

· 管理纵横 ·

化工基础数据获取新范式：机制+数据驱动^{*}

吉远辉¹ 朱家华² 穆立文² 陆小华^{2**} 周晨³ 张国俊^{3**}

1. 东南大学 化学化工学院, 南京 211189
2. 南京工业大学 材料化学工程国家重点实验室/化工学院, 南京 211816
3. 国家自然科学基金委员会 化学科学部, 北京 100085

[摘要] 基于国家自然科学基金委员会化学科学部科技活动“化工基础数据获取新范式战略研讨会”的研讨成果, 聚焦于化工基础数据获取, 面向人工智能赋能现代化工基础研究新范式的目标, 本文凝练了化工基础理论、模型化研究与人工智能方法深度融合实现化工基础数据精确获取、助力产品设计以及过程与工艺设计的关键科学问题。同时, 总结了近年来化工基础数据智能获取、化学与化工研究方法与人工智能相耦合在数据精确扩增、化工机制研究、智能设计等方面的研究进展, 并对化工基础数据获取新范式提出建议, 有望助力我国化工智能化发展的理论创新和技术突破。

[关键词] 化工智能化; 化工基础数据; 理论模型; 产品与过程设计; 人工智能算法

化工基础数据是化学工业过程的“基因”, 是化工定量研究不可缺失的重要根基。化工基础数据的获取经历了两代技术方法, 第一代方法以实验测定为主, 耗时、费力、投入大; 第二代方法以热力学模型预测为主, 使用少量实验数据建立理论模型, 将化学工程基础数据获取、产品设计、过程与工艺设计由爱迪生式的反复试验带入定量预测时代^[1]。然而, 对比美国国家标准与技术研究院(National Institute of Standards and Technology, U. S., NIST)、Aspen Tech(Advanced System for Process Engineering)等先进机构和企业, 我国的化工基础数据积累相对薄弱、缺乏规范有效的数据标准, 导致化工工业软件研发和化工生产因缺少底层基础数据的支撑而难以突破国际技术封锁。

面向化工基础数据精准获取的迫切需求, 在国家自然科学基金委员会化学科学部科技活动专项的支持下, 由南京工业大学主办、东南大学协办召开了“化工基础数据获取新范式战略研讨会”(2023年11月20—22日)。该研讨会聚焦人工智能背景下新一代化工基础数据获取新范式, 围绕“理论计算与人工智能融合的未来发展”“限域传递与限域反应中的化



张国俊 博士, 研究员, 国家自然科学基金委员会化学科学部化学五处处长兼化学工程项目主任。自2012年起在国家自然科学基金委员会化学科学部工作, 历任工业化学项目主任、能源化学项目主任、化学工程项目主任。



学奖一等奖等。

陆小华 南京工业大学教授, 材料化学工程国家重点实验室学术委员会副主任。国家杰出青年科学基金获得者, 中国化学学会首批会士, 中国化工学会化工大数据与智能设计专委会主任委员。主持973计划、国家自然科学基金重点项目和重大研究计划集成项目等, 发表论文500余篇。获国家技术发明奖二等奖、教育部自然科学奖一等奖等。



球化工杰出女科学家奖等荣誉。

吉远辉 东南大学教授。国家级青年人才, 德国洪堡学者。主要从事化工热力学及医药化工相关研究, 主持国家自然科学基金4项, 发表研究论文100余篇, 授权国家发明专利9项。任中国化工学会理事会理事、中国化工学会化工大数据与智能设计专委会秘书长、*Chinese Journal of Chemical Engineering* 期刊编委等, 获全

收稿日期: 2024-04-12; 修回日期: 2024-07-18

* 本文根据国家自然科学基金委化学科学部科技活动“化工基础数据获取新范式战略研讨会”讨论的内容整理。

** 通信作者, Email: xhlu@njtech.edu.cn; zhanggj@nscf.gov.cn

工基础数据”和“低碳目标下产品及工艺智能设计”三个议题展开深入研讨,凝练了化工基础理论、模型化研究与人工智能方法深度融合的研究方向,提出了“机制+数据驱动”的化工基础数据获取新范式,以实现化工基础数据的精确获取、评估和分析,突破因基础数据缺失而严重阻碍技术进步和发展的种种挑战。进一步总结了亟需解决的关键科学问题,为学科未来发展提供了新的战略方向和框架思路。

1 研究基础和科学意义

现代化学工业的发展已经从传统石油化工所涉及的有机小分子,不断向复杂流体(如蛋白、自组装聚集体、金属团簇、离子液体等)^[2-4]、复杂材料(如分子筛、功能高分子膜材料、复合材料、生物医用材料等)^[5-7]及复杂过程(如膜分离、多相催化反应、结晶等)^[8-10]等拓展。复杂系统涉及微纳结构、限域、分子间非对称相互作用,已有的热力学模型和传统“三传一反”理论无法定量描述复杂纳微结构以及流体与界面之间的复杂相互作用,复杂系统的化工基础数据难以通过模型精准获取,从而导致现代化工产品和工艺设计严重依赖于实验试错,从实验室小试到工业化进程采用“逐级放大法”,费时耗力,严重阻碍了化工新材料、新技术的迭代更新速度,远不能满足国家安全战略目标下新能源、新材料、新技术等重大需求。

近年来,人工智能技术的飞速发展,有望为化工基础数据的精确获取提供有效途径^[11, 12]。然而,由于现代化学工程具有很强的非线性、非平衡、多尺度等特征,仅仅简单应用人工智能方法无法实现化工基础数据的精准预测。因此,亟需发展化工基础数据的第三代获取方法,通过将数据驱动的人工智能技术和量化计算、分子模拟、热力学模型、介科学等理论方法相结合,建立“机制+数据”双驱动的化工基础数据预测模型,实现化工小数据到大数据的扩展及模型创新,从而加速推进现代化工研发及智能化发展。

2 科学目标与关键科学问题

2.1 化工基础数据获取新范式的科学目标

人工智能技术的飞速发展,引发了对新一轮工业革命的深刻思考。形成化工基础大数据获取新范式,将极大推动下一代化学工业的发展,为产业变革带来深远的影响。本文从以下三个方面具体阐述:

(1) 现代化工基础数据的精准获取、分类、甄别

及数据库建立。基于新一代化工基础理论如介科学、分子化学工程、限域传递和反应等新机制,创新“理实交融”方案,通过实验数据助力理论模型创新发展,理论模型助力实现化工小数据到大数据的扩展、分类和甄别,为化工产品和过程的开发、设计、过程集成与控制提供基础和支撑;构建基于化工过程、实验研究、文献数据等多维度、全链条、面向未来发展的化工标准数据库;制定系统、科学的方法以推动数据获取、数据存储和数据共享,加强对数据隐私的保护,为化工智能化发展提供数据基础。

(2) 人工智能助力下的化工多尺度、多机制融合的模式构建。深入理解和创新发展人工智能模型运行的内在规律和自我更新机制,基于量化计算、分子模拟和化工模型化方法,将物理、化学、化工的原理、机制和模型融入人工智能算法,针对现代化工复杂系统在多尺度、非常规条件下机制难以明晰的难题,应用并发展人工智能算法从数据中挖掘知识,完善已有机制,提出新假设,揭示新的科学规律,为化工智能化发展提供感知和决策的模型基础。

(3) 低碳目标下产品及工艺智能设计。面向现代化工研究系统的复杂性和全球低碳经济的发展需求,针对现有人工智能技术在解决多变量、多尺度、多目标的共性复杂问题时所面临的挑战,开发适用于现代化工的人工智能算法和软件平台基础,实现产品的绿色设计与工艺流程的低碳化重构。具体而言,发展人工智能算法分析大量实验和过程数据,优化化学反应路径,减少能耗和排放,高效设计出更加环保的化工产品。同时,发展人工智能技术对生产过程进行实时监控和动态优化,确保工艺流程的最优运行,最终实现低碳、节能和可持续发展的目标。

2.2 化工基础数据获取新范式的关键科学问题

面向国家安全、双碳等重大战略需求,发展“机制+数据驱动”的化工基础数据获取新范式,需要重点解决以下三个关键科学问题:(1) 数据和机制双驱动的化工大数据获取新方法:如何发展数据与机制的双驱动研究新范式,实现化工小数据到大数据的扩展及大数据的分类和甄别,为化工智能化提供感知和决策的数据基础。(2) 基于人工智能的化工多尺度新机制:如何应用并发展人工智能算法,从数据中挖掘知识,完善已有机制,提出新假设,揭示新的科学规律,为化工智能化提供感知和决策的模型基础。(3) 多变量、多尺度、多目标约束下决策的科学机制:如何开发适用于现代化工产品设计和过程设计与优化的人工智能算法,挖掘多变量、多尺度、

多目标约束下,精准提供可靠数据的科学决策机制,为化工智能化提供感知和决策的算法和软件平台基础。

3 化工基础数据智能化获取的研究进展

NIST^[13]在2016年发布的大数据定义中指出:“大数据的特征包含四个方面:大规模、多样性、动态特征和可变性”。现代化工基础数据完全符合大数据的基本特征,是最为经典的大数据范例,蕴含着丰富的内涵和巨大的挖掘潜力。首先,化工基础数据呈现出庞大的数据量,涵盖了众多不同的化学物质及其在各种反应条件下的性质和行为。其次,化工基础数据涉及的范围广泛,不仅包括了物性数据,还拓展到了分子参数、热力学相平衡和反应平衡、动力学、设备设计参数等多个层面。最后,化工是动态过程,不同条件、原料组合和环境参数都会对基础数据产生影响,基础数据会随时间和空间而动态变化。因此,现代化工基础数据的变革不仅仅是对数据获取方式和范畴的拓展,更是对数据质量和应用效果的提升。

3.1 现代化工推动的基础数据变革进展(先进技术→化工数据)

近年来,现代化工在绿色化、高端化、智能化的背景下,不断向复杂体系延伸,基础数据的获取方式和基本内涵发生重大变革。随着实验技术和测量设备的不断进步,化工数据的实验测定方式更加多元化。现代化学分析技术、光谱技术、核磁共振等技术的广泛应用,使得基础数据的实验获取变得更为高效和全面。同时,传感器技术的发展使得在生产环境中实时监测生产数据成为可能,从而实现了动态过程数据的精准监控与获取。在此基础上,发展基于传感器技术的数字孪生系统,实现虚实交互,不仅可实现对动态过程的精准监控,还可探索化学反应和分离过程,及时发现工艺流程的潜在问题。此外,大科学装置表征技术如同步辐射、正电子湮灭技术和中子衍射技术的开发,极大地推动了现代化工和材料科学的发展。它们提供了高分辨率、灵敏和无损的手段,能够深入揭示化工材料和过程的微观、介观结构和动态行为。不仅推动了新材料和新工艺的发展,也为提高工业生产效率和降低成本和减少环境污染提供了有力支持。随着计算能力的提升和数学算法的发展,基于模型预测获取基础数据已经成为一种重要手段。通过结合化工理论模型和少量实验数据验证,可以实现一些复杂体系基础数据的预

测。此外,量子化学、分子动力学模拟、人工智能等方法的应用,为在特定条件下缺乏实验数据的体系提供了可行的数据预测方案。如在金属有机骨架化合物(Metal organic Framework, MOF)材料领域,由分子动力学模拟预测的 MOF 材料吸附气体能力已经被界内认可为一种可靠且稳定的数据获取方式,被大量用于构效关系的研究。此外,机器学习的发展为计算化学提供了新的工具^[14-16],使得化工复杂系统基础数据的高质量、高精度获取成为可能,为加速高端化学品和新材料的设计应用奠定了基础。

同时,随着现代化工不断向纳微尺度延伸,对数据的需求更加多样化和复杂化。针对复杂结构系统,如微化工体系、限域传递与反应过程等,相关物性数据库以及机器学习辅助建模的研究较少。不同物质和纳微结构对数据影响巨大,甚至可造成数据数量级的变化,此外,许多较大的数据集是企业或出版商的专有数据,未能公开或免费获得,这使得获取精准的基础数据仍面临挑战。

3.2 机制和数据驱动的化工基础数据获取范式进展(机器学习+模型→拓展化工数据)

自1956年人工智能诞生以来,其发展历程历经专家系统、统计机器学习、深度学习等阶段。伴随着大数据、云计算、物联网、AlphaGo、AlphaFold2、ChatGPT大语言模型的发展,人工智能迎来新一轮浪潮^[17-19]。数据驱动为现代化工机制的发展和基础数据的获取带来新机遇,然而纯粹的数据驱动模式对于化工学科研究并不完全适用^[20]。一方面,化工过程不会轻易产生TB级的数据,缺少构建纯粹数据驱动大模型的基础支撑;另一方面,第一性原理知识可以用来减少对大量数据的依赖。基于第一性原理模型的底层物理内涵,人工智能在理论化学计算领域建树颇丰^[21]。如通过结合人工智能与第一性原理计算构建高精度的分子力场^[22];或者将人工智能和量子化学计算相结合,发现催化领域的全新描述符^[23],为计算化学提供了新的理论工具。同时,Aspen Plus研究团队^[24]于2021年推出科学指导下的人工智能混合模型,通过人工智能和理论模型的相互补充,为机制与大数据混合模型下的工业过程进一步放大提供新的参考模式,但在完善机制、发现物理规律等方面仍存在局限性。此外,面向化工装备和过程创新,人工智能方法在反应器的传递、反应动力学等数据获取和开发方面发挥了重要作用。基于实时数据更新机器学习模型,能够动态优化反应器操作参数,提高生产效率和产品质量,对于反应器

结构和操作参数的优化具有重要指导意义^[25]。利用深度学习模型可加速计算流体力学(Computational Fluid Dynamics, CFD)模拟,特别是复杂流场,如湍流和多相流的计算。深度学习模型通过处理反应器内的可视化图像,识别和重建内部流场结构,从而为传热与传质研究提供详细数据^[26]。

对于人工智能时代的化工发展而言,不仅仅需要机制驱动和数据驱动的融合,更需要通过两者的互动和迭代为复杂科学研究提供新的认知。例如,通过耦合分子热力学模型和人工智能算法,发现适用于“药物—高分子辅料”体系的氢键描述符,助力微纳结构定量构效关系的模型构建,从而快速拓展难以测定的多相系统热力学数据^[27, 28]。同时,借助智能化机器学习对物理模型偏差的训练,可以快速寻找未考虑的关键因素,提高理论框架的精度和普适性。在理论模型和科学规律的约束下,化工机制也可以为人工智能算法的训练提供高质量数据集和可解释机理,从而增强人工智能模型的可靠性,一定程度上弥补机器学习模型的物理解释性差的弊端。因此,机制与数据双轮驱动正成为新一代的化工基础数据获取新范式(图 1)。

3.3 化工智能化研究进展[数据+模型→应用(智能决策、智能制造、智能优化)]

新一代智能制造的内涵是“数字化+网络化+智能化”^[29]。化工制造形式经历了手工操作、自动化到智能化的演变。化工智能化变革旨在引领我国化工行业由传统“高能耗、高污染、高投入”的生产模式向“资源节约、环境友好、高附加值、高端化和智能化”的新模式转变。受益于云计算和人工智能等新兴技术的迅猛发展,数字化及智能化成为化工行业发展的趋势。智能化阶段则将数字技术、网络技术

和人工智能进行融合,通过感知、分析、提取“大数据”中隐含的知识,为生产系统提供优化改进方案,使得生产过程更为智能、灵活^[30, 31]。研究报道表明,自动化和机器学习的结合可以借助宏观的信息数据辅助实验进行理性化的评估筛选,充分发挥人工智能在高通量筛选、设计反应、合成及放大等方面的优势,极大地提升科学实验的效率,缩短整个化工技术的研发周期^[32, 33]。

智能化的核心在于对系统物理信息的精准感知和智能决策。在化工智能化的进程中,精准可靠的基础数据显得尤为关键。通过深刻理解化工基础数据变化的本质,并建立科学机制,形成数据驱动和机制驱动相融合范式,实现对生产过程中不确定性的高效控制。如在医药开发领域,有研究^[34]表明 AI(Artificial Intelligence)算法可以分析大量的数据集,从而识别新的药物靶标,并预测候选药物的特性和成功开发的概率,从而加速和优化药物研发的过程。此外,为提升药物制剂的开发效率,AI能够基于大数据模拟产生最优的制剂处方和工艺,并应用迁移学习等技术,从与目标化学系统相近的反应过程中抽取有用的信息和知识,以快速适应新的化学环境,从而避免繁琐的试错实验,同时确保较优的生物利用度和临床疗效^[35]。再以风电润滑系统中轴承故障判断为例,传统的预判手段主要通过传感器检测振动频率,这种方法往往在轴承已经发生严重损坏时才给出警告^[36]。从底层科学原理的角度来看,轴承磨损与润滑油粘度密切相关。然而,在轴承的实际润滑体系中,涉及高压、混合物、纳微界面等复杂因素,导致粘度变化的机制不明确,缺乏必要的粘度物性数据,进而无法为智能感知系统提供精准可靠的工作方程,风电企业面临着巨大的轴承更换成本,因此,发展风电轴承智能预警系统的重要性不言而喻。在电力系统的控制、优化和故障检测方面,近年来研究指出基于机器学习的控制技术有助于发电系统的智能化转型升级,提高系统的可见性、可操作性、灵活性、收益性以及安全性,因此,有望成为传统基于模型方法的替代方案^[37, 38]。

此外,针对化工多变量、多尺度、多目标的特点,发挥人工智能模式识别、数据挖掘和复杂系统建模等方面的优势,将热力学建模和人工智能深度融合、协同创新,实现双模驱动,有望加快化工过程的智能化进程^[39]。然而,人工智能中多种传统算法并不能完全适用于复杂化工过程,如经典的贝叶斯算法在对多尺度体系进行逆向求解时就面临着计算复杂性

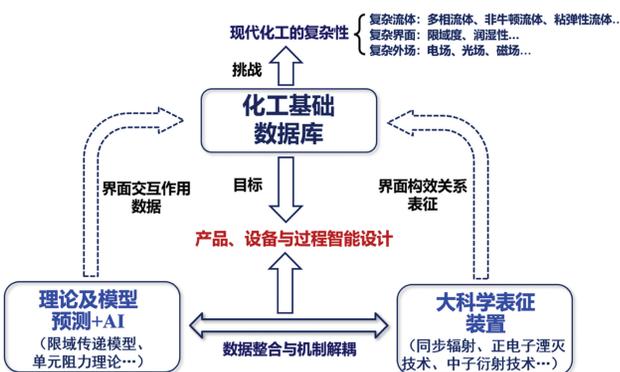


图 1 机制和数据驱动的化工基础数据获取新范式基本框架示意图

和模型适应性方面的挑战^[40]。因此,化工智能化中存在着数据缺失、机制不明、算法不足等问题,在知识融合与多尺度模型耦合方面尤其欠缺,需要在这些领域进行深入研究。

4 结论和发展建议

在化工智能化发展过程中,高质量大数据、化工多尺度大模型以及人工智能新算法和软件是关键推动因素。针对现代化工的复杂性给化工智能化发展带来的挑战,亟需发展现代化工基础数据获取新范式以实现大数据—新模型—新软件的目标。与化工数据的传统获取方式相比,由人工智能助力的化工基础数据获取有望拓展化工数据和模型的边界,极大扩增数据规模,为化工复杂机制的深入剖析提供了重要基础。围绕前述关键科学问题,建议未来在以下四个方面开展重点研究和布局。

4.1 数据和机制双驱动的化工大数据获取和验证

(1) 现代化工复杂系统中的关键机制。基于第一性原理、分子模拟、热力学模型(含基于热力学到动力学性质预测的拓展模型)和人工智能的深度融合实现高精度化工小数据到大数据的扩展,并重构精准预测物性、相平衡、反应平衡和动力学性质的模型,明辨现代化工复杂系统中的关键机制。

(2) 算法与理论模型的升级迭代。通过人工智能算法和化工理论模型的相互迭代,不断完善、改进机制,使模型能够适用于真实化工场景,更好地反映实际工业应用中的复杂现象和复杂规律,实现对化工过程静态和动态数据的分类与甄别。

(3) 拓展现代化工所需关键数据。通过化工机制模型与人工智能算法的深度融合,研究从相对容易获取的数据拓展到含介观结构复杂系统、非常规条件(如高粘、外场、限域过程等)下数据精准获取的新方法。此外,发展高通量筛选和自动化实验设备快速获取大量实验数据,通过实验、模拟和模型等多种方法的紧密互动与融合来评估和验证数据的一致性和准确性。开发和应用高精度、高灵敏度的传感器实时监控化工过程,实时捕捉关键参数,获取高质量的实时数据,从而为现代化工产品和过程设计及智能化发展提供数据基础。

4.2 基于人工智能的化工复杂机制研究

(1) 关键描述符的获取和模型构建。深入研究现代化工复杂系统(微纳结构、限域环境等导致分子间非对称相互作用)及非常规条件下系统的复杂相互作用机制,发展人工智能算法筛选基于机制的化

工关键描述符,建立更普适的化工多尺度模型。

(2) 基于人工智能算法的化工新机制研究。融合化工底层理论和人工智能算法,对已有机制进行优化,提出新的假设,发现新的物理规律。

4.3 多变量、多尺度、多目标的智能优化

(1) 经典贝叶斯算法向多尺度问题的拓展。针对经典的贝叶斯算法在解决多变量、多尺度、多目标的化工复杂系统优化问题时存在的挑战,发展适用于现代化工复杂性特点的人工智能算法。包括但不限于对多尺度信息的整合、对高维度数据的解析处理、自适应和持续学习机制以及多尺度优化逆向求解算法的创新等。

(2) 化工大数据分类、甄别及数据库的构建。基于人工智能创新算法和多尺度机制,实现化工大数据的分类、甄别及数据库构建,并实现高端化学品更快速、更精准地理性设计、工艺变革和生产放大,推进化工过程向高端化、绿色化、智能化方向发展。

4.4 基于化工大数据和模型的流程再造

(1) 流程再造所需化工大数据的获取。分析、提取、甄别数据源的有效信息,构筑化工产品性质、性能和流程再造过程中传感数据的关联;基于大数据和模型,通过对多传感器信息的融合处理,完成分类、识别、判断、决策、规划等任务,实现流程再造的智能感知。

(2) 流程再造所需多目标优化及智能化设计。面向绿氢绿电等国家战略需求,综合考虑产品性能、技术经济性、可持续性等多目标,通过对传统能源体系下化工流程与化工新技术、绿氢绿电等新过程耦合的系统分析,推进新能源目标下化工工艺和流程再造工艺的智能设计和产品的智能化制备。

人工智能与化工学科相互交叉、相互融合,共同发展。通过人工智能算法和化工理论模型的相互迭代,不断完善、改进机制,使模型逐渐趋近真实化工场景,更好地揭示实际工业应用中的复杂现象和复杂规律。通过人工智能方法获取化工关键描述符,结合新机制建立更普适的理论模型,有望加速推进现代化工的智能制造。此外,以化工热力学模型和“三传一反”理论为基础,深入理解化工系统和过程的特性、科学规律,开发更加适配、高效的人工智能算法,是变革人工智能算法以更好解决化工复杂性问题的重要契机。通过多变量、多尺度、多目标人工智能创新算法的开发,助力化工过程向高端化、绿色化、智能化方向发展。

参 考 文 献

- [1] 郭慕孙. 过程工程. 过程工程学报, 2001, 1(1): 2—7.
- [2] Zhou P, Xing RR, Li Q, et al. Steering phase-separated droplets to control fibrillar network evolution of supramolecular peptide hydrogels. *Matter*, 2023, 6(6): 1945—1963.
- [3] Tan XX, Sun XF, Han BX. Ionic liquid-based electrolytes for CO₂ electroreduction and CO₂ electroorganic transformation. *National Science Review*, 2021, 9(4): nwab022.
- [4] López Ruiz A, Xiao MY, Sathya A, et al. Influence of particle z-potential and experimental procedure on protein corona formation and multicomponent aggregation. *AICHe Journal*, 2023, 69(12): e18237.
- [5] Qiao ZH, Zhao S, Sheng ML, et al. Metal-induced ordered microporous polymers for fabricating large-area gas separation membranes. *Nature Materials*, 2019, 18(2): 163—168.
- [6] Ding L, Wei YY, Li LB, et al. MXene molecular sieving membranes for highly efficient gas separation. *Nature Communications*, 2018, 9(1): 155.
- [7] Wu S, Butt HJ. Near-infrared-sensitive materials based on upconverting nanoparticles. *Advanced Materials*, 2016, 28(6): 1208—1226.
- [8] 吉远辉, 陈俏, 翁靖云. 聚合物辅料对阿司匹林结晶动力学影响机制的非平衡热力学建模及预测. *化工学报*, 2021, 72(1): 508—520.
- [9] Zhang N, Yang XW, Wang YL, et al. A review on oil/water emulsion separation membrane material. *Journal of Environmental Chemical Engineering*, 2022, 10(2): 107257.
- [10] Liu LC, Corma A. Bimetallic sites for catalysis: from binuclear metal sites to bimetallic nanoclusters and nanoparticles. *Chemical Reviews*, 2023, 123(8): 4855—4933.
- [11] 钱锋, 钟伟民, 杜文莉. 流程工业智能优化制造的基础理论与关键技术. *工程*, 2017, 3(2): 154—160.
- [12] 张锁江. 化学工程发展战略: 高端化、绿色化、智能化. 北京: 化学工业出版社, 2023.
- [13] Change WC, Nancy G. NIST big data interoperability framework: Volume 1, Definitions, (2024-07-21)/[2024-07-22]. <https://doi.org/10.6028/NIST.SP.1500-1r2>.
- [14] 商城, 康沛林, 刘智攀. 基于机器学习势函数的原子模拟软件的开发及应用. *硅酸盐学报*, 2023, 51(2): 476—487.
- [15] Xia RX, Kais S. Quantum machine learning for electronic structure calculations. *Nature Communications*, 2018, 9(1): 4195.
- [16] Kocer E, Ko TW, Behler J. Neural network potentials: a concise overview of methods. *Annual Review of Physical Chemistry*, 2022, 73: 163—186.
- [17] Wang HC, Fu TF, Du YQ, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 2023, 620(7972): 47—60.
- [18] Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications*, 2022, 13(1): 1265.
- [19] Butler KT, Davies DW, Cartwright H, et al. Machine learning for molecular and materials science. *Nature*, 2018, 559(7715): 547—555.
- [20] Venkatasubramanian V. The promise of artificial intelligence in chemical engineering: is it here, finally? *AICHe Journal*, 2019, 65(2): 466—478.
- [21] Rajulapati L, Chinta S, Shyamala B, et al. Integration of machine learning and first principles models. *AICHe Journal*, 2022, 68(6): e17715.
- [22] Unke OT, Chmiela S, Sauceda HE, et al. Machine learning force fields. *Chemical Reviews*, 2021, 121(16): 10142—10186.
- [23] Mou LH, Han TT, Smith PES, et al. Machine learning descriptors for data-driven catalysis study. *Advanced Science*, 2023, 10(22): 2301020.
- [24] Sharma N, Liu YA. A hybrid science-guided machine learning approach for modeling chemical processes: a review. *AICHe Journal*, 2022, 68(5): e17609.
- [25] 朱礼涛, 欧阳博, 张希宝, 等. 机器学习在多相反应器中的应用进展. *化工进展*, 2021, 40(4): 1699—1714.
- [26] Zhu LT, Tang JX, Luo ZH. Machine learning to assist filtered two-fluid model development for dense gas-particle flows. *AICHe Journal*, 2020, 66(6): e16973.
- [27] Ge K, Ji YH. Novel computational approach by combining machine learning with molecular thermodynamics for predicting drug solubility in solvents. *Industrial & Engineering Chemistry Research*, 2021, 60(25): 9259—9268.
- [28] Ge K, Huang YP, Ji YH. Machine learning with active pharmaceutical ingredient/polymer interaction mechanism: prediction for complex phase behaviors of pharmaceuticals and formulations. *Chinese Journal of Chemical Engineering*, 2024, 66: 263—272.
- [29] Zhou J, Li PG, Zhou YH, et al. Toward new-generation intelligent manufacturing. *Engineering*, 2018, 4(1): 11—20.
- [30] Qian F. Editorial for special issue “artificial intelligence energizes process manufacturing”. *Engineering*, 2021, 7(9): 1193—1194.
- [31] Qian F, Zhong WM, Du WL. Fundamental theories and key technologies for smart and optimal manufacturing in the process industry. *Engineering*, 2017, 3(2): 154—160.

- [32] Zhao HT, Chen W, Huang H, et al. A robotic platform for the synthesis of colloidal nanocrystals. *Nature Synthesis*, 2023, 2: 505–514.
- [33] Xie YF, Feng S, Deng LX, et al. Inverse design of chiral functional films by a robotic AI-guided system. *Nature Communications*, 2023, 14(1): 6177.
- [34] Zhong FS, Wu XL, Yang RR, et al. Drug target inference by mining transcriptional data using a novel graph convolutional network framework. *Protein & Cell*, 2022, 13(4): 281–301.
- [35] Yang YL, Ye Z, Su Y, et al. Deep learning for *in vitro* prediction of pharmaceutical formulations. *Acta Pharmaceutica Sinica B*, 2019, 9(1): 177–185.
- [36] 张琦, 张强, 刘志伟, 等. 大型风力发电机主轴轴承故障分析及预防方法. *机械设计与制造工程*, 2022, 51(8): 127–130.
- [37] Ajagekar A, You FQ. Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems. *Applied Energy*, 2021, 303: 117628.
- [38] Zhao N, You FQ. Toward carbon-neutral electric power systems in the New York State: a novel multi-scale bottom-up optimization framework coupled with machine learning for capacity planning at hourly resolution. *ACS Sustainable Chemistry & Engineering*, 2022, 10(5): 1805–1821.
- [39] Ge K, Ji YH. A thermodynamic approach for predicting thermodynamic phase behaviors of pharmaceuticals in biorelevant media. *Chemical Engineering Science*, 2022, 261: 117973.
- [40] Ning C, You FQ. Data-driven adaptive nested robust optimization: general modeling framework and efficient computational algorithm for decision making under uncertainty. *AIChE Journal*, 2017, 63(9): 3790–3817.

New Paradigm for Determining Fundamental Chemical Engineering Data: A “Mechanism + Data Driven” Strategy

Yuanhui Ji¹ Jiahua Zhu² Liwen Mu² Xiaohua Lu^{2*} Chen Zhou³ Guojun Zhang^{3*}

1. *School of Chemistry and Chemical Engineering, Southeast University, Nanjing 211189*

2. *State Key Laboratory of Materials-Oriented Chemical Engineering, College of Chemical Engineering, Nanjing Tech University, Nanjing 211816*

3. *Department of Chemical Sciences, National Natural Science Foundation of China, Beijing 100085*

Abstract With the support from Department of Chemical Sciences, National Natural Science Foundation of China, “Symposium on New Paradigm for Determining Fundamental Chemical Engineering Data” focuses on the determination of chemical engineering data and presents suggestions addressing key scientific issues encountered in the deep integration of basic chemical engineering fundamental theories and modeling research with artificial intelligence methods. The suggestions aim to achieve accurate determination of chemical engineering fundamental data, thereby assisting in product design, process optimization, and technology development. This article also highlights recent research progress in the intelligent determination of chemical engineering data, focusing on areas such as accurate data augmentation, chemical engineering mechanism research and intelligent design. Furthermore, this article proposes a new paradigm of determining fundamental chemical engineering data, with the expectation of fostering theoretical innovation and technological breakthroughs for the intelligent development of China’s chemical industry.

Keywords intelligent chemical engineering process; fundamental chemical engineering data; modeling; product and process design; artificial intelligence algorithms

(责任编辑 刘敏 张强)

* Corresponding Authors, Email: xhlu@njtech.edu.cn; zhanggj@nsfc.gov.cn