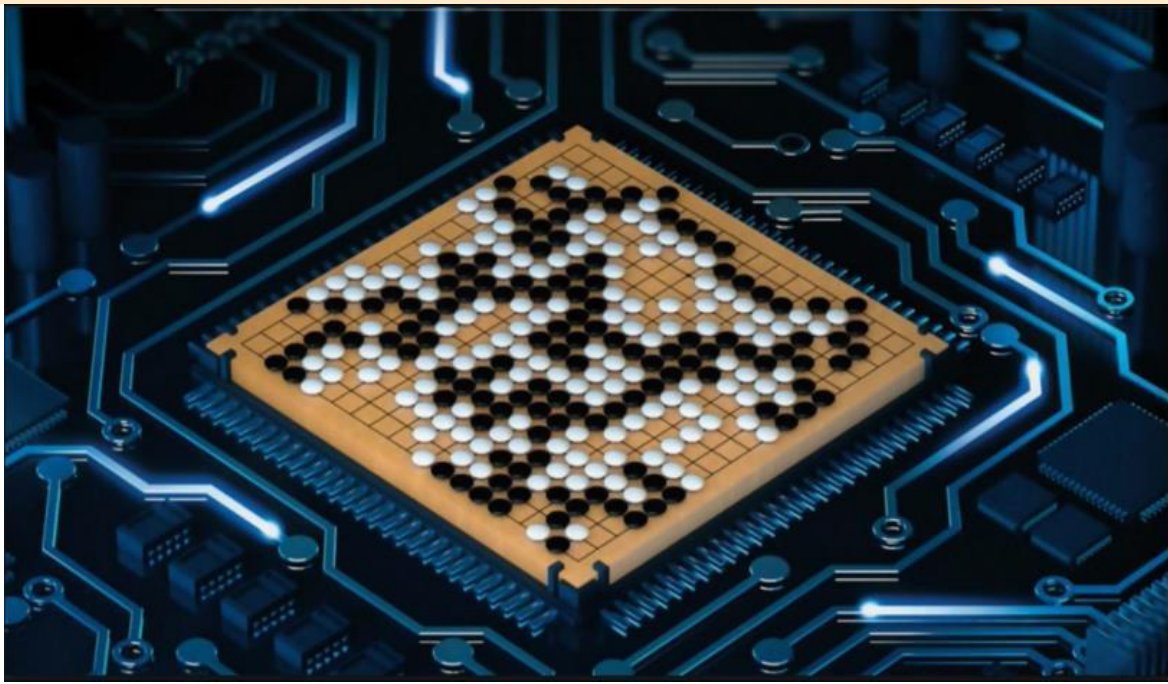


人工智慧如何用來下圍棋（二）：從 AlphaGo 到 AlphaGo Zero

2018/10/10

王庭毅 | 國立中山大學資訊工程學系

世界圍棋冠軍李世乜敗給人工智慧——「AlphaGo」，AlphaGo 的研發期程雖然僅短短幾年，但卻能有如此顯著的成就表現，其中之奧祕值得探究。在本篇文章中，我們將要介紹 AlphaGo 的發展史。



▲AlphaGo 示意圖。(圖片來源：Deepmind / 網址：
<https://deepmind.com/research/>)

2016 年 3 月，韓國籍世界圍棋冠軍李世乜敗給人工智慧——「AlphaGo」，而後 2017 年 1 月升級版的 AlphaGo Master 更是橫掃千軍，以 60 連勝之姿稱霸這項蘊含人類千年發展的智慧結晶 - 圍棋。AlphaGo 被 Google 旗下的英國公司 Deepmind 僅花 2 年研發的電腦程式，為人工智慧的新時代吹響號角。然而，AlphaGo 的研發期程雖然僅短短幾年，但是它卻能有如此顯著的成就表現，

其中之奧祕值得探究。在本篇文章中，我們將要介紹 AlphaGo 的發展史。

AlphaGo Lee (AlphaGo) 與 AlphaGo Master

初代 AlphaGo (與李世石對弈版本，又稱 AlphaGo Lee)，使用近年來高度發展的類神經網路，它是一種模擬人腦神經傳遞訊息過程的運算方式。AlphaGo 使用了兩種網路來建構它的程式，分別為：Value Networks (估值網路) 以及 Policy Networks (走棋網路)。其中 Value Networks 用以評估棋盤每個位置的優劣，而 Policy Networks 則是選擇落點步法。AlphaGo 使用了上述兩種神經網路，透過與人類專家對奕來學習現有的對奕知識，並藉由自我學習技術來提升對奕技巧，使自己變得更聰明。除此之外，AlphaGo 的程式更進一步引入一種搜尋演算法，名為「蒙地卡羅樹狀模擬 (Monte Carlo method)」，來模擬上千種隨機自我對奕結果的演算法。以上敘述的幾種技術，都是使 AlphaGo 實力達到全新高度的方法。

而在此之後所開發的 AlphaGo Master 除了軟體演算法的細節更加精進之外，加上硬體改良與更精簡指令的配合，使新版的 AlphaGo (也就是 AlphaGo Master) 更加強大。對比舊版的 AlphaGo (AlphaGo Lee)，新版本在處理計算時所消耗的電量僅為以往的十分之一，運作效率亦提升了 10 倍，亦即可以用更少的時間執行更多深遠的預測，使 AlphaGo Master 運算能力提高了三子 (意旨能多預測三步)，這就是造就 AlphaGo Master 能夠 60 連勝的原因。



▲AlphaGo Lee(左)與李世乭(右)對弈轉播畫面。(網址: gogameguru.com)

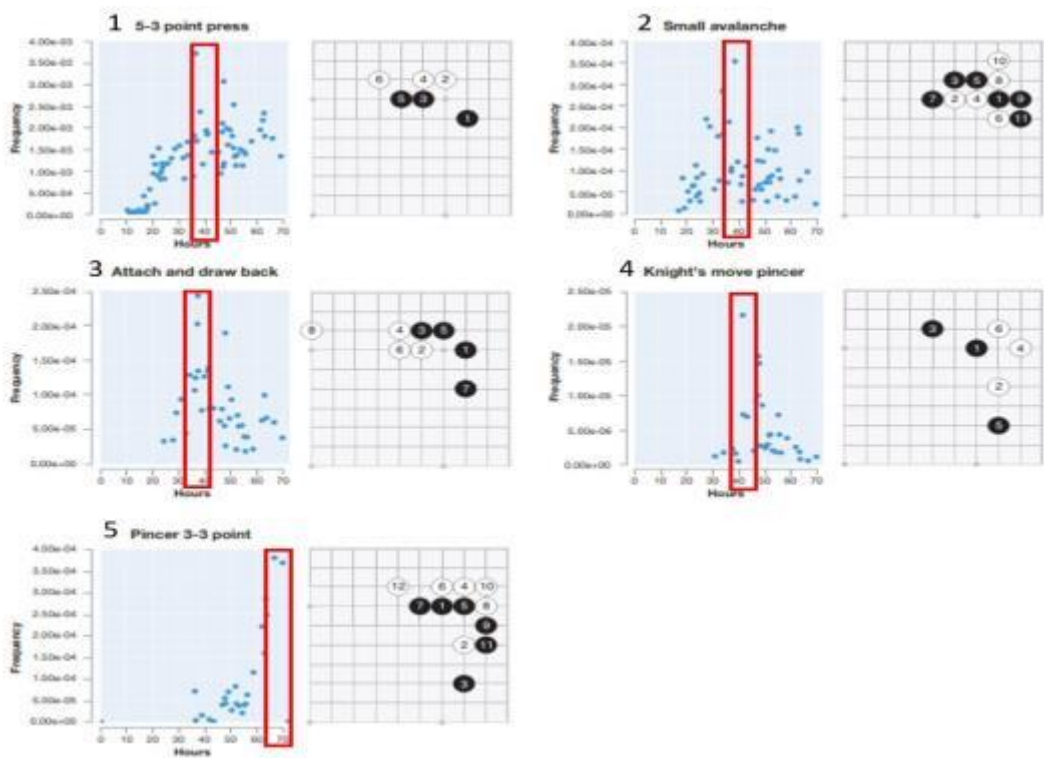
精益求精的 AlphaGo Zero

AlphaGo (AlphaGo Lee) 與 AlphaGo Master 也分別在擊敗李世乭，與取得 60 連勝後，由 Deepmind 公司宣布退役。但 Deepmind 公司，並沒有隨即終止 AlphaGo 的相關研究，反而精益求精開發了比 AlphaGo Master 更加無人能敵的 AlphaGo Zero。

AlphaGo Zero 摒棄以往使用的 Value Networks(估值網路) 和 Policy Networks (走棋網路) 演算法，而是打造全新的自學系統，並完全透過自己與自己對弈來學習圍棋，而過程中並沒有人類插手指導。令人驚訝的是，AlphaGo Zero 用僅僅約 40 小時的自學時間，就摸索出所有重要且高階的人類圍棋定式，並達到了與李世乭對戰的 AlphaGo Lee 相同的水平。3 天後，AlphaGo Zero 對戰 AlphaGo Lee 達到 100% 的勝率。40 天後，AlphaGo Zero 對戰 AlphaGo Master 也達到近 90% 勝率，成為有史以來 AlphaGo 棋力最強的版本。

AlphaGo Zero 不再需要人為的監督學習，而是在輸入圍棋規則後從零開始玩，並全靠自學下圍棋來達到無師自通的程度。而這樣的演算法並非沒有缺點，因為從無到有的探索圍棋，導致 AlphaGo Zero 在初期的學習成效極差，但也因此能夠打破所有人類思考模式，最終擊敗原已無人能敵的 AlphaGo Master，將 AlphaGo Zero 的圍棋實力提升到前所未見的層次。

儘管 AlphaGo Zero 從沒有公開對弈，但從 Google 釋出的棋譜、統計與論文，可以觀察到許多 AlphaGo Zero 的有趣現象。不同於先前版本，AlphaGo Zero 自始自終的自學，且無 Value Networks (估值網路) 監控思考過程，使得研究人員僅能以棋譜中觀察 AlphaGo Zero 的思路與成長，如圖表 1，在約 40 小時的學習時程後，也就是達到與李世石對戰的 AlphaGo 版本相同等級。由此可見 AlphaGo Zero 已摸索出許多人類常用的重要圍棋棋譜，並以極高的頻率使用在對弈過程，且於而後的學習過程中，漸漸摒棄了那些既有的定式，開始發明了一些人類未知定式，或是先使用低階的棋路，隨後再摒棄之，進而再進一步發明出更加高深莫測的棋譜。此種棋力自學的模式恰巧與金庸小說中要練成《葵花寶典》所述：「欲練神功，必先自宮」的概念不謀而合，也是科學家分析 AlphaGo Zero 思路時所發現的有趣之處！



▲圖表 1: 編號 1 - 4 為人類常定式(使用頻率都在 40 小時左右來到最高峰), 編號 5 為 AlphaGo Zero 發現的新定式。(資料來源: Deepmind / 網址: <https://deepmind.com/research/>)

1997 年 5 月, 由 IBM 公司打造的超級電腦 - 深藍 (Deep Blue) , 以 2 勝 1 敗 3 和的戰績戰勝當時國際象棋 (西洋棋) 俄羅斯籍世界冠軍加里·基莫維奇·卡斯帕洛夫 (Garry Kimovich Kasparov) , 成了首位敗倒於電腦的人類冠軍, 然而他卻在數年後出版的書籍中表示, 儘管他不喜歡輸的感覺, 但被自動化機器或看不見的演算法挑戰、超越甚至取代的感覺, 的確漸漸成為現今社會的一部份, 然而這也是人類科技進步的最有力證明。展望未來, 人工智慧科技的進步已可以預期成為推動人類社會發展的巨大助力, 並能給我們人類文明開創無限的進步空間。

副總編輯：國立中山大學資訊工程學系 陳坤志教授

總編輯：國立中山大學資訊工程學系 黃英哲教授

(本文由科技部補助「新媒體科普傳播實作計畫」執行團隊撰稿)