撰文／林守德

深度學習的深度

模擬人腦的分層次學習，讓電腦辨識能力大幅躍進！

讓電腦具備與人類相同的聽覺、視覺、閱讀甚至翻譯的能力，一直是人工智慧學家追求的目標之一。衍生的研究領域包括：能夠辨識影像中物體的電腦視覺、聽聲辨語的語音辨識，以及閱讀並翻譯文字的自然語言處理。過去半世紀，從最初的「規則式專家系統」到近年流行的「統計機器學習」，電腦在這些領域的辨識能力已有長足進步，但仍比不上可分辨貓狗的三歲孩童。

我們不禁要問，為什麼人腦的辨識能力這麼好？認知科學的研究指出，人類大腦是高度複雜的結構，能藉由許多不同層級的處理，把接收自外界的訊號轉換成人類能理解的訊息。例如，視覺系統會把可見光分階段處理再傳遞到大腦：進入眼球的光經折射後到達視網膜，其上的感光細胞先對不同的刺激（例如顏色、形狀、深度）做出反應，再經由視神經把訊息傳至外側膝狀核（LGN）進行中繼處理，接下來依序傳到初級視覺皮質V1以及之後的V2、V3、V4等層級，越後面的階段處理越高階的視覺訊息。人類辨識能力強，原因在於視覺系統不僅能夠很快接收並處理低層次的刺激（例如分辨顏色與形狀），也能迅速從事高層次的判斷（例如臉部與物品辨識），進而到更高階的認知處理（例如理解對方的動作、目的）。

於是有些機器學習研究者提倡，讓電腦也用這種「多層次」的處理方式從事辨識活動，較低的層級處理初始訊號、較高的層級辨識比較抽象的意義。這種類比人腦學習的模式稱為「深度學習」（deep learning），主要精神是假設一些基礎的訊號（例如聲音的波動）經過一些組合，可以成為較具體的物件（例如聲波組合成單字的讀音），然後具體物件的組合，又可以代表更抽象的意義（例如一組讀音成為一個有意義的詞或句子），所以電腦「學習」這些訊號的過程也應該分層次。

深度學習技術能否成為成就人工智慧的關鍵？

深度學習可看成是一種資訊的表達方式，而「多層次神經網絡」則是最適合實現這種表達方式的技術。多層次神經網絡在最底層把資料分散成單元，我們可以把每個單元想成是一個神經元，這些神經元經過組合的結果，成為第二層的輸入，第二層的輸入又經過組合產生新的結果，進而變成第三層的輸入，如此一層一層訓練出來的機器學習模組，辨識能力將越來越強。除了多層次神經網絡之外，更複雜的深度學習模型，例如多層次的機率圖型模型也陸續提出。

深度學習的觀念其實在20幾年前就出現了，但是因為容易產生「過度訓練」（overfitting）導致成效不彰，加上訓練模型速度慢，一直未受注目。近年來，過度訓練的缺點得到新解法，加上電腦運算速度增加，深度學習不再是紙上談兵。自從2009年具有記憶能力的深度學習模型「長短期記憶」（long short-term memory）模型贏得手寫辨識比賽冠軍，證明深度學習的效果之後，越來越多專家也利用類似的方法來解決問題。Google成立了Google Brain團隊，專門研究深度學習技術，已開發出利用GPU的快速運算單元來加速深度模型。目前在語音、手寫、影像辨識以及機器翻譯等領域中，深度學習的成果都能夠領先其他方法，在某些影像辨識的任務中甚至可以達到人類的水準。

模仿人類腦部多層次辨識的深度學習，為「強人工智慧」（strong AI）帶來一絲曙光。當然，它也有受質疑之處：首先，目前仍缺乏穩固的理論基礎來說明深度學習在什麼條件下能成功；其次，它無法明確表達例如因果關係與邏輯運算等更深入的智慧行為。因此，雖然深度學習的技術已把電腦辨識的能力大幅提升，但要把它當做成就人工智慧的關鍵，可能仍是過於樂觀的期待。 SA

林守德是台灣大學資訊工程系副教授。

影像來源：林守德