



塗鴉塗出了數學博士

再過幾十年，世上最偉大的發明家可能是個電腦程式。

在1980~90年代，波蘭數學家法托維茲（Siemion Fajtlowicz）設計出一個叫做「塗鴉」（Graffiti）的電腦程式，號稱能夠自動發現與圖學相關的不等式猜想，例如其中一個典型猜想是：圖學中一個圖的路徑總數會大於它的直徑減掉一。

「塗鴉」的運作原理，基本上是結合統計方法以及人工智慧中的「啟發法則」（heuristics），在現有的圖學資料庫中自動搜尋可能正確的猜想，問世至今一共自動產生了800多個猜想；更令人讚歎的是，這些結果公佈至今不僅引起許多圖學專家的注意，而為了證實這些猜想的真偽，前後更造就了超過60篇的數學會議、期刊、學位論文，其中甚至包括數學博士論文。

利用電腦高速運算的能力併以特殊的搜尋演算法，從相關資料中發現全新的知識或是現象，稱為「機器發現」，「塗鴉」不是唯一利用機器發現的電腦程式。1990年問世的ASTRA系統，能利用量子力學的知識來自動預測並解釋一些從未被發現的反應路徑，如 ${}^3\text{He} + {}^7\text{Li} \rightarrow {}^9\text{B} + n$ 即為一可以產生中子的融合反應路徑。在天文學方面，美國加州理工學院設計的系統SKICAT，能自動歸納與整合天文資料，並且能自我學習，至今已經發現了至少五個新的類星體，效率不輸一般天文研究團隊。此外，機器發現在古文解碼上也有貢獻：2005年，筆者利用期望值最佳化演算法（EM algorithm），成功以電腦解出象形盧維語（中亞出土的二維象形古文）的線性讀法，並推展至能發現任意二維文字的線性解。除了知識的發現，電腦還被用在發現特殊個體上，比如911之後，美國便極力研究能夠自動發現可疑人物的系統。

上述的機器發現系統都充份利用了電腦高速運算及資料儲存的能力，來分析一般人無法快速而精確處理的資料。然而若仔細比較每個系統的核心方法，又會發現其各不相

同。正如歷史上所有重大的發明，幾乎沒有一個是來自相同的方法、過程或背景。

電腦擁有學習的能力已經不足為奇：只要有足夠的案例，佐以分析的演算法，電腦便能自動由這些例子歸納出相關的規則、進而舉一反三。這項技術稱為「機器學習」，已被廣泛利用在如機器翻譯、語音辨識等產品上。

「學習」跟「發現」最大的不同在於「學習」的過程中有範例或老師指導，而「發現」的過程中卻沒有，所以難度遠大於機器學習。「塗鴉」這類程式把人工智慧從「學習」提升到「發現」的境界，是個重大的突破。

綜觀目前機器發現系統的發展，機器發現的程式至少應具備以下三個要素：首先，電腦必須能夠從資料中自

動找出值得注意的現象，這些現象可能不符合從既有資料分析得出的模型的常態，以激發更進一步的探究。這

塗鴉程式把人工智慧 從「學習」提升到「發現」

方面的相關研究在資料探勘領域中稱為「關注性量測」（interestingness measure）。第二，機器發現程式必須要能對領域的背景知識建立計算模型，並能在模型中從事啟發式搜尋（heuristic search），比如說筆者以雜訊通道（noisy channel）模型來模擬古文寫作過程，於是可以利用相關的演算法找出解答。最後，機器發現系統最好能有「自我評估與解釋」的能力。也就是說，其不僅要能夠發現新的知識或現象，還要解釋為什麼它做出這樣的決定。如此人們便可藉由解釋判斷結果的真偽，並評定它的價值。這樣的解釋系統通常需要有特徵擷取（feature selection）與自動語言生成（natural language generation）的功能。

網路的崛起讓我們容易取得各種資料，也給予機器更寬廣的空間從事探勘與發現。或許2050年，《時代》雜誌公佈21世紀上半影響世界最深的百大發明家時，記者得絞盡腦汁想著如何採訪那個名為「牛頓」的電腦程式。

林守德 台灣大學資訊系及資訊網路與多媒體研究所助理教授