

材料的革新對技術進步和產業發展具有非常重要的作用，但是傳統開發新材料的過程，都採用的試錯法，實驗步驟繁瑣，研發週期長，浪費資源。實驗過程中，研究人員往往達不到自己的實驗預期，而產生了很多不理想的資料。雖然這些實驗過程給我們提供了試錯經驗，但是失敗的實驗資料擺放在那裡彷彿變得並無用處；此外，目前材料表徵技術手段越來越多，對應的圖形資料以及維度也越來越複雜，依靠人力的實驗分析有時往往無法挖掘出材料性能之間的深層聯繫；再者，隨著電腦的發展，許多諸如第一性原理計算、相場模擬、有限元分析等手段隨之出現，用以進行材料的結構以及性能方面的計算，但是往往計算量大，費用大。這些都是限制材料發展與變革的重大因素。

為瞭解決上述出現的問題，結合目前人工智慧的發展潮流，科學家發現，我們可以將所有的實驗資料，計算類比資料，整合起來，無論好壞，便能形成具有一定數量的資料庫；在資料庫中，根據材料的某些屬性可以建立機器學習模型，便可快速對材料的性能進行預測，甚至是設計新材料，解決了週期長、成本高的問題。近二十年來，機器學習方法的發展為我們生活帶來許多便利，舉凡智慧網路搜索，語音識別，乃至無人超市、無人駕駛汽車等，依託於機器學習方法的新事物正迅速地在生活中普及。Alpha Go 的橫空出世更讓世界驚歎於人工智慧的潛在價值。這種利用機器學習預測新材料的方法越來越受到研究者的青睞。在科研領域，大資料的理念正在改變著科研人員對未知世界的探索方式。

本書第一章節講述 AI Chemistry, Materials, 還有 Nanoscience 的領域上，能扮演的角色、會遇到什麼挑戰以及後期希望發展的未來藍圖等等…

2018 年，在 nature 正刊上發表了一篇題為 “Machine learning for molecular and materials science.” 的綜述性文章[1]。文章詳細介紹了機器學習在指導化學合成、輔助多維材料表徵、獲取新材料設計方法等方面的重要作用，並表示新一代的電腦科學，會對材料科學產生變革性的作用。

所謂的機器學習就是賦予電腦人類的獲得知識或技能的能力，然後利用這些知識和技能解決我們所需要解決的問題的過程。

簡單來說，利用機器學習解決問題的過程為定義問題-資料收集-建立模型-評估-結果分析。如圖 1 所示[2]。就是針對於某一特定問題，建立合適的資料庫，將電腦和統計學等學科結合在一起，建立數學模型並不斷的進行評估修正，最後獲得能夠準確預測的模型。根據機器學習訓練集是否有對應的標識可以分為監督學習、無監督學習、半監督學習以及強化學習。需要注意的是，機器學習的範圍非常龐大，有些演算法很難明確歸類到某一類。而對於有些分類來說，同一分類的演算法可以針對不同類型的問題，在解決實際問題時要做具體的分析。此外，隨著機器學習的不斷發展，深度學習的概念也時常出現在我們身邊。深度學習是機器學習中神經網路演算法的擴展，它是機器學習的第二個階段—深層學習，深度學習中的多層感知機可以彌補淺層學習的不足。深度學習演算法包括迴圈神經網路（RNN）、卷積神經網路（CNN）等[3]



圖 1 機器學習的學習過程流程圖

使用計算模型和機器學習進行材料預測與設計”這一理念最早是由加州大學伯克利分校的材料

科學家 Gerbrand Ceder 教授提出。Ceder 教授指出，可以借鑒遺傳科學的方法，就像 DNA 碱基對編碼蛋白質等各種生物材料一樣，用“材料基因組”編碼各種化合物，而實現這一“編碼”的工具便是電腦的資料採擷及機器學習演算法等。這一理念受到了廣泛的關注。隨後，2011 年夏天，奧巴馬政府宣佈了“材料基因組計畫”（Materials Genome Initiative, 簡稱 MGI），該計畫在材料科學中掀起了一場革命。目前，機器學習在材料科學中已經得到了一些進展，如進行材料結構、相變及缺陷的分析[4]、篩選電催化劑[5]、以及利用“失敗實驗”資料預測新材料[6]等。

2017 年 6 月，Isayev[4]等人將 AFLOW 庫和結構-性能描述符聯繫起來建立資料庫，利用機器學習演算法對成千上萬種無機材料進行預測。首先，構建帶有屬性標注的材料片段模型（PLMF）：將材料的晶體結構分解為相互關聯的拓撲片段，表示結構的連通性；為 PMLF 圖中的頂點賦予各個原子獨有的物理和化學性能（如原子在元素週期表中的位置、電負性、摩爾體積等），以此將不同的材料區分開。然後，採用梯度提升決策樹演算法，建立了 8 個預測模型，其中之一為二分類模型，用於預測該材料是金屬還是絕緣體；另外 7 個模型為回歸模型，預測絕緣體材料的帶隙能（ E_{BG} ），體積模量（ B_{VRH} ），剪切模量（ G_{VRH} ），德拜溫度（ θ_D ），定壓熱容（ C_p ），定容熱容（ C_v ）以及熱擴散係數（ α_v ）。經過計算並驗證發現，在資料庫中的 26674 種材料中，金屬/絕緣體分類的準確度為 86%，僅僅有 2414 種材料被誤分類（圖 2）。發現極性無機材料有更大的帶隙能（圖 3），所預測的熱機械性能與實驗和計算的資料基本吻合（圖 4）

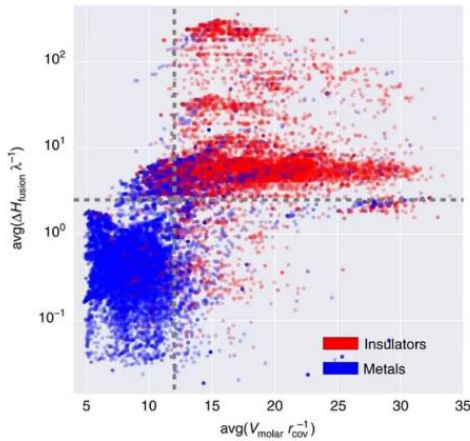


圖 2 資料集分類圖

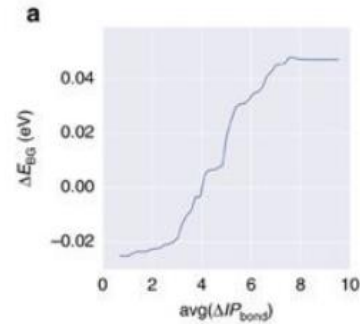


圖 3 帶隙能與電離勢關係圖

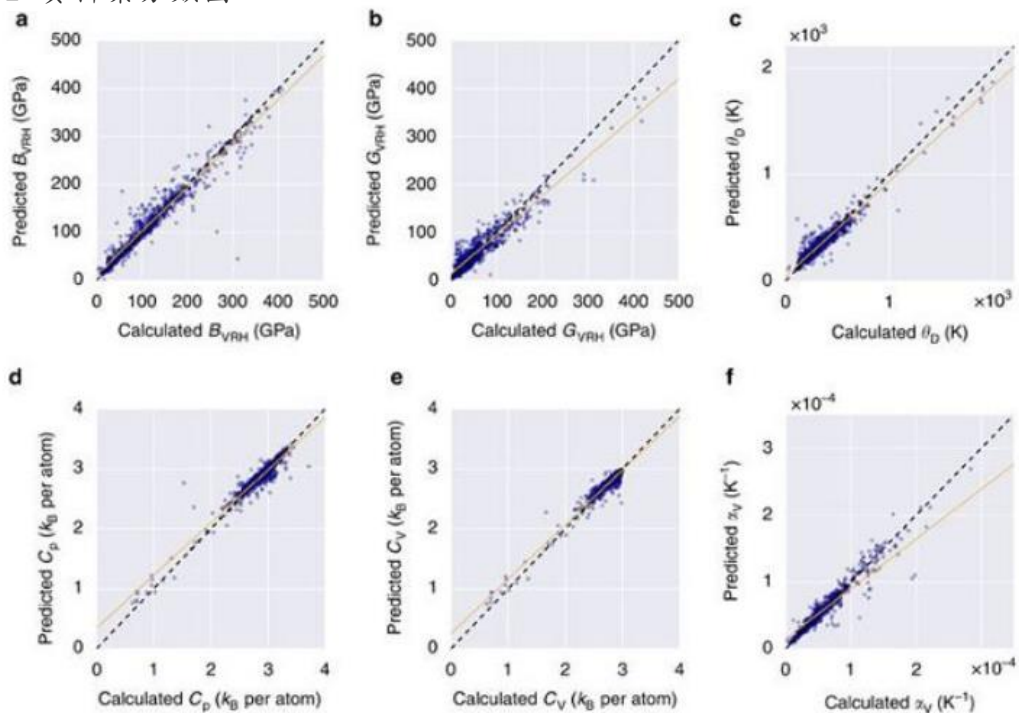


圖 4 模型預測資料與計算資料的對比曲線

卡內基梅隆大學的 Kevin Tran 和 Zachary W. Ulissi[5]設計了全自動化的篩選方法，結合機器學習和 DFT 計算，自動搜索金屬間化合物的各種表面活性位點對 CO 及 H 的吸附能，從而預測具備高反應活性的雙金屬電催化劑。作者採用圖 1 所示的工作流程，使用機器學習模型從無限大的設計空間中預測具有最優活性的金屬間化合物及相應表面活性位點，並通過自動化的 DFT 演算法計算這些位元點對 CO 和 H 的吸附能計算所得的資料存入資料集後繼續用於訓練機器學習模型，形成機器學習篩選催化劑-自動 DFT 計算驗證-機器學習模型再訓練的閉合回饋迴圈過程，產生的資料庫可以自動連續增長，無需人工干涉(圖 5)。

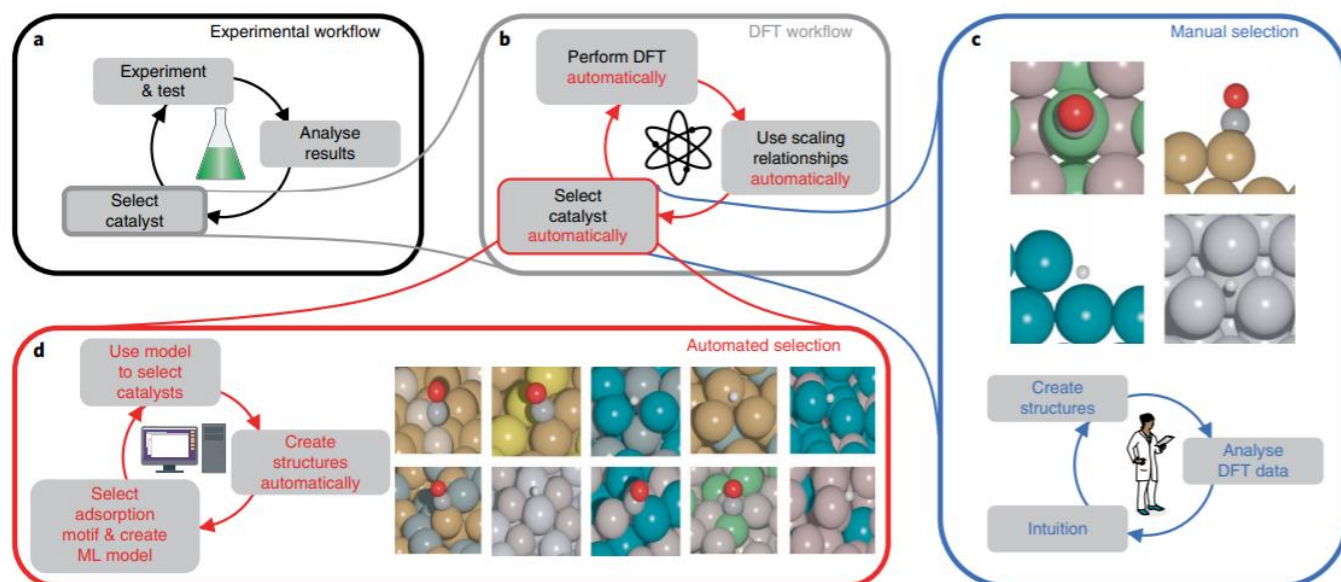


圖 5 自動化理論材料發現的工作流程

新材料的研發是一個充滿挑戰的過程，伴隨著無數次的失敗，但正所謂“失敗是成功之母”，每一次的失敗，也讓研究人員離成功更近一步。哈佛大學的 Alexander J. Norquist 團隊[6]利用實驗室未成功的水熱反應的資料訓練機器學習模型，並用得到的模型來預測新的反應，所得的模型能夠成功預測新的有機-無機材料的合成條件，合成成功率達 89%。化學領域研究人員發表的文獻通常只包括反應成功的例子，但實際上大量未被報導的失敗實驗同樣包含合成條件相關資訊，這些失敗實驗包含的資訊對預測反應成功和失敗的邊界條件也有重大價值。作者收集了大量實驗室失敗反應的資料，以反應物物理化學性能（如分子品質，元素週期表位置等）及反應條件（如反應物配比、反應溫度、環境 pH 等）為特徵，訓練了一個 SVM 模型，該模型預測其測試集的反應結果時，準確率可達 78%，對釩-亞硒酸鹽體系反應的預測準確率達 79%。通過將該 SVM 模型轉換為方便人類理解的決策樹模型，還能進一步認識反應相關機理，從而指導新的合成反應。

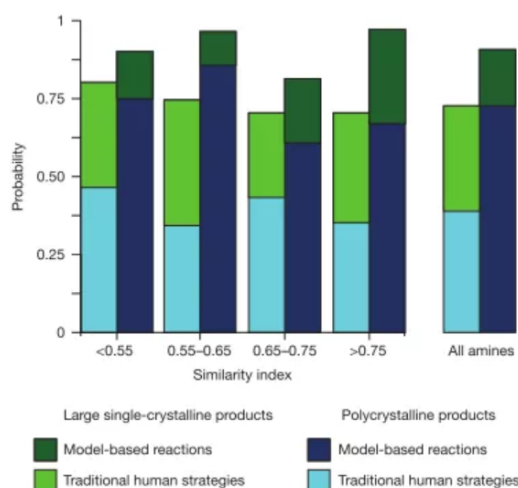


圖 6 合成實驗結果

AI 與傳統理論計算最大的不同，就是 AI 神經網路可以通過對材料資料庫的主動、深度學習獲得經驗知識，從而根據結構組分預測新材料的性質，規劃化學合成路徑，提高圖譜表徵的品質及效率，對滿足要求的材料進行全譜搜索，基於多層次神經網路建立全新的材料基因組構效關係，從而大大提高材料研發的效率和自動化程度。AI 可以提高模擬計算的效率和精度。傳統密度泛函計算（DFT）使用原子組成和晶體結構求解電子結構，從而計算電導率等物理性質，該過程不需要樣本的實驗值，而機器學習中的監督學習需要這些實驗值作為標記資料，和對應的組分結構共同組成訓練集，使模型學習到性質和組分結構之間的對應關係。當有新材料出現時，模型可以對其電導率進行預測。總的來說，未來材料資訊學的發展需要高通量實驗、高通量表徵和高通量模擬計算來產生更多客觀，全面和真實的資料，這需要基礎設施的不斷完善，包括集成化物聯網環境下的高性能計算集群和更先進的表徵合成設備等。另外，機器學習演算法也需要提高其解釋性，並降低對資料量的需求。隨著相關領域資料庫的不斷完善及演算法的不斷優化，AI 將在各種領域發揮越來越多的作用。[9]

文獻來源

- [1] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature*, 559 (2018) 547.
- [2] D.-H. Kim, T.J. Kim, X. Wang, M. Kim, Y.-J. Quan, J.W. Oh, S.-H. Min, H. Kim, B. Bhandari, I. Yang, *International Journal of Precision Engineering and Manufacturing-Green Technology*, 5 (2018) 555-568.
- [3] 周子揚, 電子世界, (2017) 72-73.
- [4] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, *Nature communications*, 8 (2017) 15679.
- [5] Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution. *Nature Catalysis*. 2018;1(9):696-703.
- [6] Tran K, Ulissi ZW. Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution. *Nature Catalysis*. 2018;1(9):696-703.
- [7] <https://reurl.cc/z8ZW20>
- [8] <https://reurl.cc/pd1WDx>
- [9] <https://reurl.cc/qd5OYD>
- [9] <https://reurl.cc/L3mpla>