國立臺灣大學工學院材料科學與工程學系

碩士論文

Department of Materials Science and Engineering

College of Engineering

National Taiwan University

Master's Thesis

主動學習高分子材料設計

Active Learning for Polymer Materials Design

張瑋哲

Wei-Che Chang

指導教授: 陳俊杉 博士

Advisor: Chuin-Shan Chen Ph.D.

中華民國 114 年 6 月

June, 2025

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

致謝

終於完成碩士論文，首先要感謝的當然是陳俊杉老師一直以來的指導。從大三加入 Chen's Group 以來，陳老師一直在推著我學習，並在迷惘時給我研究方向，因為有老師指導才能順利完成碩士學位。而陳錦文老師與游濟華老師兩位老師同樣幫助我很多，不只是研究上的指導，也常常給予我生涯規劃上的建議，超級感謝兩位老師。

接著要感謝實驗室的同伴們，國頎跟廷儒兩位學長一直幫我們很多忙，不只是研究上，也包含實驗室的管理和活動等，感謝你們撐起實驗室。也感謝陳老師的研究生們，每次聚會或是一起當助教開會時，總是讓氣氛很歡樂，研究所沒有認識太多朋友，感謝你們讓我的生活變得更加豐富。

最開始做 PolymerAI 的研究時，只有我一個人，常常覺得壓力很大，後來宗耘加入後感覺安心很多，事情交給他就可以放心，感謝超單的宗耘。再來又多了很多學弟妹，大家都很優秀，相信未來研究一定可以很順利，加油加油。

最後當然要感謝父母，給予我支柱讓我能無後顧之憂地完成碩士學位，也在我需要的時候都能及時幫助我，未來我會繼續努力，不辜負你們的養育之恩及支持。

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

摘要

高分子是生活中常見的材料，能夠透過多樣化的材料設計方式得到不同性能，以滿足各種需求，然而過往使用實驗高分子設計所需成本過高，因此有許多研究試圖以數據驅動方法解決此設計難題。現今數據驅動方法蓬勃發展，數據驅動用於高分子設計也成為熱門的研究議題，然而資料不足侷限了其發展，固在本研究中導入主動學習技術，解決此設計難題。

本研究共分為三部分，皆以主動學習進行高分子材料性質設計最佳化。包含使用主動學習進行隨機共聚物序列設計，成功減少 98% 需要標註的資料量，有望解決標註資料不足的問題。第二部分的研究則是進行單體多目標最佳化，透過主動學習驅動實驗設計，僅以 10^2 以下的標註資料量，即可在龐大的設計空間中完成高分子單體多目標最佳化。最後一部分研究將先前的方法用於純實驗數據驅動設計，並以 Vitrimer 之配方以及比例進行研究，透過模型建議，成功在設計空間中找出機械性質最佳之配方。

本研究解決了過往數據驅動高分子設計的難題，並以模擬及實驗資料分別進行驗證，此研究成果所建立的主動學習設計方法，可望用於實際工業應用，加速不同需求之高分子材料的開發。

關鍵字：高分子材料、機器學習、主動學習、材料設計、最佳化

doi:10.6342/NTU202501030

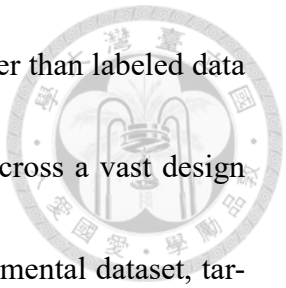
doi:10.6342/NTU202501030

Abstract

Polymers are ubiquitous materials in daily life, offering a wide range of properties through versatile material design strategies to meet diverse application requirements. However, traditional experimental approaches to polymer design often lead to high costs, which motivates recent efforts to address this challenge using data-driven methods. With the rapid advancement of data-driven technologies, their application to polymer design has emerged as a prominent research focus. However, the limited availability of labeled data remains a major bottleneck that hinders further progress. To overcome this limitation, we introduce active learning techniques to enhance polymer design processes.

This research is structured into three major parts, each leveraging active learning to optimize the design of polymer materials. The first part involves the design of random copolymer sequences using active learning, which successfully reduces the amount of required labeled data by 98%, thus addressing the problem of data scarcity. The second part focuses on the multiobjective optimization of monomer structures, where active learn-

doi:10.6342/NTU202501030



ing is employed to guide experimental design. Remarkably, with fewer than labeled data points, the method achieves effective multi-objective optimization across a vast design space. The final part applies the developed methods to a fully experimental dataset, targeting the formulation and composition optimization of vitrimers. Using model-driven suggestions, the optimal formulation with superior mechanical properties was identified within the design space.

Overall, this study addresses key challenges in data-driven polymer design by integrating active learning, validated through both simulation and experimental datasets. The proposed active learning framework holds promise for practical industrial applications, offering a pathway to accelerate the development of polymer materials tailored to specific performance requirements.

Keywords: Polymers, Machine Learning, Active Learning, Materials Design, Optimization

doi:10.6342/NTU202501030

目次

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

圖次

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

表次

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

第一章　緒論

1.1　研究背景

材料設計是人類科技發展重要的一環，在生活中遇到各種不同應用場景，往往需要透過材料設計，方能滿足各種需求。過往材料設計方式多是基於試誤法(trial and error)，透過反覆實驗找出最佳解，並從中歸納出材料參數以及性質之間的關聯。然而，在材料開發流程中，試誤法所衍生的時間以及金錢成本過高，成為待解決的問題。近年來，隨著電腦算力的高速發展，各種模擬方法以及機器學習逐漸改變各個領域，在材料設計上也有越來越多相關研究，試圖使用數據驅動方法推動材料設計，取代以往的試誤法，比如著名的材料基因體計畫 (Materials Genome)[4]，整合材料科學、模擬技術、資料科學等領域，建立材料資料庫，讓數據驅動方法替材料設計提供嶄新的可能性。目前主流研究上的數據驅動材料設計流程，主要包含資料收集-代理模型 (Surrogate model) 建立-反向設計最佳化三步驟，如圖1.1所示。

高分子材料是日常生活中最常見的材料之一，由於其經過設計後，能夠具有多樣化的性能，因此被廣泛應用在能源、航太、生醫等各種領域。近年來隨著機器學習的發展，有越來越多結合機器學習與高分子材料設計的相關研究，這些數據驅動方法雖然仍在啟蒙階段，目前有許多問題需要解決，但已經展現其極高的潛力，有機會改善並優化以往的開發流程，加速高分子材料的配方及單體優化，

doi:10.6342/NTU202501030

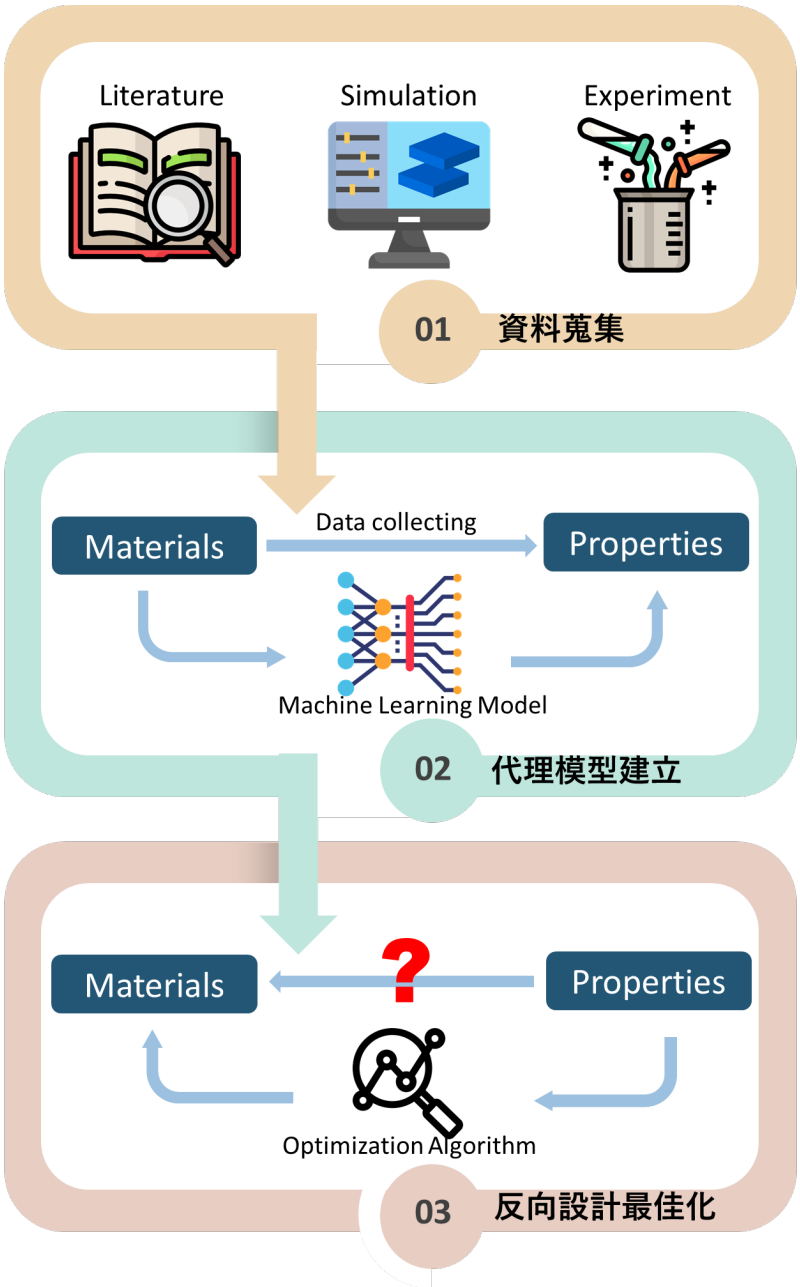


圖 1.1: 數據驅動材料設計之流程。

doi:10.6342/NTU202501030

滿足各種不同的應用場景。

1.2 研究動機

近年來數據驅動設計在各種領域逐步發展，逐漸受到重視，然而，在高分子材料設計上，進展卻較為緩慢。侷限數據驅動高分子方法發展的最大原因為缺乏指標性資料集，由於高分子材料複雜性，包含單體種類、聚合度等都會影響材料性質，也可以使用不同單體組合成共聚物 (copolymer)，這些設計複雜度能夠衍生出多樣化的設計空間，如圖1.2所示，再加上高分子材料實驗需要控制的變因多以及時間長，造成難以透過實驗建立數量足夠龐大的標註資料集，也讓數據驅動高分子設計的進展停滯不前。為了解決上述困難，目前多是根據設計目標性質以及設計高分子種類，而建立專門的資料集，然而如何減少建立資料集過程中所衍生之成本，並透過規模較小的資料集進行數據驅動材料設計，仍需更進一步研究及實驗，以上為本研究的動機。

主動學習 (Active learning) 是機器學習的一個分支，其與傳統機器學習的差異在於，主動學習的流程是會透過模型回饋，運用迭代的方式有效率地進行標註，進而達到最少的標註資料量。目前主動學習的技術，已經在某些領域被證實能夠大幅減少訓練模型所需的標註資料成本，並透過機器學習模型的引導進行實驗設計，可望快速在設計空間中找到最優解。雖然目前主動學習應用於材料設計的研究仍然不多，但是具有極大的潛力，有望克服以往難以解決的挑戰，故本研究認為應用主動學習進行材料設計，值得更深入的探討，也是本研究的主軸。

doi:10.6342/NTU202501030

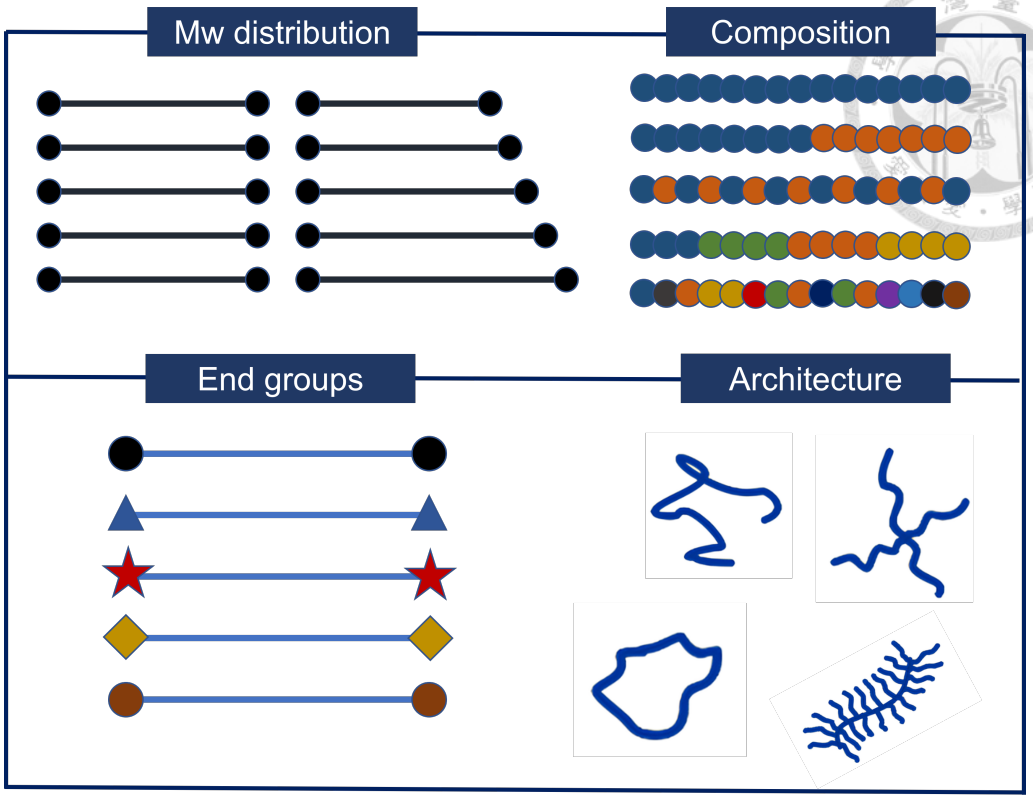


圖 1.2: 高分子材料之設計複雜度，包含單體種類、分子量、官能基等。

1.3 研究目的

本研究旨在開發主動學習應用於高分子材料設計的流程架構，研究目的為透過主動學習方法，解決過往數據驅動高分子設計時，所遇到標註資料不足的問題。在研究中會設計不同實驗，進行不同模型以及主動學習策略之間的比較，驗證主動學習是否可以如預期，有效減少所需之標註資料量，同時也會用視覺化方法呈現並進行模型以及主動學習策略解釋。本研究的實驗成果預期將能建立主動學習驅動高分子設計的標準流程，涵蓋共聚物序列設計、單體選擇的情境，且包含單目標最佳化以及多目標最佳化等現實高分子設計常見的場景，同時能解決過往之標註成本過高的問題，並將研究結果用於真實實驗數據驅動材料設計案例，展示主動學習輔助材料設計的潛能。此研究結果預期能為數據驅動工程應用帶來新穎的方法，突破過往所遇之瓶頸。

doi:10.6342/NTU202501030

1.4 論文架構

本論文其餘各章節內容如下:

第二章 [文獻探討]: 本章節中會回顧以往數據驅動高分子材料的發展及挑戰，接著進行主動學習理論及方法的探討，並整理過往主動學習應用於高分子材料的相關研究成果，以利了解目前主動學習高分子設計流程的發展、潛力以及挑戰。

第三章 [研究方法]: 本章節會詳細介紹本研究中所使用之方法，包含研究所使用的資料集、模型種類、主動學習挑選策略、資料視覺化工具、高分子合成方法等。

第四章 [結果與討論]: 本章節中會呈現本研究目前的研究成果，包含主動學習的效率、最終最佳化後的成過、使用的標註資料量以及視覺化呈現模型所挑選之標註資料等，並在最後將此流程用於 Vitrimer 高分子的配方設計案例。

第五章 [結論與未來展望]: 本章總結研究的主要發現和結論，並提出未來在相關領域進一步研究和應用的可能方向。

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

第二章　文獻探討

2.1　數據驅動高分子材料設計

在過往的文獻研究中，有許多使用機器學習於高分子材料性質預測及反向設計的案例。一般而言，此類型研究之標準流程為資料收集-建立模型-反向設計[5]。其中高分子的資料來源包含實驗、計算模擬以及文獻資料，而現今也有一些機構學者致力於建立高分子材料的公開資料集，包含由日本材料研究機構 (NIMs) 整理的 PolyInfo[6]、Kim 等人建立的 Polymer Genome[7] 等，高分子的公開資料集整理如下表2.1，這些公開資料集的建立，讓高分子機器學習演算法的研究得以發展，然而，先前提到的高分子材料複雜度等問題，造成公開資料集難以涵蓋所有高分子種類及考量所有設計因素，此問題僅依靠公開資料集所涵蓋的資料，仍有其侷限性難以解決。

表 2.1: 高分子材料公開資料集整理。

資料集	網址
PolyInfo	https://polymer.nims.go.jp/[6]
Polymer Genome	https://www.polymergenome.org/[7]
Polymer Property Predictor and Database	https://pppdb.uchicago.edu/
MatWeb - Material Property Data	https://www.matweb.com/
NanoMine	https://materialsmine.org/nm
NIST Synthetic Polymer MALDI Recipes Database	https://maldi.nist.gov/
Materials Project	https://next-gen.materialsproject.org/[4]

透過預先準備的標註資料集，就可以進行機器學習模型訓練，利用標註資料建立分子描述以及標註性質之間的關聯，然而如何讓機器能夠讀懂分子表達

doi:10.6342/NTU202501030

形式，成為另外一個挑戰。過去最主流的分子表達方式為分子指紋 (Molecular Fingerprints)，分子指紋的概念為將分子資訊向量化，將化學資訊隱含在向量中，進而讓機器可以讀懂化學資訊。建立分子指紋最常見的方法為提取分子中官能基結構，接著透過 hashing 的方式將結構轉為向量，因此向量中的每一個維度皆代表高分子單體中所含之官能基，接著便能透過此向量比較分子間相似性以及進行機器學習訓練。基於定義好之官能基的 MACCs 分子指紋 [8] 以及透過定義搜尋半徑尋找子結構的 Morgan 分子指紋 [9] 為常見的兩種分子指紋，這些分子指紋都已經整合在知名化學套件 rdkit 中 [10]。除了向量化之外，使用字串表達式是另一個機器學習常見的輸入手法，由於大語言模型 (Large Language Models, LLMs) 的發展，讓此表達方式逐漸受到重視。目前最常見的字串表達形式為簡化分子線性輸入規範 (Simplified molecular input line entry specification, SMILES)，SMILEs 透過嚴謹的規範，成為最通用的數位分子表達式 [11]。在 SMILEs 的基礎上，加入高分子重複單體的連接點位資訊，以 * 表示，就成為高分子表達形式 P-SMILEs，如下圖2.1所示，P-SMILEs 也被廣泛用在高分子機器學習研究中 [7]。透過這些輸入方式，能夠準確傳達分子中隱含的化學訊息，加上資料集中的性質標註，就能使用機器學習演算法比如 SVM[12]、ANN[13] 等，訓練高分子之機器學習模型，這些利用數據驅動取代真實複雜物理行為的模型，我們常稱之為代理模型 (Surrogate model)。

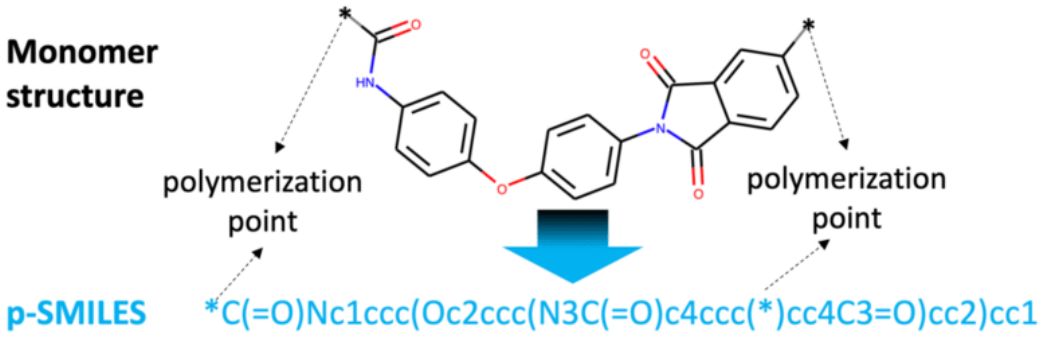


圖 2.1: P-SMILEs 範例。(摘自 [1])

doi:10.6342/NTU202501030

機器學習模型相較於實驗以及模擬方法，具有快速預測輸入輸出的優勢，透過資料建立的代理模型輔以最佳化演算法，能夠實現反向設計，有望透過數據驅動找出具有優異性能的新高分子單體或配方 [14]。過去的案例有 Meenakshisundar 等人使用數據驅動進行高分子序列設計，得到最佳化的相容劑 (Compatibilizers)[15]。而 Tao 等人則透過模擬以及機器學習，找到具有高耐熱性及強度的聚酰亞胺 (Polyimide, PI)，並實際進行合成及量測性質證實設計的結果 [16]。以上這些研究建立了標準的數據驅動設計流程，並且已經有成功合成從數據訓練模型預測，來設計的優異高分子案例，印證了數據驅動用於高分子材料設計的潛能。

2.2 高分子機械性質模擬

在本研究中，設計的目標性質主要為機械性質，並使用模擬方法進行機械性質的預測。用於高分子的主流模擬方法為分子動力學 (Molecular Dynamics, MD)，MD 的原理是透過牛頓力學的力場 (Force Fields)，計算原子之間的交互作用，並藉由統計力學，將微觀尺度的能量轉換為巨觀尺度的性質 [17]。由於電腦算力的飛速發展，比如 LAMMPS[18]、Gromacs[19] 等常用的 MD 模擬套件都開放支援 GPU，讓 MD 開始可以模擬更大的系統以及更多的原子數量，因此在生醫、材料等領域更為被廣泛使用。

然而要模擬高分子材料的機械性質，必須考慮高分子鏈在拉伸的回彈效應以及高分子鏈之間的交互作用，使用全原子模擬 (All-Atom Molecular Dynamics, AAMD) 時，由於計算量大以及積分步長的限制，即使以現今算力，仍無法達到模擬高分子鏈運動的時間及空間步長。為了解決此問題，粗粒分子動力模擬 (Coarse-Grained Molecular Dynamics) 方法被提出 [20]，CGMD 透過將複數個原子

doi:10.6342/NTU202501030

視為一個質量相等的粗粒化珠子 (Beads)，可以大幅降低系統自由度以及運算量，進而提升模擬的時間及空間尺度，使之更接近現實拉伸行為，因此相較於 AAMD 可以更加精確預測高分子的機械性質。而進行粗粒化模擬前，同樣必須要先得到粗粒化後的力場，此力場參數可以從巨觀性質進行擬和，此方法稱為 Top-down，或是從微觀角度出法，即透過預先得到的全原子分子動力學模擬結果，進行能量分布以及運動軌跡擬和，此從較小尺度出發的方法稱為 Bottom-up，Top-down 以及 Bottom-up 分別對應的模擬尺度如圖2.2所示。透過這些方法找出來的力場，可以有效增加分子動力學模擬尺度，並在降低計算量的同時模擬更大系統，避免因為模擬尺度太小造成誤差，更有機會準確預測高分子之機械性能 [21]。而目前也已有發展好的高分子粗粒化模型，將單體視為一個珠子，能夠直接用這些模型進行 CGMD 模擬。

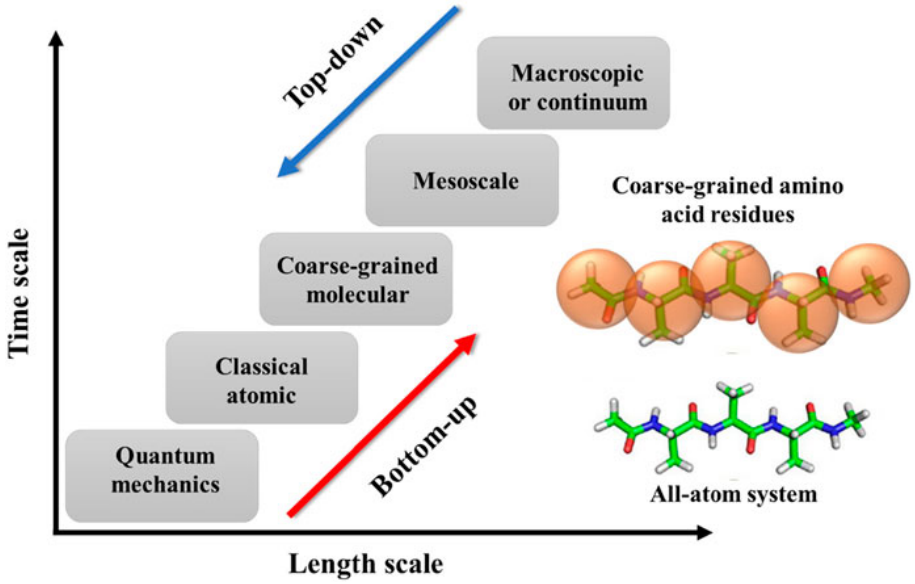


圖 2.2: 全原子以及粗粒化分子動力模擬的尺度關係。(摘自 [2]

2.3 高分子基礎模型

基礎模型 (Foundation Model) 是近期機器學習重要的研究方向，和傳統監督式學習不同，基礎模型通常會透過大量未標註資料，以半監督式學習 (Semi-

doi:10.6342/NTU202501030

Supervised Learning) 進行預訓練，通常利用隨機遮蓋 (random masking) 等方法，從未標註資料中製造模型可以學習的偽標籤 (Pseudo-label)，接著只要再透過標註資料，就可以實現適應下游 (Down Stream) 任務。由於未標註資料的取得成本遠小於標註資料，因此較容易建立大規模的資料集訓練模型。最早基礎模型的概念是 Google 在 2018 年提出的 BERT 模型 [22]，為自然語言處理 (Nature Language Processing, NLP) 領域的基礎模型，透過大量文本資料進行預訓練，而在語意分析、詞性分類、擷取問答等下游任務都取得重大成功，預訓練模型也因此改變了機器學習領域的研究走向。透過基礎模型在預訓練過程中學習到的資訊，只需少量標註資料就能在下游任務中微調而成為不同任務的專家模型，因此現今在許多領域都在發展各自之基礎模型。

高分子領域最早之基礎模型是由 Kuenneth 等人提出的 PolyBERT[23] 和 Xu 等人提出的 TransPolymer[24]，兩者皆是參考自然語言處理的基礎模型架構，將高分子 P-SMILEs 分子式視為自然語言進行預訓練，預訓練後再透過標註資料集進行下游微調。兩者的研究成果顯示，基礎模型已經成功在高分子機器學習領域取得重大突破。隨後也有更多高分子基礎模型相關研究，比如 Wang 等人提出考慮多模態輸入形式的 MMPolymer[25]，以及 Gao 等人提出使用圖 (Graph) 作為輸入方式的基礎模型 [26] 等。這些基礎模型的發展，大幅減少訓練模型所需要的標註資料量，也更有機會將模型用於不同高分子種類。

2.4 主動學習

傳統監督式學習 (Supervised Learning) 的流程為先收集足夠數量標註資料，接著使用設計好的機器學習演算法，透過資料集標註，學習輸入輸出之間的關聯 [27]，藉此建立機器學習模型，因此，在傳統機器學習訓練過程中，標註資料

doi:10.6342/NTU202501030

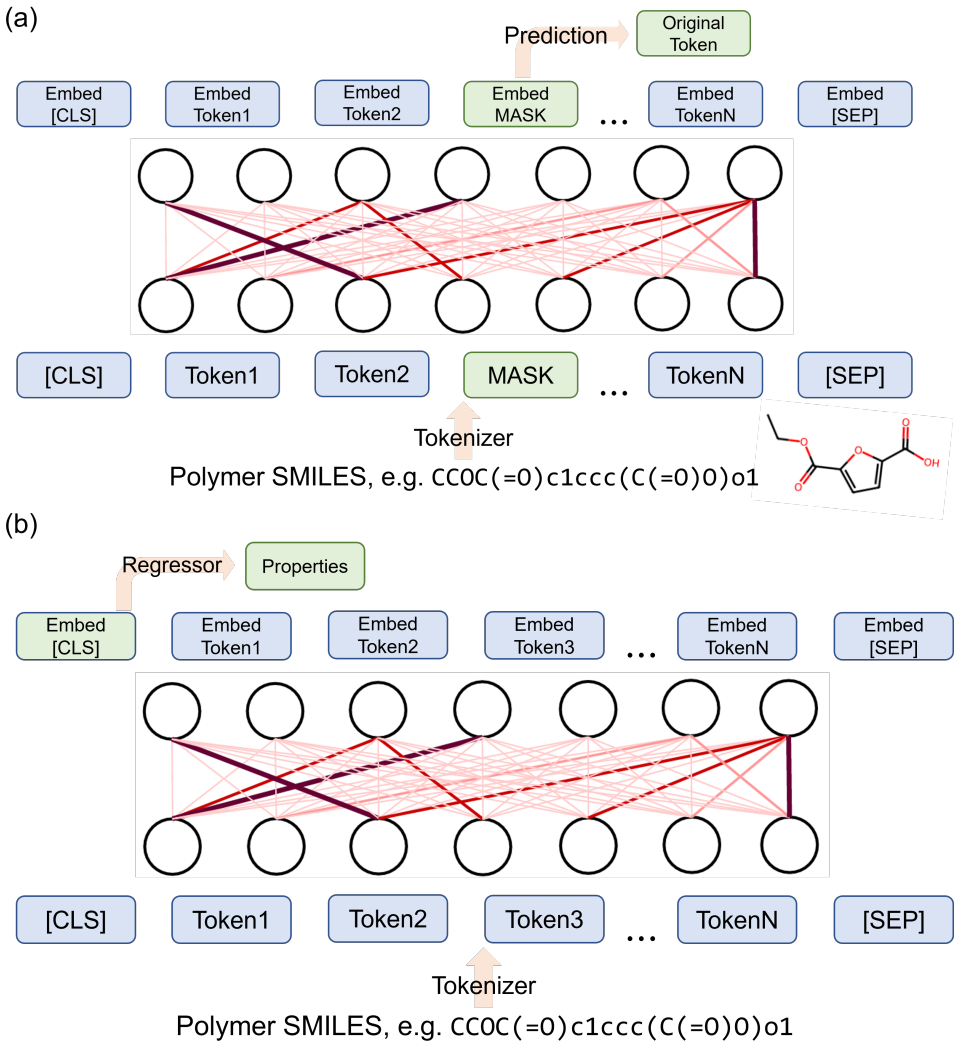


圖 2.3: 高分子基礎模型訓練方式示意圖。(a) 預訓練階段，使用隨機遮蓋進行半監督式學習。(b) 下游任務，使用標註的資料集進行基礎模型的微調。

doi:10.6342/NTU202501030

的數量以及品質會大幅影響模型表現。而主動學習方法被提出後，被視為是標註資料不足的其中一種解決方法，其核心概念是先透過少量資料建立代理模型 (Surrogate Model)，並提前定義好設計空間，透過代理模型的先備知識以及預測能力，在設計空間中進行探索，搭配設計好的採樣演算法，決定下一個需要標註的資料。主動學習在機器學習領域中被視為一種人機互動 (human-in-the-loop) 的學習方式 [28]，與傳統監督式學習不同，主動學習以迭代 (iterate) 的流程進行訓練，流程示意如圖2.4，由機器學習模型引導下一步，人類扮演標註提示的角色，透過此方式最小化所需標註資料數量，因此主動學習也被稱為最優實驗設計 (Optimal Experiment Design)[29]。

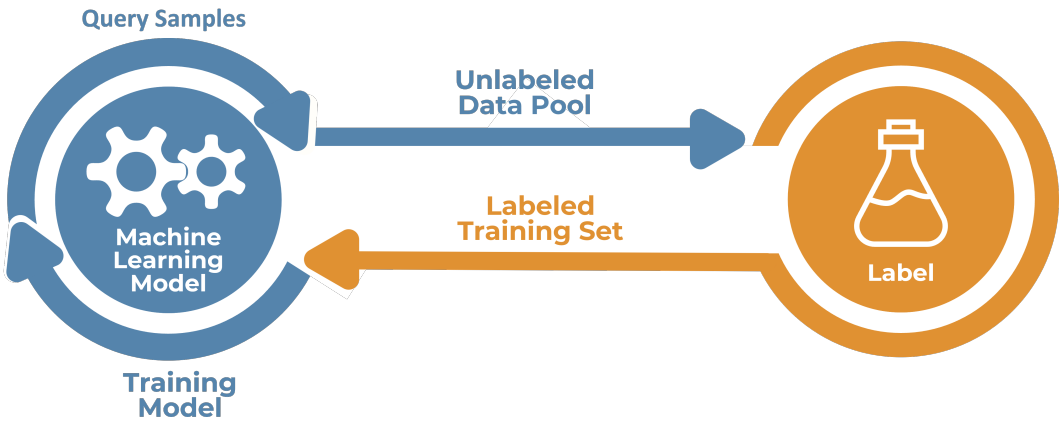


圖 2.4: 主動學習流程示意圖。透過模型在設計空間中進行探索，並引導提示下一步實驗規劃，可大幅減少訓練模型所需之資料量。

主動學習最核心的部分為採樣演算法，根據過往研究，採樣方法包含基於模型不確定性 (uncertainty-based)[30, 31]、基於資料多樣性 (diversity-based)[32]、期望模型改變 (expected model change)[33] 等，其中以基於模型不確定性的採樣策略最為常見 [34]。然而根據求解問題不同以及模型種類差異，在不同類型問題上選擇合適的採樣演算法能夠適時加速主動學習的收斂及表現。

doi:10.6342/NTU202501030

2.5 主動學習輔助高分子材料設計

在上一章節提到，主動學習使用模型回饋選取標註資料，能夠從少量資料出發，從沒有標註的資料集中，依靠模型以及演算法進行後續實驗挑選，且依照不同的挑選方式，除了可以快速提升模型表現，也可以用來進行設計空間中之材料參數最佳化。近年來也有相關研究將此技術應用於高分子材料性質設計，透過主動學習技術引導高分子材料實驗設計，使用最少的實驗及時間成本，在設計空間中尋找最佳材料參數，示意圖如圖2.5。比如 Kim 等人利用主動學習技術，在只有 5 筆資料的初始條件下，能夠在 731 筆未標註資料的設計空間中，快速找出具有高玻璃轉化溫度 (Glass Transition Temperature, Tg) 的高分子單體，證明了主動學習在分子設計上，能夠大幅減少以往數據驅動材料設計所需的標註資料量 [35]。Ramesh 等人則是使用主動學習結合分子動力學模擬 (Molecular Dynamics, MD)，針對共聚高分子的序列，快速在龐大的設計空間中進行高分子迴轉半徑 (Radius of Gyration, Rg) 的設計，利用主動學習的優勢，解決共聚高分子序列排列組合過多，過往在實驗設計時使用試誤法難以考慮的問題 [36]。Zhao 等人在近期發表的文章中，則是透過主動學習配合機械性能實驗標註，進行環氧樹脂 (Epoxy Resin) 的配方比例最佳化，研究成果顯示依靠主動學習的提示採樣，能夠成功找出最佳的原料配方，提升環氧樹脂的機械性能 [37]。表2.2中整理了目前主動學習用於高分子材料設計的文獻、資料來源以及設計的目標性質，可以證實主動學習用於高分子設計的潛力。

doi:10.6342/NTU202501030

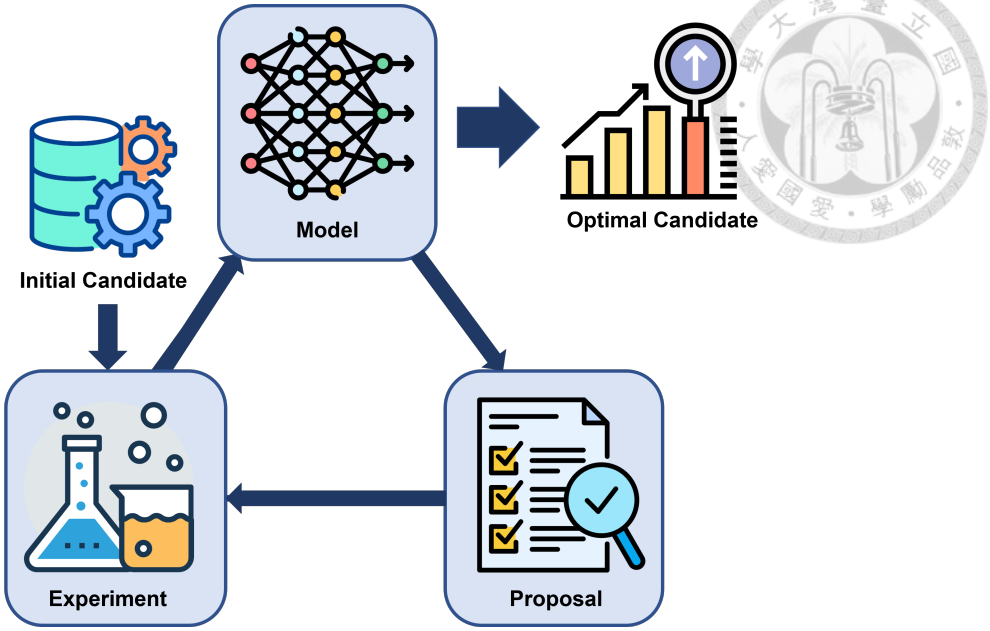


圖 2.5: 主動學習輔助高分子材料設計示意圖，透過模型提供下一步實驗規劃，再進行高分子合成及性質量測，以得到設計空間中之最佳化材料參數。(摘自 [3])

表 2.2: 主動學習用於高分子設計相關研究。

設計目標性質	資料來源	參考文獻
Tg	Literature	Kim 等人 [35]
Rg	MD simulation	Ramesh 等人 [36]
molecular weight distribution	KMC simulation	Zhou 等人 [38]
Adsorption free energy, Energy barrier, Rg	MD simulation	Jablonka 等人 [39]
Hole mobility	MD and DFT simulation	Antono 等人 [40]
Adhesive joint strength	Experiment	Pruksawa 等人 [3]
Elastic modulus, Adhesive joint strength	Experiment	Kraisornkachit 等人 [41]
Tensile strength, Elastic Modulus, Elongation	Experiment	Zhao 等人 [37]

2.6 高分子材料多目標最佳化

最佳化問題是在已定義好的設計空間中，透過設計好的演算法，尋找設計空間中最佳解的問題。為了解決此類型問題，目前已經發展出數種演算法，比如基因演算法 (Genetic Algorithm, GA)[42]、貝葉斯優化 (Bayesian Optimization, BO)[43] 等，這些知名的演算法透過不同方式，在設計空間中尋求快速搜索並達到收斂。

而多目標最佳化 (Multi-Objective Optimization, MOO) 的目的同樣是建立搜尋演算法，然而當目標性質不只一種時，讓問題變得更為複雜。一般多目標最佳化的應用場景是在目標性質彼此互斥 (trade-off) 時，互斥代表設計空間中沒有最佳

doi:10.6342/NTU202501030

解，需要在最佳化決策過程中做出取捨，方能在設計空間中做出最佳的決策。而在材料設計領域，材料性質彼此互斥是很常見的議題，比如機械強度以及延展性是常見的例子，在提升材料機械強度過程中往往會降低延展性，因此這類型最佳化的目標不再是尋找空間中單一最佳解，而是一組解的集合，通常稱之為 Pareto set，此集合中有複數解，稱為 Pareto solutions。目前這類型多目標最佳化已被廣泛用於經濟、工程等各種領域，是在實際應用時很常面對的問題。多目標最佳化問題可以定義為:

$$\min_{x \in X} \big(f_1(x), f_2(x), \ldots, f_k(x)\big) \tag{2.1}$$

其中 k 為欲優化的性質數量並滿足 $k \geq 2$，X 為設計空間。而面對先前提到的性質彼此互斥關係，過去有一些研究使用不同權重將多目標優化的優化目標加權相加，使之變為單純的單目標優化場景，然而此方法並不適用材料參數選擇的問題上，原因為我們難以找出符合所有材料應用場景的重要性權重，無法簡單將之轉為單目標最佳化問題。因此我們以最大化的最佳化問題為例，定義 Pareto set，在 Pareto set 中的每一個解都不能被設計的向量空間中其他點支配 (dominate)，假設 x_1 支配 x_2，可將數學式寫為:

$$f_i(x_2) \leq f_i(x_1),\ \forall i \in \{1, \ldots, k\} \quad \text{and} \quad \exists j \in \{1, \ldots, k\} \text{such that} f_j(x_2) < f_j(x_1) \tag{2.2}$$

意即在 Pareto set 中的所有資料點，在設計空間中，都保證不會有其他點在所有性質都同時優於它。透過找出 Pareto set，我們可以得到由 Pareto set 中所有點所構成的邊界，一般稱之為 Pareto front，如圖2.6所示，而 Pareto set 中的所有點都是多目標優化場景中的最佳解 (Pareto optimal solutions)。

高分子材料多目標最佳化的相關研究有 Mannodi-Kanakkithodi 等人透過蒙地卡羅的方法尋找帕雷托最佳解，成功在設計空間中找到同時具有高介電常數

doi:10.6342/NTU202501030

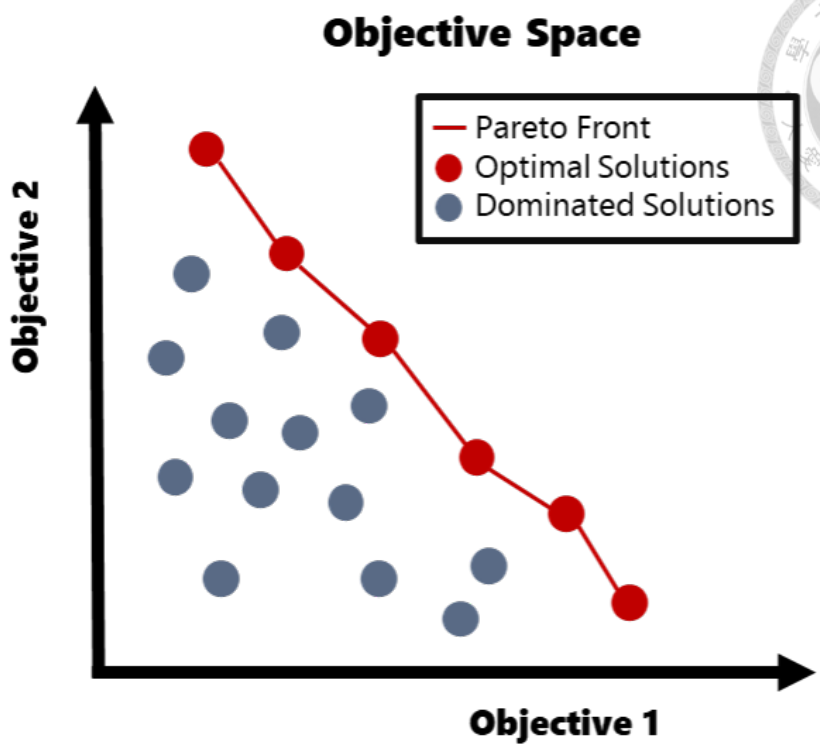


圖 2.6: Pareto front 示意圖。

(dielectric constant) 以及能隙 (bandgap) 的高分子材料 [44]。Kraisornkachit 等人則是透過數據驅動主動學習搭配帕雷托方法，成功設計出具有高彈性模數 (Elastic modulus) 以及黏著接合強度 (Adhesive joint strength) 的高分子配方以及製程參數 [41]。Jablonka 等人則是透過實驗設計 (Design of Experiment, DOE)、迭代搭配帕雷托方法，成功在龐大的設計空間中，找出能夠滿足分散劑三種目標性質的隨機共聚物序列 [39]。以上這些研究致力於解決高分子多目標最佳化問題，目前已有一些標準流程建立，然而這些研究多用於模擬所得之高分子性質，如何在現實問題中建立完整多目標最佳化流程仍待研究。

2.7 Vitrimer 材料

Vitrimer 是一種兼具熱固性高分子高機械性能與熱塑性高分子可加工性的新型材料，其結構由動態共價鍵 (Dynamic Covalent Bonds) 組成動態共價網絡

doi:10.6342/NTU202501030

(Covalent Adaptable Networks, CANs)，能在升溫時通過鍵交換實現網絡拓撲的重組，圖2.7展現了 Vitrimer 動態共價鍵交換機制。這種材料在拓撲凍結溫度 (T_v) 以下為固體，以上則為黏彈性液體，具備高強度、耐用性、自修復性及可回收性 [45, 46]。Vitrimer 最初是基於酯交換 (Transesterification) 反應實現，後來擴展至硫醇酯交換、亞胺交換等多種化學機制，並引入催化劑增加反應速率以降低加工溫度。此類材料已廣泛應用於結構性複合材料，如可回收的碳纖維增強複合材料，並在自修復材料及形狀記憶等多功能領域表現出潛力。然而，工業化挑戰仍存在，特別是在動態化學前驅物 (Dynamic Chemical Precursors) 的規模化製備及降低加工成本方面。未來，Vitrimer 作為下一代高性能可回收材料，有望在航太、風力發電葉片及電子封裝等領域發揮重要作用，同時推動循環經濟的發展 [47]。

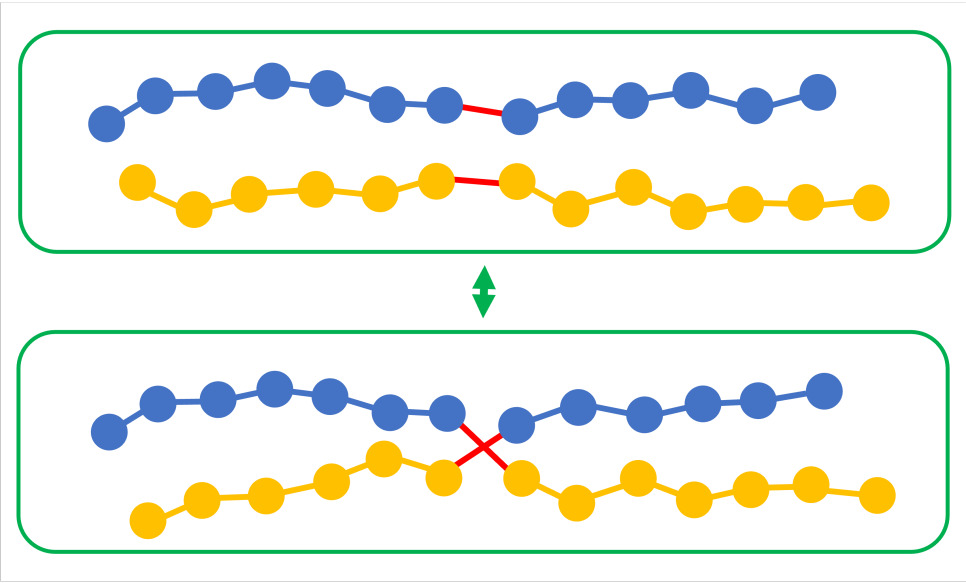


圖 2.7: 動態共價鍵示意圖。

聚酯類 Vitrimer 是透過酯交換反應達成動態共價鍵，透過環氧樹脂與羧酸或酸酐在催化劑作用下交聯固化而成。此類型的 Vitrimer 機械性能由組成比例與交聯密度主導，例如剛性鏈段比例的提高能顯著增強拉伸強度和楊氏係數，但可能導致延展性降低。聚酯類 Vitrimer 具有較低的 T_g 與適中的 T_v，使其具有較好的彈性，又能在高溫下加工與回收 [48]。同時，通過催化劑的調控 (如 $Zn(OAc)_2$)，

doi:10.6342/NTU202501030

可以精確控制化學鍵交換速率，進一步最佳化應力分佈和恢復性能。這種動態行為允許 Vitrimer 實現應力鬆弛和自我修復功能，在應力集中環境下延長材料壽命[49]。此外，Vitrimer 展示出卓越的形狀記憶與耐用性，其形狀固定率可達 99%，回復率接近 88% 至 99%，在多次回收循環後仍能保持接近初始的機械性能。這使其在高應力結構材料（如碳纖維複合材料）及電子封裝領域具備廣闊應用潛力[50]。然而，目前多數聚酯型 Vitrimer 的研究著重於鍵的交換速率以及相關衍生特性，雖然此類型研究可以提升材料的再利用性以及加工性，卻可能導致設計的 Vitrimer 強度不足。除此之外，過去也有文獻指出樹脂材料在提升強度以及耐熱性時，同時會讓黏度增加，而使加工便得困難，影響加工性以及實際商用價值。因此，面對此類型多目標且互相衝突的設計場景，如何在提升機械性能及耐熱性的同時，兼顧動態反應速率與加工性，仍是未來研究的重要挑戰。

總體而言，Vitrimer 在機械性能上的表現優異，兼具高強度、耐用性與多功能性，為環境友好型高分子材料的設計提供了新的契機，而聚酯類更有單體好取得且種類多的特性[51]，可以根據不同需求更改單體以及組成比例，透過設計以應付不同應用需求，具有極高的商業化價值，然而要能設計出兼具各種性質的 Vitrimer 仍有待研究，對於此類型的材料設計難題，機器學習用於配方最佳化是具有潛力的解答。

2.8 小結

在本章節中，整理了過去數據驅動高分子材料設計的發展，以下幾點說明本研究的重要性:

1. 數據驅動材料設計目前已成為重要的研究方向，目前在高分子材料的應用上，雖然已經有相關研究以及公開資料集，然而受限於高分子材料的設計

doi:10.6342/NTU202501030

複雜度，仍難以涵蓋到所有設計空間，加上針對不同高分子的標註資料集的成本過高，使數據驅動高分子設計的發展停滯不前。

2. 主動學習是一種新發展的機器學習技術，透過模型引導標註樣本，有望大幅減少標註所需的成本，達到模型快速進步。

3. 主動學習的概念能用於最佳化流程中，近期也有相關研究將此方法用於高分子材料最佳化上。

4. Vitrimer 是一種新型的環保高分子材料，如何找出具有優異性能的 Vitrimer 合成配方，目前是高分子領域相當熱門的研究議題。

5. 綜合以上幾點，主動學習應用於高分子材料設計是很有潛力的方法，然而目前仍未建立標準的流程以及模型比較等探討，故本研究欲透過不同實驗設計，建立完整的主動學習高分子設計流程，以兩種案例進行測試，最終再將此流程用於純實驗驅動設計，並以 Vitrimer 作為範例。

doi:10.6342/NTU202501030

第三章　研究方法

本研究共分為三個部分，分別為主動學習隨機共聚物序列設計、小數據驅動多目標最佳化以及實驗數據驅動 Vitrimer 配方最佳化，在本章節會分別介紹三部分所使用的資料集以及研究方法。

3.1　主動學習隨機共聚物序列設計

3.1.1　軟硬共聚高分子

在隨機共聚物序列設計的研究中，所使用的資料集為 Aoyagi 所建立的 ABA 軟硬共聚物資料集 [52]，此資料集中共有 1200 筆模擬資料，皆為 ABA 三段鏈區塊共聚高分子 (ABA-Triblock Copolymers) 及其以模擬得到的應力應變曲線 (stress-strain curve, ss-curve)，基於 Aoyagi 的研究基礎，本研究使用其資料集作為初始資料，並延伸到更為複雜的隨機共聚高分子序列設計。由於隨機共聚高分子的排列組合更多，會面臨統計上常見的組合爆炸問題，以最簡單的兩種 AB 單體構成的共聚物，鏈長 60 至 120 為例，構成的設計空間共有以下組合:

$$\frac{1}{2}\sum_{i=60}^{120} 2^i \simeq 1.33 \times 10^{36} \tag{3.1}$$

doi:10.6342/NTU202501030

此設計空間的規模即使以模擬方法，也難以全部標註，因此本研究中導入主動學習解決此議題。

而在此研究設定的目標材料為丁苯橡膠 (Styrene-Butadiene Rubber, SBR)，SBR 是一種具優異性能的常見彈性體材料，其中硬的單體為苯乙烯 (Styrene)，之後在本研究中以”A” 表示，而軟的單體為丁二烯 (Butadiene)，以”B” 表示，高分子鏈及單體結構式如圖3.1所示。

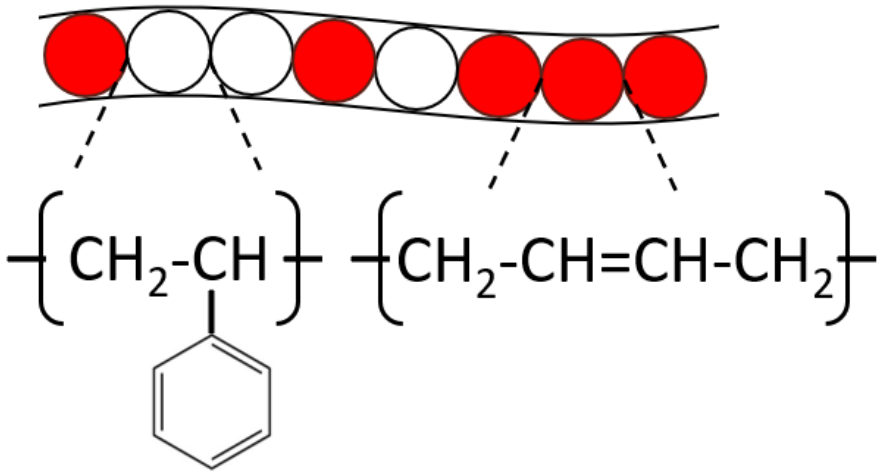


圖 3.1: SBR 示意圖。白色單體為苯乙烯，紅色單體為丁二烯。

3.1.2 模擬流程

在本研究中，會採用拉伸模擬的結果作為標註性質。軟硬共聚高分子的特性是會形成相分離結構 [53]，根據過往的研究，此相分離現象所形成的初始結構，會大幅影響表現出來的機械性質，因此在本研究的模擬中，必須先進行相分離的模擬，接著再使用分子動力學進行拉伸，才能確保模擬可靠度。模擬流程可以分為以下三個階段進行:

1. 軟硬共聚物會形成相分離結構，初始相分離結構會影響後續拉伸的行為。因此，在本研究中，我們使用自洽場方法（Self-Consistent Field Theory, SCFT）模擬方法來獲取相分離結構，SCFT 是一種量子化學的迭代計算方

doi:10.6342/NTU202501030

法，能夠透過電子分佈的計算進而得到單體密度資訊，此相分離模擬是透過軟體 SUSHI[54] 進行。

2. 使用節點密度偏置蒙地卡羅（Node Density-Biased Monte Carlo, NDBMC）方法，透過蒙地卡羅迭代方式，將分子模擬的珠子填充至模型中直到收斂，就可以將 SCFT 的相分離結果轉換為密度相等的顆粒模型。

3. 最後一步是利用粗粒化分子動力學（Coarse-Grained Molecular Dynamics, CGMD）模擬拉伸測試。從獲得的相分離結構開始進行拉伸模擬，我們使用軟體 COGNAC[55] 進行模擬，最終得到應力應變曲線。

其中本研究中的 CGMD 模擬是採用 Kremer–Grest 模型 [56]，這是一種簡化後的粗粒化高分子模型，將高分子單體簡化為珠子，單體之間的鍵結視為彈簧，僅考慮 Lennard-Jones 交互作用以及鍵結作用力，由於忽略了很多高分子的自由度，比如鍵角已經在粗粒化被忽略，因此可以提升模擬的時間及空間尺度，進而捕捉高分子在拉伸時的回彈效應，因此此模型可以有效預測高分子彈性體在拉伸時的行為。而在粗粒化的過程中，會使用無因次單位進行模擬，本研究中之參數以 σ、ϵ 以及 m 來做計算，分別代表單位長度、單位能量以及單位質量，其他模擬中之參數都可以由此三個基本量推導得到，比如模擬中重要的單位時間步長可以寫為 $\tau = \sigma(m/\epsilon)^{1/2}$。而高分子 Kremer-Grest 模型的 Lennard-Jones 作用力寫為以下形式:

$$U_{ij}^{LJ}(r) = \begin{cases} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} - \left(\left(\frac{\sigma_{ij}}{r_{ij,\text{cut}}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij,\text{cut}}}\right)^{6}\right)\right], & r \le r_{ij,\text{cut}} \\ 0, & r > r_{ij,\text{cut}} \end{cases} \tag{3.2}$$

ϵ_{ij} 和 σ_{ij} 為單體 i 和單體 j 之間的 Lennard-Jones 參數，以無因次的單位長度以及單位能量表示，而 $r_{ij,\text{cut}}$ 則是粒子模擬中的截斷距離 (cut-off distance)，根據過往文獻，將 A 粒子間的截斷距離設為 $2,5\sigma$，B 粒子間則為 $2^{1/6}\sigma$[57]，Kremer-Grest

doi:10.6342/NTU202501030

模型是透過透過改變截斷距離，區分軟硬單體的不同行為。

鍵結作用力則是寫為:

$$U^B(r) = \begin{cases} -\frac{1}{2}kR_0^2 \ln\left[1-\left(\frac{r}{R_0}\right)^2\right], & r \le R_0 \\ \infty, & r > R_0 \end{cases} \tag{3.3}$$

其中 k 設為 $30\epsilon/\sigma^2$，R_0 設為 1.5σ，兩者皆為以單位量表示的常數，另外，此鍵結作用力的形式，可以防止模擬中出現高分子鏈重疊的可能，確保高分子運動的合理性及正確性。

以上三個模擬步驟所獲得的結果分別如圖3.2所示，所有步驟皆是使用 Octa[58] 套件中的功能進行模擬，包含 SUSHI 以及 COGNAC，模擬的盒子大小設定為 16*16*16，經實驗此大小可以在精度以及模擬時長中取得平衡，本研究使用 Intel I9-13900k 進行模擬，每一次模擬約須耗時 5 至 8 小時。

3.1.3 代理模型

3.1.3.1 模型輸入

本研究使用 7 個參數描述 AB 隨機共聚高分子，分別列在下表3.1，圖3.3則展示了其中一個由 AB 隨機共聚高分子轉為 7 個參數的案例，這些參數除了高分子材料最重要的聚合度之外，也著重於使用統計方法，描述隨機共聚高分子的區域鍊段結構，且這些區域的序列特徵已被證實會影響 AB 隨機共聚高分子的相分離行為，間接影響所表現出的機械性質，故本研究選擇用以作為模型輸入，以利模型有效進行預測。

doi:10.6342/NTU202501030

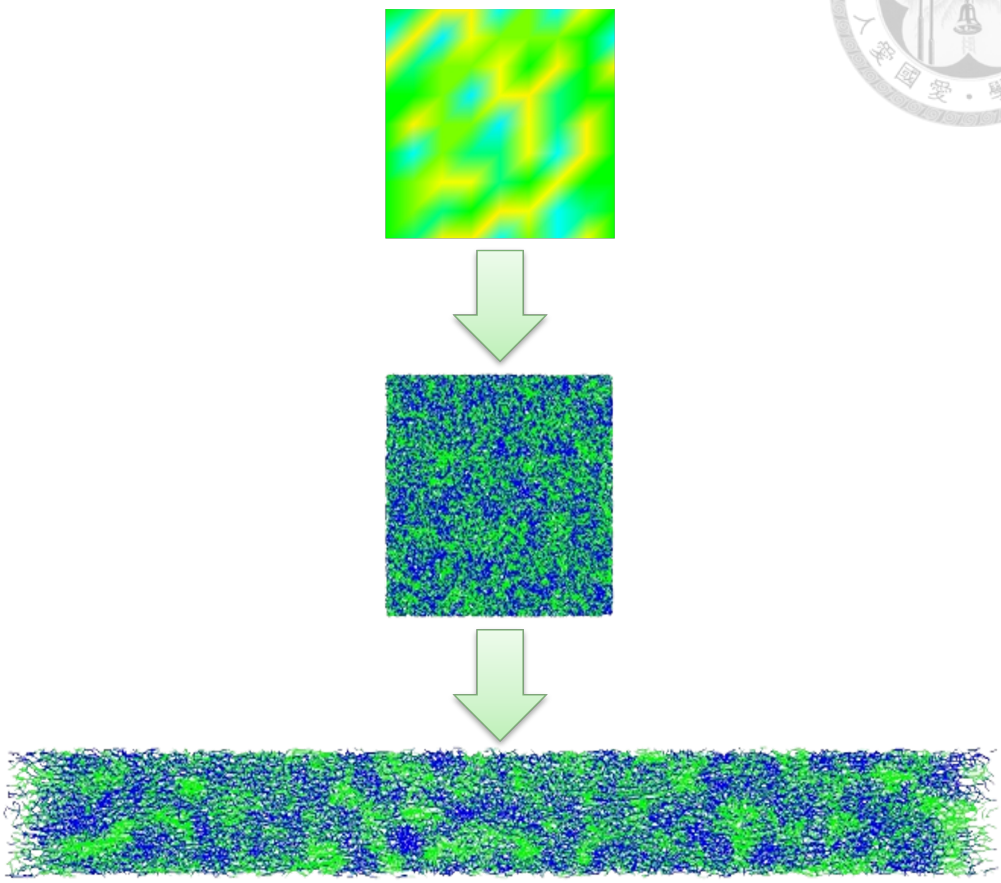


圖 3.2: 模擬三步驟流程圖，三個子圖分別為 SCFT、DBMC、CGMD 模擬後所得之計算結果。

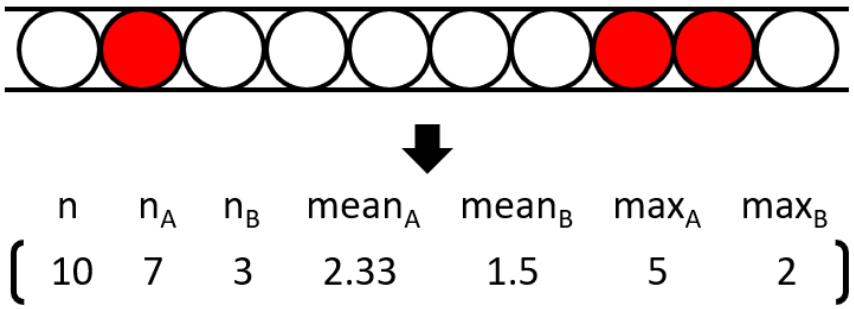


圖 3.3: AB 隨機共聚高分子的參數描述案例。

doi:10.6342/NTU202501030

表 3.1: 模型的輸入參數表。

參數名稱	參數說明
n	高分子鏈長 (聚合度)
n_A	A 單體數量
n_B	B 單體數量
$mean_A$	A 單體平均聚集長度
$mean_B$	B 單體平均聚集長度
max_A	A 單體最大聚集長度
max_B	B 單體最大聚集長度

3.1.3.2 模型選擇

本研究中所選用的模型為高斯過程回歸 (Gaussian Process Regression, GPR)，GPR 是一種統計模型，具有可解釋性高、泛化能力強等優點，被廣泛用在複雜函數擬合中，且由於 GPR 推論結果是一個機率分佈而非單一值，能夠直接用來估計資料不確定性，因此常用作為主動學習架構中的代理模型。GPR 的函數形式由平均值以及協方差組成，寫作:

$$f(x) \sim GP(m, k(x_i, x_j)) \tag{3.4}$$

其中 $f(x)$ 是真實的未知函數，GP 是高斯過程，m 表示均值函數 $m(x)$，而 k 則代表協方差函數 $k(x_i, x_j)$。在本研究中，使用七個描述高分子的參數 (表3.1) 來定義均值函數，而協方差函數則由一個核函數 (Kernel Function) 來定義。在訓練高斯過程回歸（GPR）時，核函數的選擇至關重要，因為核函數在協方差矩陣中，用以學習到輸入參數之間彼此的交互作用關係。經過在資料集上測試多種核函數後並比較結果後，選定了 Matérn 核函數用於本研究，函數形式寫為:

$$k(x_i, x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}}(\frac{\sqrt{2\nu}}{l}d(x_i, x_j))^{\nu} K_{\nu}(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)) \tag{3.5}$$

其中，$d(\cdot, \cdot)$ 為歐式距離，K_{ν} 為 Bessel function，$\Gamma(\cdot)$ 為 gamma function，l 表示核函數的長度尺度，而 ν 則是控制整體函數平滑程度的參數，在研究中將 ν 設為

doi:10.6342/NTU202501030

1.5，l 則會在訓練中透過資料標註進行參數優化。

GPR 的學習與預測過程是建立在貝葉斯定理 (Bayes' theorem) 中，關於先驗機率與後驗機率的基礎上。在獲得標註資料集後，這些數據可透過最大對數似然 (Maximum Log Likelihood) 來最佳化均值函數與協方差函數中之參數，接著透過訓練好的模型，我們便能透過聯合分佈 (joint distribution)，預測與訓練資料相關的未知數據。本研究中使用最常見的記號表示，X 為已知標籤 f 的訓練集輸入，而 X^* 則表示待預測的未知資料集。其聯合機率分佈可表示為：

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim N\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right) \tag{3.6}$$

其中，f 是訓練資料的聯合分佈，f_* 是測試資料的聯合分佈，$K = K(X, X)$ 為訓練資料之間的協方差矩陣，$K_* = K(X, X_*)$ 表示訓練資料與測試資料之間的協方差，$K_{**} = K(X_*, X_*)$ 則是測試資料之間的協方差矩陣。因此，我們可以透過從訓練資料得到的聯合分佈，進而獲得 f_* 的後驗分佈 (posterior distribution)。本次研究中使用 scikit-learn 套件進行 GPR 模型建立及訓練 [59]。

3.1.4 主動學習流程

在本研究中，是採用基於模型不確定性 (uncertainty) 的主動學習策略，由於 GPR 的推論基於貝氏機率，因此在預測時除了能給出平均值，還能同時帶有模型預測之標準差，過往主動學習研究經常採用 GPR 模型，在推論過程中給出之標準差，作為衡量資料不確定性的指標。本研究中也採用相同作法，研究流程首先會在 ABA 三段鏈共聚物資料集訓練 GPR 模型，並以未標註的隨機共聚物資料集作為設計空間，在每一次迭代中，會使用訓練好的 GPR 模型對設計空間資料進行預測，並選擇模型不確定性最高的資料進行標註，標注是使用前述的模擬方法，進

doi:10.6342/NTU202501030

行標註後會將模擬後結果加入訓練資料集中，並重新訓練模型，此過程即為一次迭代。接著重複流程，直到滿足設定的中止條件為止，此基於不確定性的主動學習方法如圖3.4。透過主動學習對於不確定性的衡量，預期能夠達到使用最少標註之成本，使模型快速進步並收斂的目的，本研究即透過此方式，達到少量標註資料訓練模型。

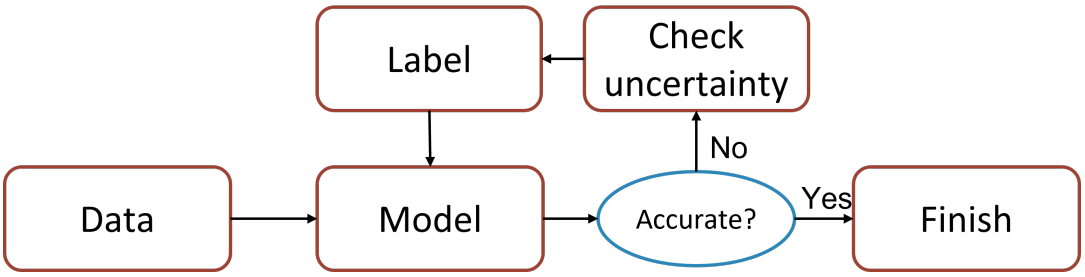


圖 3.4: 基於不確定性的主動學習流程圖。

3.1.5 資料視覺化方法

由於描述 AB 隨機共聚高分子的資料參數向量共有七個維度，難以直接進行視覺化呈現，因此在本研究中，會預先對資料進行降維處理，再繪製資料分佈圖進行視覺化，呈現資料間的關係。本研究中所使用的降維方式為 McInnes 等人所提出的 Uniform Manifold Approximation and Projection(UMAP)[60]。UMAP 是一種非線性的降維方法，透過在每個點周圍展開半徑，建立高維空間中的圖形，接著再將此高維圖形映射 (projection) 到低維空間上。跟早期線性降維方法如主成分分析 (PCA) 相比，非線性降維能夠保留更多高維空間中的結構資訊。此外，UMAP 相較於另一個主流的非線性降維方式 t-SNE[61] 速度更快，在越高維的空間此差異由為顯著。因為有上述的優點，故本研究中選擇 UMAP 作為資料視覺化過程的降維工具。

doi:10.6342/NTU202501030

3.1.6 小結

在本章節中，介紹主動學習用於隨機共聚物序列設計部分的研究方法，以下為本章節統整:

1. 本研究的目標為隨機共聚物序列設計，目標材料選定軟硬高分子 SBR，並使用 CGMD 拉伸模擬進行機械性質標註。
2. 本研究使用 GPR 作為代理模型，模型輸入為描述隨機高分子的七維向量，並透過 GPR 模型在預測時提供的標準差資訊，做為主動學習的不確定性檢驗方式，進行主動學習流程。
3. 研究中會使用 UMAP 作為降維工具，接著進行資料分佈圖繪製，透過此視覺化方法，了解主動學習過程中，模型挑選標註資料的行為。

3.2 小數據驅動多目標最佳化

3.2.1 實驗資料集

在此部分的研究中，使用 PolyInfo[6] 作為實驗資料集，如同前述提到，PolyInfo 是目前規模最大的高分子材料文獻資料庫，具有龐大數量的高分子當體及其對應的性質。由於本研究目標為多目標最佳化，因此採用兩種目標性質，分別為延展性 (Elongation) 以及彈性模數 (Elastic Modulus)，分別代表材料的延展性以及機械強度，將資料集繪製成分布圖如圖3.5所示，從圖上可以發現此兩性質之間有強烈的互斥 (trade-off) 關係，符合先前在多目標最佳化章節所提到的材料性質多目標最佳化應用場景，即材料設計時的互斥問題。而 PolyInfo 的資料所收集的皆為單體的 P-SMILEs，因此圖上每一個點即代表一種高分子單體所對應的材料

doi:10.6342/NTU202501030

強度及延展性，資料集中共有 805 筆資料。

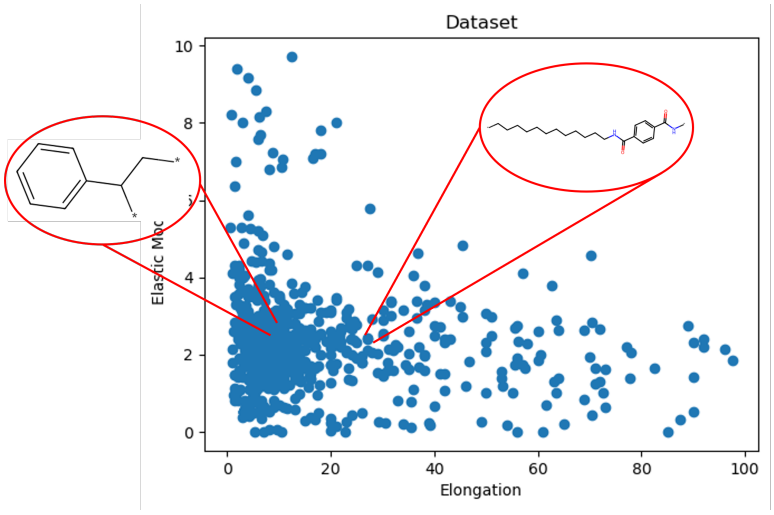


圖 3.5: PolyInfo 資料集分佈。

3.2.2 研究流程

在此部分的研究中，目的是模擬真實世界以數據驅動進行實驗設計，進而達到最佳化單體配方。因此，雖然 PolyInfo 的資料集共有 805 個有標註的資料點，為了模擬目前實驗能產生的初始資料規模，通常只有 100 筆以下標註資料，本研究僅以 30 筆資料作為初始標註資料集，其他資料作為未標註的設計空間，提供主動學習流程中模型進行搜索。接著同樣進行主動學習流程，模擬現實中的材料設計場景，透過模型引導下一步實驗設計，能夠快速達到單體最佳化之目的。

3.2.3 代理模型

本研究主要在探討如何使用小資料集驅動主動學習流程，因此與傳統機器學習相比，標註資料更少。為了使用小數據集有效建立代理模型，在本研究中測試了三種不同代理模型，分別為 GPR、Gradient Boost Regression(GBR) 以及基礎模型 TransPolymer[24]。其中 GPR 已在上一部份研究方法中提到，是常用於主動學習研究中的一種統計模型。Gredient Boosting 是常被用於小資料集訓練的機器

doi:10.6342/NTU202501030

學習演算法，其概念是透過組合好幾個不同的簡單函數，而構成的機器學習模型，因此是由簡單函數構成，讓 GBR 有不容易過擬合 (overfitting) 以及學習快速等優點，適合用於資料稀缺的場景。而將 Gradient Boosting 的概念用於回歸問題 (regression) 時，透過組合簡單模型預測之殘差 (residue)，使之越來越逼近真實解。

基礎模型 TransPolymer 則是透過未標註的高分子資料集 PI1M[62] 進行預訓練，總共透過一百萬個高分子的 P-SMILEs 進行，預訓練是採用隨機遮蓋 (random masking) 的方法進行，因此可以將 TransPolymer 視為具有高分子官能基排列先備知識的機器學習模型，僅需透過少量資料進行微調，即可進行代理模型的訓練。而本研究中的微調方法則是加入一層全連階層 (Fully Connected Layer)，再利用標註資料進行全連階層參數的訓練，研究中損失函數 (Loss Function) 使用 AdamW，學習率 (learning rate) 設為 5e-5。

3.2.4 主動學習策略

在本研究中的主動學習，目的除了增強模型表現之外，還有同步進行多目標最佳化，因此會採用兩種策略進行標註資料採樣，分別為探索 (exploration) 以及利用 (exploitation)。在最佳化以及強化學習等領域，探索是針對模型未知的區域進行採樣，以增加採樣的資料範圍，進而最大化模型對於設計空間的認知。而利用則是相反的概念，此策略是採用目前模型的學習成果，使用當下的模型預測進行最佳化採樣之策略，簡單示意如下圖3.6。由於此兩種策略的特性，在最佳化流程中，往往是會帶有互斥的效果，即探索主要目的是增加模型的學習效率，而利用則是加速多目標最佳化，然而若模型預測力不足，卻反而增加標註成本，因此需要仔細平衡兩種策略，才能在增強模型以及最佳化效率之間取得平衡。

在本研究中，探索是採用不確定性採樣，而利用則是選取預測值構成之

doi:10.6342/NTU202501030

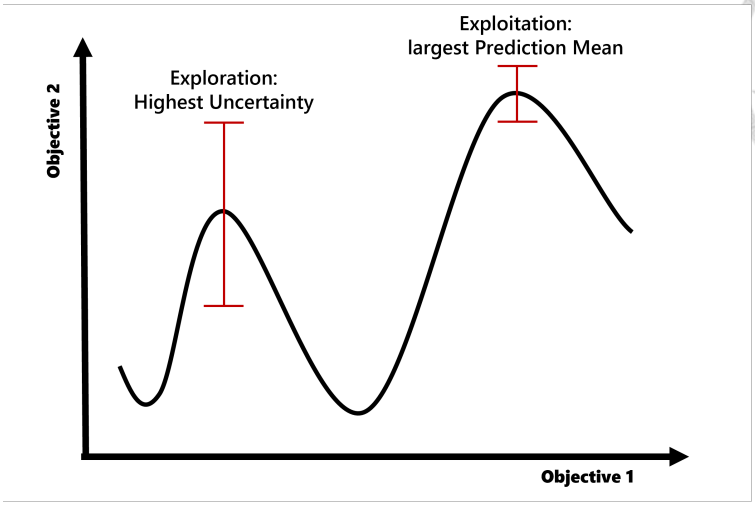


圖 3.6: 簡單函數的探索 (exploration) 以及利用 (exploitation) 示意圖。

Pareto front 上的資料點，我們將此方法稱之為 Pareto 引導主動學習 (Pareto-Guided Active Learning)，透過模型引導下一步採樣的實驗對象，可以達到快速進行多目標最佳化之目的。本研究的目標性質為二個維度，因此可以將 Pareto 引導主動學習的兩個不同策略採樣方式繪製如圖3.7，實驗中會透過比較不同策略之間的比例，比較不同比例所得到的最佳化效果，包含標註資料數量以及最佳化之結果，以作為未來實際應用此方法於實驗設計之指標。

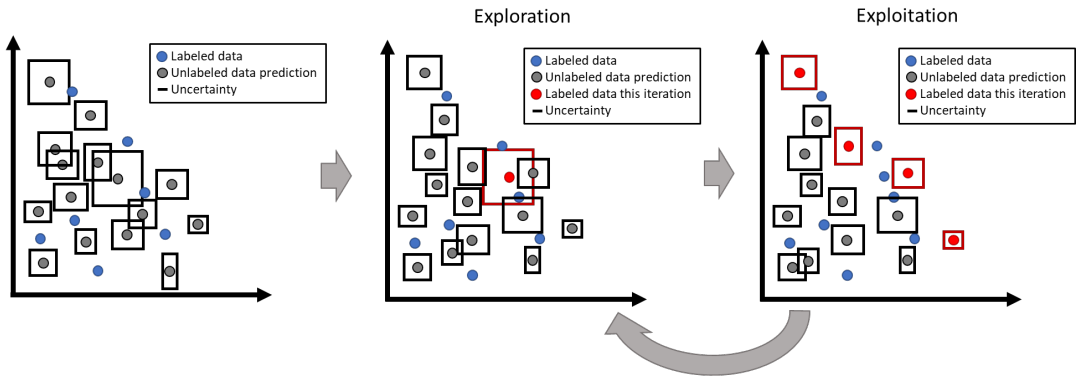


圖 3.7: 本研究中主動學習兩種策略示意圖，探索 (Exploration) 對不確定性最大的資料進行採樣，利用則是對 (Exploitation) 在 Pareto front 上的點進行採樣。

doi:10.6342/NTU202501030

3.2.5 模型不確定性

GPR 的模型在推論中會同時給出不確定性，GBR 模型在此研究中，則是透過集成 (ensumble) 的方式，透過使用不同初始模型訓練，並集合所有模型取平均值及標準差，透過此方式所得到之標準差作為衡量資料點不確定性之指標。

TransPolymer 的不確定性衡量方式較為複雜，由於目前研究中對於如何衡量神經網路預測不確定性仍有爭議，衍生出許多算法，本研究採用的是 Monte-Carlo Dropout(MCD) 的方法 [63]。MCD 方法是由 Gal 等人提出，其概念是透過隨機遮蓋神經元，讓神經網路的預測值不再是一個單一值，而是變為一個機率分佈，此分佈可表示為:

$$q(y^*|x^*) = \int p(y^*|x^*, \omega)\, q(\omega)\, d\omega \tag{3.7}$$

其中，y^* 為對應輸入 x^* 的輸出，p 表示隨機遮蓋的機率，ω 表示與模型中隨機遮蓋相關的隨機變數，這些變數來源於模型的權重。在引入 MCD 機制之後，神經網絡的權重 (weights) 可表示為：

$$W_i = M_i \cdot \text{diag}\left(\left[z_{ij}\right]_{j=1}^{K_i}\right) \tag{3.8}$$

其中，$z_{ij} \sim \text{Bernoulli}(p_i)$，對 $i = 1, \ldots, L$ 及 $j = 1, \ldots, K_{i-1}$ 成立，表示以機率 p_i 進行 Bernoulli 試驗的結果，其維度與 W_i 相符。M_i 表示參數轉換矩陣。因此，進行 T 次 MCD 推理後的模型預測結果可表示為：

$$\mathbb{E}_q(y^*|x^*)(y^*) \approx \frac{1}{T}\sum_{t=1}^{T} \hat{y}^*\left(x^*, W_1^t, \ldots, W_L^t\right) = p_{MC}(y^*|x^*) \tag{3.9}$$

其中，y^* 為對應輸入 x^* 的輸出，而 p_{MC} 表示經過 MCD 後所得預測機率分布的均值 [64]。MCD 方法視覺化如圖3.8所示。

doi:10.6342/NTU202501030

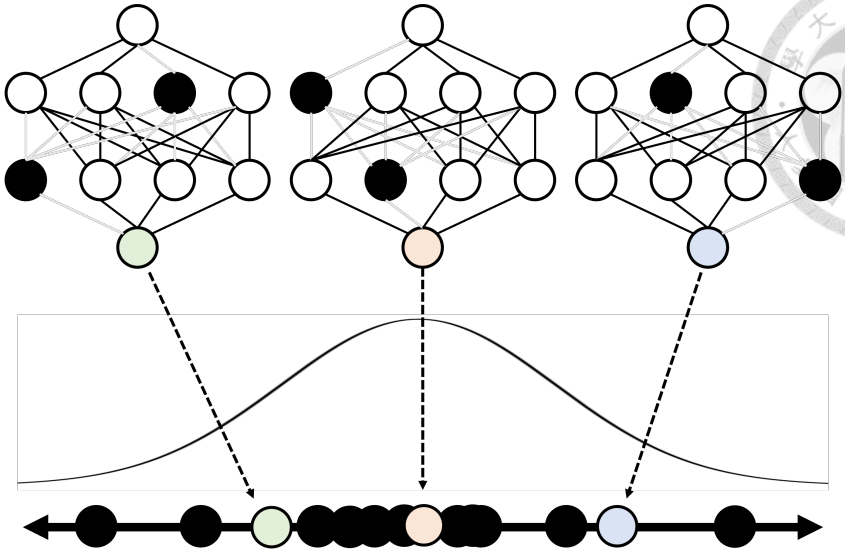


圖 3.8: MCD 方法視覺化示意圖。

3.2.6 多目標最佳化衡量指標

在多目標最佳化研究中，常用的衡量指標為超體積 (hypervolume)，超體積的算法為先定義參考點，接著計算由參考點與 Pareto front 所構成的體積。由於超體積可以同時進行收斂性、效率的衡量，並且可以簡單算出單一點所造成的貢獻，因此被廣泛用於多目標最佳化的研究中。圖3.9視覺化呈現了一個二維性質空間所構成的超體積。

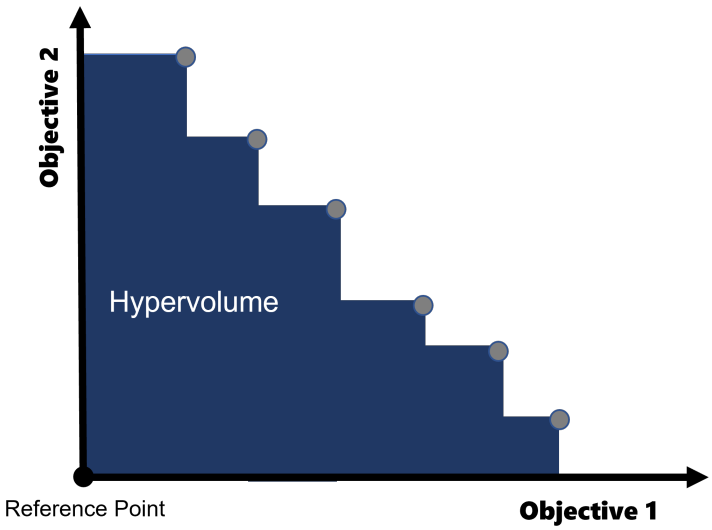


圖 3.9: 超體積視覺化示意圖。

doi:10.6342/NTU202501030

3.2.7 小結

在本章節中，介紹了小數據驅動多目標最佳化研究所用到之方法，以下為本章節統整:

1. 本研究中使用 PolyInfo 作為研究資料集，資料集中包含機械強度以及延展性兩種互斥的性質標註，並為了模擬現實場景，只會以 30 筆標註資料作為起始條件，其餘當作未標註的設計空間提供模型搜索。
2. 為了進行多目標最佳化研究，研究中會分別比較以基礎模型 TransPolymer、GBR 及 GPR 作為代理模型的最佳化表現，同時也會嘗試兩種主動學習的採樣策略。
3. 研究中會以超體積作為最佳化的衡量指標，除了檢驗最佳化得到的超體積，也會透過超體積收斂性檢驗主動學習流程應用於最佳化的穩定性。

3.3 實驗數據驅動聚酯型 Vitrimer 最佳化

3.3.1 Vitrimer 單體選擇

本研究所使用的單體共有雙酚 A 二縮水甘油醚 (Bisphenol A Diglycidyl Ether, DGEBA)、可變官能基之二羧酸、3-(羧甲基) 環戊烷-1,2,4-三羧酸 (3-(Carboxymethyl)cyclopentane-1,2,4-Tricarboxylic acid, TCAA) 與檸檬酸 (Citric Acid)，單體使用示意如圖3.10，分子式列於附錄 A，TCAA 與檸檬酸在反應中作為交聯劑，透過 DGEBA 的環氧基與二羧酸的羧基進行酯化反應形成網狀結構，並在特定條件 (如高溫或催化劑存在) 下發生動態交換重組拓撲結構，實現 Vitrimer 的酯交換反應。

doi:10.6342/NTU202501030

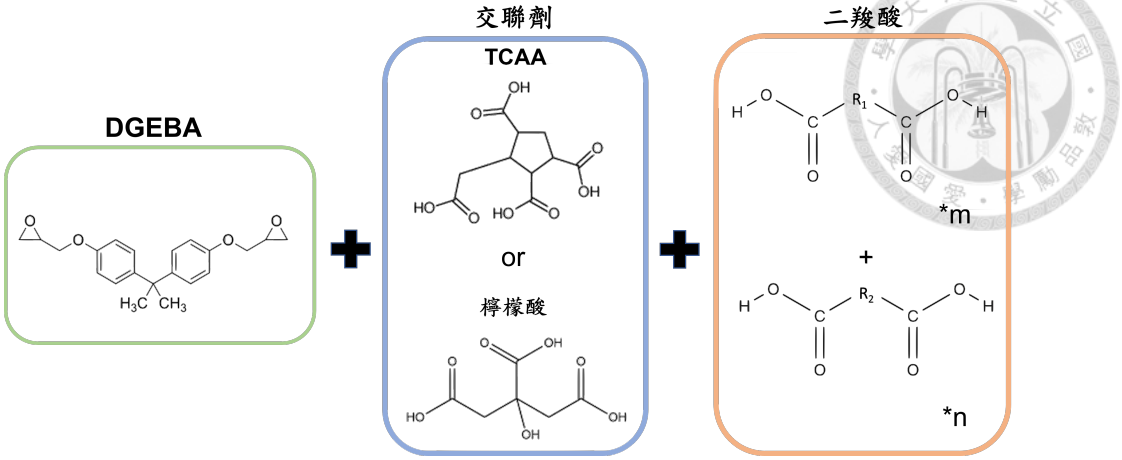


圖 3.10: 本研究中所使用單體組合。

每一次進行合成時，環氧基皆是使用 DGEBA，交聯劑使用 TCAA 或是檸檬酸 (Critic acid)，而二羧酸則是在每一次合成過程中從丁二酸 (Succinic Acid)、己二酸 (Adipic Acid)、癸二酸 (Sebacic Acid) 這三種不同長度的二羧酸單體中選擇，並在確保環氧基與羧基的當量數相同的前提下，添加不同比例做為單體進行合成。以上三種二羧酸單體各自在性質上有不同的優劣，透過兩種不同二羧酸的組合，本預期可以設計出性質更多元靈活的聚酯型 Vitrimer，以滿足各種應用上的需求，並且由於設計空間的複雜度，先前尚未有相關研究進行同時添加兩種不同二羧酸對於性質影響的探討，因此本團隊欲使用主動學習架構來協助設計，解決此一材料設計應用上的議題。

3.3.2 Vitrimer 合成及性質量測方法

此章節對 Vitrimer 的合成方法進行說明，首先會將 DGEBA、兩種二酸以及一種交聯劑 (TCAA、檸檬酸) 粉末混合，將混合後的粉末放進 PTFE 燒杯中，視不同單體將燒杯置於特定溫度的油浴中加熱並攪拌，直至完全溶解。獲得均勻的混合物後，再加入特定量的催化劑，並手動攪拌直至溶解並混合均勻。在本研究中，預定使用二正丁氨 (Di-n-butylamine) 作為催化劑。接著，將混合物倒入金屬模具中，放入熱壓機中在 110°C 下預固化 1 小時，隨後在 160°C 下固化 6 小時，就完

doi:10.6342/NTU202501030

成 Vitrimer 的合成。在研究中所有樣品的 [羧基]/[環氧基] 比例均固定為 1:1，二正丁氨的濃度按照相對於 DGEBA 莫爾比的 0.1 倍計算。

在進行 Vitrimer 的合成後，便可以進行 Vitrimer 性質量測作為標註。在本研究的配方中，所合成的 Vitrimer 為熱固型 Vitrimer，因此材料機械性質的優劣對其應用至關重要，因此，本研究選定極限抗拉強度 (Ultimate Tensile Strength, UTS) 作為設計目標。其量測方式為根據 ASTM D638 V 型標準，將樣品製備成啞鈴形狀，於室溫下以通用材料測試儀 (CometechQC-508M2F) 以 10mm/min 的速度進行拉伸測試，記錄應力-應變曲線 (stress-strain curve)，接著再從應力應變曲線中取應力最大的點得到極限抗拉強度。

3.3.3 設計空間

在本研究中，事先建構設計空間以進行主動學習搜索流程。如上一部分所述，本研究中環氧基單體使用 DGEBA，交聯劑使用 CA 和 TCAA 兩種，二羧酸單體有丁二酸、己二酸、癸二酸三種，在每次合成中，確保環氧基與羧基莫耳數相同的前提下，每次合成取出兩種，並更改兩者之間的比例，從 0.1 至 0.9，以 0.1 遞增，共分為 9 等份，比例定義為 Dicarboxylic Acid A/(Dicarboxylic Acid A + Dicarboxylic Acid B)。因此，設計空間可計算為 $1\times C_2^3\times 2\times 9 = 54$，共有 54 種組合，如表3.2所示，研究目標是從這些設計空間中透過主動學習驅動實驗設計，找出最佳化成分的 Vitrimer 單體及配方。

表 3.2: 本研究設計空間。

設計因子	性質	範圍	參數數量
Epoxy	離散	[DGEBA]	1
Types of Dicarboxylic Acids	離散	[3 Dicarboxylic Acids]	C_2^3
Cross-linker	離散	[TCAA, Citric Acid]	2
Ratio of (Dicarboxylic Acid A / Total Dicarboxylic Acid)	連續	[0.1–0.9]	9

doi:10.6342/NTU202501030

3.3.4 主動學習流程

本研究使用實驗數據驅動主動學習輔助 Vitrimer 設計，受限於 Vitrimer 實驗的時間成本限制，因此僅以三筆資料作為初始資料集，並採用實驗設計進行初始資料均勻選取，初始資料參數如下表3.3所列。

表 3.3: 初始資料集樣品選擇。

樣品	二酸 (1)	二酸 (2)	比例	交聯劑
樣品 1	Succinic Acid	Adipic Acid	0.3	CA
樣品 2	Succinic Acid	Sebacic Acid	0.5	TCAA
樣品 3	Adipic Acid	Sebacic Acid	0.8	TCAA

研究中由於資料量更小，因此代理模型選擇 GPR，Vitrimer 的描述方式使用分子指紋，而主動學習策略則是採用模型不確性引導的探索 (Exploration)，以及選取模型預測最佳解的利用 (Exploitation)，使用此兩策略交替進行設計空間的採樣，以利完成 Vitrimer 最佳化。

3.3.5 小結

在本章節中，介紹了實驗數據驅動 Vitrimer 設計所用到之方法，以下為本章節統整:

1. 本研究中將先前研究建立的設計流程用於真實實驗數據驅動高分子設計上，設計目標為 Vitrimer 機械性質。
2. 為了進行此研究，會實際進行聚酯型 Vitrimer 的合成，並將設計空間定義為二酸單體種類、彼此之間的比例以及交聯劑。
3. 本研究使用 GPR 作為代理模型，並採用探索以及利用交替的採樣策略，以利在設計空間中快速達到最佳化。

doi:10.6342/NTU202501030

第四章　結果與討論

4.1　主動學習隨機共聚物序列設計

在第一部分的研究中，目標是進行隨機共聚物的序列設計。在研究方法的章節有提到，本研究使用 Aoyagi 所建立的 ABA 三段鏈區塊共聚物資料集 [52]，並以此作為基礎，使用主動學習技術進行更複雜的隨機共聚物設計研究，以下為研究結果:

4.1.1　隨機共聚物資料挑選及資料分佈

本研究中，我們將隨機共聚物鏈長限制在 60 到 120 的範圍內，設計空間中所有隨機共聚物皆由兩種單體組成，分別以 A 和 B 表示，對應於硬段與軟段，所有的排列組合總數已於先前計算，由於組合過多難以全部進行設計，我們採用抽樣策略，並透過設計採樣方法以確保能夠在設計空間中均勻地採樣。

由於鏈長在決定聚合物性質中扮演關鍵角色，因此研究中必須確保抽樣共聚物聚合度的均勻度，我們控制聚合物鏈長在 60 至 120 之間，以間隔為 2 進行採樣。每種長度隨機抽樣 50 種，總計為 $31 \times 50 = 1550$ 種聚合物。接著，對於相同長度的聚合物，我們進一步控制其單體比例分佈的均勻性，此一因素在共聚物設計中極為重要，將 A 單體比例 0.1 開始，每種聚合物依序以 0.015 的比例遞增。

doi:10.6342/NTU202501030

最後，將每條序列隨機打亂，經過上述步驟，生成隨機共聚物設計資料集。

資料集最終包含 1550 條隨機共聚物序列，均未具備已標記的性質資訊。我們的目標是建立一個代理模型，來對應這些序列與其性質之間的關係。由於本資料集作為代理模型輸入為七維空間，直接理解其分佈情形較為困難，因此我們使用了 UMAP[60] 進行降維，用以協助視覺化隨機共聚物資料分佈。我們將維度從七維降至二維，結果如圖4.1所示。結果顯示，隨機共聚物均勻分散於整個降維後的空間中，而區塊共聚物則主要集中在資料的右側區域。

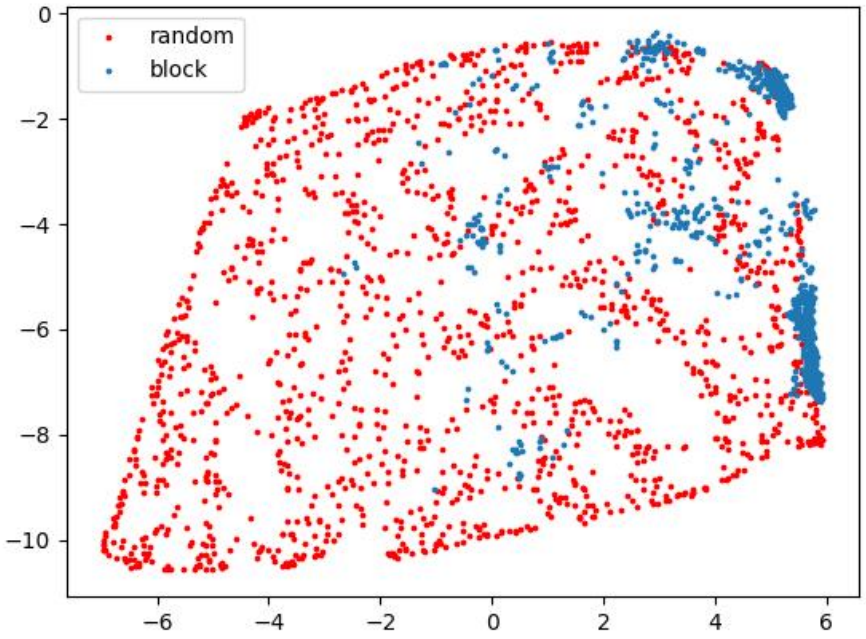


圖 4.1: 共聚物的資料分佈。紅點表示隨機共聚物，藍點表示區塊共聚物。

如預期所示，隨機共聚物的分佈明顯比區塊共聚物更為分散，先前僅有區塊共聚物的資料具有標註，而本研究的主要目標為預測隨機共聚物的應力應變曲線，因此，我們希望能透過主動學習，將區塊共聚物中的標記資訊有效地轉移至未標記的隨機共聚物。

doi:10.6342/NTU202501030

4.1.2 主動學習結果

在進行主動學習的流程之前，我們需要先建立一個以部分資料訓練的監督式學習訓練模型，才能透過模型的預測力進行採樣，在本研究中，我們使用 Aoyagi 的 ABA 三嵌段區塊共聚物資料集來訓練初始的模型。這個任務的第一步，是將區塊序列轉換為七維模型輸入，接著訓練一個 GPR 模型，本階段中使用了全部 1200 筆 ABA 三嵌段共聚物資料，經過這個步驟後，我們便得到了能夠根據區塊共聚物輸入預測其對應應力應變曲線的 GPR 模型。

第二步是將已標記共聚物的資訊轉移到未標記的資料上。在這一步中，採用了基於不確定性的主動式學習方法，利用檢驗資料不確定性有效率地挑選樣本資料，來取代隨機抽樣。主動式學習的過程是以迭代方式進行，初始 GPR 模型在區塊共聚物資料上訓練完成後，我們對所有隨機共聚物資料進行推論，以估計每一個未標註資料點的預測不確定性。接著，我們選出具有最高不確定性的資料，並對該資料進行標註。

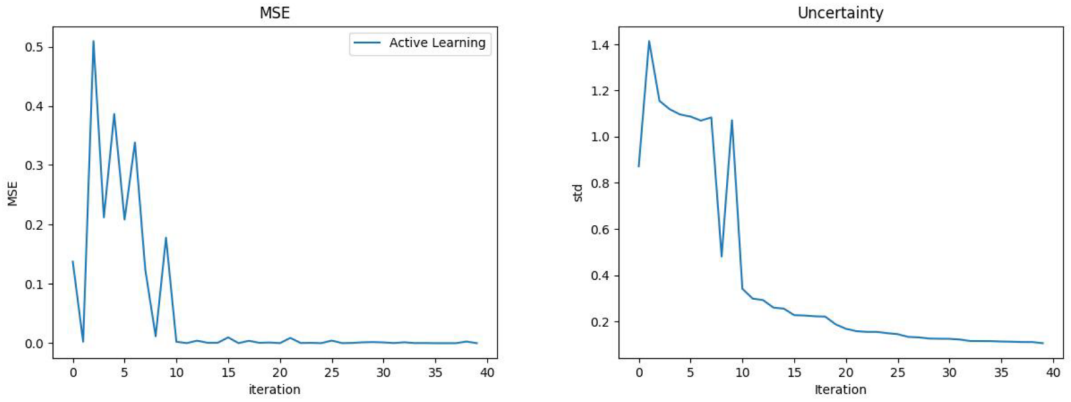
資料的標註方式是透過 CGMD 模擬來實現，模擬方法的詳細內容已在研究方法中說明，模擬完成後，我們得到了該筆高不確定性資料對應的應力應變曲線，作為其標註。我們將此模擬結果重新輸入模型進行訓練，完成一次迭代過程。接著，我們再次透過推論進行下一筆不確定性最高資料進行採樣，並重複整個流程。

主動學習流程讓代理模型能夠在每一次迭代中逐步改進預測正確率，我們透過模型預測結果與模擬結果之間的均方誤差 (Mean Square Error, MSE) 來評估改進的效果，MSE 定義如下：

$$MSE = \sum_{i=0}^{n} (\sigma_{ML}(i) - \sigma_{MD}(i))^2 \tag{4.1}$$

doi:10.6342/NTU202501030

其中，σ_{ML} 為代理模型所預測的應力，σ_{MD} 為模擬所預測的應力。n 的範圍從 0 到 101，對應的應變值從 0 到 3。此外，在每一步中監控不確定性，也能以視覺化方式呈現模型性能的提升。在本研究中，這兩種方法皆被用來評估主動式學習的成果，最後實驗結果如下圖4.2:



(a) MSE (b) Uncertainty

圖 4.2: (a) 模型預測與模擬結果之間的 MSE。(b) 主動學習過程中的不確定性收斂圖。兩張圖皆顯示模型在約第 30 次迭代後趨於收斂。

結果顯示，在十次迭代後，模型預測與模擬結果之間的誤差變得非常小，並且在三十次迭代內變化不大，同時，不確定性在三十次迭代後大幅下降，接下來的十次迭代中變化趨於穩定。這些發現顯示，主動學習過程在四十次迭代後已經趨於收斂，透過兩張圖的視覺化，呈現模型預測以及真實資料的誤差，以及模型不確定性的收斂，可以說明使用不確定性採樣的主動學習策略，可以幫助挑選標註資料，讓模型有效率的進行學習。

4.1.3 模型採樣資料分佈

在完成主動學習流程後，我們對基於不確定性的主動學習如何選擇資料進行了視覺化分析，如同研究方法段落所述，由於輸入資料共有 7 個維度 (表3.1)，難以直接進行視覺化，因此在此研究中採用 UMAP[60] 對資料進行降維處理，使之能夠以二維分布圖的方式呈現，並將 ABA 三鍛鏈區塊共聚物 (Aoyagi 資料集)、

doi:10.6342/NTU202501030

隨機共聚物 (本研究設計空間) 以及主動學習採樣的資料分別標出，如圖4.3所示。

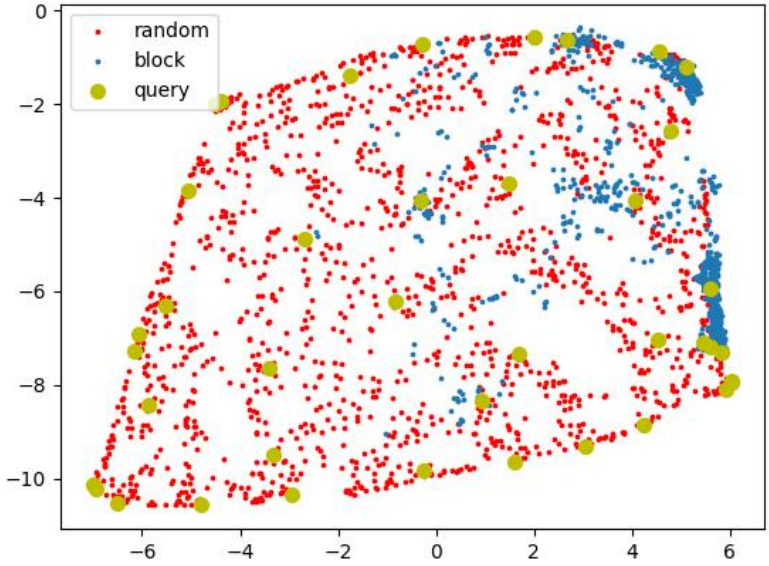


圖 4.3: 主動學習採樣視覺化呈現圖。

從圖中觀察可以得出結論：大多數被選取的資料 (圖中黃點) 位於設計空間的邊界，因此透過基於不確定性的主動學習，利用模型的預測力選取資料，能夠輕易地將此問題轉化為一個內插問題。儘管輸入資料具有七個維度，依靠實驗設計難以進行有效率選取標註，但不確定性策略依然能在龐大的設計空間中，有效找出最具價值的標記資料。這在面對組合爆炸問題或資料標記成本極高的情況下尤其具備優勢，可以透過機器學習方法挑選欲標註的標的，並顯著降低標駐所需成本。

在本研究中，透過不確定性採樣的主動學習策略，將模型學習轉為內插問題，使得我們在 1550 筆設計空間資料中，只需標記其中 40 筆資料，約只佔總設計空間中的 2%，主動學習引導實驗設計大幅降低了標記成本，展示出主動學習在標記資源有限或無法涵蓋整體資料集情境下的實用價值。

doi:10.6342/NTU202501030

4.1.4 設計空間之外採樣結果

本研究中的設計空間是採用均勻採樣選出 1550 種隨機共聚物序列，並已經透過主動學習建立代理模型，透過 MSE 以及不確定性的收斂，證實模型的預測力。在此部份實驗，則是要測試在設計空間以外的序列，透過主動學習建立起的代理模型，是否仍然保有其預測力。

我們隨機選擇了數個代理模型在主動學習流程中尚未見過的資料點，讓已訓練好的模型直接預測其應力應變曲線。所選的資料鏈長分別為 60、90、120、50 與 130。這五個樣本的組成列於表4.1中，其序列則顯示於圖4.4。前面三個樣本的聚合度在 80 到 120 的範圍中，用以評估模型是否能準確預測訓練資料長度範圍內的資料；而後兩個樣本則選擇略大於原先設計空間的鏈長限制，分別為 60 以及 130，選取此兩個樣品的目的為，用以測試模型在具備超出訓練資料長度範圍資料的清況下，是否仍能保有預測能力。

表 4.1: 研究中用於測試的五個設計空間外樣本序列組成。

	樣品 1	樣品 2	樣品 3	樣品 4	樣品 5
A 單體數量	45	45	30	30	70
B 單體數量	15	45	90	20	60

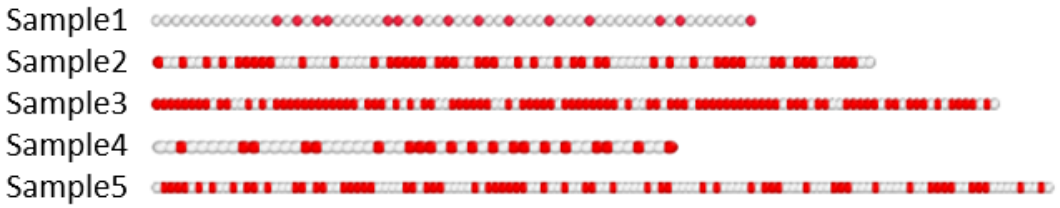
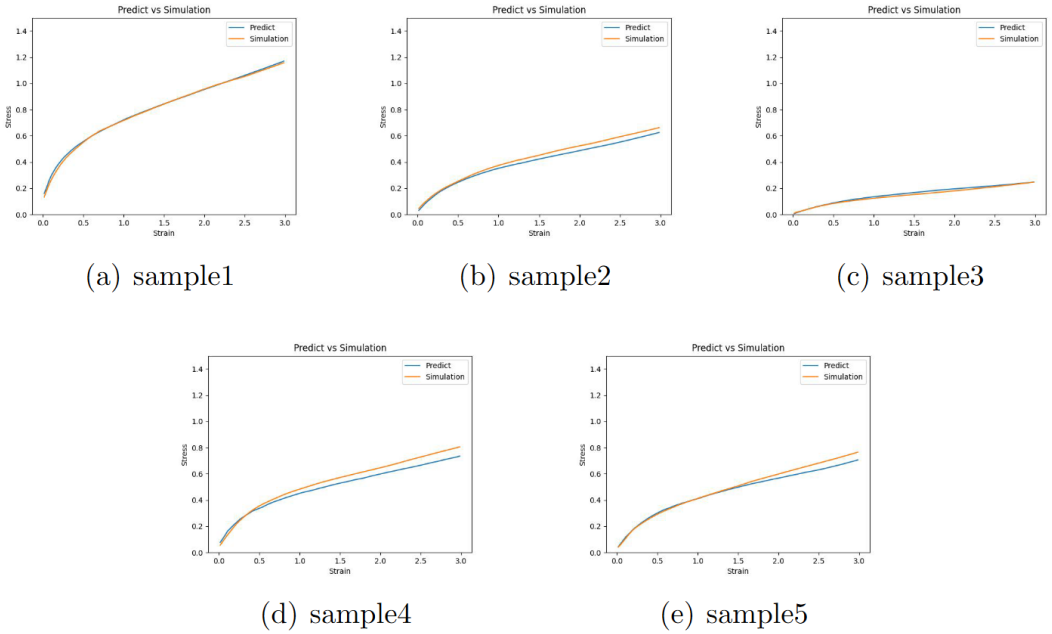


圖 4.4: 五個樣本的序列。白色代表 A 單體，紅色代表 B 單體。

隨後，我們將序列轉為七維的模型輸入格式，並使用訓練好的代理模型進行預測，同時也進行 CGMD 的計算模擬。將兩種方法所獲得的應力應變曲線同時繪製，相關結果如4.5所示，結果顯示兩者具有高度一致性。從圖中可以看出，該代理模型能夠有效預測 CGMD 模擬在不同區域中的結果，包括彈性區、降伏點及應

doi:10.6342/NTU202501030

力硬化行為，並且不僅在設計空間內的聚合物長度中表現優異，對於略微超出指定範圍的長度也同樣具有高預測力。



(a) sample1 (b) sample2 (c) sample3

(d) sample4 (e) sample5

圖 4.5: 五個樣本的序列。白色代表 A 單體，紅色代表 B 單體。

這些發現總結指出，我們已成功使用少量資料，構建了能準確預測設計空間中所有隨機共聚物序列機械性質的代理模型。透過這個代理模型，我們可以實現隨機共聚物的性質預測與序列設計。

4.1.5 小結

透過此部分的實驗，我們進行了主動學習用於序列設計的研究，相關研究結果總結如下:

1. 研究中使用均勻採樣建立大小為 1550 的設計空間，並採用主動學習進行模型訓練，透過 GPR 的不確定性採樣以及 CGMD 標註性質，以及 MSE 及模型不確定性收斂測試，最終僅透過 40 筆標註資料即完成模型訓練，減少 98% 標註資料數量。

doi:10.6342/NTU202501030

2. 透過 UMAP 降維視覺化分析並繪製成分佈圖，可以發現主動學習挑選出的標註資料幾乎都落在設計空間中的外圍，因此將模型預測轉化為內插問題，可以有效率進行學習。

3. 對設計空間外的資料點，透過實驗證明此代理模型仍然能夠保有其預測力，並且對於聚合度略大於設計空間的序列也能準確預測。

4. 此部分研究證實主動學習用於隨機共聚高分子序列設計的能力，並僅以少量標註成本就能建立隨機共聚物的代理模型，此研究結果建立的新流程，證實數據驅動方法用於高分子序列設計的潛力。

4.2 小數據驅動多目標最佳化

在小數據驅動多目標最佳化的研究中，目的是模擬現實世界中進行材料設計時資料稀缺的場景，透過主動學習引導單體實驗選擇，進行高分子材料的實驗設計並同時達到單體最佳化，本研究中使用 PolyInfo 資料集，研究結果如下:

4.2.1 代理模型比較

我們在 Pareto 引導的主動學習演算法中，測試了前述三種代理模型:GPR、GBR 和基礎模型 TransPolymer，每種模型的不確定性估計方法已在研究方法部分中詳細說明，初始條件只抽樣 30 筆資料，模擬現實世界資料不足的應用場景。為了減少初始條件對最佳化可能造成的影響，我們使用不同的隨機種子 (random seed)，抽取了五組初始資料集，每組包含 30 筆已標記資料，以此初始條件運行最佳化演算法，並以超體積進行最佳化效率評估。相關結果如圖4.6所示。

從實驗結果可以看出，使用 TransPolymer 作為代理模型的結果，明顯優於另外兩種模型，不僅超體積平均明顯優於其他兩者，顯示使用 TransPolymer 能

doi:10.6342/NTU202501030

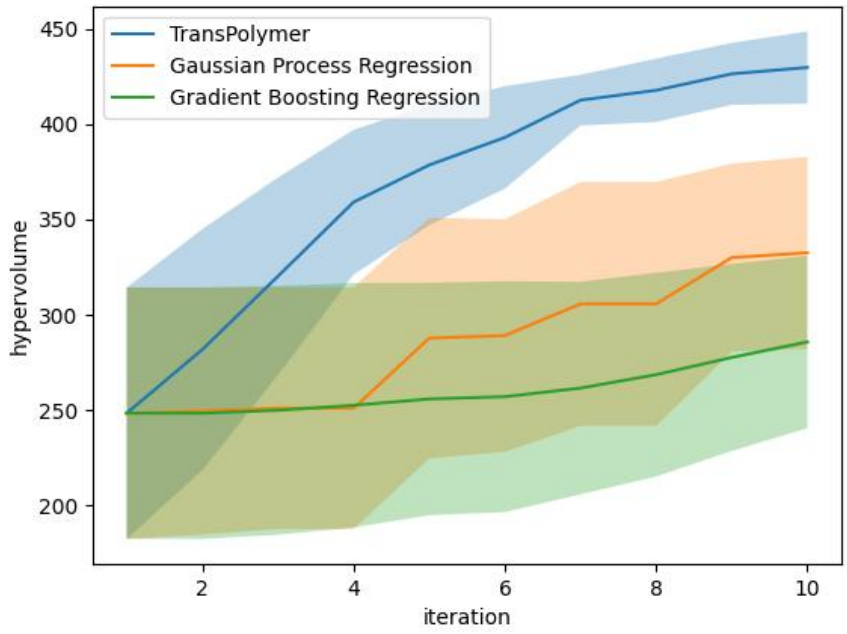


圖 4.6: 使用不同代理模型進行最佳化所得到的超體積比較，圖上包含每一次迭代的超體積平均及標準差。

更快速選取到較優的 Pareto set，並且標準差也較其他兩模型小，顯示出使用 TransPolymer 也能夠增加最佳化流程的穩定性。

由於基礎模型已經在育訓練過程中學習到先備知識，在小樣本資料中只需微調即可不用重新學習，因此相較其他模型在小資料集中更具優勢，因此其較高的預測精度，進一步提升了最佳化效率，顯著超越了其他兩者。雖然 GPR 和 GBR 在資料稀缺的情形下，過去文獻顯示具有相當不錯的表現，但預訓練模型仍取得了更佳的結果。因此，基於此實驗的結果，我們將在後續的實驗中，採用 TransPolymer 作為代理模型。

4.2.2 主動學習策略比較

本研究中，我們採用了三種策略：探索（exploration）、利用（exploitation）以及兩者結合的混合策略，混合策略中，每一次迭代都交替執行探索與利用，三種

doi:10.6342/NTU202501030

策略各自執行十次迭代。同樣為了減少初始條件的影響，我們同樣使用五組不同的隨機種子來挑選初始的 30 筆資料，再進行主動式學習的最佳化流程，以測試不同策略造成的影響。實驗結果如圖4.7所示。

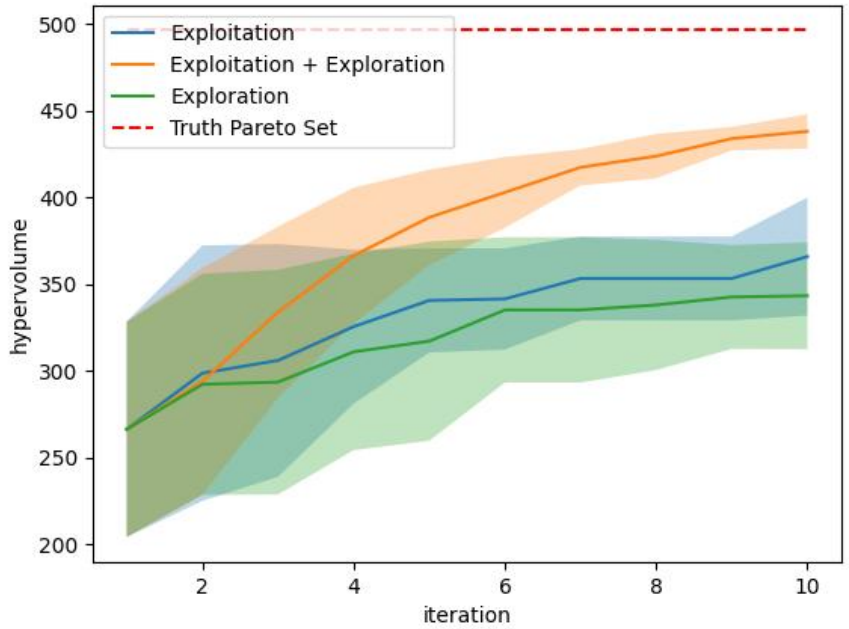


圖 4.7: 不同主動式學習採樣策略的比較。

實驗結果顯示，探索與利用策略之間取得平衡，能夠顯著優於僅使用單一策略的結果，由於初始資料集規模有限，代理模型所掌握的知識不足，導致在這種情況下僅採用利用策略的效果不佳，而採用探索雖然能快速讓模型提升預測力，但也需要適時推進 Pareto front，因此，使用利用並交替加入探索策略能顯著提升最佳化結果，相較於單一策略更加有效。此外，混合策略的變異程度也低於其他兩種策略，顯示其在不同初始條件下仍具有優異的穩定性，此結果證明，在 Pareto 引導的主動式學習過程中，適當地平衡探索與利用策略，不僅能提升演算法效能，也能增強其穩定性，因此後續實驗的主動學習策略，會採用交替探索以及利用的方式進行。

doi:10.6342/NTU202501030

4.2.3 主動學習收斂性比較

在主動學習最佳化的實際應用中，由於我們無法得知真實的最佳解，因此主動學習面臨的一項重要挑戰，是如何設定合適的停止條件，已有效平衡計算成本與最佳化效能。在本研究中，我們將「收斂」定義為停止準則，當超體積連續在三次迭代未發生變化時，就視為最佳化已經達到收斂，停止主動學習迭代，確立此停止條件後，我們測試了多組初始條件，以評估不同採樣策略的收斂行為，實驗結果如圖4.8所示。

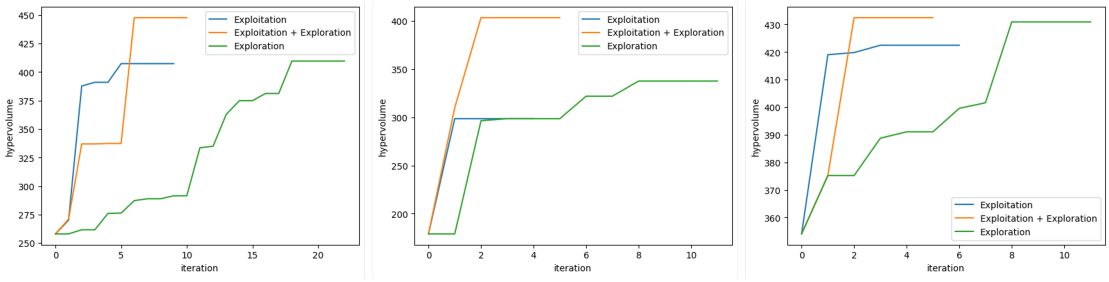


圖 4.8: 三種策略在不同初始條件下的收斂實驗結果。

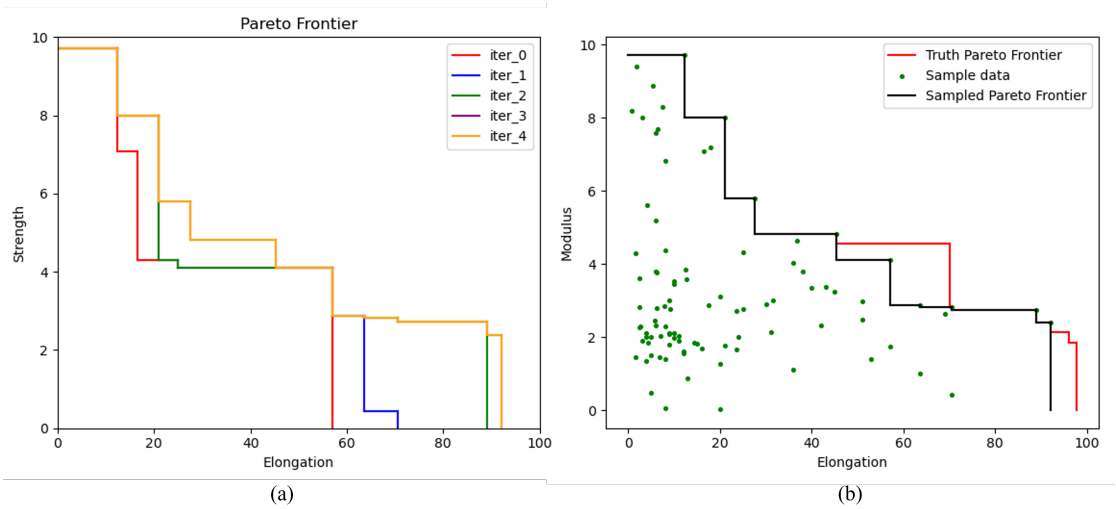
實驗結果顯示，在不同策略下皆呈現出一致的趨勢，全探索策略讓模型根據預測的不確定性探索設計空間，雖然最終能達到相對較好的超體積值，但所需標記資料最多，收斂速度也是三種策略中最慢的。相較之下，全利用策略的收斂速度明顯較快，但其達到的超體積值較低，這是因為小樣本資料所帶來的初始知識可能存在偏差，而僅依賴利用策略會限制模型透過採樣獲取新知，導致最佳化效能不佳。

混合使用探索與利用策略的交替方法則優於單一策略，透過探索步驟，模型能及時更新，有助於之後利用步驟中的更準確採樣，透過交替進行有效結合了兩種策略的優點，在準確度與效率方面都優於其他主動式學習方法。

doi:10.6342/NTU202501030

4.2.4 標註資料數量

透過以上三個實驗結果，可以確認 Pareto 引導主動學習用於多目標最佳化的完整流程，包含使用 TransPolymer 作為代理模型以及使用交替的選取標註資料策略。接下來要進行標註數量的評估，以利評估此流程在真實世界應用的可能性，並記錄每一次迭代過程標註資料與 Pareto front 的關係，實驗結果如下圖4.9所示。



(a) (b)

圖 4.9: (a) Pareto front 隨迭代次數的推進情形。(b) 最終採樣的 Pareto front 與整體資料集真實 Pareto front 的比較。

從圖4.9(a) 的結果中可以觀察到，隨著標記資料數量的增加，所採樣的 Pareto front 逐漸推進，每一次迭代包含一次利用步驟與一次探索步驟。探索過程會增加模型的預測力，而利用步驟則會根據代理模型的預測，逐步推進 Pareto front。在圖4.9(b) 中，紅色線條代表真實的 Pareto front，由 PolyInfo 資料集中全部 835 筆資料計算而得。我們從 30 筆初始資料出發，成功地透過主動學習多目標最佳化，將採樣得到的 Pareto front 推進至接近真實前緣的位置。

實驗中所標記的資料數量如表4.2所列，從表上可以看出，在 Pareto 引導主動學習過程中，僅需 85 筆已標記資料，即可將 Pareto front 推進至接近真實前緣的位置。在複雜的高分子單體多目標最佳化問題上，此研究結果大幅降低了整體設

doi:10.6342/NTU202501030

計過程中資料標記的成本，並有望用於真實的高分子單體設計上。

表 4.2: 圖 4.9(a) 實驗中所使用的標記資料點數量。

	迭代次數 0	迭代次數 1	迭代次數 2	迭代次數 3	迭代次數 4
標註資料數量	30	46	57	73	85

4.2.5 小結

透過此部分的實驗，我們進行了小數據驅動高分子多目標最佳化研究，相關結果總結如下:

1. 本研究透過 PolyInfo 的資料進行，資料集中包含 805 筆資料及有強烈互斥關係的兩標註性質，並在研究中採用 30 筆資料作為初始資料集。
2. 透過實驗結果可以發現，代理模型選擇基礎模型 TransPolymer，主動學習策略採用探索以及利用交替，此組合可以在最佳化的超體積以及收斂性上有最佳表現。
3. 在本實驗中，僅使用 85 筆資料標註即可以完成高分子單體的多目標最佳化，相較於總設計空間減少相當多的標註資料量，並且此資料規模有機會以純實驗獲得。
4. 本研究示範了主動學習用於高分子單體最佳化的流程，透過機器學習模型挑選下一步實驗標的，取代實驗設計，能夠減少標註資料及最佳化材料所需的成本，並且用於複雜且離散的設計空間，比如高分子單體選擇，機器學習模型能提供更全盤考量的下一步實驗建議。

doi:10.6342/NTU202501030

4.3 實驗數據驅動聚酯型 Vitrimer 最佳化

此部分研究使用純實驗數據驅動 Vitrimer 最佳化，將前兩部分的研究結果實際應用到材料設計中，透過模型回饋進行實驗設計，並達到最佳化，以下為研究結果:

4.3.1 初始資料集採樣及模型訓練結果

本研究最開始進行初始資料集的合成以及性質量測，合成及量測方法如前面章節所述，由於實驗包含系統誤差，對於每一個合成的實驗樣品會進行重複測試，並根據 Jha 等人的研究 [65]，採用量測之中位數作為性質標註，能夠避免極值及系統隨機誤差的影響，因此也將此實驗設定用於本研究中。經過實驗量測之後，所得結果列於表4.3，本研究案例使用此三筆資料訓練初始模型，進行主動學習流程。

表 4.3: 初始資料集測量結果。

樣品	二酸 (1)	二酸 (2)	比例	交聯劑	UTS
樣品 1	Succinic Acid	Adipic Acid	0.3	CA	43.545
樣品 2	Succinic Acid	Sebacic Acid	0.5	TCAA	60.963
樣品 3	Adipic Acid	Sebacic Acid	0.8	TCAA	38.813

使用表4.3的資料訓練 GPR 模型後，對於設計空間剩餘之資料點進行推論，所得之結果如下圖4.10，透過此推論結果，我們可以進行下一步的標註資料挑選，在此次迭代過程選擇不確定性最高的三個點進行下一次實驗。

4.3.2 第一次迭代結果

接著繼續進行下一輪實驗，此次迭代是採用探索 (exploration) 策略，由於初始資料集規模過小，因此使用此策略幫助模型快速學習。透過初始資料集訓練的

doi:10.6342/NTU202501030

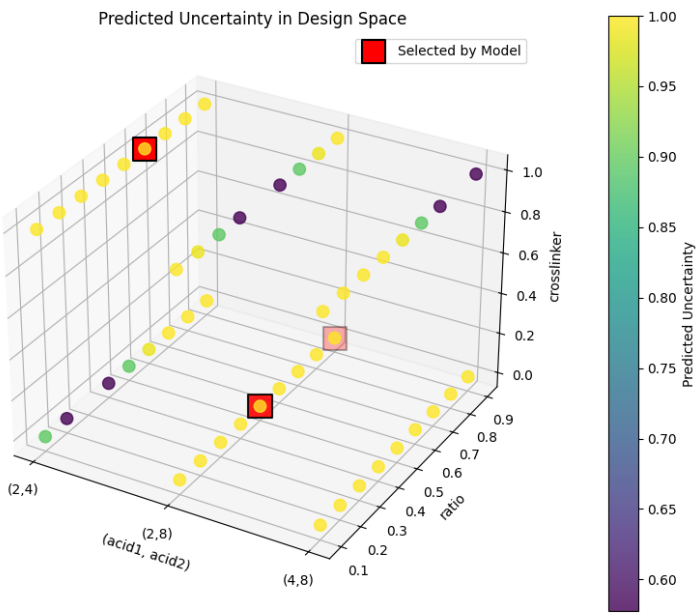


圖 4.10: 初始資料集訓練模型之推論結果。

代理模型推論後，所選出的下一輪迭代候選參數組合，同樣進行合成及性質量測，結果列於下表4.4。

表 4.4: 第一次迭代性質量測結果。

樣品	二酸 (1)	二酸 (2)	比例	交聯劑	UTS
樣品 4	Succinic Acid	Sebacic Acid	0.5	CA	55.619
樣品 5	Succinic Acid	Sebacic Acid	0.9	CA	19.507
樣品 6	Succinic Acid	Adipic Acid	0.6	TCAA	24.963

將此次迭代的實驗結果加入資料集重新訓練 GPR 代理模型，並對於剩下的未標註設計空間進行推論，所得到的結果如圖4.11所示，可以發現模型經過此輪訓練後，預測的結果差異不大，使用此階段的模型進行推論仍不具有足夠可靠度，因此本研究中仍繼續使用探索策略進行下一輪的主動學習迭代。

doi:10.6342/NTU202501030

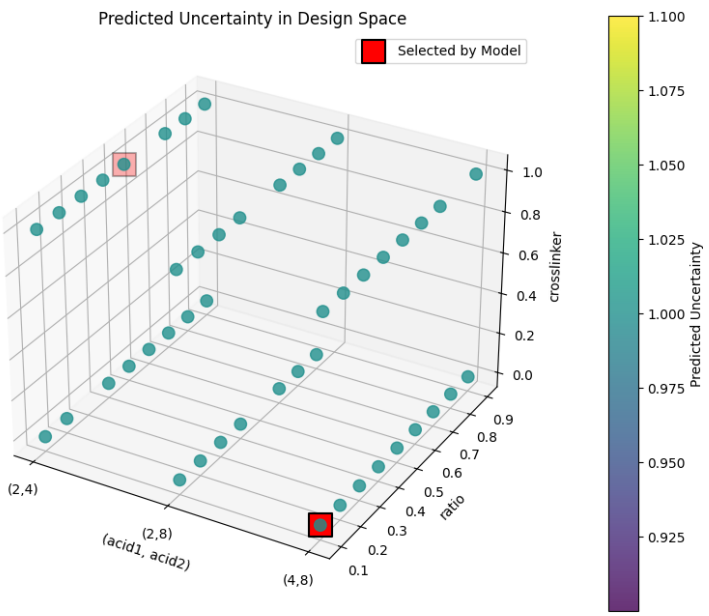


圖 4.11: 第一次迭代後模型之推論結果。

4.3.3 第二次迭代結果

將第一次迭代標註的資料加入訓練資料集之後，再繼續主動學習流程，從模型推論結果來看，經過一輪主動學習迭代後的代理模型表現依然不佳，因此在本次迭代中，仍會以探索的策略為主，在本輪中選取不確定性最高的兩個資料點，進行合成後性質量測結果如下表4.5。

表 4.5: 第二次迭代性質量測結果。

樣品	二酸 (1)	二酸 (2)	比例	交聯劑	UTS
樣品 7	Adipic Acid	Sebacic Acid	0.1	CA	17.863
樣品 8	Succinic Acid	Adipic Acid	0.5	TCAA	28.036

將此次主動學習迭代後所得之 8 筆標註資料重新訓練 GPR 模型後，所得之預測結果如下圖4.12。透過兩次的迭代流程，可以確保代理模型模型之預測力，接下來將透過模型預測進行最佳化採樣。

doi:10.6342/NTU202501030

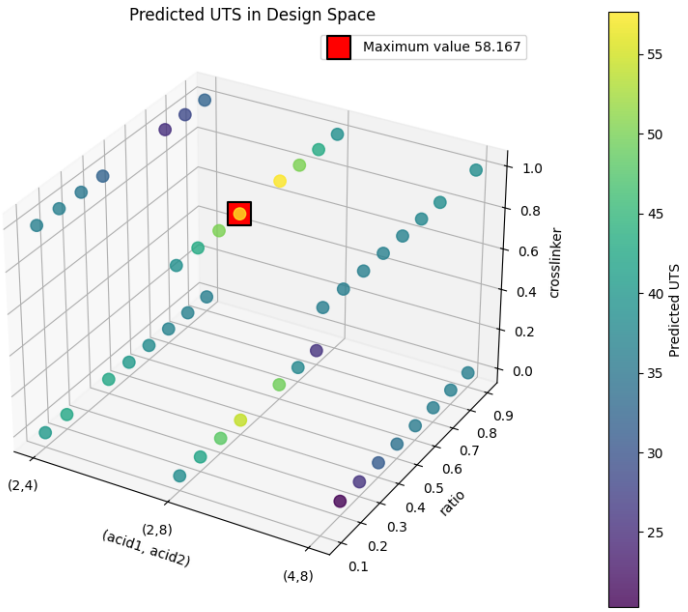


圖 4.12: 第二次迭代後模型之推論結果。

4.3.4 第三次迭代結果

在本次迭代中，採取利用 (Exploitation) 策略，直接採樣模型在設計空間預測的極值。根據模型預測，設計空間中的 UTS 最大參數組合及其合成量測結果列於下表4.6。

表 4.6: 第三次迭代性質量測結果。

樣品	二酸 (1)	二酸 (2)	比例	交聯劑	UTS
樣品 9	Succinic Acid	Sebacic Acid	0.4	TCAA	35.853

經過實驗量測性質後，此參數組合之 UTS 並無高於原先標註資料集之最大值，故視為此主動學習流程已達到收斂，設計空間中的最佳解為 Succinic Acid(SA) 與 Sebacic Acid(SeA) 各以 0.5 的比例混合，交聯劑使用 TCAA，本最佳化配方不僅來自機器學習模型預測結果，也可從二酸與交聯劑的分子結構進一步驗證其合理性。SA 為設計空間中碳鏈最短的二酸，具有較高的極性與剛性，其鏈段能形成緊密、強力的交聯結構，賦予材料優異的強度，但柔韌性不足，延展性

doi:10.6342/NTU202501030

相對較低。相對地，SeA 具有十個碳的長鏈，極性低、柔軟性高，有助於提升材料的延展性與韌性。將 SA 與 SeA 等比例混合，有助於兩者在分子層次上互補，短鏈與長鏈的隨機共聚破壞了聚酯結晶區域的形成，提高網絡非結晶性，從而兼顧強度與延展性。

在交聯劑方面，TCAA 擁有四個羧酸官能基，可與多個環氧基反應，建立三維交聯結構，其交聯點密度遠高於一般二官能酸。研究顯示，僅以 SeA 與 DGEBA 製成的 vitrimer 結構鬆散，雖延展性佳但強度不足；加入適量 TCAA 後，材料剛性與強度明顯提升，但過度交聯導致鏈段活動受限，使伸長率大幅下降[66]。

將 SA/SeA 等比混合並搭配 TCAA 所建構的交聯網絡，整合了剛性節點與柔性鏈段的優勢。TCAA 交聯樞紐結合短鏈 SA 與長鏈 SeA，形成同時具備承載能力與延展能力的結構單元。在外力作用下，短鏈段先行承受應力，長鏈段隨後變形吸能，有效分散應力並延遲裂紋擴展。這種協同效應提升了材料整體的抗拉強度與韌性，使 UTS 達到最佳，同時保有適度的斷裂伸長率。

4.3.5 小結

本部分研究為前兩部分主動學習研究用於真實實驗設計的案例，相關結果總結如下:

1. 本部分研究透過實驗數據驅動主動學習流程，進行 Vitrimer 的配方及單體設計，最後透過主動學習流程，成功在 54 筆資料的設計空間中進行 9 個實驗達到收斂。
2. 以實驗數據驅動主動學習時，會遇到實驗之系統誤差，本研究中依據過往研究，使用實驗資料點之中位數作為標註。

doi:10.6342/NTU202501030

3. 此研究示範了前兩部分的主動學習流程，如何應用於真實世界的材料開發流程，透過模型引導實驗設計，可以使用少量資料達到最佳化。然而受限於時間因素，此研究的設計空間仍有相當大的侷限性，初始資料規模也略有不足，造成模型表現欠佳，未來仍需更進一步優化流程，包含代理模型的正確性檢驗。

doi:10.6342/NTU202501030

doi:10.6342/NTU202501030

第五章　結論與未來展望

5.1　結論

本研究著重於探討主動學習用於高分子材料設計的可能性，研究共分為三部分，分別為主動學習共聚物序列設計、小數據驅動高分子單體多目標最佳化以及實驗數據驅動 Vitrimer 配方設計。前兩部份透過在已知資料集進行實驗，證實主動學習應用於分子設計的潛力，並解決過往數據驅動高分子材料所面臨標註資料不足的問題，最後以 Vitrimer 的單體配方最佳化進行案例探討，實現純實驗小數據資料集驅動材料設計，以下是本研究主要結論:

5.1.1　主動學習共聚物序列設計

1. 隨機共聚高分子在序列設計時，面臨組合爆炸問題，造成過往即使透過數據驅動，也難以達成序列設計的目標。本研究中提出均勻採樣以及主動學習的設計流程，並以模擬資料進行實驗，證實此架構的可行性。

2. 實驗結果證明，無論是在模型預測的正確性或是收斂性上，採用主動學習都可以有良好的表現。並且透過視覺化工具，也能證明主動學習挑選標註資料的合理性。最終此架構成功減少 98% 的標註資料量，證實此方法能確

doi:10.6342/NTU202501030

實解決過往序列設計遇到的挑戰。

5.1.2 小數據驅動高分子單體多目標最佳化

1. 高分子單體設計以及多目標最佳化一直是高分子設計上的重要議題，由於高分子單體的離散特性，傳統 DoE 方法或統計分析難以進行有效的實驗標的，且多目標最佳化讓單體選擇更加複雜，因此本研究中以數據驅動方法解決此問題，使用機器學習模型提供下一步實驗標的，以快速在設計空間中達到最佳化。

2. 透過 PolyInfo 資料集進行實驗，本研究成功找出主動學習流程中最佳的代理模型以及採樣策略，此最佳組合無論是在最佳化結果的超體積評估或是超體積的收斂及穩定性上，都顯著優於其他組合。

3. 觀察採樣資料和 Pareto front 的關係，可以發現在利用 (exploitation) 的策略步驟時，會逐步推進採樣的 Pareto front，且最終使用的標註資料量為 85 筆資料，相較於總設計空間的 805 筆資料，大幅減少了採樣成本。

5.1.3 實驗數據驅動聚酯型 Vitrimer 最佳化

1. Vitrimer 是新型的環保高分子，具有極高的研究以及商用價值，然而需要透過設計找出配方以及比例，才能滿足實際應用的性能需求。

doi:10.6342/NTU202501030

2. 透過主動學習的導入，在 54 種組合的設計空間中，實際僅進行 9 次合成，就找出設計空間中具有最優異機械性質的參數組合。

以上三個研究案例皆展示了主動學習用於高分子設計的潛力，本研究先以資料集證實此架構確實可以大幅減少設計成本，改善以往數據驅動高分子設計的問題。透過此主動學習架構，能夠用於各種高分子材料的設計問題，比如單體設計、序列設計、配方設計等。接著再以簡單案例示範如何將先前建立的架構用於 Vitrimer 實驗設計最佳化，這些案例示範了，如何透過機器學習靈活彈性以及主動學習減少標註資料的能力，有望將此架構實際應用到工業上，讓機器學習提供實驗設計成為高分子材料的標準設計流程。

5.2 未來展望

本研究前兩部分建立了標準的主動學習高分子設計流程，並分別示範了如何應用主動學習技術，於共聚高分子 (copolymers) 序列設計以及高分子單體多目標最佳化。本研究的貢獻在於建立主動學習標準設計流程，並透過實驗確立了在各種場景適用的代理模型以及主動學習策略，也證實可以透過主動學習架構減少數據驅動材料設計所需資料量。最後將此架構用於實驗數據驅動 Vitrimer 設計，成功示範如何導入主動學習於材料的設計開發流程。

然而，目前主動學習方法仍有缺陷，比如實驗的誤差等，可能會導致模型在做決策時，產生不同結果，而不同主動學習策略也還需進一步研究，此外，受限於時間因素，目前純實驗資料驅動的設計空間仍相當侷限，需要更進一步於複雜的設計空間中進行流程。儘管如此，仍能看出主動學習高分子設計有極大的潛力，未來希望有相關研究能夠應用相同的方法，以純實驗驅動實驗設計，透過機器學

doi:10.6342/NTU202501030

習提供的實驗建議逐步達到最佳化，並以此基礎設計出全新的材料配方，建立材料參數以及材料性質之間的關聯。而在工業上，在各種不同需求的應用場景所需的不同材料，也能透過此方法減少所需的開發時間，進而滿足各種工業應用。

doi:10.6342/NTU202501030

參考文獻

[1] Ruimin Ma, Hanfeng Zhang, and Tengfei Luo. Exploring high thermal conductivity amorphous polymers using reinforcement learning. ACS Applied Materials & Interfaces, 14(13):15587–15598, 2022.

[2] Danh Nguyen, Lei Tao, and Ying Li. Integration of machine learning and coarse-grained molecular simulations for polymer materials: physical understandings and molecular design. Frontiers in Chemistry, 9:820417, 2022.

[3] Sirawit Pruksawan, Guillaume Lambard, Sadaki Samitsu, Keitaro Sodeyama, and Masanobu Naito. Prediction and optimization of epoxy adhesive strength from a small dataset through active learning. Science and Technology of Advanced Materials, 20(1):1010–1021, 2019.

[4] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. APL materials, 1(1), 2013.

[5] Wuxin Sha, Yan Li, Shun Tang, Jie Tian, Yuming Zhao, Yaqing Guo, Weixin Zhang, Xinfang Zhang, Songfeng Lu, Yuan-Cheng Cao, et al. Machine learning in polymer informatics. InfoMat, 3(4):353–361, 2021.

doi:10.6342/NTU202501030

[6] Shingo Otsuka, Isao Kuwajima, Junko Hosoya, Yibin Xu, and Masayoshi Yamazaki. Polyinfo: Polymer database for polymeric materials design. In 2011 International Conference on Emerging Intelligent Data and Web Technologies, pages 22–29. IEEE, 2011.

[7] Chiho Kim, Anand Chandrasekaran, Tran Doan Huan, Deya Das, and Rampi Ramprasad. Polymer genome: a data-powered polymer informatics platform for property predictions. The Journal of Physical Chemistry C, 122(31):17575–17585, 2018.

[8] Joseph L Durant, Burton A Leland, Douglas R Henry, and James G Nourse. Reoptimization of mdl keys for use in drug discovery. Journal of Chemical Information and Computer Sciences, 42(6):1273–1280, 2002.

[9] Harry L Morgan. The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. Journal of Chemical Documentation, 5(2):107–113, 1965.

[10] A Patrícia Bento, Anne Hersey, Eloy Félix, Greg Landrum, Anna Gaulton, Francis Atkinson, Louisa J Bellis, Marleen De Veij, and Andrew R Leach. An open source chemical structure curation pipeline using rdkit. Journal of Cheminformatics, 12:1–16, 2020.

[11] David Weininger. SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences, 28(1):31–36, 1988.

[12] Hossein Ziaee, Seyyed Mohsen Hosseini, Abdolmajid Sharafpoor, Mohammad Fazavi, Mohammad Mahdi Ghiasi, and Alireza Bahadori. Prediction of solubility

doi:10.6342/NTU202501030

of carbon dioxide in different polymers using support vector machine algorithm. Journal of the Taiwan Institute of Chemical Engineers, 46:205–213, 2015.

[13] Z Zhang, N-M Barkoula, J Karger-Kocsis, and K Friedrich. Artificial neural network predictions on erosive wear of polymers. Wear, 255(1-6):708–713, 2003.

[14] Tarak K Patra. Data-driven methods for accelerating polymer design. ACS Polymers Au, 2(1):8–26, 2021.

[15] Venkatesh Meenakshisundaram, Jui-Hsiang Hung, Tarak K Patra, and David S Simmons. Designing sequence-specific copolymer compatibilizers using a molecular-dynamics-simulation-based genetic algorithm. Macromolecules, 50(3):1155–1166, 2017.

[16] Lei Tao, Jinlong He, Nuwayo Eric Munyaneza, Vikas Varshney, Wei Chen, Guoliang Liu, and Ying Li. Discovery of multi-functional polyimides through high-throughput screening using explainable machine learning. Chemical Engineering Journal, 465:142949, 2023.

[17] Michael P Allen. Introduction to molecular dynamics simulation. Computational Soft Matter: From Synthetic Polymers to Proteins, 23(1):1–28, 2004.

[18] Aidan P Thompson, H Metin Aktulga, Richard Berger, Dan S Bolintineanu, W Michael Brown, Paul S Crozier, Pieter J In't Veld, Axel Kohlmeyer, Stan G Moore, Trung Dac Nguyen, et al. LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. Computer Physics Communications, 271:108171, 2022.

[19] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E Mark, and

doi:10.6342/NTU202501030

Herman JC Berendsen. Gromacs: fast, flexible, and free. Journal of Computational Chemistry, 26(16):1701–1718, 2005.

[20] Rui Shi, Hu-Jun Qian, and Zhong-Yuan Lu. Coarse-grained molecular dynamics simulation of polymers: Structures and dynamics. Wiley Interdisciplinary Reviews: Computational Molecular Science, 13(6):e1683, 2023.

[21] Shaorui Yang and Jianmin Qu. Coarse-grained molecular dynamics simulations of the tensile behavior of a thermosetting polymer. Physical Review E, 90(1):012601, 2014.

[22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers), pages 4171–4186, 2019.

[23] Christopher Kuenneth and Rampi Ramprasad. polyBERT: a chemical language model to enable fully machine-driven ultrafast polymer informatics. Nature Communications, 14(1):4099, 2023.

[24] Changwen Xu, Yuyang Wang, and Amir Barati Farimani. Transpolymer: a transformer-based language model for polymer property predictions. npj Computational Materials, 9(1):64, 2023.

[25] Fanmeng Wang, Wentao Guo, Minjie Cheng, Shen Yuan, Hongteng Xu, and Zhifeng Gao. MMpolymer: A multimodal multitask pretraining framework for polymer property prediction. pages 2336–2346, 2024.

doi:10.6342/NTU202501030

[26] Qinghe Gao, Tammo Dukker, Artur M Schweidtmann, and Jana M Weber. Self-supervised graph neural networks for polymer property prediction. Molecular Systems Design & Engineering, 9(11):1130–1143, 2024.

[27] Issam El Naqa and Martin J Murphy. What is machine learning? In Machine Learning in Radiation Oncology: Theory and Applications, pages 3–11. Springer, 2015.

[28] Eduardo Mosqueira-Rey, Elena Hernández-Pereira, David Alonso-Ríos, José Bobes-Bascarán, and Ángel Fernández-Leal. Human-in-the-loop machine learning: a state of the art. Artificial Intelligence Review, 56(4):3005–3054, 2023.

[29] Burr Settles. From theories to queries: Active learning in practice. In Active learning and experimental design workshop in conjunction with AISTATS 2010, pages 1–18. JMLR Workshop and Conference Proceedings, 2011.

[30] William H Beluch, Tim Genewein, Andreas Nürnberger, and Jan M Köhler. The power of ensembles for active learning in image classification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 9368–9377, 2018.

[31] Ajay J Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 2372–2379. IEEE, 2009.

[32] Mustafa Bilgic and Lise Getoor. Link-based active learning. In NIPS Workshop on Analyzing Networks and Learning With Graphs, volume 4, page 9, 2009.

[33] Alexander Freytag, Erik Rodner, and Joachim Denzler. Selecting influential examples: Active learning with expected model output changes. In Computer

doi:10.6342/NTU202501030

Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13, pages 562–577. Springer, 2014.

[34] Yazhou Yang and Marco Loog. Active learning using uncertainty information. In 2016 23rd International Conference on Pattern Recognition (ICPR), pages 2646–2651. IEEE, 2016.

[35] Chiho Kim, Anand Chandrasekaran, Anurag Jha, and Rampi Ramprasad. Active-learning and materials design: the example of high glass transition temperature polymers. MRS Communications, 9(3):860–866, 2019.

[36] Praneeth S Ramesh and Tarak K Patra. Polymer sequence design via molecular simulation-based active learning. Soft Matter, 19(2):282–294, 2023.

[37] Wenlin Zhao, Xuemeng Fu, Xinyao Xu, Liangshun Zhang, Liquan Wang, Jiaping Lin, Yaxi Hu, Liang Gao, Lei Du, and Xiaohui Tian. Design of multicomponent thermosetting polymers with enhanced tensile properties through active learning. Composites Science and Technology, 256:110779, 2024.

[38] Haifan Zhou, Yue Fang, and Hanyu Gao. Using active learning for the computational design of polymer molecular weight distributions. ACS Engineering Au, 4(2):231–240, 2023.

[39] Kevin Maik Jablonka, Giriprasad Melpatti Jothiappan, Shefang Wang, Berend Smit, and Brian Yoo. Bias free multiobjective active learning for materials design and discovery. Nature Communications, 12(1):2312, 2021.

[40] Erin Antono, Nobuyuki N Matsuzawa, Julia Ling, James Edward Saal, Hideyuki Arai, Masaru Sasago, and Eiji Fujii. Machine-learning guided quantum chemical and

doi:10.6342/NTU202501030

molecular dynamics calculations to design novel hole-conducting organic materials. The Journal of Physical Chemistry A, 124(40):8330–8340, 2020.

[41] Paripat Kraisornkachit, Masanobu Naito, Chao Kang, and Chiaki Sato. Multi-objective optimization of adhesive joint strength and elastic modulus of adhesive epoxy with active learning. Materials, 17(12):2866, 2024.

[42] Seyedali Mirjalili and Seyedali Mirjalili. Genetic algorithm. Evolutionary Algorithms and Neural Networks: Theory and Applications, pages 43–55, 2019.

[43] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. Proceedings of the IEEE, 104(1):148–175, 2015.

[44] Arun Mannodi-Kanakkithodi, Ghanshyam Pilania, Rampi Ramprasad, Turab Lookman, and James E Gubernatis. Multi-objective optimization techniques to design the pareto front of organic dielectric polymers. Computational Materials Science, 125:92–99, 2016.

[45] Nathan J Van Zee and Renaud Nicolaÿ. Vitrimers: Permanently crosslinked polymers with dynamic network topology. Progress in Polymer Science, 104:101233, 2020.

[46] Melania Bednarek and Przemysław Kubisa. Reversible networks of degradable polyesters containing weak covalent bonds. Polymer Chemistry, 10(15):1848–1872, 2019.

[47] Yong Zheng, Tingting Liu, Haodong He, Zilu Lv, Jiayun Xu, Dayong Ding, Lin Dai, Zhanhua Huang, and Chuanling Si. Lignin-based epoxy composite vitrimers with

doi:10.6342/NTU202501030

light-controlled remoldability. Advanced Composites and Hybrid Materials, 6(1):53, 2023.

[48] Suman Debnath, Swaraj Kaushal, and Umaprasana Ojha. Catalyst-free partially bio-based polyester vitrimers. ACS Applied Polymer Materials, 2(2):1006–1013, 2020.

[49] Marc Guerre, Christian Taplan, Johan M Winne, and Filip E Du Prez. Vitrimers: directing chemical reactivity to control material properties. Chemical science, 11(19):4855–4870, 2020.

[50] Vincent Schenk, Karine Labastie, Mathias Destarac, Philippe Olivier, and Marc Guerre. Vitrimer composites: current status and future challenges. Materials Advances, 3(22):8012–8029, 2022.

[51] Shao-Yu Chien, Jane Wang, and Ying-Ling Liu. Biodegradable polyester-based vitrimers exhibiting transesterification-induced topography isomerization under recycling. ACS Applied Polymer Materials, 6(15):9191–9199, 2024.

[52] Takeshi Aoyagi. Optimization of the elastic properties of block copolymers using coarse-grained simulation and an artificial neural network. Computational Materials Science, 207:111286, 2022.

[53] Shao-Min Mai, Withawat Mingvanish, Simon C Turner, Chiraphon Chaibundit, J Patrick A Fairclough, Frank Heatley, Mark W Matsen, Anthony J Ryan, and Colin Booth. Microphase-separation behavior of triblock copolymer melts. comparison with diblock copolymer melts. Macromolecules, 33(14):5124–5130, 2000.

[54] Takashi Honda. SUSHI: Density functional theory simulator. Computer Simulation of Polymeric Materials: Applications of the OCTA System, pages 67–100, 2016.

doi:10.6342/NTU202501030

[55] Takeshi Aoyagi. COGNAC: Coarse-grained molecular dynamics simulator. Computer Simulation of Polymeric Materials: Applications of the OCTA System, pages 29–65, 2016.

[56] Kurt Kremer and Gary S Grest. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. The Journal of Chemical Physics, 92(8):5057–5086, 1990.

[57] Takeshi Aoyagi, Takashi Honda, and Masao Doi. Microstructural study of mechanical properties of the aba triblock copolymer using self-consistent field and molecular dynamics. The Journal of Chemical Physics, 117(17):8153–8161, 2002.

[58] Masao Doi. Octa (open computational tool for advanced material technology). In Macromolecular Symposia, volume 195, pages 101–108. Wiley Online Library, 2003.

[59] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. the Journal of Machine Learning Research, 12:2825–2830, 2011.

[60] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426, 2018.

[61] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. Journal of Machine Learning Research, 9(11), 2008.

[62] Ruimin Ma and Tengfei Luo. PI1M: a benchmark database for polymer informatics. Journal of Chemical Information and Modeling, 60(10):4684–4690, 2020.

doi:10.6342/NTU202501030

[63] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In International Conference on Machine Learning, pages 1050–1059. PMLR, 2016.

[64] Jhonatan Contreras and Thomas Bocklitz. Enhancing decision confidence in ai using monte carlo dropout for raman spectra classification. Analytica Chimica Acta, 1332:343346, 2024.

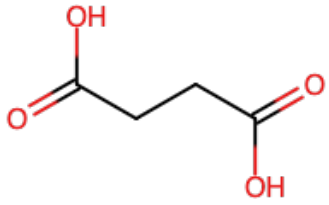
[65] Anurag Jha, Anand Chandrasekaran, Chiho Kim, and Rampi Ramprasad. Impact of dataset uncertainties on machine learning model predictions: the example of polymer glass transition temperatures. Modelling and Simulation in Materials Science and Engineering, 27(2):024002, 2019.

[66] Hsu-I Mao, Jun-Yuan Hu, Jia-Wei Shiu, Syang-Peng Rwei, and Chin-Wen Chen. Sustainability and repeatedly recycled epoxy-based vitrimer electromagnetic shielding composite material. Polymer Testing, 127:108200, 2023.

doi:10.6342/NTU202501030

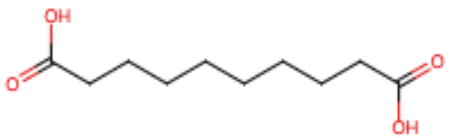
附錄 A — 實驗藥品

- **Succinic Acid（SA）**



分子式	$C_4H_6O_4$
分子量	118.09
CAS No	110-15-6
供應商	Anhui Sunsing Chemicals Co., Ltd.

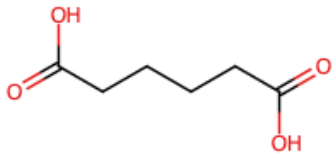
- **Sebacic Acid（SeA）**



doi:10.6342/NTU202501030

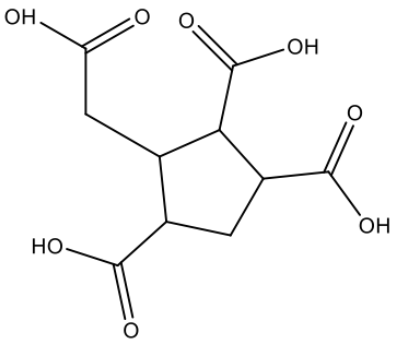
分子式	$C_{10}H_{18}O_4$
分子量	202.244 g/mol
CAS No	111-20-6
供應商	Anhui Sunsing Chemicals Co., Ltd.

- **Adipic Acid（AA）**



分子式	$C_6H_{10}O_4$
分子量	146.14
CAS No	124-04-9
供應商	Anhui Sunsing Chemicals Co., Ltd.

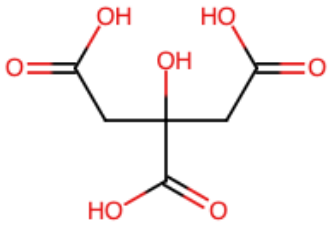
- **3-(carboxymethyl)cyclopentane-1,2,4-tricarboxylic acid（TCAA）**



分子式	$C_{10}H_{20}O_8$
分子量	268.262 g/mol
CAS No	24434-90-0
供應商	LCY Chemical Corporation

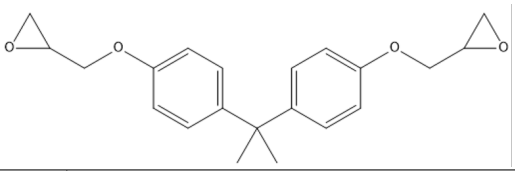
doi:10.6342/NTU202501030

- **Citric Acid（CA）**



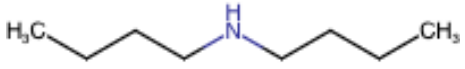
分子式	$C_6H_8O_7$
分子量	192.12 g/mol
CAS No	77-92-9
供應商	景明化工股份有限公司

- **Bisphenol A diglycidyl ether（DGEBA）**



分子式	$C_{21}H_{24}O_4$
分子量	340.42 g/mol
CAS No	1675-54-3
供應商	Sigma-Aldrich

- **Di-n-butylamine（二正丁胺）**



doi:10.6342/NTU202501030

分子式	$C_8H_{19}N$
分子量	129.24 g/mol
CAS No	111-92-2
供應商	UNI-ONWARD Corporation

doi:10.6342/NTU202501030