知不覺中我們所查詢的內容，卻成為了揭開內心另一面的關鍵。

伊德的理論是這樣的，他是心理學史上第一個對人格進行全面而深刻研究的心理學家，他的研究深入了人的內心世界，這較以往的心理學對人的內心世界的認識更加深刻。

佛洛伊德主張，人們在網路搜尋的關鍵字，是反應最真的自我內心想法，但根據大數據的統計，顯示出的只不過是人們打錯字或是拼錯音所產生出的結果，並非內心真的意識。另一方面，人們造訪色情網站的次數相當的高，為滿足好奇心，進而發現了人們不分男女，普遍也會迷戀色情網站。

人們經常說謊，是為了隱藏內心的真實聲音，不想被其他人貼上標籤，但卻為了想要留下好印象。

大數據運用早就已經成為我們生活上不可或缺的一部分；其實在我們生活周遭不管是看到的資訊或是一些預測或統計數字等等都是大數據的運用，例如：本次總統大選投票預測、世界衛生組織 WHO 提出有關武漢肺炎的流行疫情的預測或是氣象局針對颱風預報、天氣、空氣品質變化也好，甚至用更宏觀的視野來預測全球氣候的暖化及氣候變遷等等…都是大數據運用的一部分，這也顯示我們對大數據有相當的依賴。

再者我們幾乎每天都會使用電腦或平板及手機，我們都知道未來是一個雲端，全球的電腦伺服器都將聯結在一起，再加上每個人的行動裝置，手機、平板或桌上型電腦，然後有許多軟體以及應用程式（Application」就是 APP 的運用，將巨量資料也將為人類生活帶來前所未有的量化面向，進而使生活全面改觀及進化。另外，政府機關也政府機關也積極建立遠距醫療、遠距教育、無人車等數位經濟的雲端化政府，透過使用大數據方法精密解析，串連相關資料，加強更全面性、更全方位之政策規劃能力，如人工智能、文字探勘、深度學習、推薦系統等需要透過極大量數據分析，不論在相關儲存、數據規模及分析方式或運算技術，大數據運用都帶起了一波典範轉移，同時也讓資料帶進了科學中的其他領域，未來隨著資料收集與儲存越來越多元容易且更便宜，才能整合訓練出好模型的領域可以迅速的發展並取得了極大的商業價值，數據不再只被大企業壟斷使用。然而在此同時，人才也是另外一個兵家必爭的戰場，唯有團隊中有硬底子又有軟實力的人才，才能保證技術與服務永遠走在這場數據戰爭的最前端。

為結合更多不同的技術以發掘用戶各種需求，並提供更豐富且舒適的使用者體驗或改善工業製程以提升良率等方向前進，大數據分析的戰場也將轉為比較誰能真正有效且能正確的運用資料，只要有心，人人都可以透過資料分析挖掘到金礦。