

# 調查研究—— 「大數據」的方法與應用

陶振超\* 張仁和\*\*

大數據 (big data)、資料科學 (data science)、人工智慧 (Artificial intelligence)、機器學習 (machine learning) 等概念與技術自 2010 年前後蓬勃發展，不只在資訊科學中蔚為風潮，近期更廣泛應用在各社會科學領域之中。以 Annual Review 出版社近期於各社會科學領域中的發表來看，在社會學 (Couper 2017)、政治學 (Brady 2019)、經濟學 (Bok et al. 2018)、心理學 (Mohr et al. 2017)、公共衛生 (Mooney and Pejaver 2018) 等，皆有大型的回顧與討論。此外，著名學術出版社 Nature Publisher Group 也於 2019 年建立新期刊 *Nature Machine Intelligence*，從跨領域的多元視角探討資料科學與人工智慧相關理論和應用。在臺灣，前中央研究院資訊所陳昇瑋<sup>1</sup> 研究員也於 2014 年舉辦「台灣資料科學年會」，後續於 2018 年建立台灣人工智慧學校，且積極

---

\* 本期「大數據與資料科學特刊」主編，交通大學傳播與科技學系。

\*\* 本期「大數據與資料科學特刊」主編，中央研究院民族學研究所臺灣大學心理學系。

1 陳昇瑋研究員對於大數據與資料科學在臺灣的發展有極高的貢獻，並且深入與社會科學領域進行應用實踐和對話反思，他先後成立台灣資料科學年會、台灣人工智慧年會、台灣人工智慧學校等。然而，編輯群在籌備此特刊時，不幸得知他已於今年初（2020 年）逝世，於此深表遺憾。

與臺灣社會科學進行對話跟應用(江彥生、陳昇瑋 2016)。如此面對大數據與資料科學的浪潮下,勢必對傳統的調查方法有極大的影響,誠如前述 Couper (2017) 於 *Annual Review of Sociology* 所言,這會挑戰到社會科學典型的研究方法,尤其以傳統自陳報告之調查方式,但此挑戰並非是消極地捨棄既有的研究取徑,而是透過新式數位工具及分析技術,有更多延伸的可能,進而對個人與社會有更全面的理解。

然而,目前臺灣社會科學領域之學術期刊對此議題之討論尚為有限,較缺乏系統性的專題探討。據此,本期刊《調查研究—方法與應用》進行大數據與資料科學特刊,尤其聚焦於調查研究上的應用,我們也以「調查研究—『大數據』的方法與應用」為文作序,期望讀者能對此特刊內容有初步理解,並藉由閱讀本期論文,而從各個角度理解大數據與資料科學之實踐、應用與反思。我們自去年初(2019年)開始徵稿,在經過嚴格的匿名審查後,最終有五篇論文收錄於本期特刊,涵蓋綜論性質的回顧與前瞻性探討(黃從仁、江庭瑋等)、文詞調查指標之建立與應用(林瑋芳等、王賢)、以及對大數據與資料科學的學術反思(余貞誼)。

首先,黃從仁一文系統性地檢視近十年(2010-2019)整體大數據在社會科學的影響歷程,進而探討小數據與大數據、解釋模型與預測模型、乃至於研究重現與不可重現的交互關係。江庭瑋、陳思宇、林煜軒則聚焦手機在數位足跡上的應用可能,並且詳盡介紹在手機成癮、睡眠作息與工作過勞三種行為指標之建立跟實作。之後,林瑋芳等則建立最新版的中文化語言探索與字詞計算詞典,該工具能對中文進行自動化的心理屬性類別分類,是較從由上而下(top-down)的角度理解語文材料中所蘊含的特性跟表徵。相對於此,王賢一文則是採由下而上(bottom-up)的方式,針對「批踢踢實業坊」(PTT)中的

「公職板」，挑選提問性的文章約兩萬篇，進行文詞探勘之主題模型分析，並由此了解公務人員所關切的議題屬性。最後，余貞誼則以後設角度檢視大數據在社會科學中之本體論、認識論和方法論的內在意涵，由此探討其研究倫理之風險管理，以及在技術革新與反思之間的拿捏平衡。

整體而言，這五篇文章從不同面向與角度探討大數據與資料科學之技術應用與影響，並以臺灣在地實徵資料進行應用。我們很榮幸地擔任此特刊主編，誠摯地跟您推薦這五篇大作，期待有拋磚引玉之效，同時冀望本特刊能讓臺灣在大數據與資料科學的研究質量有更豐富而積極的發展。

## 參考文獻

- 江彥生、陳昇璋，2016，〈簡介「計算社會學」：一個結合電腦與數位科技的新興社會學研究〉。《台灣社會學》32: 171-201。(Chiang, Yen-Sheng and Sheng-Wei Chen, 2016, "Computational Sociology: A New Sociological Area Empowered by Modern Computational and Digital Technology." *Taiwanese Sociology* 32: 171-201)
- Bok, Brandyn, Daniele Caratelli, Domenico Giannone, Argia M. Sbordone, and Andrea Tambalotti, 2018, "Macroeconomic Nowcasting and Forecasting with Big Data." *Annual Review of Economics* 10: 615-643.
- Brady, Henry E., 2019, "The Challenge of Big Data and Data Science." *Annual Review of Political Science* 22: 297-323.
- Couper, Mick P., 2017, "New Developments in Survey Data Collection." *Annual Review of Sociology* 43: 121-145.
- Mohr, David C., Mi Zhang, and Stephen M. Schueller, 2017, "Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning." *Annual Review of Clinical Psychology* 13: 23-47.
- Mooney, Stephen J., and Vikas Pejaver, 2018, "Big Data in Public Health: Terminology, Machine Learning, and Privacy." *Annual Review of Public Health* 39: 95-112.