



ISSN 1001-5965
CODEN BHHDE8

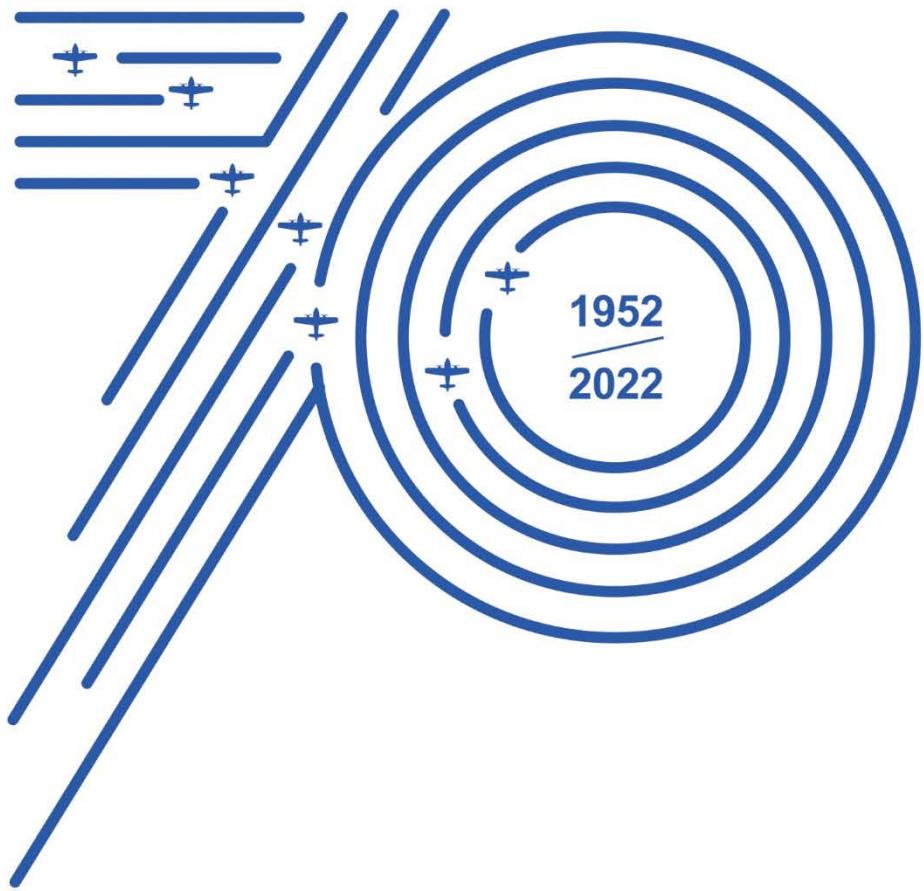
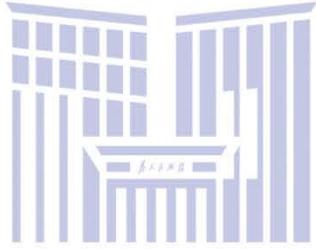
北京航空航天大學 学报

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS



70周年校庆专刊

2022-09
Vol.48 No.9



七十载空天报国
新时代逐梦一流

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn

DOI: 10.13700/j.bh.1001-5965.2022.0318

材料基因工程加速新材料设计与研发

孙志梅^{1,*}, 王冠杰^{1,2}, 张烜广¹, 周健¹

(1. 北京航空航天大学 材料科学与工程学院, 北京 100083;

2. 北京航空航天大学 集成电路科学与工程学院, 北京 100083)

摘要: 在未知材料化学成分和性能关系的情况下, 通过传统的“试错-纠错”方法研发具有特定功能的新材料成本高且经常失败。随着人工智能和数据驱动的第四科学范式的发展, 材料基因工程(MGE)已经成为材料设计与研发的新模式。综述了材料基因工程中高通量计算、材料数据库和人工智能方法的研究进展。介绍了材料高通量计算常用的框架和方法; 阐述了材料数据库在材料数据类型和数据标准两方面的发展现状和有待解决的难题; 总结了人工智能方法在材料关键基础问题中的应用。从高通量可视化计算方法、材料多类型数据库和可视化机器学习框架三方面重点阐述了自主开发的多尺度集成可视化的高通量自动计算和数据管理智能平台 ALKEMIE。展望了材料基因工程未来的发展趋势。

关键词: 材料基因工程(MGE); 高通量计算; 材料数据库; 机器学习; 智算平台

中图分类号: TB39

文献标志码: A

文章编号: 1001-5965(2022)09-1575-14

先进材料是国家工业的支柱, 在传统“试错-纠错”研发模式下, 材料学家基于自身知识储备和认知能力, 通过反复迭代的试错-纠错改进材料性能, 实现新材料的设计与研发。随着新一轮工业革命和互联网时代的到来, 新材料的研发速度严重滞后于材料性能需求速度, 按需逆向设计和精准控制性能已成为新材料设计的必然趋势。20世纪末, 美国兴起组合材料学(combinatorial materials science, CMS)^[1], 通过并行合成和高通量表征技术, 实现了新材料的快速制备和筛选; 21世纪初, 美国和欧洲部分国家提出的集成计算材料工程(integrated computational materials engineering, ICME)^[2]将不同时间尺度和空间尺度的多种材料模拟方法相结合, 在新材料设计领域取得了突破。如今, 随着计算机和人工智能技术的飞速

发展, 材料基因工程(materials genome engineering, MGE)被视为实现材料科学技术飞跃和新材料高效研发与设计的基础, 是新材料研发的加速器。

材料基因工程是受人类基因组计划(human genome project, HGP)的启发而建立的。在生物学中, 基因是一组编码信息, 被视为生物体生长和发育的蓝图, 而在材料领域, 基因可被看作决定其宏观性能的微观特征单元。作为基于数据驱动的科学发展第四范式, 材料基因工程将高通量计算和设计、高通量制备、高通量表征、材料数据库和人工智能相结合, 大大缩短了材料研发周期、降低了研发成本, 从而快速研发出满足日益增长的性能需求的新材料。

2002年, 美国宾夕法尼亚州立大学 Liu^[3-4]首次提出材料基因的概念, 并于2005年设立材料基

收稿日期: 2022-05-06; 录用日期: 2022-05-18; 网络出版时间: 2022-06-01 15:07

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220531.1837.003.html

基金项目: 国家自然科学基金(51872017); 中国博士后科学基金(2022TQ0019); 北航高性能计算平台

*通信作者: E-mail: zmsun@buaa.edu.cn

引用格式: 孙志梅, 王冠杰, 张烜广, 等. 材料基因工程加速新材料设计与研发[J]. 北京航空航天大学学报, 2022, 48(9): 1575-1588. SUN Z M, WANG G J, ZHANG X G, et al. Novel material design and development accelerated by materials genome engineering[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(9): 1575-1588 (in Chinese).

因基金会。2011 年,美国正式发布提升美国全球竞争力的材料基因组计划 (materials genome initiative, MGI)^[5],确立了面向未来的集成计算、实验和数据库的材料研发新模式。MGI 构建了包含上百万条先进能源材料的计算模拟结果材料数据库,并提出了计算机辅助材料研发、模块化的材料模拟体系、开放式的材料高性能数据库以及多尺度计算融合等研究方向。欧盟也相继提出了“新材料发现 NOMAD”计划、德国推出了工业 4.0 战略、俄罗斯推出“2030 年前材料与技术发展战略”、中国提出了“材料基因工程”等一系列政策将新材料探索和材料创新设计与研发作为首要发展目标。

本文首先介绍了国内外材料基因工程领域常用的高通量计算模拟软件和框架。其次,从材料数据来源、多类型数据库和数据标准方面介绍了目前常用的材料数据库。然后,总结了机器学习方法在材料学中的热点应用,重点介绍了笔者团队自主开发的多尺度集成可视化的高通量自动计算和数据管理智能平台 ALKEMIE 研究进展。最后,总结提出了材料基因工程未来的发展方向。

1 材料高通量计算方法

科学发展经历了如图 1 所示的 4 个过程:从文艺复兴时期实验主导的经验范式、以经典力学和热力学为主导的理论模型范式、基于分子动力学和密度泛函理论的计算科学范式,到如今的大数据驱动的科学研究范式^[6]。近年来,美国、欧

洲、日本和中国等国家科研人员在数据驱动研发模式的推动下,开发了一系列用于材料计算基础设施的高通量计算框架和“即插即用”功能完善的高通量计算软件,如表 1 所示。2011 年,在美国国家科学技术委员会、能源部和教育部支持下,加利福尼亚大学伯克利分校劳伦斯伯克利国家实验室 Jain 等^[7]主导开展了材料基因组项目 (materials project, MP),该项目开发了 4 款分别用于材料建模、材料计算模拟、自动纠错和服务器部署的高通量计算分析软件。其中,FireWorks 用于构建材料高通量计算模拟框架,解决高通量材料计算过程中多任务间的依赖关系和任务间的参数及数据传递;Custodian 用于高通量计算过程中的自动纠错;Pymatgen 通过抽象的 Python 对象解析材料结构对称性,自动分析不同尺度计算模拟软件的输入参数和结果,工作原理如图 2 所示;Atomate 则实现了完整的高通量流程及数据存储和服务器的配置,如图 3 中能带计算工作流所示。该项目以其独特的命令行操作方式为高通量计算奠定了软件基础。

2012 年,同属美国材料基因项目的杜克大学 Curtarolo 等^[8]基于 Python2 开发了适用于第一性原理计算的高通量计算软件 AFLOW-π。该软件针对高通量第一性原理计算,集成了数据实时反馈、错误控制、数据管理和归档等功能,可用于实现能带结构、态密度、声子色散、弹性特性、复介电常数、扩散系数等高通量计算,并针对性地优化了紧束缚的哈密顿量 (tight-binding Hamiltonians, TBH) 计算和数据分析流程。

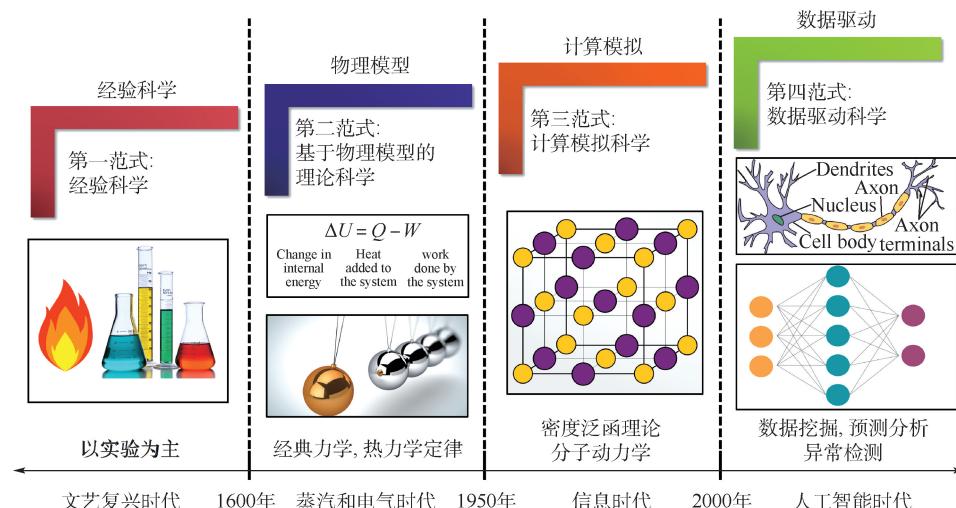


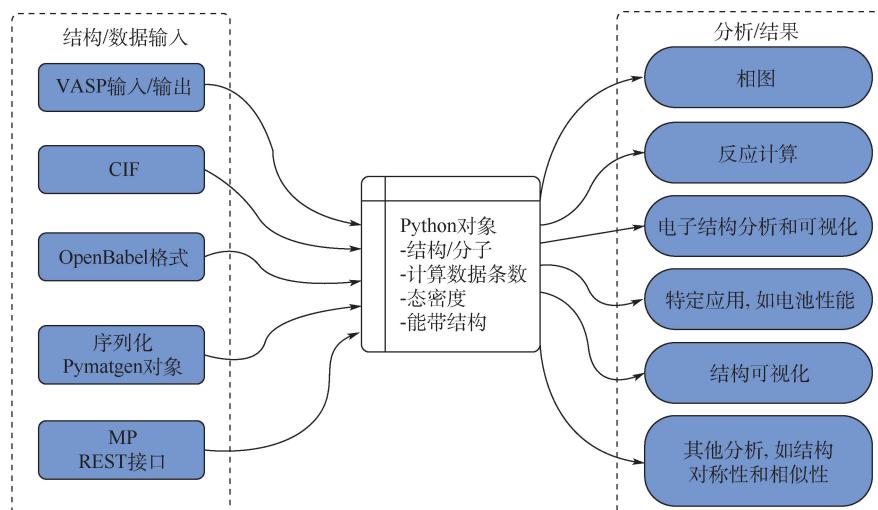
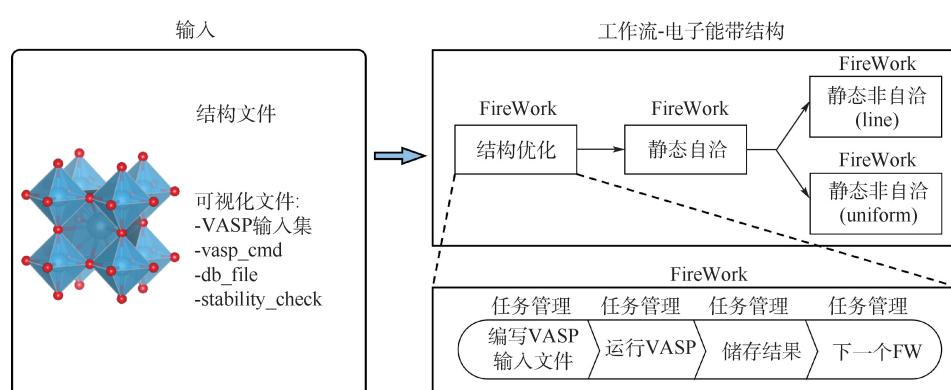
图 1 科学发展的四个范式^[6]

Fig. 1 Four paradigms of science^[6]

表 1 材料高通计算软件和框架发展现状

Table 1 Software and frameworks of material high-throughput calculation

名称	说明	网址
ALKEMIE	高通量计算框架和软件	https://alkemine.org/
Pymatgen	高通量计算框架	http://pymatgen.org/
FireWorks	高通量计算框架	https://materialsproject.github.io/fireworks/
Atomate	高通量计算框架和软件	https://atomate.org/
AFLOW-π	高通量计算软件	http://aflowlib.org/scr/aflowpi/
ASE	高通量计算框架和软件	http://wiki.fysik.dtu.dk/ase/
AiiDA	高通量计算框架和软件	http://aiida.net/
Abipy	高通量计算软件	https://github.com/abinit/abipy
MPInterfaces	高通量计算框架	http://henniggroup.github.io/MPInterfaces/
Imeall	高通量计算软件	https://github.com/Montmorency/imeall
MedeA	高通量计算软件	https://www.materialsdesign.com/
IprPy	高通量计算框架	https://www.ctcms.nist.gov/potentials/iprPy/
Pylada	高通量计算框架	http://pylada.github.io/pylada
MIP	高通量计算软件	http://www.mip3d.org/
MatCloud	高通量计算软件	http://matcloud.cnic.cn/
JAMIP	高通量计算框架	http://jamip-code.com
MatAi	高通量计算框架	https://www.mat.ai/
中国材料基因工程高通量计算平台	高通量计算框架	http://mathc.nscc-tj.cn/

图 2 Pymatgen 高通量软件工作原理^[7]Fig. 2 Principle of Pymatgen high-throughput software^[7]图 3 Atomate 中高通量能带计算流程^[7]Fig. 3 High-throughput computational workflow of band structure in Atomate^[7]

丹麦科技大学 Larsen 团队^[9]开发了材料批量化计算平台原子模拟环境 (atom simulation environment, ASE), 该软件由于缺少工作流程、计算参数和结果的自动纠错功能, 并非完整意义上的高通量计算。而随着版本的迭代, 科研人员为其进一步添加了可视化的用户界面、多个材料软件计算器、多种算法的分子动力学计算软件和多种晶体结构优化算法及边界条件, 可以满足不同用户不同功能的计算需求。目前, 该平台包含了 ABINIT、CASTEP、CP2K、VASP 和 LAMMPS 等常用的材料计算模拟软件。

2016 年, 瑞士洛桑联邦理工大学 Pizzi 等^[10]开发了高通量计算引擎 (automated interactive infrastructure and database for computational science, AiiDA), 该软件系统基于自动化、数据库和开源共享理念, 开发了支持数万个材料计算任务并发运行的高通量算法。材料科学家不仅关注计算模拟的输入和输出, 更关注计算模拟过程中的精度及构型的变化是否准确, 因此, 该软件保存了材料高通量计算中的子任务依赖关系, 并自动跟踪记录所有计算和工作流程的输入、输出和中间元数据, 以便在其开放式数据库 Materials Cloud 中查询数据。

基于 MP 发展的材料高通量计算基础框架, 比利时天主教鲁汶大学 Gonze 等^[11]开发了基于多体微扰论的第一性原理高通量计算软件 Abipy; 美国佛罗里达大学 Mathew 等^[12]发展了针对二维材料表面和异质结的高通量计算流程 MPIInterfaces(见图 4); 英国伦敦国王学院 Lambert 等^[13]发展了用于原子晶界的高通量计算框架 Imeall 等。

相比于国外材料基因工程的研究成果, 中国高通量计算起步较晚, 但在 2016 年《中华人民共和国国民经济和社会发展第十三个五年规划纲要》中材料基因国家重点专项支持下, 中国也涌现了多个成熟的材料高通量计算框架和软件。

多尺度集成可视化的高通量自动计算和数据管理智能平台 (artificial learning and knowledge enhanced materials informatics engineering, ALKEMIE) 是由笔者团队 Wang 等^[14-15]基于 Python 开源框架自主开发的中国第一个高通量自动流程可视化计算和数据管理智能平台。该平台从设计出发吸取国外材料基因相关软件的先进理念, 克服了计算过程中可能遇到的兼容性差、接口不统一和功能拓展困难等问题, 开发了包含材料高通量自动计算模拟、材料数据库及数据管理、基于人工

智能和机器学习的材料数据挖掘 3 个核心理念的智能平台。ALKEMIE 平台适用于数据驱动的材料研发, 详细内容见第 4 节。

MIP (materials information platform) 是由上海大学 Yang 等^[16]开发的适用于热电材料高通量筛选的高通量计算软件; MatCloud 是由北京迈高材云科技有限公司 Yang 等^[17]开发的第一性原理计算引擎, 目前支持 VASP 和 ABINIT 等第一性原理高通量计算软件; JAMIP 是由吉林大学 Zhao 等^[18]开发的开源高通量集成软件, 该软件利用人工智能算法在高通量计算的海量材料数据中智能寻求新材料和新原理。中国材料基因工程高通量计算平台 (CNMGE) 则是由国家超级计算天津中心开发的网页版高通量集成计算平台, 该平台可以集成多种不同的高通量计算软件, 包括 ALKEMIE 高通量智算平台、含能材料分子专用高通量筛选系统 EM-Studio 和用于无机骨架材料的晶体结构解析与预测软件 (framework generator, FraGen)。

上述高通量计算软件和框架, 一方面为科研人员提供了快捷方便的自动计算工作流, 可以在高性能超算中实现高效并行计算, 显著提升计算效率; 另一方面避免了传统试错-纠错法中可能出现的人为误差, 使得研究人员有更多的精力关注材料科学问题本身, 而非高通量所涉及的技术难题。近年来, 高通量计算模拟已经在材料构型预测、材料结构稳定性和相稳定性预测、能源材料能量转化效率、最优掺杂元素分级筛选等方面获得了广泛应用。举例来说, Curtarolo 等^[19-20]通过高

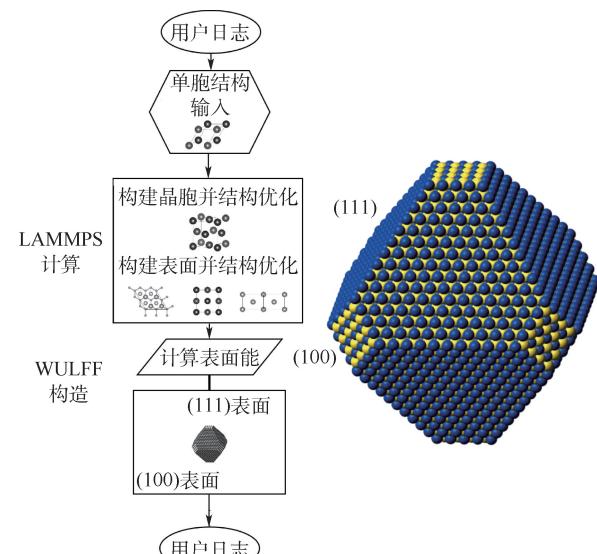


图 4 二维材料表面和异质结的高通量计算软件

MPIInterfaces 工作流程^[12]

Fig. 4 Workflow for 2D material surfaces and heterojunctions in MPIInterfaces software^[12]

通量方法从 435 个含 d 电子的二元金属间化合物中筛选出了 283 个能量稳定化合物,其中 273 个(96.5%)化合物最终获得了实验验证;Hu 等^[21]基于第一原理计算的层级筛选,从 29 个过渡金属元素中筛选出可以提高相变材料 Sb_xTe_{3-x} 性能的最佳掺杂元素 Y、Sc 和 Hg,除了有毒的 Hg 元素,Y 和 Sc 均获得了实验验证;Gan 等^[22]基于第一原理计算的高通量分级筛选,通过结构能量、声子谱、力学稳定性和转换效率等 4 个筛选标准从 21 060 个候选光电材料中筛选出了 78 个稳定化合物,且其中 22 个化合物的性能优于当前太阳能电池材料 GaAs。

2 材料数据库研究进展

高通量计算大幅提高了计算模拟的效率,并产生了海量的数据,这些数据中既包含了有用的材料性质数据,又包含了大量重复的无效数据。而材料学中由于所研究的材料体系、成分、结构等的不同,材料测试、制备工艺和流程也不尽相同。对于不同用途材料,所关注的材料性能和关键指标也有差异,因此,在数据库的构建过程中面临数据存储类型、数据库的兼容性和泛化能力等一系列问题。

本文根据材料数据来源,将数据大致分为材料计算数据、材料测试数据和已发表的文本数据三大类。本节将介绍目前国内常用的材料数据库和数据共享标准。

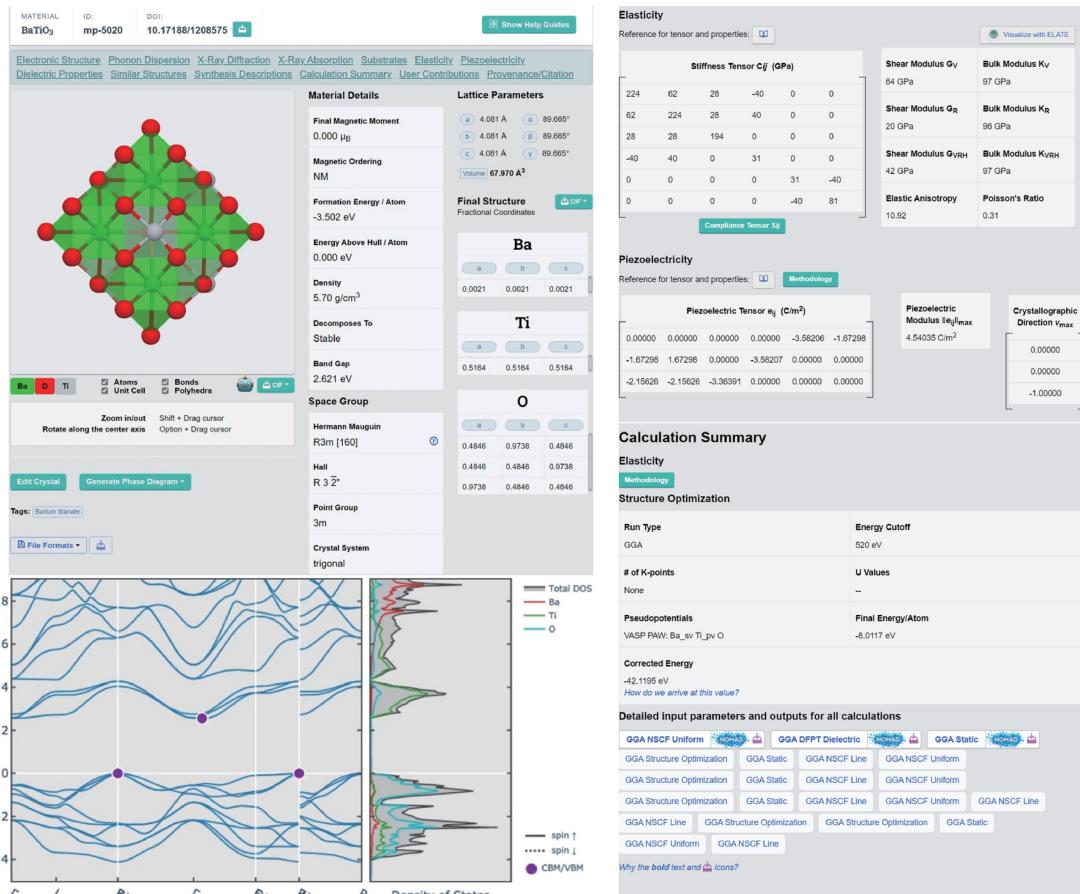
2.1 材料多类型数据库

目前常用的材料数据库如表 2 所示。ICSD 是由莱布尼茨信息基础设施研究所 Belsky 等^[23]构建的无机材料结构数据库,收录了自 1913 年以来在 1 600 个学术期刊发表的超过 8 万篇论文中的共计 21 万条晶体结构数据,覆盖金属、有机物、同素异形体等各种形态的材料体系;COD (crystallography open database) 是由英国剑桥大学 Quirós 等^[24]开发的包含超过 700 万个有机、无机、金属有机化合物和矿物的晶体结构数据库;MP Database 是由美国加利福尼亚大学伯克利分校劳伦斯伯克利国家实验室 Jain 等^[7]建立的材料计算模拟专用数据平台(见图 5),不仅收录了材料结构数据,也收录了元素性质、电子结构、弹性张量和能源转换电极性能等数据;AFLOWLIB (automatic flow lib) 是由美国杜克大学 Curtarolo 等^[19]基于 AFLOW-π 高通量软件开发的材料计算数据库,数据库中收录了 6 400 余条热力学相

表 2 材料多类型数据库的发展

Table 2 Development of multi-type databases of materials

名称	说明	网址
ALKEMIE-DB	通用材料数据库	https://alkemine.net/
ICSD	晶体结构数据库	http://www2.fiz-karlsruhe.de/icsd_home.html
COD	晶体结构数据库	http://crystallography.net
AFLOWLIB	材料计算数据库	http://aflowlib.org/
MP Database	材料计算数据库	http://materialsproject.org/
NOMAD	材料计算数据库	https://repository.nomad-coe.eu/
OQDM	材料计算数据库	http://oqmd.org/
CMR	材料计算数据库	https://cmr.fysik.dtu.dk/
OMDB	电子结构数据库	https://omdb.mathub.io/
JARVIS-DFT	材料势能数据库	https://ctems.nist.gov/-knc6/JVASP.html
C2DB	二维材料数据库	https://cmr.fysik.dtu.dk/c2db/c2db.html
AMSD	矿物材料数据库	http://rruff.geo.arizona.edu/AMS/amcsd.php
ASM	合金材料数据库	https://mio.asminternational.org/ac
CatApp	催化材料数据库	https://suncat.stanford.edu/theory/it-facilities
HSMD	储氢材料数据库	https://hydrogenmaterialssearch.govtools.us
Thermocalc	热力学数据库	https://thermocalc.com/products/databases/
Citrination	通用材料数据库	https://citrination.com
MatNavi	通用材料数据库	https://mits.nims.go.jp/en
NIST	通用材料数据库	https://www.nist.gov/programs-projects
Materials Web	通用材料数据库	https://materialsweb.org/twodmaterials
Materials Cloud	通用材料数据库	https://materialscloud.org/discover
MSDSN	通用材料数据库	http://www.materdata.cn/
Atomly	通用材料数据库	https://atomly.net/
MCDC	腐蚀防护数据库	https://www.corrdata.org.cn/
MGED	材料计算数据库	https://www.mgedata.cn/

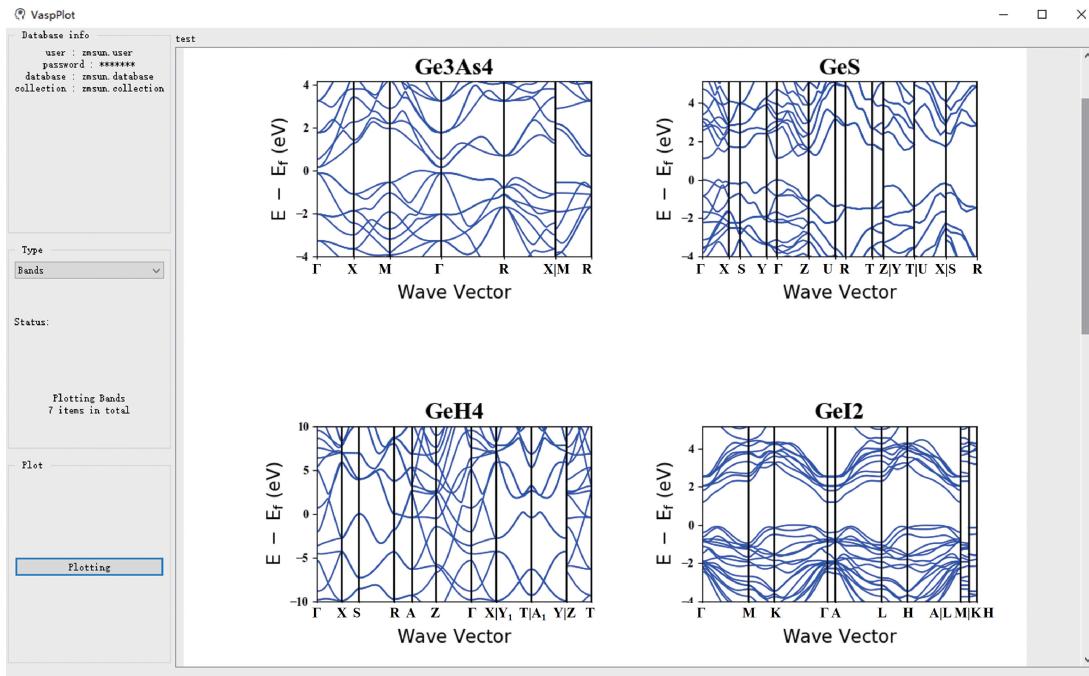
图 5 Materials Project Database 数据库概况^[7]Fig. 5 Snapshot of Materials Project Database^[7]

图数据和超过 45 万个四元混合物的材料性质数据;Materials Cloud 是由瑞士洛桑联邦理工大学 Pizzi 等^[10]开发的第一性原理计算元数据的数据库,包括超过 752 万条第一性原理结构弛豫流程和纳米多孔材料吸收和扩散相关的材料性质数据;OQDM (open quantum database for materials) 是由美国西北大学 Saal 等^[25]开发的第一性原理计算热力学数据库,包含了数万个二元、三元和四元相图;NOMAD(novel materials discovery)是由欧洲马克斯·普朗克学会 Draxl 等^[26]开发的欧洲最大的新材料共享数据库,该数据库包含了 49 TB 的各类材料数据;MatNavi 是由日本国家材料科学研究所 Ogata 等^[27]开发的多种材料数据的集合,包含聚合物数据库(化学结构、加工、物理性质、NMR 光谱数据)、无机材料数据库(晶体结构、相图、物理性质)、金属材料数据库(密度、弹性模量、蠕变性质、疲劳特性)、电子结构计算数据库等。

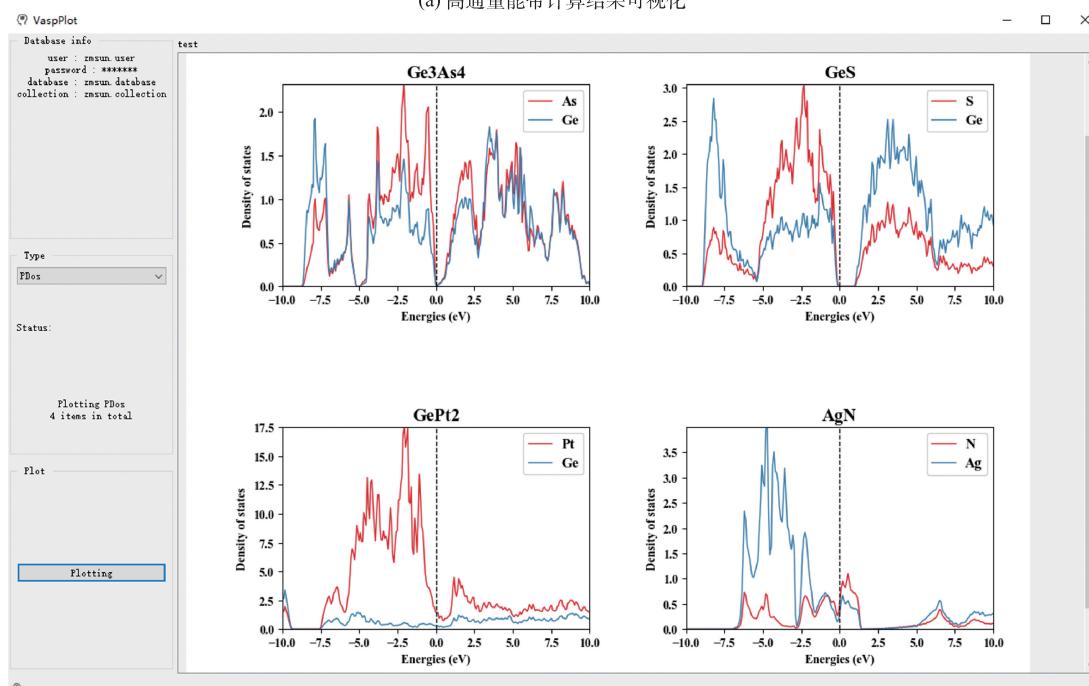
除了上述通用的多类型材料数据库,还有众多针对材料某个特定领域的特色数据库。例如,美国佛罗里达大学开发的材料表面界面数据库 Materials Web;美国国家标准技术研究所

Choudhary 等^[28]开发的赝势数据库 JARVIS-DFT;丹麦科技大学 Larsen 等^[9]开发的二维材料、硒化物和硫化物数据库(computational materials repository, CMR);北欧理论物理研究所 Borysov 等^[29]开发的三维有机晶体材料电子结构和带隙数据库(organic materials database, OMDB);美国斯坦福大学 Hummelshøj 等^[30]开发的催化材料活化能数据库 CatApp;此外,还有 ASM 出版社开发的热处理数据表及应力应变和蠕变曲线材料数据库(ASM alloy center database, ASMDB)、矿物材料数据库(American mineralogist crystal structure database, AMSD)和储氢材料数据库(hydrogen storage materials database, HSMD)等。

中国目前也发展了多个大型材料数据库共享平台。ALKEMIE-DB 是由笔者团队 Wang 等^[14-15]基于高通量智能计算平台 ALKEMIE 开发的多类型材料数据库(见图 6),该数据库分为隐私数据库和共享数据库两大类,根据数据类型进一步细分为含 60 余万组数据的晶体结构数据库、含 1 万余条声子能带的声子谱数据库、含 20 余万组数据的深度学习赝势数据库、高通量计算工作流数据库等。MSDSN 是由国家统筹建设、北京科技大学



(a) 高通量能带计算结果可视化



(b) 高通量态密度计算结果可视化

图 6 ALKEMIE-DB 数据库中高通量能带和态密度计算结果可视化^[14]Fig. 6 Visualization of high-throughput band structures and density of states calculations in ALKEMIE-DB databases^[14]

实施完成的材料科学数据共享平台,主要分为包含有色金属材料和特种合金和微观组织模拟的实验数据库、热力学和动力学相关的计算模拟数据库等。Atomly 是由中国科学院物理研究所松山湖材料实验室和怀柔材料基因平台共建的材料计算数据库,目前包含 20 万条材料结构数据和 5 万条相图数据等。MCDC(national materials corrosion and protection data center)是由国家科技部门建设的腐蚀防护数据平台,包含环境数据、腐蚀数据、

腐蚀检测和腐蚀预测等。

2.2 材料数据标准

随着高通量计算、高通量实验和超级计算机计算能力的发展,材料数据形成的数据海(data ocean)面临着 5 个重要的挑战(“5V”特性):

1) 速率(velocity)。新数据产生速率和旧数据更新速率越来越快,对数据的格式化存储和快速读写提出了更高的要求。

2) 数据量(volume)。材料数据以 TB 量级不

断增加,需要可靠的数据存储、高效的数据检索和开放共享的可重复利用。

3) 多样性(variety)。材料数据的存储形式、材料体系、测试和计算方法,以及数据蕴含的物理意义更加多样化。

4) 真实性(veracity)。数据的不确定性和可靠性决定了数据是否真实有效,进一步决定了数据挖掘和机器学习模型的精度和泛化能力。

5) 低价值密度(value)。数据价值密度高低与数据总量大小成反比,数据量越大,数据价值密度越低。

如何在海量数据中分析预测数据隐藏的真实意义和价值是材料基因工程方法努力探索的主要方向。

为了解决上述问题,欧洲马克斯·普朗克学会 Draxl 等^[26]提出了材料数据库建设的 FAIR Data 准则,即可发现(findable)、可获取(accessible)、可互操作(interoperable)和可再利用(reusable),来提升材料数据的开源共享性。欧洲用于材料设计的开放式数据库集成团队 Andersen 等^[31]提出了材料数据共享标准 OPTIMADE(见图 7),该标准通过 JSON 格式定义了材料数据共享的统一标准,目前大多数材料数据库 MP Database、AFLOWLIB、NOMAD、Materials Cloud 和 ALKEMIE-DB 等均提供了 OPTIMADE 通用接口支持。

中国中关村材料试验技术联盟于 2019 年也提出了材料基因工程数据通则,将材料科学数据分为样品信息、源数据(未经处理的数据)与衍生数据(经分析处理得到的数据)三大类,并从宏观上定义了材料数据的通用性与专用性;北京科技大学材料基因工程北京市重点实验室在 2020 年 MSDSN 数据库中定义了材料科研数据 DOI 编码规则:“10.11961/classification. project. date. sequence”,编码中包含了材料数据分类号、项目支

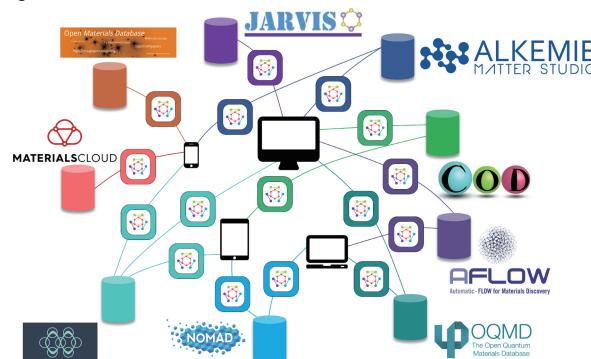


图 7 材料数据共享标准 OPTIMADE 概况^[31]

Fig. 7 Overview of materials data sharing standard of OPTIMADE^[31]

撑信息、注册日期和流水号;北京航空航天大学 ALKEMIE-DB 数据库也发展了用于材料数据共享的唯一标识符:“alkemie. date. classification/user_defined_label. number.”,其中 alkemie 为数据库社区唯一标识,date 代表数据创建日期(精确到 μ s),classification 表示数据类别,user_defined_label 为用户自定义字段,number 为数据唯一索引序号。

3 机器学习在材料设计中的应用

机器学习传统上分为监督学习和非监督学习两大类。监督学习是指给算法一个数据集,对于数据集中的每一个样本,都给出对应的映射(即标签),算法的目的是给出更多的映射,得到更多的答案;而非监督学习常常被用于在大量无标签数据中探索并发现规律。进一步,监督学习分为处理连续值的回归问题和处理离散值的分类问题。目前,常用的机器学习算法已经有很多相关综述,本文不再赘述。随着计算机科学和人工智能的发展,机器学习在材料结构设计、材料性能预测和材料分析图像识别等领域扮演着越来越重要的角色。本节主要介绍机器学习在材料性能预测、材料数据文本挖掘和机器学习原子间作用势。

3.1 材料性能预测

机器学习在材料学中的应用主要包含格式化材料数据、机器学习模型训练和高效材料性能预测 3 个步骤,并在二维、光伏、催化、合金和热电等材料中均获得了显著的研究成果。近年来,瑞士洛桑联邦理工大学 Lin 等^[32]通过高通量筛选和机器学习算法,从 ICSD 和 COD 数据库的 108 423 个三维晶体构型中,通过对称性和几何算法筛选了 1 036 个容易合成和 789 个具有潜在可能性的层状二维材料混合物;东南大学 Lu 等^[33]通过高通量筛选和机器学习描述符,从 5 158 个候选材料中筛选了能量转化效率高且带隙在 0.9~1.6 eV 之间有机无机结合的钙钛矿光伏材料(HOIPS),并构建了可以高效预测带隙值的结构描述符和机器学习模型,进一步将模型拓展到双层 HOIPS 中,成功从 11 370 个混合物中预测了 204 个无毒且稳定的光伏材料^[34];北京科技大学 Zhang 等^[35]在铜合金体系中通过贝叶斯优化迭代方法,分别构建了误差小于 7% 的硬度模型和误差小于 9% 的电导率模型,并通过迭代优化设计了兼具优异力学和电学性能的 Cu-Ni-Co-Si-Mg 合金。

3.2 材料数据文本挖掘

2018 年,瑞士联邦理工学院 Villars 等^[36]通

过计算机视觉和自然语言方法从已经正式发表的论文中自动识别有效材料结构和数据,并通过数据挖掘探索数据背后隐藏的物理模型和机理,构建了MPDS(materials platform for data science),近年来,该团队首次从35 000篇论文中解析了15 500个化学成分,并基于其中的2 330个二元体系构建了机器学习分类模型;进一步从超过8 000篇发表的论文中构建了2 800个二元混合物的原子配位环境多面体分析算法(atomic environment types,AETs);随后,从超过50 000篇已发表论文中分析了290 000个原子配位环境数据,将该算法模型拓展到了多元无机混合物中。2019年,Tshitoyan等^[37]发展了Word2vec非监督机器学习模型,成功从330万材料文本中筛选了超过7 000个候选的热电材料。除了材料计算模拟数据,在实验合成方式上,Kononova等^[38]通过文本挖掘和自然语言处理方式从53 538篇科学文献数据中筛选了19 488个无机金属合成方式,包括材料成分、制备条件、化学平衡方程和反应过程,该数据库为实验中无机材料的制备过程提供了有力的数据支持。

3.3 机器学习原子间作用势

材料计算模拟根据模拟时长和体系大小分为原子尺度、分子尺度、介观尺度和宏观尺度模拟,尺度越小模拟精度越高,尺度越大越接近真实体系,但是不同的模拟尺度采用的物理模型和近似原理不同,数据耦合非常困难,而数据驱动的机器学习方法被视为材料多尺度模拟的耦合剂。经典大规模分子动力学常被用来模拟近似真实材料体系的服役性能,但可靠、精确的原子间势函数的匮乏限制了其广泛应用。基于密度泛函原理(DFT)的第一性原理模拟具有精确的赝势库,但求解本征值所需的巨大计算量限制了该方法在大的原子尺度和时间尺度上的模拟,常用的VASP仅限数百原子的体系。因此,简单方便地获得适用于经典分子动力学的可靠势函数非常重要。随着计算机技术、计算机视觉和材料基因理念的快速发展,通过机器学习结合大数据、高通量计算的方法拟合可靠的适用于经典分子动力学模拟的势函数成为了研究热点。

机器学习势函数的发展主要经历了原子个数受限的低维度势函数和泛化能力强的高维度神经网络势函数2个发展过程。1995年,Blank等^[39]开发了第一个基于统计学的势函数模型,用于研究氢原子的分子动力学模拟;2009年,Malshe等^[40]进一步提出了通过神经网络预测经典多体势

方程参数的模型。但是,上述模型均不能改变输入的原子个数,因此限制了机器学习势函数的应用。

2011年,Behler^[41]提出了原子中心对称函数,通过数学方程解析原子局域环境,构建了输入原子个数不受限的高维度神经网络模型。2018年,Gastegger等^[42]发展了权重相关的对称函数(wACSF),通过卷积神经网络提升了模型的精度和实用性,但是由于局域近似,无法包含超过截断半径的原子长程相互作用。2018年,Yao等^[43]提出了包含长程静电作用和散射作用的HDNNP方法,但是该方法并未获得广泛应用,一方面由于物理学中超过6~10 Å的静电作用通常对体系的整体影响较小,另一方面添加长程作用会急剧增加模型训练成本,与其对精度的微小提升相比有待进一步优化。笔者团队Wang等^[14]开发了适用于相变材料Sb单质的跨尺度机器学习势函数PotentialMind,该势函数模型与DFT比较,对能量预测的精度达到98%,平均到每个原子上的能量误差值小于0.045 eV/atom,对力的预测精度达到89%,该算法具有很强的扩展性和通用性,易于扩展到多元材料体系中。

机器学习势函数方法一方面实现了具有第一原理精度且更大原子数体系和更长时间尺度的大规模分子动力学模拟,另一方面通过替代求解复杂多体薛定谔方程本征值,使得模拟速度提升2~3个数量级,目前,该方法已经在多个材料体系中获得了应用。例如,Sosso等^[44]发展了适用于二元相变存储材料GeTe的人工神经网络势函数,并实现了具有第一原理精度的4 096个原子体系的大规模分子动力学模拟,通过模拟相变材料的多个淬火过程(100~300 ps),探究了淬火速度和模拟体系大小对GeTe非晶结构的影响;Artrith和Urban^[45]基于Fortran语言开发了适用于钙钛矿TiO₂的神经网络势函数软件(atomic energy network,AENET),并面向科研人员开源使用,加速了机器学习势函数方法在能源材料中的应用;Mocanu等^[46]通过高斯近似方法构建了三元相变材料Ge₂Sb₂Te₅的势函数模型,实现了含7 200个原子的非晶体系的大规模分子动力学模拟,揭示了模型大小和原子个数对非晶局域结构的影响,并阐明了非晶构型中化学键和晶化过程中的微观结构演化;Zhang等^[47]开发了适用于高性能并行计算的深度神经网络势函数方法DeePMD,实现了模拟体系超过1亿原子、模拟时长超过1 ns的大规模分子动力学模拟,显著加速了新材料设计与研发。

4 多尺度集成可视化的高通量自动计算和数据管理智能平台 ALKEMIE

2016 年,笔者团队在国家重点研发计划材料基因工程专项的支持下,基于 Python 开源框架自主开发了一套多尺度集成可视化的高通量自动计算和数据管理智能平台 ALKEMIE,主要包含高通量自动工作流 ALKEMIE Matter Studio(MS)、数据管理及材料数据库 ALKEMIE Database(DB)、基于机器学习的材料数据挖掘 ALKEMIE Potential Mind(PM)三部分^[14-15]。

4.1 高通量智能计算 ALKEMIE-MS

ALKEMIE 基于 AMDIV 设计理念,解决了材料高通量智能模拟过程中 5 个核心问题:自动化计算(automation)、模块化拓展(modular)、材料数据库(database)、人工智能方法(intelligence)和可视化界面(visualization),可实现从建模、运行到数据分析,全程自动无人工干预。

ALKEMIE 中,多尺度集成的高通量自动计算可以通过不同模块间以搭积木的方式实现自动耦合并完成计算。不同模块的连接方式如图 8 所

示。首先,由建模模块控制输入,通过多种建模方式将材料构型导入高通量预处理器,科学计算模块控制任务的计算顺序和纠错(可进行电子尺度、原子尺度、分子尺度和介观尺度的高通量自动流程)。然后,服务器用来协调计算资源,配置远程节点,实现本地与远程服务器通信并提交任务,数据存储系统负责保存整个流程中所有的元数据,将计算结果保存在不同类型的数据库中。最后,通过数据分析和人工智能进行数据挖掘。

目前,该软件可以实现超过 10^4 量级的高通量并发计算,包含第一性原理计算 VASP 、 QE 和 ABINIT ,分子动力学模拟 LAMMPS 和 ASE ,热力学计算软件 Gibbs ,动态蒙特卡罗模拟(KMC),相场相图模拟 OpenPhase 和 OpenCalphd 等多尺度模拟软件,可通过参数传递的方式实现跨尺度计算。目前,该软件系统已部署在 9 家超算中心,包括 4 家国家超算、 4 家高校超算和 1 家企业超算。软件可移植性和可拓展性强,适用于对材料模拟掌握程度从初级到专业的所有材料研究人员,可提供材料建模、高通量智能计算、数据挖掘和人工智能一体化的新材料设计方案。

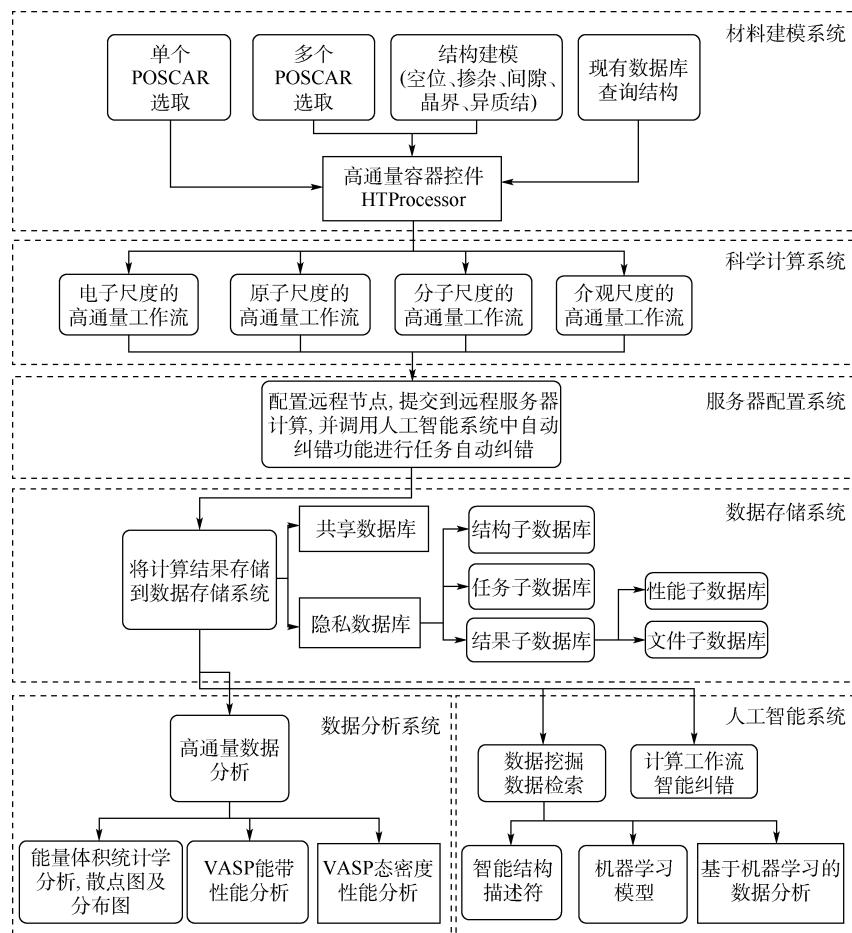


图 8 多尺度集成可视化的高通量自动计算和数据管理智能平台 ALKEMIE 计算模块概况

Fig. 8 Overview of platform with multi-scale integration of visualized automatic high-throughput calculation and intelligent data management ALKEMIE

4.2 材料多类型数据库 ALKEMIE-DB

ALKEMIE-DB 材料多类型数据库分为五大类:材料结构数据库、工作流源数据库、材料性能数据库、机器学习描述符数据库和论文数据库。目前,收录了超过 64 万条材料结构数据、296 条相

图数据、1 万条声子能带数据、1 418 条赝势数据和 20 余万条机器学习描述符数据。通过 ALKEMIE 中 JSON 格式的 API 和国际通用的 OPTIMADE 格式的 API 实现数据高效查询检索,并通过 Finder 控件实现结构的数据可视化,如图 9 所示^[14]。

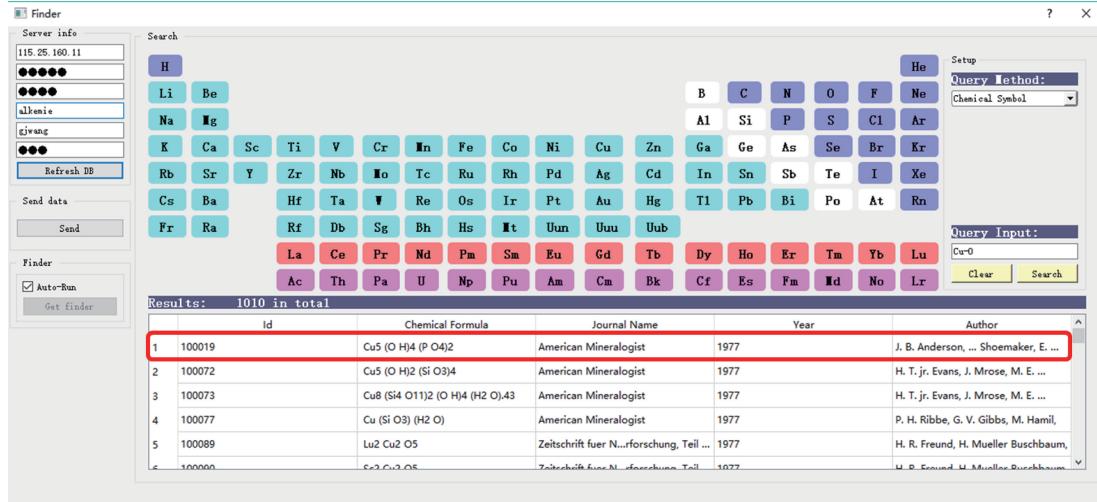


图 9 ALKEMIE-DB 材料结构数据库^[14]

Fig. 9 Material structure database of ALKEMIE-DB^[14]

4.3 材料学机器学习应用框架 ALKEMIE-PM

机器学习在材料学中通常被视为未知的黑盒模型,而材料研究人员相比于机器学习方法更关注材料性能、成分和工艺问题,因此,如何构建简单通用的可视化机器学习框架至关重要。ALKEMIE 通过抽象凝练高级 API 及规范化和格式化机器学习的每个具体步骤开发了一套通用的可视化机器学习的流程。Datasets 模块给定了数据集及特征的输入格式,Model 模块定义了不同的机器学习算法,Evaluate 模块给出了机器学习训练过程的收敛情况及模型在测试集或模型在部署过程中的应用情况,Plotter 模块中多种分析方法提供了将训练过程及其结果可视化的功能,如图 10

所示。

目前,笔者团队基于 ALKEMIE 可视化机器学习框架发展了一系列高效机器学习模型。Chen 等^[48]通过团簇展开法和高通量第一性原理计算搜索了二维过渡金属硫化物的单层和双层无序掺杂结构,得到了稳定掺杂结构和能量之间的对应关系,通过机器学习挖掘了影响半导体-金属转变(SMT)的 2 个关键特征,即范德华间隙内掺杂氧原子的浓度差和 Mo-S/O 键角正切的平均值($\tan \theta$) ; Peng 等^[49]通过线性回归方法构建了适用于 MXene 材料单原子催化剂结构与氧还原反应(ORR)性能构效关系的组分描述符,基于简单的元素性质可以高效快速预测材料的催化活性;

模型泛化能力和预测精度不达标,将其添加到原始数据中重新训练模型

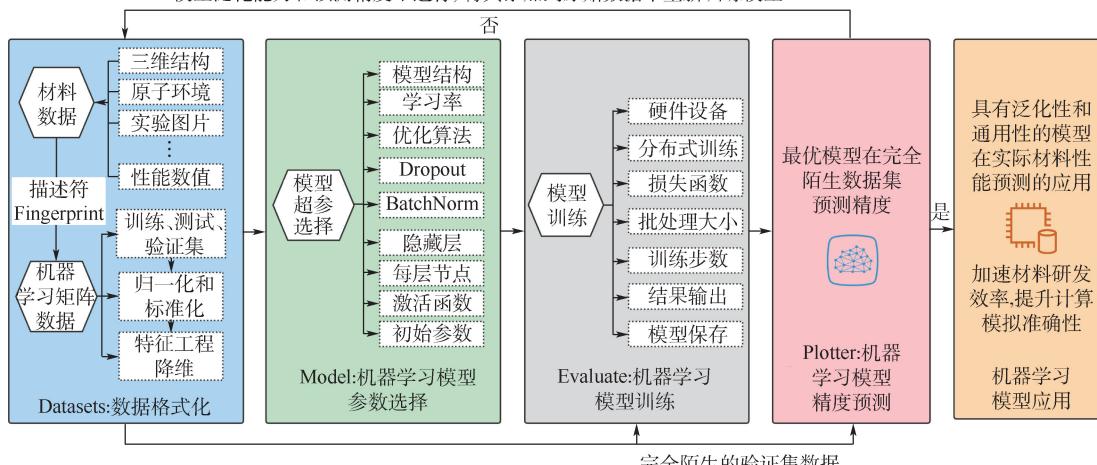


图 10 ALKEMIE-PM 机器学习框架

Fig. 10 Framework of ALKEMIE-PM machine learning

Gan 等^[50]通过高通量筛选和可视化机器学习框架,开发了 2 个精度分别为 90.90% 和 91.67% 的机器学习模型预测层状 IV-VI 族半导体不同温度下的最大热电优值(ZT_{max})和实现 ZT_{max} 所对应的最佳掺杂类型,并成功从 840 个候选成分中筛选出数种具有优良潜力的热电材料。

4.4 未来的发展方向

综上所述,ALKEMIE 已经研发了集可视化高通量自动计算流程、材料多类型数据库和人工智能方法于一体的新材料设计研发智能平台,但是未来仍有亟须发展的新方向和新方法。

在高通量计算方面,开发从原子、分子、介观到器件的跨尺度模拟方法是目前极具挑战且具有广阔应用前景的热点问题。Martin Karplus、Michael Levitt 和 Arieh Warshel 三位科学家在分子领域凭借量子力学和分子动力学跨尺度模拟方法(QM/MM)获得了 2013 年诺贝尔化学奖。而在周期性材料的研究方面,由于体系周期性边界条件和原子局域环境的复杂性,使得跨尺度模拟的精度非常难以控制,发展高通量跨尺度高并发、自动纠错及数据耦合方法,通过机器学习数据挖掘等算法进一步提升跨尺度模拟精度是未来的研究热点之一。

材料数据库方面,应该保持开源和共享的发展理念,基于 FAIR 原则,构建包含材料计算和实验元数据及中间数据的高效数据库,发展数据规模更大、种类更丰富且具有航空特色的共享数据平台,完善更加通用兼容的数据标准和共享标识均是未来重要的研究方向。

在机器学习领域,材料学中数据集的构建非常困难,因此,研发基于小数据集的高效机器学习模型训练算法至关重要;由于机器学习模型的黑盒特性,探索可解释的机器学习模型,阐明模型背后隐藏的物理意义,实现逆向材料成分和结构设计也是未来的热门研究领域。

5 结束语

材料基因工程颠覆了传统的“试错-纠错”材料研发模式,通过数据驱动的高通量方法和人工智能模型加速新材料的研发与设计。本文系统总结了国内外知名的材料高通量计算框架、常用的针对多种材料体系的大型材料数据库和机器学习方法,并概述了多尺度集成可视化的高通量自动计算和数据管理智能平台 ALKEMIE 的研究进展,提出了未来发展的研究方向,为实现按需逆向设计新材料提供参考。

参考文献 (References)

- [1] XIANG X, SUN X, BRICENO G, et al. A combinatorial approach to materials discovery [J]. Science, 1995, 268 (5218): 1738-1740.
- [2] ALLISON J. Integrated computational materials engineering: A perspective on progress and future steps [J]. JOM, 2011, 63 (4): 15-18.
- [3] LIU Z. High throughput CALPHAD modeling and the materials genome [J]. Technology, 2010, 617: 622.
- [4] LIU Z. View and comments on the data ecosystem: Ocean of data [J]. Engineering, 2020, 6 (6): 604-608.
- [5] WHITE A. The materials genome initiative: One year on [J]. MRS Bulletin, 2012, 37 (8): 715-716.
- [6] AGRAWAL A, CHOUDHARY A. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science [J]. APL Materials, 2016, 4 (5): 053208.
- [7] JAIN A, ONG S, HAUTIER G, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation [J]. APL Materials, 2013, 1 (1): 011002.
- [8] CURTAROLO S, SETYAWAN W, HART G, et al. AFLOW: An automatic framework for high-throughput materials discovery [J]. Computational Materials Science, 2012, 58: 218-226.
- [9] LARSEN A, MORTENSEN J, BLOMQVIST J, et al. The atomic simulation environment-A Python library for working with atoms [J]. Journal of Physics: Condensed Matter, 2017, 29 (27): 273002.
- [10] PIZZI G, CEPELLOTTI A, SABATINI R, et al. AiiDA: Automated interactive infrastructure and database for computational science [J]. Computational Materials Science, 2016, 111: 218-230.
- [11] GONZE X, AMADON B, ANGLADE P M, et al. ABINIT: First-principles approach to material and nanosystem properties [J]. Computer Physics Communications, 2009, 180 (12): 2582-2615.
- [12] MATHEW K, SINGH A K, GABRIEL J J, et al. MPInterfaces: A materials project based Python tool for high-throughput computational screening of interfacial systems [J]. Computational Materials Science, 2016, 122: 183-190.
- [13] LAMBERT H, FEKETE A, KERMODE J R, et al. Imeall: A computational framework for the calculation of the atomistic properties of grain boundaries [J]. Computer Physics Communications, 2018, 232: 256-263.
- [14] WANG G J, PENG L, LI K, et al. ALKEMIE: An intelligent computational platform for accelerating materials discovery and design [J]. Computational Materials Science, 2021, 186: 110064.
- [15] WANG G J, LI K, PENG L, et al. High-throughput automatic integrated material calculations and data management intelligent platform and the application in novel alloys [J]. Acta Metallurgica Sinica, 2021, 58 (1): 75-88.
- [16] YANG J, LI H, WU T, et al. Evaluation of half-Heusler compounds as thermoelectric materials based on the calculated electrical transport properties [J]. Advanced Functional Materials,

- 2008,18(19):2880-2888.
- [17] YANG X,WANG Z,ZHAO X,et al. MatCloud: A high-throughput computational infrastructure for integrated management of materials simulation, data and resources[J]. Computational Materials Science,2018,146:319-333.
- [18] ZHAO X G,ZHOU K,XING B,et al. JAMIP: An artificial-intelligence aided data-driven infrastructure for computational materials informatics [J]. Science Bulletin, 2021, 66 (19) : 1973-1985.
- [19] CURTAROLO S,SETYAWAN W,WANG S,et al. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations [J]. Computational Materials Science,2012,58:227-235.
- [20] CURTAROLO S,HART G L,NARDELLI M B,et al. The high-throughput highway to computational materials design [J]. Nature Materials,2013,12(3):191-201.
- [21] HU S,LIU B,LI Z,et al. Identifying optimal dopants for Sb_2Te_3 phase-change material by high-throughput ab initio calculations with experiments [J]. Computational Materials Science, 2019, 165:51-58.
- [22] GAN Y,MIAO N,LAN P,et al. Robust design of high-performance optoelectronic chalcogenide crystals from high-throughput computation [J]. Journal of the American Chemical Society, 2022,144(13):5878-5886.
- [23] BELSKY A,HELLENBRANDT M,KAREN V L,et al. New developments in the inorganic crystal structure database (ICSD): Accessibility in support of materials research and design [J]. Acta Crystallographica,2002,58(3):364-369.
- [24] QUIRÓS M,GRAŽULIS S,GIRDZIJAUSKAITÉ S,et al. Using SMILES strings for the description of chemical connectivity in the crystallography open database [J]. Journal of Cheminformatics,2018,10(1):1-17.
- [25] SAAL J E,KIRKLIN S,AYKOL M,et al. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD) [J]. JOM, 2013, 65(11):1501-1509.
- [26] DRAXL C,SCHEFFLER M. The NOMAD laboratory: From data sharing to artificial intelligence [J]. Journal of Physics: Materials,2019,2(3):036001.
- [27] OGATA T,YAMAZAKI M. New stage of MatNavi, materials database at NIMS[C]// ASME Pressure Vessels and Piping Conference. New York: ASME,2009:1561-1568.
- [28] CHOUDHARY K,DECOST B,TAVAZZA F. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape [J]. Physical Review Materials, 2018,2(8):083801.
- [29] BORYSOV S S,GEILHUFÉ R M,BALATSKY A V. Organic materials database: An open-access online database for data mining[J]. PLOS One,2017,12(2):171501.
- [30] HUMMELSHØJ J S,ABILD-PEDERSEN F,STUDT F,et al. CatApp: A Web application for surface chemistry and heterogeneous catalysis [J]. Angewandte Chemie International Edition, 2012,51(1):272-274.
- [31] ANDERSEN C W,ARMIENTO R,BLOKHIN E,et al. OPTIMADE, an API for exchanging materials data[J]. Scientific Data,2021,8(1):217.
- [32] LIN X,YANG W,WANG K L,et al. Two-dimensional spintronics for low-power electronics [J]. Nature Electronics, 2019, 2(7):274-283.
- [33] LU S,ZHOU Q,OUYANG Y,et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning [J]. Nature Communications,2018,9(1):3405.
- [34] WU Y,LU S,JU M G,et al. Accelerated design of promising mixed lead-free double halide organic-inorganic perovskites for photovoltaics using machine learning [J]. Nanoscale, 2021, 13(28):12250-12259.
- [35] ZHANG H,FU H,ZHU S,et al. Machine learning assisted composition effective design for precipitation strengthened copper alloys[J]. Acta Materialia,2021,215:117118.
- [36] VILLARS P,CENZUAL K,GLADYSHEVSKII R,et al. PAULING FILE-towards a holistic view [J]. Chemistry of Metals and Alloys,2018,11(3):43-76.
- [37] TSHITOYAN V,DAGDELEN J,WESTON L,et al. Unsupervised word embeddings capture latent knowledge from materials science literature [J]. Nature,2019,571(7763):95-98.
- [38] KONOVOVA O,HUO H,HE T,et al. Text-mined dataset of inorganic materials synthesis recipes [J]. Scientific Data,2019,6(1):203.
- [39] BLANK T B,BROWN S D,CALHOUN A W,et al. Neural network models of potential energy surfaces [J]. The Journal of Chemical Physics,1995,103(10):4129-4137.
- [40] MALSHE M,NARULKAR R,RAFF L M,et al. Development of generalized potential-energy surfaces using many-body expansions, neural networks, and moiety energy approximations [J]. The Journal of Chemical Physics,2009,130(18):184102.
- [41] BEHLER J. Atom-centered symmetry functions for constructing high-dimensional neural network potentials [J]. The Journal of Chemical Physics,2011,134(7):074106.
- [42] GASTEGGER M,SCHWIEDRZIK L,BITTERMANN M,et al. wACSF: Weighted atom-centered symmetry functions as descriptors in machine learning potentials [J]. The Journal of Chemical Physics,2018,148(24):241709.
- [43] YAO K,HERR J E,TOTH D W,et al. The TensorMol-0.1 model chemistry: A neural network augmented with long-range physics [J]. Chemical Science,2018,9(8):2261-2269.
- [44] SOSSO G C,MICELI G,CARAVATI S,et al. Neural network interatomic potential for the phase change material GeTe [J]. Physical Review B,2012,85(17):174103.
- [45] ARTRITH N,URBAN A. An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for TiO_2 [J]. Computational Materials Science,2016, 114:135-150.
- [46] MOCANU F,KONSTANTINOU K,LEE T,et al. Modeling the phase-change memory material, $Ge_2Sb_2Te_5$, with a machine-learned interatomic potential [J]. The Journal of Physical Chemistry B,2018,122(38):8998-9006.
- [47] ZHANG L,HAN J,WANG H,et al. Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics [J]. Physical Review Letters,2018,120(14):143001.
- [48] CHEN Q,CHEN M,ZHU L,et al. Composition gradient media-

- ted semiconductor metal transition in ternary transition metal dichalcogenide bilayers [J]. ACS Applied Materials & Interfaces, 2020, 12(40):45184-45191.
- [49] PENG Q, ZHOU J, CHEN J, et al. Cu single atoms on Ti_2CO_2 as a highly efficient oxygen reduction catalyst in a proton exchange membrane fuel cell [J]. Journal of Materials Chemistry A, 2019, 7(45):26062-26070.
- [50] GAN Y, WANG G, ZHOU J, et al. Prediction of thermoelectric performance for layered IV-VI semiconductors by high-throughput ab initio calculations and machine learning [J]. NPJ Computational Materials, 2021, 7(1):176.

Novel material design and development accelerated by materials genome engineering

SUN Zhimei^{1,*}, WANG Guanjie^{1,2}, ZHANG Xuanguang¹, ZHOU Jian¹

(1. School of Materials Science and Engineering, Beihang University, Beijing 100083, China;

2. School of Integrated Circuit Science and Engineering, Beihang University, Beijing 100083, China)

Abstract: Without knowing the relationship between the chemical composition and properties of the material, the development of new materials with desired properties based on conventional trial-and-error methods is cost inefficient and sometimes ends fruitlessly. With the development of artificial intelligence and the fourth science paradigm, data-driven science, materials genetic engineering (MGE) has become a new approach for designing novel materials. In this paper, recent advances in high-throughput computation, materials database, and artificial intelligence methods in MGE are reviewed. First, the framework and software for high-throughput computations of materials are introduced. Then, the progress and critical problems of materials databases are presented in terms of types and standard interfaces of materials data. Next, we summarize the applications of artificial intelligence methods in critical issues of materials science. In particular, from the aspects of visualized high-throughput calculation methods, multi-type materials database and visualized machine learning framework, we emphasize an in-house developed platform ALKMIE, featured by multi-scale integration of automatic high-throughput calculation, visualization, and intelligent data management. Finally, we highlight the future directions of MGE.

Keywords: materials genome engineering (MGE); high-throughput calculation; materials database; machine learning; artificial intelligence calculation platform

Received: 2022-05-06; **Accepted:** 2022-05-18; **Published online:** 2022-06-01 15:07

URL: kns.cnki.net/kcms/detail/11.2625.V.20220531.1837.003.html

Foundation items: National Natural Science Foundation of China (51872017); China Postdoctoral Science Foundation (2022TQ0019); The High Performance Computing (HPC) Resources at Beihang University

* **Corresponding author.** E-mail: zmsun@buaa.edu.cn