

判別數字人文的兩個準則*

邱偉雲

在數字人文萌芽之初,我們需對其定義採取開放性態度,暫不進行“劃定疆界”的定義工作,以免使其內涵過於貧乏。而如今,劃定疆界的時刻已經到了,因為從正面對數字人文下定義的工作已持續了將近十年,這一概念中已經填滿了各種可能的內涵,以至於其獨特性與合法性開始被質疑。為避免扼殺數字人文發展的其他可能,我們在劃定疆界時須採取一種寬泛的態度,嘗試提出判別數字人文研究的準則。採用較有彈性的劃界方式有兩點好處,一是能調整數字人文研究的方向,確立其合法性,二是能繼續保有為數字人文概念填充內涵的空間。這正是筆者討論“什麼不是數字人文”這一問題的基本態度。

一、不具人文性的研究不是數字人文

我們必須給數字人文勾勒出一個輪廓,即其與眾不同、無可取代的合法性根據,人們可以在其中持續填充新內涵而不逾越基本疆界。筆者認為,這個輪廓就是“人文性”(humanity)。所謂的“人文性”,可以嘗試透過人性、人類、仁、慈愛、文學教育、文學等概念去掌握,其共同指涉就是“人的情感”,不論是大我或小我的情感。數字人文研究中的“人文性”,就是指在研究中必須要有人的情感和思考方式,亦即算法必須密切結合人文理論(由人的思考而成的理論),而對數據的解讀也必須要有人文價值(能夠回答人的問題),從而在具有“人的情感”的溫度下進行數字人文研究。反過來說,如果不具“人文性”,那麼就不是數字人文研究成果。

例如之前看到一些數字人文研究成果,雖也是運用數字技術進行的人文研究,但細觀其內容,雖然在算法上運用了最新的處理技術,但卻未對數據賦予“人文性”的解釋,未分析數據背後的人文信息及其背後的人文情感,沒有深入到人文與歷史的脈絡中去。這樣的數字人文研究就只是半成品,而不是真正的數字人文研究。在進行數字人文的算法研究時,不能只專注於算法,而必須要結合人文理論,思考算法在人文研究上可帶來的人文意義與價值;而對數據的解讀也必須從人文角度出發。不具“人文性”的數字人文研究,只是點綴了人文的計算機算法研究而已。目前海內外學界對數字人文最多見的批評之一,就是重編程而輕人文。筆者從 2010 年開始從事數字人文研究,至今將近十年,深知數字人文研究絕不應重編程而輕人文,而應兼重數字算法與人文詮釋。令人擔憂的是,現在的數字人文研究論文中,大量充斥著重編程而輕人文的研究成果。這一方面會導致局外人(非數字人文研究者)以重編程而輕人文的刻板印象來污名化數字人文研究;另一方面則會引

* 本文係教育部人文社會科學研究規劃基金項目“數字人文視野下中國近代‘道’的概念的轉型與變遷”(項目號:19YJA770012)的階段性成果。

發局內人(數字人文研究者)的爭論。

從以上論述可見,討論“什麼不是數字人文”是有對話對象的,那就是認為數字人文研究就是簡單的“數字+人文”研究模式的群體。這樣的研究看似有數字人文的外貌,但卻沒有數字人文的精神,在研究結果中往往看不出任何人文價值、意義與關懷,人文多成為點綴和算法處理的“對象”。在這一框架下,容易以算法為主,這時算法的優劣與否就無法從人文角度去判斷其召回率與準確率,因而失去算法的判據。例如某些研究者以為只要找到一批數據以及一種算法,就可以將兩者結合計算出來的數據強加解釋。這看似有道理,但卻可能形成為數據而論述的帶有偏見的結論。這樣沒有人文溫度的“數字+人文”研究,就不是好的數字人文研究。在數字人文研究中,不管算法和人文分析的比重孰高孰低,關鍵是要以“人文性”為最終關懷。只有採用傳統人文研究的思考方式去考察文本研究對象,自覺地從“人文性”出發,才能自覺意識到“數據”也可能帶有錯誤與需要修正之處,而能對算法從“人文性”角度提出修改建議。

從知識社會學視角看,“數字+人文”模式的研究將招致計算機學界和人文學界兩個知識群體的夾擊。計算機學界會認為這類研究雖然在算法上精進,但卻看不到算法研發背後所要處理的議題的價值和意義;而人文學界則會認為這種以算法為主的研究,並未處理與解答任何重要的人文問題,因此還不如傳統的人文研究方式。“數字+人文”模式是兩面不討好的。

二、太輕易完成的研究不是數字人文

在許多人的想象中,數字人文就是用些軟件跑出數據並進行解釋,整個研究大概不到一個月就能完成。這種研究太過簡易,得出的結論也未必比得上長期浸淫於史料的傳統人文學者。作為長期投入數字人文研究的工作者,筆者實有義務去揭示數字人文研究中的辛苦,並就此再提出一項判別數字人文研究的標準:太輕易完成的研究不是數字人文!

偶而可聽到某些人文學者很不謙遜地說,數字人文中的數據就是用電腦跑一跑就會出來,重要的是人文學者對數據的詮釋。這類學者的發言與態度,筆者認為也絕非數字人文的真精神。因為一個“跑”字,實則蘊藏著對數據處理工作的無知與輕視,殊不知“跑”字背後是大量繁瑣的數據清理工作、算法從無到有的構思工作,以及無數的無效實驗,最後才能給出一個展現在人文學者面前的算法與數據結果。真正的數字人文研究項目,絕不是簡單地運用一些既有程序,按幾個按鍵,畫出幾張漂亮的圖案,然後對視覺圖像數據解釋幾句就萬事大吉了。

DHer(數字人文學者)橫跨在人文與計算的交界處,可以同時體會到計算機與人文兩個學界對數字人文研究的誤解。兩個學界應彼此虛心學習與互相欣賞,了解對方的辛苦,這樣才能使數字人文研究健康發展。因為,誰也不願意跟無法共情的人協同研究,誰也不願意成為對方的“工具人”(有些人文學者把算法當做一種炫技的工具或帶有定量光輝的裝飾,而把計算機學者當做跑程式的工具人)或是“拾人牙慧者”(有些計算機學者對以人文學者為主的數字人文研究進行批評,認為其重人文而輕編程,甚至說這類研究是拾人牙慧,即用最簡單的算法作出的非常基本的研究)。那麼,到底在數字人文研究中,各自的辛苦在哪兒呢?在數字計算方面,數據清理和數據挖掘這兩項工作都非常辛苦。雖然可以透過電腦自動清理,但通常還需人工過濾清理得更為乾淨後方能達到研究的標準。以筆者的經驗,每每問學生是否願意從事數據清理工作時,學生都會說這工作太枯燥乏味。從人文方面來說,假若認為人文詮釋一點也不辛苦,只是靠嘴皮子,那也不是數字人文研究者該有的態度。因為人文學者在拿到數據之後,得比過去在選精集粹視野下的研究看範圍更廣的

文獻,才能理解與合理解釋數據背後代表的歷史意涵。這樣的解釋除憑借人文學者長年積累外,也需依靠對史料與數據的對應與重讀,這工作是非常辛苦的。總而言之,數字人文研究一定比過去單一學科下的研究過程要更辛苦。

對算法重要性的無知,或認為算法至上的傲慢,這兩種態度都不該出現在數字人文研究中。人文學者該體會“跑”字背後數十個小時的清理數據與構思、測試算法的辛苦,而計算機學者也該體會“詮釋”背後數十個小時的語料閱讀與數據分析工作,不論是“跑”或是“詮釋”都是不容易的,只有對數字計算和人文詮釋兩方面工作各自的辛苦產生共情,才能讓數字人文有更長久的發展。

三、退而遠瞻數字人文的未來發展

金觀濤先生曾說:“反思的本質是人可以從他當下生活的意義世界中跳出來,或退而遠瞻,或進入他所陌生不喜歡的價值系統,以獲得對意義世界的新認識。”(金觀濤:《數位人文研究的理論基礎》,收入項潔主編:《數位人文研究的新視野:基礎與想像》,台北:台灣大學出版中心,2011年,第11~24頁。)上述主張中的“退而遠瞻”,正是筆者認為當下應對數字人文發展進行的重要工作。退而遠瞻,才能抽離於數字人文之外去審視數字人文的發展方向,藉以調整與修正,進而重新出發。為了退而遠瞻,可以從知識社會學的視角對“數字人文”作一個危機的情境報告,一方面讓參與過數字人文研究過程者(局內人)能透視該項過程,另一方面也讓未參與過數字人文研究過程者(局外人)對這一過程有更深入全面的理解。藉助知識社會學的退而遠瞻,能把局內人與局外人的觀點綜合討論,以避免盲點。

就知識社會學來看,有兩種看起來很相似的研究,一是人文關懷為主的數字人文研究,一是算法為主的數字人文研究。這兩種研究成果都被視為“數字人文”研究,但以人文為主的研究更為側重人文解釋,算法可能相對簡單;以算法為主的研究側重算法的更新迭代,而所解決的人文問題則往往比較單薄甚至被懸置。當數字人文研究出現這種貌同神異的情況,且以算法為主的論文又能更快發表時,就會造成以算法為主的研究成果籠罩整個數字人文研究的現象,這就導致人文學者對數字人文研究產生距離感和隔膜,最終致使人文關懷為主的數字人文研究的萎縮和算法為主的數字人文泡沫式的繁榮,造成人文學者對數字人文重編程而輕人文的刻板印象,甚至形成一種對立結構。這種對立結構是令人擔憂的,因為數字技術引入人文學界,絕不是要讓算法凌駕於人文研究之上,而是要幫助人文學界拓展與解決過去未能處理的研究視野與問題。因此,需要改變數字人文研究領域以算法為主的研究成果一家獨大的現狀,而給以人文為主的研究成果更多的空間,只有這兩種數字人文研究成果共存與互動,數字人文研究才能良性發展。因此,劃定“人文性”作為數字人文的疆界,並強調計算機學者與人文學者之間的互相體諒與理解,在當前這個歷史時刻有其緊迫性和必要性。這對於調整數字人文研究的結構,邁向下一個發展階段,將產生重要的影響。

很慶幸如今我們能夠開始談“什麼不是數字人文”這個問題,這個問題的提出揭示著數字人文研究發展的第一階段,即從正面定義提出的階段已經接近結束,從而進入一個新的階段。正是在這樣一個新的階段裡,才會產生上述退而遠瞻的提問。這樣的提問其實也揭示出數字人文研究第二階段發展的方向。在關於“什麼不是數字人文”這個問題的討論與思辯後,希望數字人文研究能夠對重編程而輕人文的現狀有所修正,向一個兼重數字算法與人文詮釋的健全方向邁進。

(作者係山東大學歷史文化學院副研究員,博士)

[責任編輯 桑海]