机器学习在聚烯烃催化剂领域的应用与展望

杨文泓^{1,2*}, 孙文华^{2,3*}

1. 中国石油天然气股份有限公司石油化工研究院, 新材料生物化工研究所, 北京 102206;

2. 中国科学院化学研究所, 北京分子科学国家实验室, 工程塑料重点实验室, 北京 100190;

3. 中国科学院大学化学科学学院, 北京 100049

* 联系人, E-mail: whyang@iccas.ac.cn; whsun@iccas.ac.cn

2021-12-27 收稿, 2022-02-06 修回, 2022-02-14 接受, 2022-02-18 网络版发表

中国科学院化学研究所创新培育项目(CXPY19)和国家自然科学基金(21871275)资助

摘要 催化剂是决定聚烯烃的工业效率以及实现聚烯烃高端化的核心。传统开发催化剂的过程采用试错法, 不仅实验步骤多、研发周期长, 且催化性能的研究需要消耗大量资源。单纯依靠实验的分析方法很难挖掘出催化剂结构与聚合性能之间的内在关系。高水平的量子化学计算可以准确地获取反应机理, 但针对宏量的实验数据, 昂贵的计算成本是其局限。大数据时代, 人工智能的发展势不可挡。机器学习作为人工智能的核心策略表现出强大的预测能力, 并在科学、技术以及工业等各个领域获得了广泛的应用与发展。本文主要介绍机器学习在聚烯烃催化剂中的最新研究进展, 并简要评述机器学习应用于烯烃催化中面临的机遇与挑战。

关键词 聚烯烃催化剂, 催化性能, 分子描述符, 机器学习, 定量构效

聚烯烃是合成树脂中产量最大, 用途最广的高分子材料。目前, 国内聚乙烯的研究基础仍然较弱, 中低端产品过剩, 而高端产品主要依赖进口。聚烯烃的高端化、差异化已成为我国聚烯烃产业的主要升级方向^[1]。催化剂决定了催化烯烃聚合的工业效率以及所得聚烯烃树脂的微观结构和宏观性能, 是实现聚烯烃高端化的核心关键^[2]。因此, 如何设计合成高性能催化剂是聚烯烃领域的重点研究目标。传统开发催化剂的过程常常采用试错法, 所需的实验量较大, 研发周期较长, 并且催化性能的实验研究成本很高, 需要消耗大量的资源。此外, 针对当前种类繁多、结构迥异的聚烯烃催化剂, 单纯依靠人力的实验分析很难挖掘出催化剂分子结构与催化性能之间的深层联系。

计算化学作为研究物质世界的重要方法, 可以提供体系在分子、原子以及电子水平的行为与性质。随着计算机软、硬件的迅猛发展, 计算化学在揭示催化

反应中的作用越来越重要^[3]。高水平的量子化学计算可以准确地获取催化反应过程的路径与机理, 为研究催化反应过程提供微观层面的探索, 但昂贵的计算成本是其局限。最近十年, 网络与计算设备在收集、存储以及传输海量数据方面的能力迅猛增长。人工智能(artificial intelligence, AI)在大数据挖掘和数据分析方面, 渗透于科学技术的各个领域, 发展势不可挡。作为人工智能的核心策略——机器学习(machine learning, ML)在图像、语音以及模式识别等很多领域表现出强大的预测能力。相比之下, 机器学习在催化领域仍处于起步的早期发展阶段^[4]。2018年, *Nature Catalysis*连续刊登3篇观点文章, 评述了机器学习在催化领域的重要作用以及广阔的发展前景^[5-7]。目前的研究结果表明, 机器学习可以通过算法从大数据中学习, 建立更好的模型, 帮助人们了解催化反应机理, 并为催化剂分子设计提供一种新的见解与思路^[8-10]。得益于新算法和新理论

引用格式: 杨文泓, 孙文华. 机器学习在聚烯烃催化剂领域的应用与展望. 科学通报, 2022, 67: 1870-1880

Yang W H, Sun W H. Application and prospect of machine learning in polyolefin catalysts (in Chinese). Chin Sci Bull, 2022, 67: 1870-1880, doi: 10.1360/TB-2021-1354

的发展、大量易得的实验数据以及较低的计算成本,可以预见机器学习在催化领域将发挥异常重要的作用.本文围绕机器学习的方法以及其在催化领域,特别是聚烯烃催化剂方面的一些重要研究结果展开,在深入探讨机器学习的发展现状基础上,对目前机器学习应用于催化领域所面临的挑战与前景进行展望.

1 机器学习方法概述

机器学习是计算机科学、统计学和数据科学等各学科相结合的一门交叉学科.它关注的是自动化学习过程,在不确定的情况下也能随着时间的推移进行决策.机器学习并不是最近刚出现的研究学科,一些开创性的工作报道可以追溯到20世纪40年代,发展至今已经涌现出很多不同的计算方法.按照学习方式的不同,可以把机器学习分为监督学习与非监督学习两大类,两者的区别主要在于是否有标记数据用于训练模型.相比之下,监督学习方法应用更广泛,比如深度学习的许多成功案例都是从挖掘预测模型的监督学习方法发展而来的.按照算法的不同,监督学习方法又可以进一步分为回归与分类两种类型,包括线性回归、人工神经网络(artificial neural network, ANN)、支持向量机(support vector machine, SVM)、随机森林(random forest, RF),以及各种形式的决策树(decision tree, DT)等方法^[11~18].在此,我们选择了3种常用方法以演示不同算法的计算原理,如图1所示.

多元线性回归分析(multiple linear regression, MLR)是机器学习中最基本,也是最简单的一类方法,如图1(a)所示,它是通过描述符(X_i)的线性组合以预测目标属性值(Y).由于自变量和因变量之间的关系简单明了,因此模型具有易于解释的优点.人工神经网络,顾名思义,是一种模仿人类大脑中生物神经元工作原理的技术.该方法是目前比较流行的非线性机器学习

算法,在回归分析和分类分析中均得到了广泛的应用.它通常包括输入层、隐藏层和输出层3部分(图1(b)).每一层由一定数量的神经元组成(h_i),接收来自输入层的信号,并在不同层之间传递响应,经过非线性激活函数的处理,最终得到可靠的输出层以预测目标属性值.随机森林(RF)的原理是在构建训练集和预测树的基础上,将给定变量的观测映射出来,从而获得对目标属性值准确而稳定的预测.它对数据点进行分类,从树的顶部开始到不同的分支,再向下通过几个节点,最后到达终端叶,如图1(c)中绿色的决策流所示.相对于人工神经网络,随机森林简单、清晰,因此更容易解释.此外,该方法可以更深入地了解描述符对目标属性值的影响.

不同的机器学习方法有各自的优缺点,方法的选择取决于数据库的大小和特征.一般来说,对于小数据集,线性回归和分类方法可以处理这类任务,除了多元线性回归方法,还包括逻辑回归(logistic regression)和朴素贝叶斯(Naive Bayes)等方法^[19,20].对于大数据集则更适合采用非线性方法,如ANN和K最近邻(K-nearest neighbor, KNN)方法^[21].其中人工神经网络的方法易于学习,通常对大数据会得到更好的模型性能.此外,也有一些方法可以同时用于线性和非线性分析,如SVM,它比人工神经网络更容易解释.不过,它对参数和函数的选择也更敏感.对于无监督学习,聚类是最常用的技术,而K均值(K-means)方法是一个很好的选择,它可以简单、快速地处理大数据集,并发现未标记数据中的隐藏模式.其缺点是对初始值比较敏感,不适合差异较大的聚类^[22].

2 机器学习在过渡金属聚烯烃催化剂中的应用

应用于烯烃催化聚合的过渡金属配合物催化剂,主要包括Ziegler-Natta催化剂、铬系催化剂、茂金属催化剂以及后过渡金属配合物催化剂.虽然机器学习

图1 常用监督类型机器学习方法的示意图。(a) 线性回归; (b) 人工神经网络; (c) 随机森林

Figure 1 The schematic diagram for common machine learning methods. (a) Linear regression; (b) artificial neural network; (c) random forest

的早期研究工作在半个多世纪以前就有报道,但一些重要的研究应用最近几年才出现在化学领域的有机化学和无机化学分支中^[23-25].机器学习在聚烯烃催化剂领域的研究报道相对较少,且其中大部分工作利用的是线性回归方法.

2.1 机器学习在Ziegler-Natta催化剂中的应用

Ziegler-Natta催化剂是以IV-VIIB族过渡金属为主催化剂, I-III A族烷基金属为助催化剂,用于 α -烯烃聚合的一类高效催化剂. 1953年,德国有机金属化学家Ziegler等人^[26]发现利用 $\text{TiCl}_4/\text{AlEt}_3$ 催化剂可以催化乙烯聚合,随后意大利化学家Natta等人^[27,28]很快启动丙烯聚合研究,利用该催化剂体系实现了工业聚丙烯的生产.由于此类催化剂的发现,民用聚烯烃产业得到了快速推广,产生了巨大的社会影响力和市场价值.因此,1963年,Ziegler和Natta共同荣获了当年的诺贝尔化学奖. Ziegler-Natta催化剂作为石化工业中极为重要的催化剂体系,对大规模生产聚烯烃具有重要意义.

尽管Ziegler-Natta催化剂已经发展了60多年,但由于其活性位点结构的不确定性,许多特性还在研究中,尤其是分子水平上对其结构与性质的真正了解依然欠缺.在过去的20年中,有很多研究利用密度泛函理论(density functional theory, DFT)的方法考察Ziegler-Natta催化剂的催化聚合反应机理. Cavallo研究组^[29]对非均相Ziegler-Natta催化剂的DFT研究进展进行了详细的综述,包括 α -烯烃聚合中立体选择性的起源、反应链终止的机理以及杂质在Ziegler-Natta催化剂中的作用等.相比之下,机器学习方法在Ziegler-Natta催化剂中的应用报道不多.

$\text{TiCl}_4\text{-MgCl}_2$ 纳米片作为Ziegler-Natta催化剂的主要组成部分,在催化剂的结构中存在一定的分布,而这种分布很可能是所得聚合物具有多分散性特征的主要原因. Taniike研究组^[30]提出了一种基于构象空间全局搜索的遗传算法(genetic algorithm, GA)与局部几何优化的DFT算法相结合的非经验结构方法,用于确定Ziegler-Natta催化剂初级颗粒中 $\text{TiCl}_4\text{-MgCl}_2$ 纳米片的结构(图2).利用该程序,即便没有相关领域背景知识的科研人员,也可成功地得到 $\text{TiCl}_4\text{-MgCl}_2$ 纳米片的稳定结构.同时,研究发现 TiCl_4 在 MgCl_2 载体上的吸附过程会诱导 $\text{TiCl}_4\text{-MgCl}_2$ 纳米片的重构,致使其结构与预期的理想表面不同.吸附过程还可以使 $\text{TiCl}_4\text{-MgCl}_2$ 体系的结构具有多样化,即在很窄的能量跨度内可以存在

图2 结合遗传算法和密度泛函计算所得的非经验结构方法用于确定 $\text{TiCl}_4\text{-MgCl}_2$ 纳米片结构^[30]. Copyright © 2019 American Chemical Society

Figure 2 The nonempirical structure determination of TiCl_4 -capped MgCl_2 nanoplates by the combination of a genetic algorithm and density functional calculations^[30]. Copyright © 2019 American Chemical Society

十几个结构不同的亚稳态.进一步,Taniike研究组^[31]利用改进的非经验结构程序产生了大量不同尺寸与化学组成的 $\text{TiCl}_4\text{-MgCl}_2$ 纳米片结构,以获取 TiCl_4 在 MgCl_2 表面吸附的结构信息.结果表明,大多数 TiCl_4 吸附在 MgCl_2 的{110}表面,少量吸附在{100}表面.同时,在相同的单个纳米片结构中, TiCl_4 的空间环境和电荷态也是有分布的.而不同的电荷分布会影响乙烯插入的反应能垒,进而影响所得聚合物的一级结构.最近,该研究组^[32]利用定量构效关系研究了以 $\text{Mg}(\text{OEt})_2$ 为载体对Ziegler-Natta催化剂结构与聚合性能的影响.以实验分析测定的催化剂结构参数为分子描述符,对16个 $\text{Mg}(\text{OEt})_2$ 负载的催化剂进行预测.结果发现,利用单一的描述符很难预测催化剂的结构以及对催化性能的影响,但综合考虑2个或3个描述符则不仅可以很好地预测催化剂结构,还可以预测催化乙烯的聚合活性以及乙烯与1-己烯的共聚效率.分析描述符的贡献发现, $\text{Mg}(\text{OEt})_2$ 颗粒的致密性对介孔和大孔体积及其分布有重要影响,其中10 nm以下的中孔与活性以及1-己烯共聚效率呈负相关.

2.2 机器学习在铬系催化剂中的应用

均相铬系催化剂可以对乙烯进行选择性齐聚制得1-己烯和1-辛烯等产物,这类线性 α -烯烃是重要的化工原料,也是高附加值的共聚单体^[33-35].2003年,雪佛龙菲利普斯化学公司(Chevron Phillips Chemical)首次实现工业化乙烯选择性三聚制备1-己烯体系^[36].南非石化公司Sasol后来发展了选择性乙烯齐聚技术,使用铬-双膦胺催化剂生产1-己烯和1-辛烯.这些催化剂的催化

性能主要受配体结构的影响, 包括空间位阻效应与电子效应. 原则上, 通过调控配体结构可以得到高选择性乙烯齐聚的催化剂. 利用DFT方法对铬系催化剂选择性齐聚的机理研究有很多报道^[37-40], 这些工作计算的过渡态模型很准确, 且具有实用性, 并且也通过实验验证了新的配体. 但是单纯依靠反应能垒的计算, 很难提供理论指导以帮助设计高选择性催化剂, 即依然缺乏一套通用的催化剂分子设计原理.

Liu研究组^[41,42]对19种双(二苯基膦基)胺铬催化剂的乙烯选择性三聚/四聚进行了定量研究. 基于软件以及自定义计算的描述符, 利用启发式方法(Heuristic method, HM)和最佳多元线性回归(best multiple linear regression, BMLR)建立的模型, 揭示了催化剂分子中与齐聚选择性密切相关的结构因素. 结果显示, 有良好平整度和对称性的PNP-Cr体系骨架结构是实现1-辛烯催化选择性的关键, 而对N原子具有较大空间位阻效应的PNP-Cr骨架结构有利于乙烯三聚.

最近, Bischof和Ess研究组^[43]针对105个Cr(P,N)催化剂, 计算了催化剂分子结构的14个描述符, 以DFT方法计算的催化乙烯齐聚生成1-己烯与1-辛烯的过渡态能量差为输出变量, 建立了预测齐聚选择性的随机森林回归模型. 分析描述符的重要性可知, 配体结构中Cr-N距离以及中心金属Cr周围相关的空间环境与乙烯的齐聚选择性密切相关, 如图3所示. 随后, 他们设计了新一代Cr(P,N)催化剂配体, 计算结果显示对1-辛烯具有超高的选择性(>95%). 该研究工作将过渡态模型与数据科学方法相结合, 通过发现潜在的化学特征信息, 以达到提高齐聚选择性的目的.

2.3 机器学习在茂金属催化剂中的应用

早在20世纪50年代, Natta和Breslow等人就发现金

图3 由随机森林模型确定的分子描述符的重要性^[43]. Copyright © 2020 Royal Society of Chemistry

Figure 3 The feature importance determined from random forest model^[43]. Copyright © 2020 Royal Society of Chemistry

属钛与环戊二烯配位在烷基铝($\text{Cp}_2\text{TiCl}_2/\text{AlEt}_2\text{Cl}$)作用下可以催化乙烯进行聚合, 但当时的催化活性较低, 并未受到重视. 由于Ziegler-Natta催化体系可以与水发生强烈的反应形成水解产物, 而失去催化聚合活性, 因此长期以来, 人们一直认为水对该催化体系具有“毒性”. 1973年, Reichert和Meyer^[44]偶然发现, 向 $\text{Cp}_2\text{TiCl}_2/\text{AlEt}_2\text{Cl}$ 催化体系中加入少量水不但没有使催化剂“中毒”失去活性, 反而大大增加了催化乙烯聚合的活性. 随后, 在1976年, 德国化学家Kaminsky^[45]发现烷基铝与水反应形成的是一种结构特殊的化合物, 称之为甲基铝氧烷(methylaluminoxane, MAO). 该化合物可以作为助催化剂与 Cp_2ZrCl_2 催化体系结合, 对乙烯和丙烯催化表现出惊人的聚合活性^[46,47]. 该结果在聚烯烃中具有里程碑意义, 是非常重要的发现. 后来, 研究者将此类由过渡金属(如钛、锆、钪)或稀土金属和至少一个环戊二烯或环戊二烯衍生物(如取代环戊二烯、茚基、茱基)配位组成的有机金属配合物, 统称为茂金属催化剂. 茂金属催化剂的成功开发使聚烯烃工业发生了革命性的变革, 引发了茂金属催化烯烃聚合的集中研究与产业化发展, 并且成为近20年里高性能聚烯烃树脂研究与发展的推动力.

由于二茂锆化合物及其衍生物和甲基铝氧烷组成的催化体系与其同族的钛、钪催化体系相比, 在催化乙烯和丙烯时的聚合活性高出两个数量级, 因此成为研究最多的茂配合物催化剂体系. Möhring和Coville^[48,49]较早利用数据分析方法来定量评价茂锆配合物催化剂体系, 他们选择Hammett常数(F)与圆锥角(θ)作为催化剂分子的结构描述符, 建立的多元线性回归模型可以很好地预催化剂结构中 $(\text{CpR})_2\text{ZrCl}_2$ 取代基R对催化活性的影响. Yao等人^[50]对茂锆配合物和氟硼酸 $[\text{B}(\text{Ph-F})_4]^-$ 共催化剂体系进行了定量构效的研究. 结果显示, 催化剂与助催化剂之间的相互作用能和空间距离是决定催化活性的主要因素.

Cruz研究组^[51-55]将三维定量构效关系(three dimensional-quantitative structure-property relationship, 3D-QSPR)方法引入到茂金属催化剂体系中, 开展了一系列的研究工作. 针对双环戊二烯基Zr(IV)、双二甲基硅烷-茚三烯基Zr(IV)以及双茚三烯-Hf(IV)等体系, 建立了预测催化活性以及聚合产物分子量的模型. 结果显示, 催化活性主要受芳香配体周围的电荷分布和活性位点周围的空间位阻影响, 而聚合产物分子量则主要受单体插入反应的空间位阻影响.

客观地讲, 与Ziegler-Natta催化剂相比, 茂金属配合物催化剂在乙烯聚合应用上, 不论是催化活性还是所得聚合物结构, 并不占明显优势. 而其可以发挥到极致的特性主要在于对丙烯的立构可控聚合以及乙烯与其他烯烃的共聚, 如 α -烯烃或者环烯烃.

Cavallo研究组^[56]提出, 利用埋藏体积百分比(% V_{Bur})的分子描述符, 可以表征桥连茂金属催化剂的空间位阻变化, 并且与实验观测的立体选择性之间存在一定的相关性. 进一步, 针对钛、锆和铪三类茂金属催化剂, 基于催化剂前驱体结构计算的% V_{Bur} 与DFT方法计算的链增长和链终止反应的能量差之间存在线性相关, 说明利用该描述符可以解释不同催化剂催化丙烯聚合所得产物的分子量^[57]. 最近, Cavallo研究组^[58]结合多元线性回归(MLR)和DFT方法预测了茂金属催化剂催化乙烯与 α -烯烃共聚时的共聚单体掺入率(图4). 值得一提的是, 该工作中的模型分子数量并不多, 一共有19个茂金属催化剂作为训练集, 而且这些实验数据来自不同的实验室. 此外, 共聚单体不仅有丙烯还有1-己烯. 尽管如此, 所得ML模型的预测与验证能力仍然不错. 并且, 利用所得模型除了对与训练集中结构相似的催化剂共聚单体插入率进行预测外, 还可以对结构不同的催化剂进行可靠的预测.

Ehm等人^[59]也基于% V_{Bur} 描述符, 通过计算催化剂前驱体的封闭和开放象限之间的% V_{Bur} 差异, 对高通量实验中合成的19个茂金属催化剂建立了定量构效关系, 所得模型可以很好地预测催化剂的立构选择性. 他们设计了取代基为邻甲苯基的催化剂, 预测其作为一种理想的高立体选择性及高聚合物分子量催化剂, 该预

图4 多元线性回归方法预测锆茂金属催化剂对乙烯与 α -烯烃共聚的插入率^[58]. Copyright © 2021 American Chemical Society
Figure 4 Multiple linear regression analysis to predict ethene/1-olefin copolymerization statistics promoted by *ansa*-zirconocene catalysts^[58]. Copyright © 2021 American Chemical Society

测被实验所证实. 进一步, 对高通量实验合成的40个硅桥联 C_2 对称的茂金属催化剂催化乙烯/己烯共聚反应的单体插入率进行研究. 研究发现, 模型中最重要的描述符依然是% V_{Bur} 这一空间位阻描述符^[60]. 具有热稳定性的茂金属催化剂一直是工业界亟需的体系, Ehm等人^[61]针对实验合成的21个茂金属配合物催化剂, 结合埋藏体积的位阻效应描述符与链异构化的电子效应描述符, 对催化丙烯聚合的立构规整度进行了定量预测(图5(a)). 研究发现, 催化剂分子骨架的刚性有利于催化性能以及耐热性. 据此, 实验合成了两种超硬的三蝶烯催化剂^[62], 如图5(b)所示, 测定该催化体系即便在120°C高温下, 催化丙烯所得聚合物仍具有很好的立构规整度.

2.4 机器学习在后过渡金属催化剂中的应用

20世纪90年代中叶, 英国学者Brookhart和Gibson^[63-65]发现2,6-二亚胺吡啶的铁、钴配合物以及 α -二亚胺的镍、钯配合物, 利用该催化体系可以实现乙烯

图5 多元线性回归方法研究锆茂金属催化剂催化丙烯聚合中反应温度对立构等规的影响^[61](a), Copyright © 2020 American Chemical Society, 和基于模型设计并合成的催化剂结构(b)^[62]
Figure 5 Multiple linear regression analysis to predict ethene/1-olefin copolymerization statistics promoted by *ansa*-zirconocene catalysts^[61](a), Copyright © 2020 American Chemical Society, and the designed and synthesized *ansa*-zirconocene catalysts with a triptycene motif (b)^[62]

聚合制备高分子量聚乙烯。由于后过渡金属配合物制备简单且稳定性高,很快地吸引了学术界与产业界的广泛关注,图6中A和B所示为两个研究最多和获得衍生物最为丰富的模型结构。后来,基于相同的理论基础,研究者发展了很多高效新型金属配合物模型(图6中C、D和E)^[66-68],呈现了更具实用价值的研究结果和新型聚合物材料。20世纪70年代,第三代Ziegler-Natta催化剂在催化活性上已经满足聚烯烃生产的需求,后续新型Ziegler-Natta催化剂和茂金属催化剂都是基于所得聚烯烃品质开展研究并用于改进生产。后过渡金属配合物催化烯烃聚合研究的突破口和意义主要集中在所得材料的结构与性能上。

Cruz研究组^[69]除了对茂金属催化剂进行定量构效的研究外,对吡啶亚胺铁(II)配合物催化剂也进行了研究,所得3D-QSAR模型对催化活性的预测并不好,但可以较好地预测聚合物分子量的变化,模型的相关系数(R^2)为0.95,交叉验证系数(Q^2)为0.8。Fayet等人^[70]研究了22个基于双(芳基)吡啶亚胺铁(II)催化剂结构与其催化乙烯活性间的关系,建立了包含4个分子描述符的多元线性回归模型,对训练集进行了预测($R^2=0.98$, $Q^2=0.92$)。

Yang等人^[71-74]在后过渡金属配合物催化剂领域中进行了一系列的工作,比较详细、系统地研究了催化剂分子结构与催化乙烯聚合反应性能之间的定量关系。首先,从催化剂结构的电子效应角度出发,发展了中心金属有效净电荷(Q_{eff})^[71]、催化剂与乙烯分子间HOMO-LUMO轨道能级差($\Delta\epsilon_1$, $\Delta\epsilon_2$),以及不同自旋态间的能量差(ΔE)^[72]等分子描述符。对铁、钴、镍等配合物催化剂的研究发现,各描述符与催化活性之间均存在一定的线性相关。中心金属有效净电荷通常与催化活性呈正相关性,这是由于烯烃的配位聚合通常是亲电反应。因此,中心金属所带电荷越多,催化乙烯的反应越容易进行,对应的活性也就越高。基于Tolman^[73]

提出的圆锥角概念,并结合后过渡金属配合物催化剂的结构特点,Yang等人^[74,75]提出利用开放圆锥角(θ)来表征催化反应活性位点周围可以容纳乙烯进行插入配位聚合的空间。 θ 越大,提供给乙烯进行配位聚合反应的空间越大,所得催化活性也就越高。综合考虑电子与空间位阻效应两方面,建立的MLR模型可以很好地预测铁、钴、镍等配合物催化剂中取代基变化对催化活性的影响,相关系数(R^2)均在0.9以上。

考虑到存在过渡拟合的风险,Yang等人^[76-78]逐步删减了描述符的个数,通过尝试不同的个数以及组合方式,从多个备选描述符中得到用于建模的最优描述符组合。针对相同配体、不同中心金属的配合物体系,研究者分析发现,利用 Q_{eff} 与 θ 两个描述符就可以很好地预测并解释配合物类似物之间活性变化的趋势与原因。进一步,用此方法对镍配合物中由于取代基和骨架的变化,以及与中心金属配位卤素的不同所引起的活性或聚合物分子量的变化,也在分子水平上给予了很好的解释^[76-78]。

为了考察更多催化剂分子,利用二维定量构效关系方法(two dimensional-quantitative structure-property relationship, 2D-QSPR)来考察催化性能^[79]。对58种2-苯基咪唑/恶唑基吡啶二亚胺类催化剂,利用PLS-VIP筛选描述符建立的模型具有很好的预测与验证能力($R^2>0.9$, $Q^2>0.8$)。结果显示,增加中心金属的有效电荷以及苯环取代基邻位的位阻,可以提高催化剂的反应活性。实验上绝大部分的后过渡金属配合物催化剂,通过改变结构使其催化活性增加时,往往伴随着聚合物分子量的下降。为了解释背后的原因,对55种二甲苯基吡啶二亚胺类(铁、钴)催化剂进行了2D-/3D-QSPR的研究^[80]。结果显示,在预测催化活性与聚合物分子量的模型中,同时存在相同的描述符,但其贡献是相反的;与催化活性呈正相关的描述符往往与聚合产物分子量呈负相关(图7)。这就解释了同一催化剂分子很难同时

图6 典型后过渡金属催化剂模型

Figure 6 The classic models of late-transition metal complex catalysts

图7 三维定量构效关系方法所得的空间位阻场三维等高图^[79]. (a) 催化活性; (b) 聚合物分子量

Figure 7 The contour maps of the steric field obtained by the three-dimensional quantitative structure-property relationship method^[79]. (a) Catalytic activity; (b) molecular weight of polymer

兼具高催化活性与高聚合分子量两种性能的原因。

基于经典的吡啶二亚胺分子骨架，实验上开发了一系列含稠环烷基的类似物，而稠环的尺寸可以明显地调控聚合性能。利用2D-/3D-QSPR对49种环烷基双芳基吡啶二亚胺类(铁、钴)催化剂的研究发现^[81]，配合物中芳香键的数量对催化活性的贡献在50%以上，说明除了环张力外的影响，共轭程度是决定催化活性的主要因素。这也解释了铁配合物中稠合七元环的衍生物的活性为什么比六元环类似物的更高。当中心金属为镍时，研究发现决定活性的主要结构因素为配合物分子中碳、氮原子的电荷分布^[82]。

尽管这些工作中的催化剂个数已经增加到50多个，但要得到更普适的模型还需要更大的数据集^[83]。最近，基于实验合成报道的294种吡啶二亚胺类配合物催化剂，利用ANN对催化乙烯的聚合活性进行了预测与解释^[84]。通过计算描述符的重要值发现，催化剂分子中的共轭程度以及大位阻取代基与催化活性密切相关。这是首次利用非线性的神经网络方法研究后过渡金属聚烯烃催化剂的工作，表明机器学习可以成为研究均相聚烯烃催化剂体系的很好工具。

3 总结与展望

机器学习是近几年非常热门的一个研究领域，其应用也呈现爆炸式的增长。得益于其强大的预测能力，机器学习在很多方面取得了巨大的成功，这其中也包括聚烯烃催化剂领域。利用机器学习可以建立预测催化性能的模型，并对催化剂结构与性能的关系进行解

释，包括Ziegler-Natta催化剂的结构与活性位点、均相铬系催化剂催化乙烯齐聚的选择性、茂金属催化剂催化丙烯的立构选择性与催化乙烯和共聚单体的插入率，以及后过渡金属配合物催化剂的活性与分子量等。

尽管如此，机器学习在催化领域中的应用仍然是一个年轻的学科，像任何新事物的出现一样，其中有许多有待探索与成长的地方。追踪近些年的研究进展，相比其他领域，机器学习在聚烯烃催化剂中的研究并不多。其中一个重要的原因是，催化聚合性能的实验研究费时费力，需要大量的资源，从动力学研究建立大数据集是一大挑战。而准确的模型是基于一定的数据集，尤其在采用非线性算法时，没有成百上千的数据量，很难得到可靠的模型。因此，如何建立大数据集或许是目前机器学习方法应用于聚烯烃催化剂领域面临的主要难题。不过，最近Cavallo研究组^[58]的工作提示，即便训练集中的数据并非来自同一实验室，依然可以得到很好的预测与验证结果。这将有利于不同研究团队通过贡献各自的实验数据以建立大数据库，并在相互之间进行分享。另一方面，如果可以采用非监督学习方法，则可以大大降低对实验数据的依赖度，这将可能成为今后该领域发展的新方向。

机器学习这一技术工具为催化领域的设计和开发提供了一个新平台，通过降低开发催化体系的相关成本，有望为聚烯烃催化剂的研究提供崭新角度的设计范式。随着应用研究的不断增加，可以预见，机器学习方法在不久的将来会成为催化领域计算工具的标志。

参考文献

- 1 Zhao W M. Development status quo and prediction for China high-end polyolefin resins (in Chinese). *Chem Ind*, 2017, 35: 46–58 [赵文明. 高端聚烯烃树脂产业发展现状及市场预测. *化学工业*, 2017, 35: 46–58]
- 2 Vasile C, Seymour R B. *Handbook of Polyolefins*. New York: Marcel Dekker, 1993
- 3 Vogiatzis K D, Polynski M V, Kirkland J K, et al. Computational approach to molecular catalysis by 3D transition metals: Challenges and opportunities. *Chem Rev*, 2019, 119: 2453–2523
- 4 Jordan M I, Mitchell T M. Machine learning: Trends, perspectives, and prospects. *Science*, 2015, 349: 255–260
- 5 Kitchin J R. Machine learning in catalysis. *Nat Catal*, 2018, 1: 230–232
- 6 Jones G. Industrial computational catalysis and its relation to the digital revolution. *Nat Catal*, 2018, 1: 311–313
- 7 Li Z, Wang S, Xin H. Toward artificial intelligence in catalysis. *Nat Catal*, 2018, 1: 641–642
- 8 Foscatto M, Jensen V R. Automated in silico design of homogeneous catalysts. *ACS Catal*, 2020, 10: 2354–2377
- 9 Funes-Ardoiz I, Schoenebeck F. Established and emerging computational tools to study homogeneous catalysis—From quantum mechanics to machine learning. *Chem*, 2020, 6: 1904–1913
- 10 Nandy A, Duan C, Taylor M G, et al. Computational discovery of transition-metal complexes: From high-throughput screening to machine learning. *Chem Rev*, 2021, 121: 9927–10000
- 11 Clarke B, Fokoue E, Zhang H H. *Principles and Theory for Data Mining and Machine Learning*. Berlin: Springer, 2009
- 12 Alexopoulos E C. Introduction to multivariate regression analysis. *Hippokratia*, 2010, 14: 23–28
- 13 Rencher A C, Christensen W F. *Method of Multivariate Analysis*. 3rd ed. New Jersey: Wiley, 2012
- 14 Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. *Nature*, 1986, 323: 533–536
- 15 Crammer K, Singer Y. On the algorithmic implementation of multiclass kernel-based vector machines. *J Mach Learn Res*, 2001, 2: 265–292
- 16 Smola A J, Schölkopf B. A tutorial on support vector regression. *Stat Comput*, 2004, 14: 199–222
- 17 Breiman L. Random forests. *Mach Learn*, 2001, 45: 5–32
- 18 Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*, 2006, 63: 3–42
- 19 Bishop C M. *Pattern Recognition and Machine Learning*. Berlin: Springer, 2006
- 20 Tipping M E. Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res*, 2001, 1: 211–244
- 21 Omohundro S M. *Five Balltree Construction Algorithms*. Berkeley: International Computer Science Institute, 1989. 1–22
- 22 Arthur D, Vassilvitskii S. k-means++: The advantages of careful seeding. In: *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2007. 1027–1035
- 23 Peiretti F, Brunel J M. Artificial intelligence: The future for organic chemistry? *ACS Omega*, 2018, 3: 13263–13266
- 24 Janet J P, Liu F, Nandy A, et al. Designing in the face of uncertainty: Exploiting electronic structure and machine learning models for discovery in inorganic chemistry. *Inorg Chem*, 2019, 58: 10592–10606
- 25 Mater A C, Coote M L. Deep learning in chemistry. *J Chem Inf Model*, 2019, 59: 2545–2559
- 26 Ziegler K, Holzkamp E, Breil H, et al. Das mülheimer normaldruck-polyäthylen-verfahren. *Angew Chem*, 1955, 67: 541–547
- 27 Natta G, Porri L, Zanini G, et al. Stereospecific polymerization of conjugated diolefins. IV. Preparation of syndiotactic 1,2-polybutadiene. *Chim Ind*, 1959, 41: 526–532
- 28 Natta G, Corradini P, Ganis P. Chain conformation of polypropylenes having a regular structure. *Makromol Chem*, 1960, 39: 238–242
- 29 Bahri-Laleh N, Hanifpour A, Mirmohammadi S A, et al. Computational modeling of heterogeneous Ziegler-Natta catalysts for olefins polymerization. *Prog Polym Sci*, 2018, 84: 89–114
- 30 Takasao G, Wada T, Thakur A, et al. Machine learning-aided structure determination for TiCl₄-capped MgCl₂ nanoplate of heterogeneous Ziegler-Natta catalyst. *ACS Catal*, 2019, 9: 2599–2609
- 31 Takasao G, Wada T, Thakur A, et al. Insight into structural distribution of heterogeneous Ziegler-Natta catalyst from non-empirical structure determination. *J Catal*, 2021, 394: 299–306
- 32 Wada T, Funako T, Chammingkwan P, et al. Structure-performance relationship of Mg(OEt)₂-based Ziegler-Natta catalysts. *J Catal*, 2020, 389: 525–532
- 33 Agapie T. Selective ethylene oligomerization: Recent advances in chromium catalysis and mechanistic investigations. *Coord Chem Rev*, 2011, 255: 861–880
- 34 McGuinness D S. Olefin oligomerization via metallacycles: Dimerization, trimerization, tetramerization, and beyond. *Chem Rev*, 2011, 111: 2321–2341
- 35 van Leeuwen P W N M, Clément N D, Tschan M J L. New processes for the selective production of 1-octene. *Coord Chem Rev*, 2011, 255: 1499–

1517

- 36 Knudsen R D, Kreischer B E, Abbott R G, et al. Method of preparation of an olefin oligomerization catalyst. US Patent, US7384886B2, 2008-06-10
- 37 Kozłowski M C, Dixon S L, Panda M, et al. Quantum mechanical models correlating structure with selectivity: Predicting the enantioselectivity of β -amino alcohol catalysts in aldehyde alkylation. *J Am Chem Soc*, 2003, 125: 6614–6615
- 38 Ianni J C, Annamalai V, Phuan P W, et al. A priori theoretical prediction of selectivity in asymmetric catalysis: Design of chiral catalysts by using quantum molecular interaction fields. *Angew Chem Int Edit*, 2006, 45: 5502–5505
- 39 Mitsumori S, Zhang H, Ha-Yeon Cheong P, et al. Direct asymmetric anti-Mannich-type reactions catalyzed by a designed amino acid. *J Am Chem Soc*, 2006, 128: 1040–1041
- 40 Jindal G, Sunoj R B. Rational design of catalysts for asymmetric diamination reaction using transition state modeling. *Org Biomol Chem*, 2014, 12: 2745–2753
- 41 Tang S, Liu Z, Zhan X, et al. 2D-QSPR/DFT studies of aryl-substituted PNP-Cr-based catalyst systems for highly selective ethylene oligomerization. *J Mol Model*, 2014, 20: 2129
- 42 Tang S Y, Liu Z, Zhan X W, et al. 2D-QSPR studies on substituting PNP-Cr catalyst systems for highly selective ethylene oligomerization (in Chinese). *CIESC J*, 2014, 65: 131–142 [唐思扬, 刘振, 占兴稳, 等. 乙烯选择性齐聚取代PNP-Cr催化体系的二维定量构效关系. *化工学报*, 2014, 65: 131–142]
- 43 Maley S M, Kwon D H, Rollins N, et al. Quantum-mechanical transition-state model combined with machine learning provides catalyst design features for selective Cr olefin oligomerization. *Chem Sci*, 2020, 11: 9665–9674
- 44 Reichert K H, Meyer K R. Zur kinetik der niederdruckpolymerisation von äthylen mit löslichen Ziegler-katalysatoren. *Macromol Chem Phys*, 1973, 169: 163–176
- 45 Andresen A, Cordes H G, Herwig J, et al. Halogenfreie lösliche Ziegler-Katalysatoren für die Ethylen-Polymerisation. Regelung des Molekulargewichtes durch Wahl der Reaktionstemperatur. *Angew Chem*, 1976, 88: 689–690
- 46 Sinn H, Kaminsky W, Vollmer H J, et al. “Living polymers” on polymerization with extremely productive Ziegler catalysts. *Angew Chem Int Edit*, 1980, 19: 390–392
- 47 Kaminsky W. New polymers by metallocene catalysis. *Macromol Chem Phys*, 1996, 197: 3907–3945
- 48 Möhring P C, Coville N J. Quantification of the influence of steric and electronic parameters on the ethylene polymerisation activity of $(\text{CpR})_2\text{ZrCl}_2$ /ethylaluminoxane Ziegler-Natta catalysts. *J Mol Catal*, 1992, 77: 41–50
- 49 Möhring P C, Coville N J. The influence of cyclopentadienyl ring substituent steric and electronic effects on the ethylene- α -olefin copolymerisation behaviour of $(\text{CpR})_2\text{ZrCl}_2$ ethylaluminoxane catalysts. *J Mol Catal A-Chem*, 1995, 96: 181–195
- 50 Yao S, Shoji T, Iwamoto Y, et al. Consideration of an activity of the metallocene catalyst by using molecular mechanics, molecular dynamics and QSAR. *Comput Theor Polym Sci*, 1999, 9: 41–46
- 51 Cruz V, Ramos J, Muñoz-Escalona A, et al. 3D-QSAR analysis of metallocene-based catalysts used in ethylene polymerisation. *Polymer*, 2004, 45: 2061–2072
- 52 Cruz V L, Ramos J, Martinez S, et al. Structure-activity relationship study of the metallocene catalyst activity in ethylene polymerization. *Organometallics*, 2005, 24: 5095–5102
- 53 Cruz V L, Martinez S, Martinez-Salazar J, et al. 3D-QSAR study of ansa-metallocene catalytic behavior in ethylene polymerization. *Polymer*, 2007, 48: 4663–4674
- 54 Martínez S, Cruz V L, Ramos J, et al. Polymerization activity prediction of zirconocene single-site catalysts using 3D quantitative structure-activity relationship modeling. *Organometallics*, 2012, 31: 1673–1679
- 55 Cruz V L, Martinez S, Ramos J, et al. 3D-QSAR as a tool for understanding and improving single-site polymerization catalysts. A review. *Organometallics*, 2014, 33: 2944–2959
- 56 Poater A, Cavallo L. Comparing families of olefin polymerization precatalysts using the percentage of buried volume. *Dalton Trans*, 2009, 41: 8878–8883
- 57 Falivene L, Cavallo L, Talarico G. Buried volume analysis for propene polymerization catalysis promoted by group 4 metals: A tool for molecular mass prediction. *ACS Catal*, 2015, 5: 6815–6822
- 58 Maity B, Cao Z, Kumawat J, et al. A multivariate linear regression approach to predict ethene/1-olefin copolymerization statistics promoted by group 4 catalysts. *ACS Catal*, 2021, 11: 4061–4070
- 59 Ehm C, Vittoria A, Goryunov G P, et al. Connection of stereoselectivity, regioselectivity, and molecular weight capability in *rac*- $\text{R}'_2\text{Si}(2\text{-Me-4-R-indenyl})_2\text{ZrCl}_2$ type catalysts. *Macromolecules*, 2018, 51: 8073–8083
- 60 Ehm C, Vittoria A, Goryunov G P, et al. On the limits of tuning comonomer affinity of “Spaleck-type” *ansa*-zirconocenes in ethene/1-hexene copolymerization: A high-throughput experimentation/QSAR approach. *Dalton Trans*, 2020, 49: 10162–10172
- 61 Ehm C, Vittoria A, Goryunov G P, et al. A systematic study of the temperature-induced performance decline of *ansa*-metallocenes for iPP.

1878

- [Macromolecules](#), 2020, 53: 9325–9336
- 62 Kulyabin P S, Goryunov G P, Sharikov M I, et al. *ansa*-Zirconocene catalysts for isotactic-selective propene polymerization at high temperature: A long story finds a happy ending. [J Am Chem Soc](#), 2021, 143: 7641–7647
- 63 Johnson L K, Killian C M, Brookhart M. New Pd(II)- and Ni(II)-based catalysts for polymerization of ethylene and α -olefins. [J Am Chem Soc](#), 1995, 117: 6414–6415
- 64 Small B L, Brookhart M, Bennett A M A. Highly active iron and cobalt catalysts for the polymerization of ethylene. [J Am Chem Soc](#), 1998, 120: 4049–4050
- 65 Britovsek G J P, Gibson V C, McTavish S J, et al. Novel olefin polymerization catalysts based on iron and cobalt. [Chem Commun](#), 1998, 7: 849–850
- 66 Small B L. Discovery and development of pyridine-bis(imine) and related catalysts for olefin polymerization and oligomerization. [Acc Chem Res](#), 2015, 48: 2599–2611
- 67 Wang Z, Liu Q, Solan G A, et al. Recent advances in Ni-mediated ethylene chain growth: Nimine-donor ligand effects on catalytic activity, thermal stability and oligo-/polymer structure. [Coord Chem Rev](#), 2017, 350: 68–83
- 68 Wang Z, Solan G A, Zhang W, et al. Carbocyclic-fused N,N,N-pincer ligands as ring-strain adjustable supports for iron and cobalt catalysts in ethylene oligo-/polymerization. [Coord Chem Rev](#), 2018, 363: 92–108
- 69 Cruz V L, Martinez J, Martinez-Salazar J, et al. QSAR model for ethylene polymerisation catalysed by supported bis(imino)pyridine iron complexes. [Polymer](#), 2007, 48: 7672–7678
- 70 Fayet G, Raybaud P, Toulhoat H, et al. Iron bis(arylimino)pyridine precursors activated to catalyze ethylene oligomerization as studied by DFT and QSAR approaches. [J Mol Struct-Theochem](#), 2009, 903: 100–107
- 71 Yang W, Chen Y, Sun W H. Assessing catalytic activities through modeling net charges of iron complex precatalysts. [Macromol Chem Phys](#), 2014, 215: 1810–1817
- 72 Yang W, Yi J, Sun W H. Revisiting benzylidenequinolinylnickel catalysts through the electronic effects on catalytic activity by DFT studies. [Macromol Chem Phys](#), 2015, 216: 1125–1133
- 73 Tolman C A. Steric effects of phosphorus ligands in organometallic chemistry and homogeneous catalysis. [Chem Rev](#), 1977, 77: 313–348
- 74 Yi J, Yang W, Sun W H. Quantitative investigation of the electronic and steric influences on ethylene oligo/polymerization by 2-azacycyl-6-aryliminopyridylmetal (Fe, Co, and Cr) complexes. [Macromol Chem Phys](#), 2016, 217: 757–764
- 75 Yang W, Ma Z, Sun W H. Modeling study on the catalytic activities of 2-imino-1,10-phenanthrolylmetal (Fe, Co, and Ni) precatalysts in ethylene oligomerization. [RSC Adv](#), 2016, 6: 79335–79342
- 76 Ahmed S, Yang W, Ma Z, et al. Catalytic activities of bis(pentamethylene)pyridyl(Fe/Co) complex analogues in ethylene polymerization by modeling method. [J Phys Chem A](#), 2018, 122: 9637–9644
- 77 Malik A A, Yang W, Ma Z, et al. Quantitative investigation of the catalytic activities of nickel complex systems in ethylene polymerization. [Catalysts](#), 2019, 9: e520
- 78 Yang W, Meraz M, Fidelis T T, et al. The quantitative influence of coordinated halogen atoms on the catalytic performance of bisiminoacenaphthynickel complexes in ethylene polymerization. [ChemPhysChem](#), 2021, 22: 585–592
- 79 Yang W, Yi J, Ma Z, et al. 2D-QSAR modeling on the catalytic activities of 2-azacycyl-6-aryliminopyridylmetal precatalysts in ethylene oligomerization. [Catal Commun](#), 2017, 101: 40–43
- 80 Yang W, Ma Z, Yi J, et al. Catalytic performance of bis(imino)pyridine Fe/Co complexes toward ethylene polymerization by 2D-/3D-QSPR modeling. [J Comput Chem](#), 2019, 40: 1374–1386
- 81 Yang W, Ahmed S, Fidelis T T, et al. Effect of cycloalkyl-fused ring on the catalytic performance of bis(imino)pyridine metal complexes by QSPR modeling. [Catal Commun](#), 2019, 132: 105820–105824
- 82 Meraz M M, Malik A A, Yang W, et al. Catalytic performance of cycloalkyl-fused aryliminopyridyl nickel complexes toward ethylene polymerization by QSPR modeling. [Catalysts](#), 2021, 11: e920
- 83 Yang W, Fidelis T T, Sun W H. Machine learning in catalysis, from proposal to practicing. [ACS Omega](#), 2020, 5: 83–88
- 84 Yang W, Fidelis T T, Sun W H. Prediction of catalytic activities of bis(imino)pyridine metal complexes by machine learning. [J Comput Chem](#), 2020, 41: 1064–1067

Summary for “机器学习在聚烯烃催化剂领域的应用与展望”

Application and prospect of machine learning in polyolefin catalysts

Wenhong Yang^{1,2*} & Wen-Hua Sun^{2,3*}¹ Advanced Materials Research Center, Petrochemical Research Institute, China National Petroleum Corporation, Beijing 102206, China;² Key Laboratory of Engineering Plastics, Beijing National Laboratory for Molecular Science, Institute of Chemistry, Chinese Academy of Sciences, Beijing 100190, China;³ School of Chemical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China* Corresponding authors, E-mail: whyang@iccas.ac.cn; whsun@iccas.ac.cn

Industrial and academic research has been extensively inspired by the ever-growing demand for polyolefin with high performance due to its special physical and mechanical properties, which has been widely applied in the area of engineering plastics, elastomer and high grade lubricants. Transition metal complex catalysts, which can make the olefin polymerization reaction feasible, have been one of the key techniques to produce polyolefin with various structures and properties. Although many fruitful reports are available describing different attempts on enhancing the performance of polyolefin catalyst by the means of the alteration of the ligand frameworks, shuffling the substituents as well as introducing new ligands. Nevertheless, the traditional process of catalyst development, using the trial-and-error method, usually needs long experimental steps and periods. Meanwhile, the measurement of catalytic performance is high cost and needs a lot of resources as well. Machine learning, as the core strategy of artificial intelligence, has shown strong predictive power in many fields of science and technology. However, the application in chemistry, especially in catalysis, is still in its infancy. Relying on the rapid development of different algorithms and computer hardware, it is the right time to harvest the potential of machine learning in the field of catalysis across academy and industry. Herein, we discuss the recent advances in the field of polyolefin catalysts by using machine learning methods, including Ziegler-Natta catalysts, phosphine monocyclic imine Cr(P,N) catalysts, *ansa* zirconocene catalysts, and late-transition metal complex catalysts. The catalytic performance is well predicted, providing insight into the underlying mechanism of the relationship between the micro-structure of the catalyst and its macro-performance at the molecule level.

Tracing the recent progress, the report of machine learning in polyolefin catalysts is relatively few. One of the main reasons is that the experimental study of catalytic performance is time-consuming and labor-intensive and it is difficult to establish a big data set from kinetic studies. However, the recent work by Cavallo suggests that the model with good prediction and validation results can still be obtained, even though the model catalysts in data set are selected from different laboratories. This may inspire researchers from different groups to contribute their experimental data together and share with each other. On the other hand, compared with supervised machine learning, the unsupervised methods have the advantage of low dependence on experimental data and may effectively solve the problem of building large data set. To the best of our knowledge, modern statistical learning techniques will be a strong tool for computational optimization and discovery. This perspective provides a platform for an integrated machine learning technique for new design in the field of polyolefin catalysts, which is expected to change the traditional way by lowering the cost associated with the initial discovery. It is anticipated that machine learning will become the signal of a high-level computational tool in the catalytic community in the near future.

polyolefin catalyst, catalytic performance, molecular descriptor, machine learning, quantitative structure-property relationship

doi: [10.1360/TB-2021-1354](https://doi.org/10.1360/TB-2021-1354)