

# 大數據時代的解決方案

人工智能時代的激活數據學

# 塊數據4.0

藍天書  
圖

大數據戰略重點實驗室 著  
連玉明 主編

## 大數據時代的解決方案

人類智慧與機器智能的高度融合

激活數據應用新場景，開啟數字文明新紀元



塊數據

4.0

人工智能時代的  
激活數據學

大數據戰略重點實驗室◎著

連玉明◎主編

Business 044

作者：大數據戰略重點實驗室

編輯：藍天圖書編輯組

出版：紅出版（藍天圖書）

地址：香港灣仔道 133 號卓凌中心 11 樓

出版計劃查詢電話：(852) 2540 7517

電郵：editor@red-publish.com

網址：http://www.red-publish.com

香港總經銷：香港聯合書刊物流有限公司

台灣總經銷：貿騰發賣股份有限公司

地址：新北市中和區中正路 880 號 14 樓

電話：(866) 2-8227-5988

網址：http://www.namode.com

出版日期：2020 年 2 月

圖書分類：資訊管理／商業策劃

ISBN：978-988-8568-75-8

定價：港幣 148 元正／新台幣 596 圓正

塊數據 4.0 by 大數據戰略重點實驗室

中文繁體字版 2019 由香港紅投資有限公司

本作品中文繁體版經中信出版集團股份有限公司授予香港紅投資有限公司獨家發行，非經書面同意，不得以任何形式，任意重製轉載。

## 編撰委員會

總顧問：陳剛 閔傲霜 李再勇

編委會主任：李再勇

編委會常務副主任：許強 陳晏

編委會副主任：聶雪松 徐昊 連玉明

主編：連玉明

副主編：朱穎慧 武建忠

執行副主編：宋希賢 宋青 胡海榮

主要撰稿人：連玉明 朱穎慧 武建忠 宋青

胡海榮 宋希賢 張俊立 張龍翔

范賢昱 龍榮遠 黃倩 鄒濤

翟斌 鄭婷 陳威 何露

姜璠 陳鵬 胡亞男 田翠梅

學術秘書：李瑞香 江岸 龍婉玲

大數據是什麼並不重要，重要的是大數據改變了人們對世界的認識。大數據不僅改變了貴州和貴陽對世界的認識，更重要的是也改變了世界對貴州和貴陽的認識。貴州和貴陽已經成為中國大數據發展的戰略策源地，並正在成為引領全球大數據發展的重要風向標。這個戰略策源地和重要風向標的主要標誌就是以塊數據為核心的理論創新，搶佔了大數據發展的理論制高點和國際話語權。

應該肯定，塊數據是貴州和貴陽發展大數據理論創新和實踐探索的產物。如果說，大數據第一次讓貴州和貴陽站在了世界面前，那麼，在這面前的正是標誌著大數據時代真正到來的塊數據。塊數據對未來文明的改變和推動已經超出所有偉大預言家的預測，這一史詩般劇碼的序幕正在拉開或者剛剛拉開。事實上，對塊數據的認識和研究從來沒有停歇過，我們建構了一個從1.0到5.0的塊數據理論體系，試圖探索和揭示塊數據的本質、規律和價值。

正如我在《塊數據 2.0：大數據時代的範式革命》一書主編序中描述的那樣：「這是一場由科技引發的社會變革，它將改變我們的思維方式和生活方式，改變世界上物質與意識的構成，改變我們的世界觀、價值觀和方法論。我們不得不更多地關注它、研究它和把握它，因為我們每個人都置身其中。這也是我們必須對塊數據肅然起敬，並持續探尋的根本動因。」

不僅如此，我們的研究發現，塊數據具有更加深遠的意義：

**第一，塊數據是大數據時代真正到來的標誌。**新一輪科技革命和產業變革正處於重要交匯期。隨著信息技術和人類生產生活交匯融合，我們已進入以大數據為標誌的信息化發展新階段。人類將以塊數據為標誌，真正步入大數據時代。塊數據就是把各個分散的點數據和分割的條數據匯聚在一個特定平台上並使之發生持續的聚合效應。這種聚合效應通過數據多維融合與關聯分析對事物作出更加快速、更加全面、更加精準和更加有效的研判和預測，從而揭示事物的本質和規律，推動秩序的進化和文明的增長。簡言之，塊數據就是研究數據運動規律的數據哲學。數據是運動的，數據運動是有規律的，數據運動所揭示的是數字文明時代秩序的增長。我們認為，塊數據是大數據發展的高級形態，是大數據融合的核心價值，是大數據時代的解決方案。

**第二，塊數據是數據、算法、場景融合應用的價值體系。**數據、算法、場景是治理科技的三大核心要素。其中，數據是基礎，算法是手段，場景是目的。塊數據價值鏈是實現超越資源稟賦的價值整合，是以全產業鏈、全服務鏈和全治理鏈為核心的價值體

系。通過數據、算法、場景的疊加效應，在塊數據系統架構下減量化的數據採集、數據傳輸、數據存儲、數據分析和數據應用的數據觀和方法論，為我們建構一個融合技術流、物質流、資金流、人才流、服務流的價值系統。以發現塊數據內海量複雜數據的潛在關聯和預測未來為目標，以複雜理論的系統思想為主要範式，實現對不確定性和不可預知性更加精準的預測。從條數據到塊數據的融合，人類社會的思維模式和行為範式將產生顛覆性變革。這種變革的文化力量就是以人為原點的數據哲學，它不僅革新了我們的世界觀、價值觀和方法論，而且開啟了我們的新時代、新生活和新未來。

**第三，塊數據引領和催生新的組織模式，成為改變未來的新力量。**塊數據既是一種經濟模式，也是一種技術革新，更是一種新的世界觀、價值觀和方法論，引領和催生新的組織模式。物物互聯、業業互聯、人人互聯，構成一個自組織、自激活的數據實時流動、共用、聚合的數據化網路生態圈。這個生態圈通過外部資源強化自身戰略地位，實現組織的再平衡。塊數據組織是一個資源分享、平台無限擴展的超穩定組織結構源，預示著組織發展的新方向。無邊界組織、自組織、雲組織等，都可以看作一種正在萌芽和生長著的塊數據組織。利他主義的數據文化是構成塊數據組織的理論基石，其出發點是數據人假設。塊數據組織中，數據力上升為組織的核心競爭力。數據力與數據關係影響著社會關係，這將引發整個社會發展模式前所未有的變革和重構。所有這一切，都預示著塊數據組織「平台化、關聯度和聚合力」三位一

體所帶來的強大組織勢能，實現組織的自激活和自適應環境變化，是組織存續與發展的根本動力所在，並最終形成共用型組織新範式。

**第四，啟動數據學成為人工智能時代大數據發展新的解決方案。**數據無序增長預示著超數據時代的來臨。小數據時代，數據越大，價值越大；大數據時代，數據越大，價值越小。人類對未來的把握和勝算，在於數據被有效存儲的同時又被迅速激活。我們沒有把大數據僅僅看作所謂的「大」的數據，而是把大數據看作一種「活」的數據，因為只有激活，大數據才有生命，才能成為未來世界人們賴以生存與發展的土壤和空氣。我們探索用生命科學的方法解決數據擁堵問題，以數據社會學的思維打開沉澱的數據寶藏。塊數據就是數據通過算法作用於場景，這種作用的動力就是激活數據學，它為我們尋找這小部分的有效數據提供了解決方案。作為一種理論假說，激活數據學就像一座朝向深邃的大數據宇宙的「天眼」。它是未來人類進入雲腦時代的預報，是關於混沌的數據世界的跳出決定論和概率論的非此即彼、亦此亦彼的複雜理論的大數據思維範式革命。

**第五，塊數據提出的數據進化論、數據資本論、數據博弈論或將成為數字文明的「新三論」。**美國學者塞薩爾·伊達爾戈的《增長的本質》一書被譽為「21世紀經濟增長理論的重要里程碑」。因為本書提出了一個重要觀點，經濟增長的本質是信息的增長，或者說秩序的增長。他認為，善於促進信息增長的國家會更昌盛。我們提出的數據進化論、數據資本論和數據博弈論，正

是重構數位文明時代人與技術、人與經濟、人與社會的新秩序。我們的研究認為，數字文明時代增長的本質不是GDP的增長，而是文明的增長和秩序的增長。新「三論」的提出，對社會結構、經濟機能、組織形態、價值世界進行了再塑造，對以自然人、機器人、基因人為主體的未來人類社會構成進行了再定義，對以數據為關鍵要素的新型權利範式和權力敘事進行了再分配。這既是研究未來生活的宏大構想，也是研究未來文明增長和秩序進化的重大發現。

我們對塊數據的研究才剛剛開始。數據驅動、算法驅動、場景驅動下的雲腦時代正在到來，激蕩著我們對於未來所有的憧憬和渴望。

我們永遠在路上。

連玉明

大數據戰略重點實驗室主任

2019年12月

眼看乾坤一局棋，滿枰黑白子離離。  
鏗然一子成何劫，唯有蒼蒼妙手知。

這是被稱為「波斯李白」的詩人奧馬爾·海亞姆（1048-1122）的《魯拜集》中的詩句。作為一名精通天文和數學的大學者，奧馬爾認為，宇宙的規律是可以探知的，並可以用嚴密而美妙的數學方式表示出來。前定與隨機，必然與偶然，向來是人文科學中長期爭論不休的命題。自然科學理論始終受實驗和觀測的檢驗，而它的每一個重大發現又都會反饋到文化和社會的層面，對人的哲學和歷史觀有所啟示。

決定性和概率性一直被當作數學、物理等學科對自然界的描述方式。在牛頓創立古典力學之後的 250 年間，直至 20 世紀 20 年代，決定論長期處於主導地位，基於概率論的統計描述或者說數據的描述，則一直屬於不得已情況下所採用的輔助手段。決定

性的牛頓力學從計算和預測的觀點來看，實際上也具有內秉隨機性，這就是微觀層次上的混沌運動。大量隱藏在暗數據背後的某些看似簡單原因所導致的複雜後果，則漸漸成為混沌研究的重要信息。混沌不是無序和混亂。與人們習以為常的週期排列或對稱形狀的數據相比，大自然和人類社會中的很多數據其實就是一種沒有週期性次序的混沌。在理想模型中，它可能包含著無窮的內在層次，層次之間存在自相似性或不盡相似。在觀察手段和技術的分辨率不高時，只能看到每一個層次或某一種類型的結構。但技術條件改變或提高後，在遠離不能識別之處就會出現更小尺度上的結構。零維的點、一維的線、二維的面、三維的體和四維的時空，是人們現在所能認知的數據空間。如果在不遠的將來，我們真的進入一個超數據時代，現有的技術和描述手段也許就無法對這種高度無序數據的混沌運動進行分形，而關於相變和臨界現象理論的框架也需要一個新的重構。

這時我們不由得想到那個著名的洛倫茲「蝴蝶效應」理論，其實和這個理論相聯繫的還有一個被稱為「湍流發生機制」的觀點，認為向湍流的轉變由少數自由度決定，經過兩三次突變，運動就到了維數不高的奇怪吸引子上。這裡所謂的吸引子是指運動軌跡長時間之後的終極形態，它可能是穩定的平衡點或週期性的軌道，也可能是繼續不斷變化、沒有明顯規則或次序的許多回轉曲線。無論是蝴蝶效應還是湍流發生機制，其實都是對我們現在正在研究的激活數據學的一種理論上的關照和呼應。事實上，大數據乃至超數據時代的數據運動，就是這樣一種處於混沌和分形

之下的對數據運動軌跡及其規律的研究。許多看起來雜亂無章、隨機起伏的數據變化或時空穿越，可能造成的就是類似亞馬遜級別的數據風暴。如果說上述蝴蝶效應粉碎的是本就無法實現的長期天氣預報的幻夢，那麼緊接著的奇怪吸引子告訴我們的是，人類對於天氣的實際預報能力並沒有因那隻蝴蝶的翅膀而受到任何影響，相反，卻因對於更加混沌的數據的研究而提高了。激活數據學就是一種基於複雜理論及混沌研究的關於未來大數據乃至超數據時代的理論假說，就像上面講的天氣預報，但它所關心的並不是下個星期的晴雨冷熱，而可能是未來10年耕種季節的平均降水量和平均氣溫。激活數據學研究使以往根據統計原則所做的預報上升為數據動力學的預報，也就是應用了似是隨機現象的內在規律，從而提高了預測單個軌道近期行為的精確度，並豐富了長期預報的辦法。

同樣，我們還可以考察一個似靜實動的模型。讓沙子從一個漏斗孔中緩緩落到桌面上，形成漸漸變大的沙堆，總有最後新添加的某一粒沙子會在整個沙堆勉強維持平衡的錐面上導致一次「雪崩」，使一撮沙子滑到堆底，雪崩留下的小窪地會被後續的沙流填平，直到下一次更大的雪崩。在我們收集所有這些雪崩的數據後，可以發現它們的大小和間隔遵循某些數據動力規律，而沙堆模型無疑也啟發了我們對於數據激活狀態中的相變和突變的研究。無論數據的平衡態的相變或非平衡的臨界多麼不確定，可以確定的是，在搜索、融合、激活和碰撞等一系列狀態下，數據在某一個臨界點附近的擾動必然會導致某種全域性後果。當然，

是否存在可以被「激活」的「數據蝴蝶」或「數據吸引子」，還需要我們進一步探尋，但這並不否定我們的所願，而僅僅需要我們從實際數據的研究和挖掘中進一步加以發現。

作為一種理論假說，激活數據學就像一座朝向深邃的大數據宇宙的「天眼」。它是未來人類進入雲腦時代的預報，是關於混沌的數據世界的跳出決定論和概率論的非此即彼、亦此亦彼的複雜理論的大數據思維範式的革命。從一定意義上來說，大數據就是面向未來社會人類需要破譯的「基因」。正如因發明一種DNA（脫氧核糖核酸）快速測序方法而獲得1980年諾貝爾化學獎的吉爾伯特針對生物學研究範式的變化指出的，「正在興起的新的範式在於，所有的基因將被知曉，今後生物學研究項目的起點將是理論的。一位科學家將從理論的假設開始，然後才轉向實驗室去檢驗該假設」。是的，借助日漸深入的人工智能的發展，大數據的理論研究正在激發人類的新的假想和猜測。正是這種假想和猜測，讓我們以某種「對稱破缺」的方式去探知深邃未知的數據海洋，發現諸多社會發展法則背後產生影響甚至支配的物質和數字的力量。

人生是一種快變量，語言是一種慢變量，而數據將是一種突變量。雖然「未知」依然是現實的一部分，但是身處海量數據大爆發時代，人們堅信，未來已來！從「塊數據1.0」到現在的「塊數據4.0」，我們一直在持續探討這個已來的未來，尤其是基於對「以人為原點的數據社會學的範式革命」的認知。事實上，從一開始，我們就沒有把大數據僅僅看作所謂的「大」的數據，而是把大數據看作一種「活」的數據，因為只有激活，大數據才有生

命，才有社會屬性，才能成為未來世界人們賴以生存與發展的土壤和空氣。最後，套用《爆裂》一書中關於現代世界生存的九大原則中「系統優於個體」的表述：真正具有競爭性的是一個系統，而非一個特別強大的個體；是一套能夠保證不斷成功的制度，而不是一個天才個人的行為。同樣，激活數據學就是這樣一個思想的系統，就是要為我們身處的這個大數據時代找到一個解決方案，這個方案可以構建一個融合數據、計算和場景的系統，讓我們在大數據的時空中真正「思考和行動」起來。世界正處於根本結構性變革中，我們必須具備這樣一種能力，即下意識地適應和發現因不適應我們的舊習慣而被忽視的事情。

連玉明

大數據戰略重點實驗室主任

2018年4月3日於北京

## 目錄

總序 V

主編序 XI

緒論 大數據時代的解決方案 001

### 第一章 超數據時代的數據擁堵

第一節 小數據時代、大數據時代和超數據時代 012

(一) 小數據時代 012

(二) 大數據時代 016

(三) 超數據時代 021

第二節 奇點來臨：數據大爆炸 026

(一) 數據連接型社會：數據量化世界 026

(二) 數據大爆炸：海量、複雜與失控 032

(三) 數據失真、數據依賴與數據安全 035

第三節 數據擁堵與數據治理 038

(一) 數據擁堵的由來 038

(二) 從生命週期視角思考數據擁堵 040

(三) 數據擁堵的治理範式 044

## 第二章 激活數據學：基於塊數據理論的解決方案

### 第一節 複雜理論與塊數據 052

- (一) 複雜性的湧現 052
- (二) 塊數據的數據觀 055
- (三) 數據學與數據科學 058

### 第二節 激活數據學的提出 061

- (一) 激活數據學的由來 061
- (二) 激活數據學的理论框架 062
- (三) 激活數據學的時代價值 068

### 第三節 激活數據學與數據激活機理 071

- (一) 數據搜索：智能感知 071
- (二) 關聯融合：智能聚合 072
- (三) 自激活：智能決策 074
- (四) 熱點減量化：智能篩選 076
- (五) 群體智能：智能碰撞 077

## 第三章 數據搜索：智能感知

### 第一節 智能感知與交互 082

- (一) 生物感知 082
- (二) 機器感知 085
- (三) 交互識別 088

### 第二節 搜索引擎：連接人與信息 091

- (一) 從「尋物」到「搜數」 091
- (二) 谷歌搜索：讓流動的信息產生智能 095
- (三) 搜索引擎的工作原理 098

### 第三節 搜索引擎到人工智能的終極演進 103

- (一) 全局化範圍搜索 103
- (二) 智能化目標識別 106
- (三) 無界化協同感知 110

## 第四章 關聯融合：智能聚合

### 第一節 人腦信息的處理與融合 118

- (一) 對象感知 118
- (二) 情景關聯 120
- (三) 信息融合 122

### 第二節 智能數據處理 124

- (一) 大數據融合處理模式 124
- (二) 數據融合處理局限 127
- (三) 基於人腦模式的數據關聯融合 129

### 第三節 數據融合：構建新型數據關係 134

- (一) 降維去噪 135
- (二) 關聯識別 140
- (三) 融合重構 144

## 第五章 自激活：智能決策

### 第一節 腦認知與類腦計算 154

- (一) 神經元與神經網絡 154
- (二) 從學習到決策 157
- (三) 人腦智能決策對機器學習的啟示 163

## 第二節 讓機器像人一樣思考 166

- (一) 從「深藍」到「阿爾法元」 166
- (二) 構造人工神經網絡 169
- (三) 深度學習驅動機器智能決策 171

## 第三節 智能判斷與決策 175

- (一) 提取特徵 175
- (二) 構建模型 177
- (三) 決策輸出 181

## 第六章 熱點減量化：智能篩選

### 第一節 遺忘，是為了更好的記憶 187

- (一) 人腦的記憶存儲極限 187
- (二) 記憶的選擇性封存 192
- (三) 遺忘也是一種學習 196

### 第二節 刪除，數據取舍之道 199

- (一) 數字記憶是生物記憶的延伸 199
- (二) 全面數字存儲下的信息失控 203
- (三) 數字記憶與信息取舍 206

### 第三節 篩選，選擇最優決策 210

- (一) 數據匹配與簡約 210
- (二) 優化算力配置 214
- (三) 選擇最優算法 218

## 第七章 群體智能：智能碰撞

### 第一節 頭腦風暴：發現好想法和做出好決策 226

- (一) 創造力是發現好想法的源泉 226
- (二) 群體合作與互動 228
- (三) 群體決策與判斷 230

### 第二節 群體學習：從個體智能到群體智能 234

- (一) 個體智能的局限 234
- (二) 從生物群體到機器人群體 237
- (三) 群體機器人的行為協作 242

### 第三節 群體空間：人腦智慧和機器智能的交互 244

- (一) 人機優勢互補 244
- (二) 機器智能進階 246
- (三) 人機社會化協作 250

## 第八章 激活數據學的應用場景

### 第一節 激活數據學下的自動駕駛 258

- (一) 智能駕駛引領新一輪工業革命 258
- (二) 激活數據學在無人駕駛中的應用場景 261
- (三) 激活數據學為智能駕駛提供理論依據 264

### 第二節 激活數據學下的城市大腦 267

- (一) 城市大腦：城市的數據智能中樞 267
- (二) 激活數據學優化城市大腦的系統應用 269
- (三) 激活數據學讓城市大腦更智慧 273

### 第三節 激活數據學下的醫療影像 274

- (一) 人工智能賦能醫療影像 274
- (二) 激活數據學在醫療影像中的應用策略 276
- (三) 激活數據學提升醫療影像價值 280

- 第四節 激活數據學下的智能語音 282
- (一) 智能語音交互：進階的交互模式 282
  - (二) 智能語音技術提升的路徑選擇 284
  - (三) 激活數據學開啟語音交互新時代 287

## 第九章 雲腦時代：開啟數字文明新紀元

- 第一節 驅動雲腦時代的「三駕馬車」 294
- (一) 數據驅動 294
  - (二) 計算驅動 297
  - (三) 場景驅動 301
- 第二節 區塊鏈：人工智能任性發展的「保險閥」 303
- (一) 哲學視域下的人工智能風險 303
  - (二) 區塊鏈與秩序互聯網 309
  - (三) 區塊鏈重塑人工智能時代新生態 312
- 第三節 數權法與數字文明新時代 315
- (一) 雲腦時代的制度安排與法律規制 315
  - (二) 數權法構建數字文明新秩序 319
  - (三) 構建網絡空間人類命運共同體 322

參考文獻 327

術語索引 341

後記 349

緒論

### 大數據時代的解決方案

本書探討的主題是大數據時代激活數據學的提出、運行機理及場景應用。激活數據學是以充分發揮人機群體智能為核心，綜合運用數據科學、生命科學和社會科學提出的海量數據存儲、處理的解決方案。激活數據學將確立一個新的觀察人類智能和機器智能的視角，引導人們重新審視數據無限膨脹可能造成的人類認知障礙，重新思考維持一個健康、安全和有效的數字社會的根本辦法，建立與人類智能複雜性同步的人工智能系統，開啓用複雜性系統思維認識未來世界和改造未來世界之旅。

### 大數據時代面臨的問題與挑戰

在人類文明的伊始，人與人的第一聲交流即意味著「連接」的開始。語言使人與人連接，並促使用於記載事物的文字、數字

符號產生，這樣的「連接」便產生了「數據」，並演化為人類文明最初的信息與知識。在漫長的農耕文明時代，「連接」主要以語言溝通和書面文字溝通的形式存在。進入工業文明時代，「連接」開始通過無線電台、電報、電視的形式存在，但這樣的「連接」產生的信息往往是單向性且缺乏互動的。互聯網時代，人和人開始通過網絡進行複雜交錯的互動連接。社交網站、電子郵件、搜索引擎、聊天工具……人類建立連接的方式趨於多樣化、多維化，人類社會產生的數據也因而大量積累。與此同時，數據的價值越來越受到人們的重視。數據深刻作用於政治、經濟、文化等領域，帶來更多的創新機會，從生產、生活到科研，一個大數據時代正在開啟。

在美麗的貴州省黔南州布依族苗族自治州平塘縣，被稱為「中國天眼」的世界上最大的單口徑射電望遠鏡——FAST（500米口徑球面射電望遠鏡）已於2016年9月25日落成啟用。FAST的計算速度需達到每秒200萬億次以上，存儲容量需達到10PB<sup>1</sup>以上。這一世界級的工程將幫助人們捕捉到更多來自宇宙的信息，它的背後是「天文級」的海量數據存儲和複雜的計算。

隨著時間的推移、科學任務的深入，以及數據的大量採集，未來對計算速度和存儲容量的需求將爆炸式增長，數據量和計算量都將「大得驚人」。

數據是沒有邊際的，而計算力、存儲力始終存在物理極限。

---

1 1PB=2<sup>50</sup>B。

在過去很長一段時間裡，科技日新月異的發展依賴摩爾定律。當摩爾定律逐漸失效時，數據的膨脹是否將引發世界性的數字爆炸？

人們寄希望於量子計算，量子計算將有可能使計算機的計算能力遠遠超過今天的計算機，但仍然存在很多障礙。如何長時間地保持足夠多的量子比特的量子相干性，同時又能夠在這個時間段之內做出足夠多的具有超高精度的量子邏輯操作，提高所需量子裝置的準確性，還面臨許多困難。

人類作為自然界中最偉大的智能體已經進化了上百萬年，人類所具有的智能是大自然賦予人類最高級的禮物。今後，人類智能仍將是人造系統模仿和研究的對象。人類未來對數據的處理，要更多地模仿人腦對數據的處理方式，人腦就是世界上最好的「量子計算機」。

目前人工智能的發展水平可以用三個維度來描述，即強度、擴展性和能力。強度是指人工智能系統的智能化程度，擴展性是指人工智能系統可以解決的問題的範圍，能力是指人工智能系統所能提供的平均解決方案的質量。雖然我們已經在專用人工智能領域取得了突破性進展，但目前人工智能水平仍然停留在「有智能沒智慧，有智商沒情商，會計算不會算計，有專才無通才」的水平。其特點是能力單一，只能在某個特定領域內發揮作用。在人工智能2.0時代，我們期待一個新的框架：首先，在人的邏輯層次可辯明；其次，可以處理大規模數據；最後，可以基於一小組標記數據進行學習。我們認為，將數據驅動的機器學習方法與知

識引導方法相結合，將為人工智能的未來推開一扇新的大門。

### 探索解決之道——激活數據學

受人腦處理數據方式的啟發，本書提出了一種新的處理海量數據的一般性框架——激活數據學。

激活數據學從複雜理論出發，將產生智能所依賴的數據、軟件硬件環境、交互規則等視作一個具有「簡單性、局部性、全域性、內聚力、動態性」的開放複雜系統，智能產生的過程即開放複雜系統中各個具有局部目標和行為的自主體通過自主行為及動態環境相互作用達到整體的全域目標所湧現的群體智能，這種湧現的群體智能能夠高效求解問題。

激活數據學中採用塊數據模型作為數據模型。塊數據把各種分散的點數據和分割的條數據匯聚在一個特定平台上並使之發生持續的聚合效應。塊數據的聚合打破了傳統信息不對稱和物理區域、行業領域對數據流動的限制，高度關聯的各類數據可以在沒有任何障礙和限制的條件下自由流動、相互作用，為智能的產生創造了條件。

激活數據學的核心是將人類認知能力與計算機快速運算和海量存儲的能力結合起來。一方面，提高智能系統的置信度，避免人工智能技術的局限性所造成的風險甚至危害；另一方面，建立人類參與的混合增強智能，將海量數據通過人機混合增強智能實現最佳釋放。這種形態是人工智能可行的、重要的成長模式。

數據搜索、關聯融合、自激活、熱點減量化、群體智能是激活數據學的五個運行階段，構成了激活數據學模型化運行的完整流程（如圖 0-1 所示）。

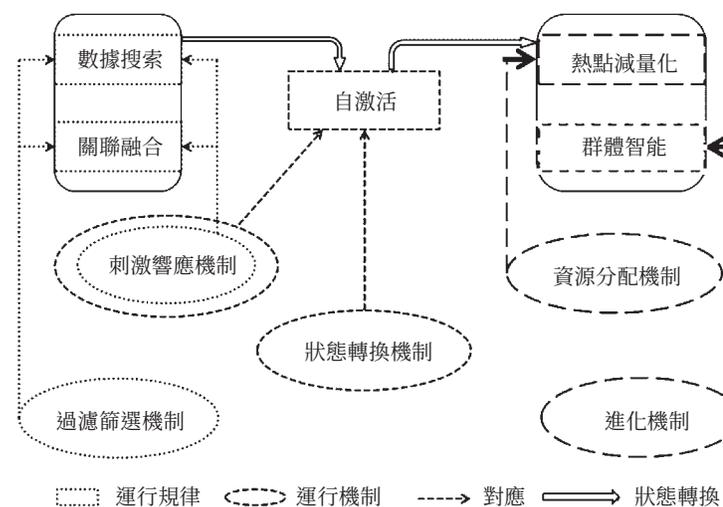


圖 0-1 激活數據學模型

**數據搜索**是激活數據學的數據準備階段，通過對數據的主動搜索和自我的深度學習，透徹感知、精確描述現實物理空間的多模態場景信息，構建更加豐富的數據特徵維度，更快、更精準地搜索到目標主體，為機器學習準備大量數據作為「燃料來源」和「助推器」。

**關聯融合**是激活數據學的數據預處理階段，對異構、多源、多模態感知數據進行信息融合，打破數據壁壘，讓不同領域、不同類型的數據成為高度關聯的整體，為智能的產生創造條件。

**自激活**是激活數據學的智能產生階段，是從「數據」躍遷為

「數聚」的階段，智能個體通過深度神經網絡學習將數據內化為知識，積累或學習經驗和知識，並可根據各自的意圖與其他智能體進行交互，修改自己的行為以適應新環境。

**熱點減量化**是激活數據學的智能篩選階段，對自激活階段產生的無效的、會對預判結果產生干擾的數據進行過濾和篩選，實現數據處理資源的優化配置，提高數據處理效率，為下一階段更好、更快的決策提供資源和環境保障。

**群體智能**是激活數據學的智能放大階段，通過人機協作強化，突破各自的局限和在某種程度上達到協同作用，釋放出超越個體智能的群體智能，提升決策的準確性和穩定性。

## 雲腦——人工智能的未來

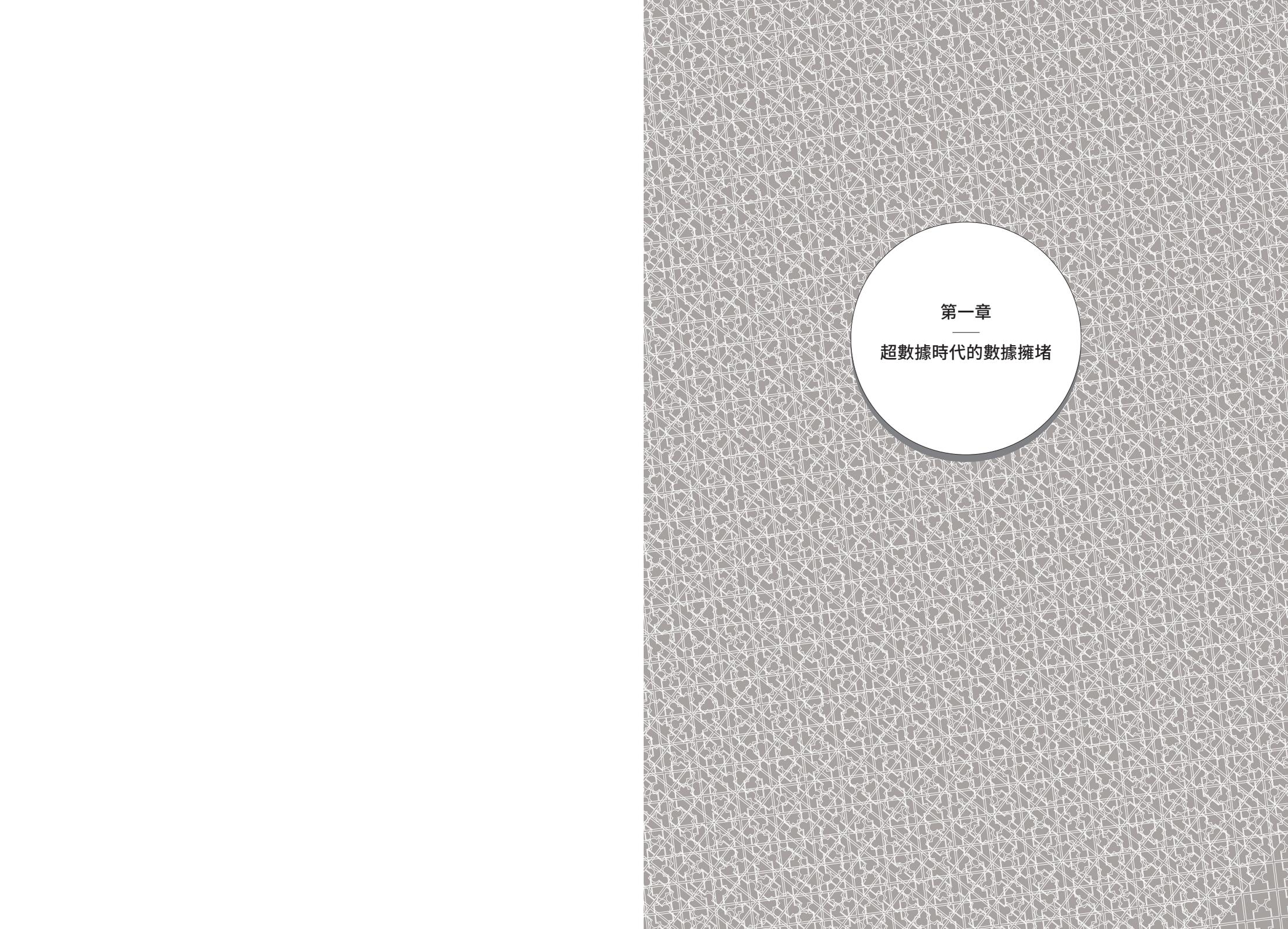
由於人類面臨的許多問題具有不確定性、脆弱性和開放性，任何智能程度的機器都無法完全取代人類。即使為人工智能系統提供了足夠或無限的數據資源，人類干預也不能排除在智能系統之外。人工智能中有許多問題需要解決，比如如何在人機交互系統中理解人類語言的細微差別和模糊性，特別是如何避免人工智能技術的局限性所造成的風險甚至危害。為了解決這些問題，一些重要的應用，如工業風險控制、醫療診斷和刑事司法系統，必須引入人的監督、互動和參與。

目前人工智能的學習方式已經挑戰了人類現有大數據處理能力的極限，動用了大量社會資源。比如讓計算機使用深度學習模

型觀看了1,000萬個視頻片段，才讓計算機學會如何在視頻中識別貓臉，這是一歲嬰兒幾天就可以學會的簡單的事情。真正的「強人工智能」則能靈活自主地學習自己想要學會的內容，能獨立思考，具備處理多類型任務和突發情況的能力，它可以在各種環境中解決複雜問題，更接近人類的思維，未來發展也更具價值。

未來已來，這將是一個電腦智能與人腦智能相互融合的時期——雲腦時代，這種融合將開創一個全新的世界，在這個新世界中，虛擬與現實、人類與機器的界限將變得模糊，人類與計算機的相互賦能或將永久改變人類的命運，在一個萬聯網、人、機所構成的未來世界，必將催生新的世界觀、新的科技秩序與道德秩序。

人工智能浪潮席捲而至，誰都無法阻擋。



第一章

超數據時代的數據擁堵

5G（第五代移動通信技術）、物聯網、人工智能等新一代信息技術與經濟社會各個領域快速發展和深度融合，數字、文字、圖像、音頻、視頻等數據形態大大豐富了數據種類，數據正以前所未有的速度極速增長和積累。這一深刻變化不僅促成了數據量及處理能力遵循摩爾定律爆發式增長，而且徹底顛覆了數據的傳統形態，將人類社會帶入了數據急劇膨脹的超數據時代。

在數據相對匱乏的小數據時代，數據的採集、搜索等技術手段的落後導致人類由於可獲得數據的有限性，難以對事物做出準確的判斷和預測，如同處在黑暗中，辨不清方向。進入超數據時代，由數據短缺變為數據過剩。信息爆炸與數據爆炸帶來海量信息、垃圾數據泛濫，使得人類被無邊界的數據層層包裹，最終形成認知障礙。我們把這種問題和困境稱為「數據擁堵」。

治理數據擁堵是超數據時代的重大議題。科學的問題要用科學的辦法來解決。在塊數據的思維結構下，融入數據科學、生命科學、社會科學、智能科學等前沿科學理論，在不明顯增加成本

的前提下盡可能地對垃圾數據、冗餘數據進行熱點減量，把超數據從「厚」做到「薄」，從「大」做到「小」，為人類社會清除認知障礙，平衡利益矛盾，提供可行方案。

## 第一節 小數據時代、大數據時代和超數據時代

人類原始時代早期所創造的數的概念、數的方法和數的科學，為東西方文化的發展提供了共同的智慧財富。人類對數據價值的認識可以粗略地分為三個階段：一是以計算機為基礎，追求數據精細化的小數據時代；二是以系統性數據資源為基礎，深入挖掘數據關係的大數據時代；三是以數據大爆炸為標誌，實現數據擁堵「治理」的超數據時代。

### （一）小數據時代

**數、數字和數據。**數的概念從萌芽到誕生經歷了極其漫長的歲月，可以追溯到人類蒙昧時期。數的概念始於人類在採集、狩獵等生產活動中，通過對不同事物的比較，逐漸認識到事物之間存在某種共通的特徵，即在同類事物中存在最小事物個體，即事物的單位性。同時，意識到非同類事物之間數量的其他共同特點，如在數量上相互間可以構成對應的關係，這種非同類事物所共有的數量的抽象性質，就是數。<sup>1</sup>數字是數的具體表現形式。公元前

1 劉紅，胡新和。數據革命：從數到大數據的歷史考察 [J]。自然辯證法通訊，2013，35（6）：33-35。

四五千年，尼羅河流域的古埃及人創造了十進制象形文數字；遠古時代，生活在中美洲的瑪雅人創造了瑪雅數字；黃河流域的中華民族創造了以商代的甲骨文數字和西周的鐘鼎數字為代表的中國數字，到唐代前後形成漢字數字。後來，隨著數學的發展，也逐步產生了不同於文字數字的符號數字，如羅馬數字、阿拉伯數字等。「數據」一詞最早源於拉丁語，「數」的概念是在「量」的基礎上進一步擴展而建立起來的，量成為數據的基本單位。在計算機普及化的今天，數字化已成為現實，因此數據形式已經擴展到量之外，數及可以轉換為數字的圖形、表格、文字都可成為數據的組成部分。數據不僅限於表徵事物特定屬性，更為重要的是成為推演事物運動、變化規律的依據和基礎。

**小數據時代是以經驗科學為基礎判斷數據價值的時代。**早在數千年以前，人類就開始計量數據、運用數據和分析數據。人類有記載的最早的計數發生在公元前 8000 年。《易經·繫辭下》有記載：「上古結繩而治，後世聖人易之以書契。」《易九家言》記載為：「事大，大結其繩；事小，小結其繩。結之多少，隨物眾寡。」在西方，自聖經時代開始，政府就通過人口普查建立大型的國民數據庫。同樣，在古代波斯也有結繩記事的記載。據說波斯王大流士給他的指揮官一根打了 60 個結的繩子，並對他們說：「愛奧尼亞的男子漢們，從你們看見我出征斯基泰那天起，每天解開繩子的一個結，到解完最後一個結那天，要是我沒回來，就收拾你們的東西，自己開船回去。」從古人結繩記事起，人類數十萬年依靠數量概念和數量科學推動著社會經濟與人類自身的

發展。人類的先知憑藉自身經驗和觀察所得，發現了數據對於自然界的物質生產、社會界的精神生產及人類自身的自我生產、存在與發展的重要價值。

**小數據時代，人類收集、挖掘和使用樣本數據的能力處於較低水平**，也沒有收集和處理大規模數據的能力，就利用整體與部分之間的關係發明一種方法來用部分的數據證實盡可能重大的發現，這就是統計學的由來。這種方法利用了整體與部分的關係，部分具有整體的一些特徵，可以從部分認識整體。在統計大規模數據時，人們會選取其中的一部分作為樣本，對樣本進行分析，從而達到對整體數據的推算，這就是採樣分析法。後來，人們意識到採集的樣本並不是越多越好，而是隨機性與精確性成正比，樣本選擇的隨機性比樣本數量更重要。從此，隨機採樣變得非常重要，成為現代測繪和現代測量領域的主心骨，它可以通過少量的隨機樣本，對整體數據進行推算和統計，而且擁有很高的精確率，這無疑是在小數據時代不可收集和分析全部數據的情況下所能達到的最完美的結果。<sup>2</sup>

人口普查是一種國家層次的重要的「數據指標行動」。據有關資料記載，中國是世界上最早統計人口的國家之一。相傳最早在公元前 210 年前的夏禹時代就有過人口統計。中國古代封建王朝設立「戶部」，戶部主管戶口、賦稅等，是負責統計人口的機構。西周的人口統計不但有公開的人口調查，而且有專司人口

統計的官吏，稱為「司民」。《周禮·秋官》載：「司民掌登萬民之數。自生齒以上，皆書於版。辨其中國，與其都鄙，及其郊野，異其男女，歲登下其死生。」這裡，我們不難看出，周朝時人口普查就已經初步設立了年齡、「國別」、城鄉、性別、生死等人口的重要指標。東漢時期的戶口調查進一步制度化，稱為「案比」，即案驗、比較，在每年的八月進行。中國魏晉時期皇甫謐著的《帝王世紀》記載：「禹平水土，還為九州，今《禹貢》是也。是以其時九州之地，凡二千四百三十萬八千二十四頃，定墾者九百三十萬六千二十四頃，不定墾者千五百萬二千頃。民口千三百五十五萬三千九百二十三人。」南朝宋范曄的《後漢書》與宋元之際馬端臨的《文獻通考》，都有同樣記載。有的統計學者認為這是「我國最早的統計數字資料」。在數千年的農業社會中，人類不斷探索新的科學技術，但是對統計數據的收集、挖掘和使用始終處於較低水平。

**計算機的誕生和發展促使小數據時代的分析方法由經驗化向精細化轉變**。在計算機出現之前，人類的經濟和政治生活根本就不是以數據為基礎的，數據僅僅作為一種計量單位而存在。自人類第一台計算機 ENIAC（埃尼阿克）問世以來，以計算機技術為主導的信息化、數字化時代的不斷發展，為數據收集、整理、分析和使用提供了前所未有的便利：數據收集更加便捷，數據整理更加科學，數據分析更加深入，數據使用更加廣泛。但是，這一階段計算機技術的主要應用範圍局限在數值領域，追求數據的豐富性和精細化。自 20 世紀 90 年代起，網絡技術、數碼技術和電

2 王浩。大數據時代下的思維方式變革 [D]。上海：東華大學，2015。

子信息系統的發展，推動計算機技術從數值領域發展到非數值領域。數據技術經歷了一次革命性的變化，多媒體技術將文字、圖形、影像、音響和動畫技術融為一體，數據的生產、複製和存儲能力急驟增強。世界各國相繼實施和推進數字化戰略，數字城市、數字社區和數字家庭不斷湧現。從磁盤、光盤到互聯網、傳統媒體數字化轉型，以手機帶動的新型移動傳輸方式的發展，大大提升了大規模數據傳輸速度。小數據時代逐漸向大數據時代過渡。

## （二）大數據時代

美國人邁克爾·考克斯和戴維·埃爾斯沃思被認為是最早提出「大數據」概念的工程師。1997年10月，邁克爾·考克斯和戴維·埃爾斯沃思在第八屆美國電氣電子工程師協會（IEEE）關於可視化的會議論集中發表了題為「為外存模型可視化而應用控制程序請求頁面調度」的文章。「可視化對計算機系統提出了一個有趣的挑戰：通常情況下數據集相當大，耗盡了主存儲器、本地磁盤甚至遠程磁盤的存儲容量。我們將這個問題稱為『大數據』。當主存儲器無法容納數據集，或者當本地磁盤都無法容納數據集的時候，最常用的解決辦法就是獲取更多的資源。」該文是在美國計算機學會的數字圖書館中第一篇使用「大數據」這一術語的文章。

2000-2010年被視為「大數據時代」奠基的10年。中國互聯網數據中心估計，2002年世界產生了5EB<sup>3</sup>新數據，2006年產生了161EB新數據，2006-2010年，每年為數字宇宙所增加的信息將是

3 1EB=2<sup>60</sup>B。

2006年的六倍多，達到988EB，或者說每18個月就翻一番。實際上，據2010年和2011年同項研究所發佈的信息，每年全球所創造的數字化數據總量超過了這個預測，2010年達到了1,200EB，2011年增長到了1,800EB。<sup>4</sup>

2012年，全球數據量已經從TB<sup>5</sup>級別躍升到PB、EB乃至ZB<sup>6</sup>級別。國際數據公司（IDC）的研究結果表明，2008年全球產生的數據量為0.49ZB，2009年的數據量為0.8ZB，2010年的數據量為1.2ZB，2011年的數據量更是高達1.82ZB，相當於全球每人至少產生200GB<sup>7</sup>的數據。到2012年為止，人類生產的所有印刷材料的數據量是200PB，全人類歷史上說過的所有話的數據量大約是5EB。美國IBM（國際商用機器公司）的研究稱，在整個人類文明所獲得的全部數據中，有90%是過去兩年內產生的。到2020年，全世界所產生的數據規模將達到今天的44倍。經過10年的發展，新的數據標準、規則更加成熟，無線通信新技術在企業生產、市場流通與大眾消費領域日益擴大。在雲計算普及化及信息環境更加完善的前提下，越來越多的企業、社區和家庭使用更高級別的數據標準，各種層次和各種功能的數據中心如雨後春筍般應運而生，數字城市、智能網絡和數據系統不斷湧現。

大數據時代與小數據時代的對比見表1-1。與小數據時代的

4 高書國。大數據時代的數據困惑——教育研究的數據困境[J]。教育科學研究，2015（1）：24-30。

5 1TB=2<sup>40</sup>B。

6 1ZB=2<sup>70</sup>B。

7 1GB=2<sup>30</sup>B。

數據觀重視數據的因果關係有所不同，在大數據時代，數據處理變得簡單快捷。利用所有的數據，而不再僅僅依靠一小部分，並且允許所有數據不精確性的存在，「不再渴求數據的因果關係，而去關注數據的相關關係，不需要知道為什麼，只需要知道是什麼」<sup>8</sup>。「思維的轉變顛覆了人類千百年的思維習慣，全新挑戰了人類的認知和與世界交流的方式。它們是相互聯繫和相互作用的。」<sup>9</sup>

表 1-1 大數據時代與小數據時代的對比

對比內容	小數據時代	大數據時代
數據量	數據量小，以 MB <sup>10</sup> 、GB、TB 為存儲單位	數據量大，以 PB、EB、ZB 為存儲單位
數據庫	結構化數據庫存儲數據	結構化及非結構化數據庫存儲數據
存儲設備	數據存儲設備價格高昂	數據存儲設備價格相對低廉
數據來源	數據來源簡單	數據來源繁多複雜
產生速度	數據產生、變化速度慢	數據產生、變化速度快
數據結構	數據結構簡單	數據結構多樣

**全部數據取代隨機樣本。**在信息處理能力差的時代，隨機樣本讓大量數據的分析成為可能，但當完整的數碼歌曲被截取成多個採樣的文件時，很多信息也會隨之丟失。因此，只有擁有全部的數據或幾乎全部的數據，人們才能進行新的分析，才能正確地

8 李軍。大數據從海量到精準 [M]。北京：清華大學出版社，2014：41。

9 維克托·邁爾—舍恩伯格，肯尼思·庫克耶。大數據時代 [M]。盛楊燕，周濤，譯。杭州：浙江人民出版社，2013：28。

10 1MB=2<sup>10</sup>B。

考察細節，才能用大數據在任何細微的層面論證新的假設。舍恩伯格曾說：「在某些特定的狀態下，我們還是可以使用樣本分析法，但它已經不是分析數據的主要方式。在未來，我們會完全拋棄不再使用。」<sup>11</sup>統計抽樣方法適用於技術受限的時期，是為解決一些特定問題而存在的。如今，技術環境已大大改善，在大數據時代進行抽樣分析就好比在汽車時代騎馬一樣。

**混雜性取代精確性。**在小數據時代，為了保證質量，人們往往把減少錯誤作為最基本、最重要的要求。那是因為收集的信息是有限的，細微的錯誤會因數據量小而被放大，甚至會影響整個結果的準確性。在當今的信息時代，我們已經從依賴清晰、準確的領域脫身，能夠容忍模糊和不確定性的出現。只要一個事物有一個更完整的概念，我們就能接受模糊和不確定性。好比印象派的畫風，近看畫中每一筆都感覺很亂，但後退一步就能發現這幅作品的偉大，因為整幅畫的整體思路在後退一步的時候就能顯現出來。也就是說，當視野局限在可以分析的數據和能夠確定的數據上時，我們對世界的整體理解可能會發生錯誤和偏差，不但會喪失盡力收集數據的動力，而且會失去從不同角度觀察事物的權利。「大數據要求我們必須能夠接受混亂和不確定性。一旦我們承認甚至擁護這個事實的話，離真相就又近了一步。」<sup>12</sup>

11 維克托·邁爾—舍恩伯格，肯尼思·庫克耶。大數據時代 [M]。盛楊燕，周濤，譯。杭州：浙江人民出版社，2013：43。

12 維克托·邁爾—舍恩伯格，肯尼思·庫克耶。大數據時代 [M]。盛楊燕，周濤，譯。杭州：浙江人民出版社，2013：66。

**相關關係取代因果關係。**在小數據時代，因為數據很少且收集數據費時費力，所以相關關係的應用也很少。但在大數據背景下，我們能夠更便捷、更容易和更清楚地分析事物，能夠量化兩個數據之間的數理關係，因此可以更好地運用數據的相關關係。比如谷歌（Google）預測的流感趨勢是通過人們在谷歌搜索的流感詞條確定的，搜索多的特定地區被認為有更多的人患了流感。相反，個人鞋碼與幸福指數根本扯不上任何關係。「在大數據時代，我們只需要知道『是什麼』，不需要知道『為什麼』。不必尋求現象背後的原因。」<sup>13</sup> 這個時代要釋放出的巨大價值是通往未來的必然改變，使我們選擇大數據的理念和方法不再是一種權衡。大數據的趨勢就是越來越多的事物都以數據的形式存在。

綜合各方對大數據時代的分析，我們認為大數據時代具有以下五個典型特徵。第一個特徵是數據量大。大數據的起始計量單位至少是 PB、EB 或者 ZB。第二個特徵是數據類型多樣化。大數據時代的數據類型和表現形式多樣，包括調查數據、網絡日誌、音頻、視頻、圖片和地理位置信息等，數據與數據之間的聯繫被數據的多樣性沖淡，多種類型的數據對數據的處理能力提出了更高的要求。第三個特徵是數據價值密度相對較低。隨著物聯網的廣泛應用，信息感知無處不在，信息海量，但價值密度較低。如何通過強大的機器算法更迅速地完成數據的價值「提純」，是大數據時代亟待解決的難題。第四個特徵是處理速度快。在數據收

集速度加快的同時，數據壽命明顯縮短，對數據挖掘的時效性要求日益提高。這是大數據區分於傳統數據挖掘最顯著的特徵。第五個特徵是關鍵數據仍是稀缺「資源」。由於網絡、視頻、掃描等數據採集工具不斷豐富，應用範圍日益擴大，流量數據十分龐大，並且佔據了大量的人力、物力和財力。數據量的增加，並沒有滿足人們對於數據質量的要求，對研究有用的關鍵數據依然稀缺。

### （三）超數據時代

5G、物聯網、人工智能、雲計算等技術的快速發展，以及視頻監控、智能終端、應用商店的快速普及，光學觀測、光學監控、健康醫護、傳感器和數據服務公司及供應鏈系統等都產生了海量、複雜、多元的數據，對這些數據更恰當的描述應該是「無限」的數據，數據都是源源不斷生成的，我們將這種數據稱為「超數據」。超數據是大數據發展的後半場，是大數據時代全面採集、存儲、積累的必然結果。隨著時間的推移，超數據的數據體量還在不斷膨脹，不論人類是否願意，都將邁入超數據時代。

「各種經濟時代的區別，不在於生產什麼，而在於怎樣生產，用什麼勞動資料生產。勞動資料不僅是人類勞動力發展的測量器，而且是勞動藉以進行的社會關係的指示器。」<sup>14</sup> 馬克思按照勞動資料或勞動工具的標準，把人類社會發展階段分別稱為石器時代、

13 張蘭廷。大數據的社會價值與戰略選擇 [D]。北京：中共中央黨校，2014。

14 馬克思，恩格斯。馬克思恩格斯全集：第 44 卷 [M]。北京：人民出版社，2002：210。

青銅時代、鐵器時代、大機器時代。馬克思沒有看到信息時代的到來，但當信息技術成為非常重要的生產資料或者生產工具的時候，我們還是依據馬克思的理論，稱這個時代為信息時代。如今超數據出現，作為新的生產資料，不斷體現出其在生產活動中的巨大作用，我們自然而然地應該去思考超數據時代的到來。「從事物的本性可以得出，人的勞動能力的發展特別表現在勞動資料或者說生產工具的發展上。」<sup>15</sup> 工具是生產力發展水平的重要標準，而生產力發展水平所決定的生產關係則是一個時代的本質特徵。超數據是作為一種新的生產資料出現的，對生產力的發展有著直接的推動作用，這也是超數據時代會被稱為一個時代的原因。超數據時代，人們的思維方式也發生了巨大的變革，總體呈現預測性趨勢、模糊性趨勢和複雜性趨勢。

**預測性趨勢：用數據看未來。**超數據時代帶給我們巨量的數據和先進的數據分析技術，以及二者的結合帶來的我們最為關心的一項能力——預測。大量的傳感器將我們身邊的一切物體納入物聯網，使一切事物的動態、變化都變成大量的數據流不斷進入負責監控的計算機。基於雲計算技術的強大數據分析能力則將對這些數據進行分析處理，得出的結果有助於人類把握事物現時的情況，同時預測其下一步的發展。超數據不但可以預測事物的發展狀況，而且可以預測人類行為。艾伯特—拉斯洛·巴拉巴西教

授在《爆發》一書中表示，人類行為的93%是可以預測的。<sup>16</sup> 超數據時代的數據監測可以將人們的行為轉化為數據，然後通過人類行為預測模型對其進行處理，就可以預測人們的行為。但是預測並非預言，超數據能做到的是對短期內影響因素較少的事物的發展進行預測，這種預測有著極大的限制，並非臆想中的無所不能、無所不知。儘管如此，超數據的預測能力還是為人們看向未來開了一扇窗。在超數據的幫助下，人們不再是摸著石頭過河，而是可以站得高一些，稍稍看清前方的路了。這種轉變對人類來說是非常重要且意義重大的。人們對於未來不再是徬徨無措、一無所知的，而是可以通過超數據的能力對未來進行推測，這是人類思維方式變革的一個大方向。

**模糊性趨勢：用概率來表達。**當「精確」不能解決全部問題時，我們就需要換一個角度，考慮試試「模糊」這條道路行不行得通。1965年美國數學家扎德發表了論文《模糊集合》，一門叫作模糊數學的學科誕生，同經典數學不同，它是研究模糊現象的一門數學。經典數學以精確為準，面對許多不能精確定義的事物，經典數學很難對其進行研究，這就催生了模糊數學。模糊數學的產生說明世界上的許多事物是不能用精確來解決的，過去科技不發達，認為是不夠精確，現在發現事物本身就存在模糊性，用精確的手段自然不能對其進行解釋和處理。在超數據時代，我們發現了更多的模糊性事物，我們的思維方式也必須從過去的精確性思維方

---

15 馬克思，恩格斯。馬克思恩格斯全集：第47卷[M]。北京：人民出版社，2002：57。

---

16 亞伯特—拉斯洛·巴拉巴西。爆發：大數據時代預見未來的新思維[M]。馬慧，譯。北京：中國人民大學出版社，2012：109。

式向模糊性思維方式轉變，學會用概率和數據說話，這樣我們才能更好地適應和推動科技的進步與社會的發展。另外，數據的模糊性還來自數據的生長性，超數據時代大多數的數據不是靜態的，而是不斷生成、不斷變化的動態數據，對於這種具備生長性的數據，很難做到精確地、簡單地定性，所以需要我們用模糊的和概率的數據來表達。因此，超數據時代，接受了錯誤和混雜，認識到數據的動態變化，我們的思維方式必將展現出一種模糊性的變化趨勢。

**複雜性趨勢：用數據來跨界。**超數據時代的研究範式打破了傳統的機械思維和還原方法論的統治，同複雜性科學研究方法類似，可以說超數據時代的研究方法本身就是一種複雜性科學，而這種複雜性科學也代表了超數據時代人類思維方式向複雜性發展的趨勢。複雜性科學將一切對象都看作有生命、會演化的系統，幾個最簡單的要素通過非線性的相互作用，也有可能湧現出複雜的行為，我們不能根據簡單的因果關係推導系統的行為。超數據時代的研究範式恰恰就是通過數據之間的關係研究事物之間非線性的相互作用。超數據時代對複雜性科學將起到巨大的推進作用，也會形成人類思維方式的複雜性變化趨勢，人們眼中的世界將不再是簡單的、可以被分割的一個個獨立的個體，而是互相有聯繫的複雜的系統，而且這個複雜的系統是動態的，時刻都在變化。過去的數據是某個時間採集到的靜態數據，這種數據是靜態的、有時滯性的；超數據時代的數據都是不斷變化的、隨時隨地都可以採集到的動態數據，可以直接反映當前的動態和行為。超數據

時代的數據在採集、存儲、傳輸、處理和使用中不斷產生和獲得最新數據，這種動態的技術手段對人們的思維方式也將產生巨大的影響。在人們的思維領域，「現在」的概念將被放大，「現在」就將是現在，不是一天前，不是一小時前，甚至不是一分鐘前，而是說話的此時此刻。

超數據時代還會將事物的變化放大，使人們認識到世界上沒有什麼是一成不變的，要用動態的眼光看待世界。恩格斯指出：「當我們深思熟慮地考察自然界或人類歷史或我們自己的精神活動的時候，首先呈現在我們眼前的，是一幅由種種聯繫和相互作用無窮無盡地交織起來的畫面。」<sup>17</sup>馬克思和恩格斯認為這個世界的一切都是聯繫和發展的，而事物之間的聯繫則是通過「中介」進行的，那麼在超數據時代，數據是不是世界聯繫的重要「中介」之一，一個容易被發現、容易被捕捉的「中介」，我們通過這個「中介」可以研究許多之前研究不了的事物間的關係？另外，數據的動態變化監測能力能夠降低我們研究世界的發展變化的難度。從馬克思的理論我們可以看出，超數據時代的研究正在朝著正確的方向進發，不斷將這個世界清晰地還原到人腦之中。在超數據時代，複雜性的、動態的思維方式將被樹立，人們的思維方式也將呈現複雜性的變化趨勢。

---

17 馬克思，恩格斯。馬克思恩格斯選集：第3卷[M]。北京：人民出版社，1995：521。

## 第二節 奇點來臨：數據大爆炸

### （一）數據連接型社會：數據量化世界

超數據時代，5G、物聯網、人工智能等新技術、新工具和新應用的不斷發展與應用，知識和信息的製造、組織、發佈與交流形式似乎每過一段時間就會被改寫，私人生活和公共生活正在被重塑，一切皆可量化，萬物皆為數據。數據驅動著人們的日常生活，改變著政治決策、公共輿論、政府監管和社會治理方式，數據連接型社會孕育而生。

普惠泛在的信息網絡體系<sup>18</sup>是構建數據連接型社會的基礎。繼美國的「智慧地球」、日本的「U-Japan」<sup>19</sup>、韓國的「U-Korea」<sup>20</sup>、歐盟的「物聯網行動計劃」後，中國提出「感知中國」的信息技術發展戰略。<sup>21</sup>「感知中國」戰略的核心是抓住信息技術躍變的機遇，提升自主創新和可持續發展能力，使我國全面進入信息社會。信息成為中國經濟和社會發展最重要的資源，中國社會信息化總體上接近發達國家。信息社會的發展可粗略地分為 e 社會和 u 社

18 普惠泛在的信息網絡體系，即 U-INS 體系（Universal, User-Oriented, Ubiquitous Information Network System）。Universal，全民普及、惠及全民；User-Oriented，面向用戶和以用戶為中心；Ubiquitous，網絡通信和服務無處不在。

19 U-Japan，是日本於 2004 年推出的基於物聯網的國家信息化戰略。這裡的「U」代指英文單詞「ubiquitous」。——編者注

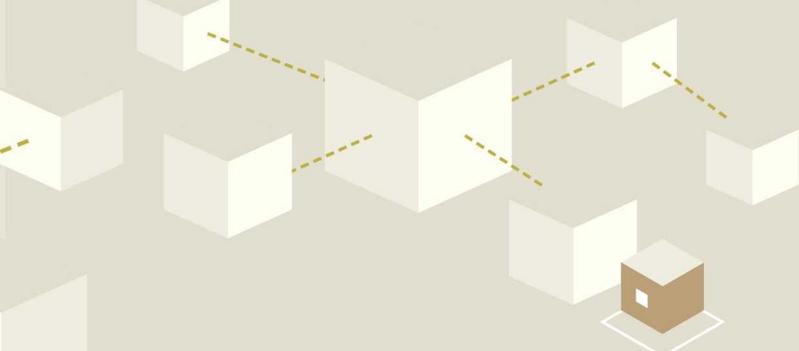
20 U-Korea，是韓國於 2004 年推出的基於物聯網的國家信息化戰略。——編者注

21 許金葉，袁樹民。基於人機物三元世界成本信息系統的數字鴻溝 [J]。會計之友，2012（2）：25。

會，e 社會是信息社會的初級階段，u 社會是信息社會的高級階段（見表 1-2）。2020 年以前我國要為邁向信息社會奠定堅實的基礎，稱為 e 社會；2020 年以後的目標是向 u 社會過渡。u 社會也就是我們所暢想的數據連接型社會，其構建關鍵在於普惠泛在的信息網絡體系的建設。普惠泛在的信息網絡體系包含六個方面的內容：一是支撐網絡、具有變革性的器件與系統，二是面向大眾、普及全民的網絡系統，三是安全可靠、個性化的網絡服務技術，四是支持產業升級和發展的數據知識產業，五是網絡科學與新的信息科學，六是國家與社會信息網絡安全體系。普惠泛在的信息網絡體系實現了與空間、地面、接入等網絡的全面融合，實現了人與人、機器與機器、人與機器之間任何時間、任何地點的通信聯絡，網絡通信無所不在且有可靠的服務保證，通信成本極低，為構建數據連接型社會奠定了良好的網絡基礎（見表 1-3）。

表 1-2 信息社會的初級階段（e 社會）和高級階段（u 社會）

		2020 年（e 社會）	2050 年（u 社會）
技術普及度	電腦普及	電腦擁有量超過 4 億台，電腦普及率達到 28%	電腦擁有量超過 8 億台，電腦普及率超過 50%
	網絡普及	網民數超過 5 億（含手機上網），網絡普及率超過 35%	網民數超過 12 億（含手機上網），網絡普及率超過 80%
	簡便易用	很多人會用電腦	絕大多數人會用電腦



塊數據強調的是數據、算法、場景融合應用的價值體系，是大數據時代的價值觀和方法論。數據、算法、場景是治理科技的三大核心要素。

——日本《日中商報》

共享是新一輪科技革命和產業變革的關鍵力量，借助共享，人類文明必將走向更高階段。可以說，基於數據力與數據關係提出的「共享價值理論」，是繼剩餘價值理論之後頗具革命性的理論。

——南非《非洲時報》

互聯網帶來了超越空間的資訊傳遞、共享與價值交換、增值，卻面臨著從無界、無價、無序走向有界、有價、有序的挑戰。《塊數據 3.0》提出的主權區塊鏈，為從信息互聯網、價值互聯網向秩序互聯網的演進帶來了暢想空間。

——英國《華商報》

面對超數據時代帶來的數據擁堵難題，《塊數據 4.0》運用激活數據學有效挖掘、分析大數據背後人的運行規律，把超數據從「厚」做到「薄」，從「大」做到「小」，為大數據時代清除認知障礙，平衡利益矛盾，讓不確定性對抗確定性成為可能。

——西班牙《僑聲報》

《塊數據 5.0》圍繞構建以人為原點的數據社會學理論與方法，創新性地提出數據進化論、數據資本論、數據博弈論的理論體系，研究和探索人與技術、人與經濟、人與社會的內在關係，以此分析人的行為、把握人的規律、預測人的未來。

——美國《中美郵報》

ISBN 978-988-8568-75-8



9 789888 568758 >



專業出版 國際銷售

紅出版文化平台

加入我們：[www.red-publish.com](http://www.red-publish.com)

ModE.

上架建議：資訊管理

定價：港幣 148 元正 / 新台幣 596 圓正