

大數據 VS 大數律

廖述源

一、前言

大數據(Big Data)，又稱為「巨量資料」，大數據源起於美國。早在 1969 年，全球著名零售商沃爾瑪公司利用大量數據對消費者之購物行為進行分析，結果發現在星期五下班時段，男性顧客在購買嬰兒尿布同時，常常會順便搭配買幾瓶啤酒來犒賞自己，於是沃爾瑪公司將一般人很難想像毫無關聯之「啤酒」與「尿布」兩種物品，混搭推出「啤酒+尿布」結合銷售之促銷手法，居然為沃爾瑪公司創造出意想不到亮麗銷售業績。至今此一經由數據分析結果之「啤酒+尿布」成功銷售策略，早已成為人們在解說「大數據」最佳經典詮釋案例。

其次，2009 年出現新型流感 H1N1 疫情，造成全世界大恐慌，此種結合禽流感與豬流感所產生之新病毒，令人聯想起 1918 年西班牙所爆發流感肆虐疫情極為相似，當時全球被感染人數約五億人，最後奪走數千萬人寶貴性命。如今在面對 H1N1 可能爆發疫情，各國在尚未研發出新疫苗前，唯一能努力者，是減緩新流感蔓延速度。為了達此目的，則必須先預判新流感即將感染地區與蔓延程度。因此美國疾病管制局(CDC)要求醫生遇有新流感病例，就必須立即通報；即便如此，通報速度依舊

比病毒傳染擴散速度落後兩個星期以上。當時 Google 工程師即利用「巨量資料」，透過利用四億五千萬個不同數學模型，努力去推算 H1N1 疫情相關性(Correlation)，成功預測未來即將在美國境內可能爆發疫情地區及時間，大數據對新流感 H1N1 疫情可能發展，提供非常大助益及貢獻。至於 Google 這套方法之可貴有二，其一，在於不需要到各處去採集檢體；其二，在於不需要去造訪各地醫院及診所；而 Google 工程師僅僅是利用「巨量資料」(即大數據)資訊，創造出對人類巨大之貢獻價值。

基上，或謂大數據可視為「第五波科技浪潮」，一點亦不為過，(按：第一波是大型電腦；第二波是個人電腦；第三波是網際網路；第四波是社群媒體)。如今，大數據早已成為人類生活上不可分割之一部分。其實在人們生活周遭，不管是接觸到任何資訊，或是得到若干預測或統計數字等，小至股、匯市金融市場預測、颱風來臨預報，或者大至全球暖化及氣候變遷等，都是經由於大數據所提供生活中重要價值之證據。至於在保險經營領域方面，基於「大數律」(即大數法則)是保險經營之重要理論基礎，至於「大數據」與「大數律」兩者之關聯性，實有予以探討之必要。

二、大數據之主要內容

(一) 大數據之意義

大數據係指由巨型數據集所組成，這些數據集之大小，通常會超出人類在可接受時間下之收集、使用、管理、及處理能力等。由於大數據之大小，並非固定不變，往往單一數據集之大小，可能從數太位元組(TB)至數十兆億位元組(PB)不等。因此大數據即是規模非常龐大之數位資訊，這些資料量巨大到無法藉人工與現有科技來儲存、傳送、及分析，進而促使人們必須研發出更高階資料儲存設備與科技。

再者，大數據亦指無法在短暫時間內用常規軟體工具，對其內容進行掘取、管理、及處理之數據集合。基本上，大數據之技術，係指從各式各樣類型數據中，快速獲得有價值資訊之能力。通常適用於大數據技術，主要包括大規模並行處理(MPP)資料庫、數據挖掘電網、分散式文件系統、分散式資料庫、雲計算平臺、互聯網、及可擴展之存儲系統等。

(二) 大數據之基本特性

基本上，大數據之基本特性，通常可用3V來表示，亦即：(1)大量化(Volume)、(2)多樣化(Variety)、及(3)快速化(Velocity)，茲依序扼要說明如下：

1. 大量化(Volume)：通常大數據之數據體量十分巨大，基本單位至少為TB、PB等級，因此不但要考量其巨大數據量外，對於巨大數據之收集及儲存成

本，亦是大數據應用時必須思考之重點。例如當需要提供數據超過1.5PB(1PB=1024TB)時，這些數據列印成A4所需要紙張，將會超過5千億張以上。

2. 多樣化(Variety)：通常傳統數據有明確結構性，且選項比較少，諸如年齡、性別等。惟現今大數據之數據類型，不僅有文本形式外，還包括更多圖片、視頻、音頻、地理位置資訊等多元類型數據，且個性化數據又占絕對極大比重。
3. 快速化(Velocity)：由於大數據量十分巨大，其必須在最短的時間內產生分析結果，因此對於數據處理應以快速化，故有數據處理「1秒定律」，意即在短暫時間內從各種類型數據中快速獲得高價值之資訊。

(三) 大數據處理流程

基本上，大數據之處理流程，通常包括下列五大流程：

1. 數據收集(Data Collection)：大數據採集係指利用多個資料庫來接收發自客戶端(Web、App 或者感測器形式等)數據，並且用戶可以通過這些資料庫來進行簡單查詢處理工作。
2. 數據清理(Data Clearing)：雖然採集端本身有很多資料庫，如果要對這些巨量數據進行有效分析，應將這些來自前端數據導入到一個集中大型分散式資料庫進行簡單清理工作，使數據本身呈現其應有之運用價值。

3. 數據儲存管理 (Data Storage and Management)：數據巨量化與快增長是大數據主要特徵，如何有效對巨量數據妥善儲存，實乃面臨挑戰問題，其中當以降低數據儲存管理成本為最重要考量因素之一。
4. 數據運算 (Data Operation)：數據運算需要用到統計分析來處理，諸如可視化工具、及若干結構算法模型等，加以分類匯總滿足數據分析需求。本項流程必須按照一定規則去分類匯總，才能得到有效分析結果。
5. 數據挖掘與應用 (Data Mining and Application)：數據挖掘與應用是大數據處理流程中最為關鍵步驟。在全部數據中，可能僅有非常小部分(約占數據量 1%) 數值型數據得到挖掘，對於語音、圖片、視頻等非結構化數據，則較難進行挖掘分析。

三、大數據與大數律之比較

謹此將大數據與大數律兩者差異性加以比較詳如表 1 所示，至於其中較重要之差異者如下：

- (一) 大數律屬於「理論學說」；大數據則屬於「資料處理」。
- (二) 大數律屬於「因果關係」；大數據則屬於「互聯關係」。
- (三) 大數律屬於「質量並重」；大數據則屬於「量重於質」。

表 1 大數據與大數律之比較

比較項目	大數據	大數律
(1)時間	21 世紀初期	18 世紀初期
(2)屬性	資料處理	理論學說
(3)觀察群體	樣本	母體
(4)數據關係	互聯關係	因果關係
(5)強調重點	量重於質	質量並重
(6)觀察期間	短期	長期
(7)核心價值	原則	法則
(8)資料性質	龐雜	單純
(9)統計結果	變數	常數
(10)可信度	較低	較高
(11)使用成本	較高	較低
(12)使用單位	各行業	保險業

四、大數據導入保險業

關於保險業如何導入大數據，今依大數據之基本特性：大量化(Volume)、多樣化(Variety)、快速化(Velocity)與保險經營功能：商品設計、費率釐訂、行銷、核保、再保、理賠、財務投資等相互結合，探討大數據與保險經營之關聯性，詳如表 2 所示：

(一) 大數據與保險商品設計

保險業在保險商品設計時，可藉由大數據分析結果，鎖定潛在目標客群，精準研發以「顧客需求為導向」之保險商品，有助於提高保險消費者投保意願，進而增加保險業者銷售業績。

表 2 大數據與保險經營之關聯

保險經營 功能面 (Functions)	大數據三大基本特性		
	大量化 (Volume)	多樣化 (Variety)	快速化 (Velocity)
保險商品設計	○	◎	○
保險費率釐訂	◎	○	○
保險行銷	○	◎	○
保險核保	○	◎	○
保險再保	○	◎	○
保險理賠	○	○	◎
責任準備金估算	◎	○	○
保險財務投資	○	◎	○

◎：顯著

○：不顯著

(二) 大數據與保險費率釐訂

保險業在保險費率釐訂時，可藉由大量數據之彙集分析，使危險發生率與保險費率形成公平合理對應關係，進而使保險費率釐訂更精細化，達成保險費率公平性目標，同時亦符合保險費率自由化之時代潮流趨勢。

(三) 大數據與保險行銷

基於保險消費者對保險需求互有不同，如何對潛在客戶做最有效率之行銷，實有賴保險業者藉由大數據來針對消費行為予以分析，依據分析所得結果再採行不同行銷通路與策略，進而達成保險經營之行銷目標。

(四) 大數據與保險核保

所謂保險核保，即是對要保業務之危險篩選，作出承保與拒保之決擇。保險業在保險核保時，可藉由大量數據之彙集分析，選擇危險性較低之良質業務予以承保；反之，對於危險性較高之劣質業務，則予以拒保。

(五) 大數據與保險再保

由於近年來保險標的日益鉅額化，外加保險消費者求償意識高漲，保險人理賠壓力倍於往昔，保險人藉由再保險來分散承擔危險，則屬司空見慣之作法。保險人惟在擬訂再保計畫時，對於再保合約之自留容量等重要決策，實可藉由大量數據予以有效分析。

(六) 大數據與保險理賠

保險業在保險理賠時，可藉由大量數據之彙集分析，使保險理賠更趨於合理化，避免理賠浮濫或刁難情事發生；此外，針對若干涉及保險詐欺之理賠案件，亦可藉由大數據予以有效偵測。

(七) 大數據與保險責任準備金估算

通常保險責任準備金包括：責任準備金、未滿期保費準備金、賠款準備金、特別準備金等，對於各種保險責任準備金估算，更深具專業技術性與複雜性，其中以賠款準備金、特別準備金為最，需藉由大數據來確保責任準備金估算能符於實際。

(八) 大數據與保險財務投資

保險業收取保險費在先，承負保險賠款在後，兩者存有時間落差，給予保險業資金運用良好時機。基此，在龐雜投資標的中，保險人如何選取最佳投資標的獲得最大投資收益率，則必須仰賴大數據之精準分析。

五、建議(代結論)

關於本文(大數據 VS 大數據律)之建議事項，茲條陳分述如下：

(一) 因應大數據時代來臨，各行各業應廣泛蒐集所有資料

過去在資料不足小量資料年代，由於很難掌握到全體資料，只好抽取樣本利用統計技巧去推估母體概況。如今在大數據

時代掌握巨量資料已屬可行。各行各業應廣泛蒐集所有資料，相互支援使用，充分發揮大數據應有之基本功能。

(二) 資料「數量」較資料「品質」更為重要

基本上，大數據之可貴，在於數據之蒐集龐雜，可謂是巨細靡遺。隨後經由數據清理立即呈現數據之應用價值，此種由先量化而質化，亦屬大數據之主要特徵。因此只要擁有巨量資料，即便資料有瑕疵，亦可透過過濾萃取出資料精華，呈現數據應有之價值。

(三) 大數據僅能顯示「互關關係」而非「因果關係」

基本上，經由大數據分析結果，僅能找出相關性之「互關關係」，無法如大數據律呈現事件之「因果關係」，此亦是大數據力有未逮之處。雖是如此，亦無損於大數據其他重要功能之發揮。

(四) 積極投入大數據處理成本與相關人才培育

由於大數據處理流程，需要大量經費與各方面科技人才投入，諸如：IDC 及 Hardware 之投入；對於數據資料採礦及挖掘(Data Mining)建立、Data Storage、Data Modeling、及 Pattern 模擬等，均必須投入大量經費；另外，涉及大數據處理各種科技人才，更需要及早長期培育。

(五) 個資保護應由「個人同意制」轉變成資料使用者「使用責任制」

雖然大數據呈報多元價值，無可諱言，亦有其黑暗面，諸如：資訊蒐集對隱私造成極大之威脅。基此，茲為加強保護受害者權益，應將個資保護由原來「個人

同意制」，改變為資料使用者之「使用責任制」。

本文作者：

淡江大學保險學系教授

地址異動
需通知保險公司
以避免中斷投保而受罰

貼心提醒！地址異動記得以書面通知強制險投保之保險公司！

保險公司寄送強制汽車責任保險續保通知或重新投保通知，是依要保人最後所留存於保險公司之地址，倘要保人因搬家或轉換工作地點等原因異動地址，一定要記得以書面通知保險公司，以免發生中斷投保情事，而遭公路監理機關舉發裁罰。



強制汽車責任保險
COMPULSORY AUTOMOBILE LIABILITY INSURANCE

專屬網站：www.cali.org.tw
免費服務專線 0800221783

強制汽車責任保險粉絲團



廣告