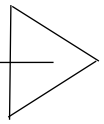




社群大數據的分析與應用

口述作者 ■林啟耀 / 源大數據科技有限公司營運總監

文字整理 ■趙崑睿 / 國立政治大學政治學系



大數據充斥我們生活之中，例如每天搭乘大眾運輸工具的「人次」、每個人搭乘捷運的「消費行為」……都是常見的大數據，而相關單位就會利用這些資訊來決定票價（例如：為何是一站 20、為何吃到飽是 1280）以及優惠程度（例如：轉乘優惠、十公里免費）等公共政策。資訊科技不斷發展，我們的日常生活與大數據密不可分，我們儼然身處在「大數據」時代中。

數據的類型

相對於上述這種被政府與民間單位應用已久的數字型大數據，也有另外一種文字型大數據在這個社群媒體時代開始受各行各業所重視。這種所謂的文字型大數據，包括數不清的政府單位或公司行號在臉書粉專上的貼文、鄉民在 PTT 或 Dcard 等論壇中的閒聊文，甚至廢文、網紅或女神在 Youtube 及 IG 上的業配文等，每天都大量及迅速的累積。當然，上述這種社群文章也具備不少數字型資訊，例如：貼文的表情數、回應數、分享數等互動指

標。不過，更讓各行各業開始重視這類型大數據的原因，不僅是社群上的互動數字，更因為這些社群文章的內容本身。相較於以前我們許多的消費行為依賴傳統媒體提供的廣告或型錄資訊，社群媒體時代下的消費者更重視彼此的意見。舉例來說，今天如果你想要換手機，決定你買哪一隻手機的決定性因素可能不是你在電視上看的廣告，而是你在社群上詢問其他網友的看法或使用經驗。這些詢問文章及回應的留言內容，可能包括 Iphone、Asus、Samsung 等品牌資訊、還有拍照、玩遊戲、聽歌、看影片等功能資訊、以及對上述資訊的評價。對於行銷人員或業者而言，網友問拍照比較多還是問玩遊戲比較多，可能就代表著大家消費時更關心的因素排序。同樣的，哪個品牌最常被詢問或提及，也反映出該品牌的社群討論度（或大家常見的說法：網路聲量）。這些訊息累積而成的結果，已經成為行銷人員或業者在產品研發或行銷時寶貴的參考資訊。例如：網友過去半年最常問到手機的

夜拍功能，那產品研發或廣告自然更需要注重及強調產品的這項功能。

掌握這類文字型大數據所提供的資訊，其實跟我們平常做民意調查的目的相當類似。我們的日常生活實際上存在各種各樣的民意調查，例如：在餐廳用餐後的客戶滿意度調查、上完一學期課程後的教師教學意見調查等，其實都是民意調查的一種。而分析社群媒體上的文字型大數據所得到的資訊，其實也具備類似民意調查的功能，只不過方法以及優缺點上有點不同。在方法上，傳統的民意調查可能是問卷調查（自填或訪員詢答、網路問卷或紙本問卷等都包含在內）也可能是座談、焦點團體或深度訪談等方式；相對的，分析社群媒體中的文字大數據，我們主要應用的是內容分析法。把內容分析法應用在媒體報導或文章的分析，是學界的主流研究方法之一。我們之所以會認為國內的媒體可能有「藍媒」或「綠媒」之分，其中一個原因就在於過去有學者應用內容分析法分析這些媒體的報導內容後發現他們可能偏向特定的意識形態。

在優缺點的比較上，我們就可以更進一步了解到為何分析社群媒體的文字型大數據所得到的民意結果，會越來越受到大家所重視。以餐廳的顧客滿意度問卷調查為例，顧客對餐廳意見的反應受到問卷題目與選項所影響或決定，如果問卷問服務生態度、問餐點滿意

度、問環境清潔度，顧客就只能針對這些面向回答。但是，消費者在社群媒體上對某某餐廳的意見反映，內容上可能更加豐富，例如不僅可以反映餐廳環境是否整潔，更可以表達餐廳布置與餐點是否搭配等資訊。因此，分析社群文字型大數據的第一個優點就在於這些意見的表達不受問卷問題與選項所限制。再來，我們以教學滿意度調查為例。雖然我們都知道教學滿意度調查是匿名調查，但是難免會有同學顧慮填答結果是否會被授課老師所知曉，進而遭到成績上的報復，因此填答上可能會有所保留。類似的問題，也會出現在政治人物或政府的施政滿意度調查，畢竟我們沒辦法確認受訪者是否擔心自己被查水表或秋後算帳，進而在回答相關問題時有所保留。相對的，網友在社群上表達對餐廳、對老師、對政府的意見時，沒有想到這些意見會被蒐集，因此會更加直接與毫不保留。因此，從側面蒐集大家意見想法的社群數據分析，可能會更有效，尤其是在問題的敏感度相當高時；第三個可能的優缺點，則在於這兩種方法蒐集民意的速度與數量上的落差。以電話訪問為例，如果我們要以電訪蒐集民眾對於美豬美牛議題的看法，從設計問卷到招好訪員，再到蒐集到 1100 份的有效問卷結果，可能需要 2-3 天的時間。相對的，在蔡英文總統宣佈相關政策的當天，社群上已經有超過 10 萬筆的數據反映

各類意見。因此，更快速更大量的蒐集網友意見，也是分析社群上的文字型大數據的優勢所在。

不過，透過分析社群媒體的文字型大數據所蒐集到的資訊，也存在一些缺陷。最主要的，莫過於這類資訊缺乏可辨識的個體資料，因此較難以進行各種人口學變數的分析。例如：不知道在社群上抱怨老師授課方式的是男同學較多還是女同學較多，不知道在社群上反對美豬美牛政策的是北部還是南部民眾等。而缺乏個體資訊的另外一個問題，就是無法驗證所蒐集到資訊的代表性，我們無法確認這些訊息是否能代表我們所關心群眾的主要意見。

網路聲量與好感度

上述對文字型大數據的分析方法，最常見的應用情境還是在社群媒體的資料分析。也由於對社群資訊的重視程度越來越高，因此有一些指標也開始受到各行各業所重視。其中，最主要的就是「網路聲量」以及「正負面聲量」。網路聲量最簡單直接的定義就是網路上的討論度，不過媒體在引用這個概念的時候較少說明他的測量方式與測量單位。就測量方式而言，可以區分為廣義網路聲量與狹義網路聲量：廣義網路聲量指的是只要文章有提到某個概念（包括人、單位、組織、議題等），這篇文章下所有的討論都算是這個概念

的聲量；狹義網路聲量則是指文章提到某個概念，文章下的討論中有提到這個概念的留言才算是這個概念的聲量。目前國內兩種方法都有採用，不過還是以廣義為大宗，主要就是因為計算較為方便與直觀。其缺點就在於，有些概念的聲量容易被膨風。舉例來說，某篇網路新聞同時提到 A 市長、B 立委、C 議員，文章下的留言大多的討論都集中在 A 市長，其次是 B 立委，C 議員幾乎沒人討論，但是在廣義網路聲量的計算下，A、B、C 從這文章得到的聲量是一樣的，所以 C 的網路聲量明顯被膨脹。

另外一個常在媒體上看到的概念是正負面聲量，也是分析文字型大數據的產物。所謂的正負面聲量，就是指提到某個概念時，該聲量（也就是文字內容）的屬性是正面、負面或者中立。在技術上，這也是一種內容分析，因為我們是依據詞庫的語詞屬性，去計算不同句子或文章的正負面屬性結果。所以，當某個品牌的社群負面聲量很高，代表討論該品牌時所用到的文字屬性偏向負面較多。對於許多公私部門而言，這個技術也讓他們快速掌握自己、組織、活動、產品或政策被網友討論的內容偏向正面或負面較多，正面較多可以加強宣傳，負面較多可能需要進行修正甚至危機處理。不過，這個看似很酷或很炫的技術，實際上當然也存在不少問題。首先，是準確度

的問題，也就是語詞屬性計算判斷的結果是
否正確。由於語詞在不同情境下可能會有不
同的屬性，因此機器計算的結果當然有誤判
的可能性。例如：辛苦一詞，如果是描述勞工
工作情形，可能是負面的語句，但如果辛苦一
詞出現在政治人物的粉專文章留言，大多卻
是體恤政治人物辛勞的正面聲量。另一種準
確度問題，則是因為廣義網路聲量的定義方
式所產生的問題。如同上面提出過的 ABC 民
代例子，C 因為 A 與 B 的關係而膨脹了他的
網路聲量。同樣的，該篇文章下的留言如果都
是對 A 與 B 的批評，自然也會膨脹 C 的負面
聲量。所以，這也就是為何很多時候政治人物
或品牌商品的正負面聲量與實際支持度有落
差的原因。第二則是大家對正負面聲量的誤
解，也就是正負面聲量不一定代表支持度或
滿意度。舉個常見的例子：死刑，台灣每逢有
令人激憤的社會事件發生時，死刑總會被拿
出來討論是否有其必要。如果把這些文章與
留言進行正負面情感分析，得出負面聲量明
顯比正面情感高，請問這是支持死刑的聲量
比較多還是反對死刑的聲量比較多呢？恐怕
兩者都不是。同樣的道理，我們用這個方法拿
來討論其他的政策，例如：美豬美牛、防疫等，
正負面聲量就代表對這些政策的支持或反對
嗎？當然，也不是。

不過，這邊也必須提醒的是，雖然正負面聲

量不代表支持度或滿意度，但是不代表文字
型大數據的分析不能測量對品牌、商品、政治
人物、政策的支持度或滿意度。如同前面所述，
文字型大數據分析應用的是內容分析法，自
然可以用來建構支持度或滿意度的編碼架構。
以剛剛提到的死刑為例，網友在社群上的討
論可能包括：應報觀點（例如：殺人償命、殺
人有報應、給死者交代等）、成本觀點（浪費
納稅人的稅金、不想花錢養死囚等）、嚇阻觀
點（例如：犯人才會怕）、死刑不能解決問題。
前面三種說法反映的就是對死刑的支持，這
些聲量的加總反映的就是支持死刑的聲量，
同時我們也可以知道網友支持或反對死刑的
原因。

結論

分析文字型大數據雖然提供我們許多創新
性與可能性，不過同時我們也注意到這個方
法的許多侷限。如果忽視這些侷限，我們往往
就會提出許多似似而非的觀點，例如用網路
聲量或正負面聲量來衡量品牌或人物的支持
度。實際上，文字型大數據的分析提供我們在
進行內容分析時更多的便利性，以人工智慧
取代人工，不過要完全取代人工，恐怕仍有很
長的一段路要走。因此，我們如果要善用文字
型大數據分析方法提供的豐富成果，人為的
參與，對機器演算的結果進行較正與修正，目
前仍是最適途徑，而且沒有之一。



作者簡介

林啟耀博士為國立政治大學政治學系博士，其主要的專長領域為民意調查與數據分析，目前擔任源大數據科技有限公司營運總監。林博士除了有豐富的業界經驗外，亦曾在《台灣政治學刊》與《選舉研究》發表數篇學術論文。