



# 巨量資料量測當下

社群網路與資訊科學攜手合作，實現當下量測的願望。

英文的 forecast 可拆解成兩個字根，fore 是「事前」跟「提早」之意，cast 是「計算」，合起來是事前計算，也就是預測或預報。混沌理論告訴我們：世界是一個複雜的非線性系統，某處的蝴蝶一振翅，可能會導致未來遠方某地發生颶風。可見預測未來之困難，必須考慮許多細微且互相影響的因素。

既然預測未來非常困難，我們是否能夠退一步、觀測分析「當下」發生了什麼事？於是，經濟學者提出了「當下量測」(nowcasting) 這個新詞，nowcasting 從字面上來看，是從 forecasting 衍生而來，目的是在量測「當下」，而非預測未來。

經濟學上有很多指標，常常需藉由普查(如問卷調查)才能獲得，例如失業率、經濟成長率、金融商品的銷售量等。但是進行這些調查的時間與金錢成本所費不貲，調查結束後還需要分析，往往導致得出結果時與原先欲預測之「當下」已經有時間上的差距。歐元區的生產毛額即為一例，歐元區生產毛額一季調查一次，光是調查過程就約六週，造成經濟學家或金融學家無法正確對市場即時決斷。

有鑑於上述困境，資訊學家開始思考能否利用資訊技術，既快又準確地從事「當下量測」。近年來，因巨量資料興起，結合大量資料與快速運算

的機器學習模型，使得當下量測可能實現。歐洲中央銀行2010年發表的一篇論文，詳述他們如何利用較頻繁調查的指標(如歐洲工業生產量)來估算一些較不頻繁調查的指標(如歐元區生產毛額)，有很不錯的成果。

網路界巨擘 Google 的首席經濟學家韋瑞安(Hal Varian)亦為當下量測的高手，他提出可利用 Google 搜尋趨勢(Google Trend)的結果來進行當下量測。Google 搜尋趨勢是一個公開

## Forecasting? Nowcasting!

服務，使用者鍵入某個關鍵字就可得到2004年至今這些關鍵字每天、每週、每月、每年被搜尋的次數。韋瑞安團隊發現，這些搜尋行為的統計值跟某些對應的經濟指標很相關。例如，特定時間內"car"及"car purchase"這類關鍵字的搜尋，與當時汽車銷售量高度相關；"job search"、"job hunting"等關鍵字可用來估計當下的失業率。針對每個相關的經濟指標，Google 列出數十甚至數百個相關關鍵字，蒐集它們在 Google 搜尋趨勢的查詢頻率，再利用多種機器學習及統計方法從事當下量測，達到很高的精確值。Google 甚至可記錄資料的 IP 位址，得知該搜尋源自哪個地區，進

而從事小範圍的預測。比如利用"gun store"關鍵字在各地的搜尋頻率，來比較美國各州槍枝泛濫的情形。

當下量測的技術不只被運用來估算經濟指標，近年來更有資訊學家利用社群網路平台，如推特(Twitter)或臉書(Facebook)的資料，來判斷群眾對於社會議題或事件的觀感。筆者所帶領的研究團隊，在2011年發表的一篇論文中，利用一些事件的名稱當成關鍵詞，在這些社群網路平台上搜尋，找出偵測發文者情緒的方法。利用機器學習模型自動判讀每一篇發文者的正負意見或情緒反應，再把所有作者的情緒累加起來，就可以知道目前大眾對於某個議題的反應。例如以某體育賽事為關鍵字，透過情緒偵測便可得知社會大眾對於賽事的狂熱程度；以某齣戲劇名稱為關鍵字，就可以搜尋並分析這齣戲在觀眾心目中是正向還是負向；以某個政治議題為關鍵字，亦可以了解社會大眾對該議題的觀感與支持程度。

網路及社群媒體的普及，讓以往許多需要透過費時費力的調查方可得到的指標，現今透過資訊科學的方法便可即時量測。若政府或公司能夠好好運用當下量測的技術，相信能夠提出更符合公眾利益、「以群眾意見為意見」的政策。

SA

林守德是台灣大學資訊工程系副教授。