

引用格式: 冯建发, 王畅畅, 苏航, 等. 材料数据库的现状与未来: AI 技术引领的创新应用前景[J]. 中国材料进展, 2026, 45(2): 89-101.

FENG J F, WANG C C, SU H, *et al.* Current Status and Future Prospects of Material Databases: Innovative Applications Led by AI Technology[J]. Materials China, 2026, 45(2): 89-101.

材料数据库的现状与未来: AI 技术引领的创新应用前景

冯建发^{1,2,3}, 王畅畅², 苏航^{2,3}, 宿彦京¹

(1. 北京科技大学 新材料技术研究院, 北京 100083)

(2. 北京新材道数智科技有限公司, 北京 100081)

(3. 中国钢研科技集团有限公司数字化研发中心, 北京 100081)

摘要: 随着人工智能技术的不断进步, 材料数据库在材料科学研究中扮演着日益重要的角色。旨在探讨材料数据库如何通过 AI 技术的融合, 扩展其应用范围并提升其核心价值。通过文献综述的方法, 系统地分析了材料数据库的当前分类, 包括材料基础数据库、生产加工数据库、应用服役数据库等, 并概述了支撑技术如机器学习、深度学习、数据标准化技术的应用情况。尽管国际上材料数据库的发展呈现出智能化、网络化、资产化、去中心化的趋势, 但在数据质量、数据共享、知识产权、市场运维等方面仍面临挑战。未来材料数据库的发展将受益于与新兴技术如材料数据工厂、区块链、隐私计算、AI 大模型的结合, 这将为新材料的研发和应用提供创新的手段和场景工具。

关键词: 材料数据库; 大数据技术; AI; 机器学习; 大模型

中图分类号: TB30; TP18 文献标识码: A 文章编号: 1674-3962(2026)02-0089-13

Current Status and Future Prospects of Material Databases: Innovative Applications Led by AI Technology

FENG Jianfa^{1,2,3}, WANG Changchang², SU Hang^{2,3}, SU Yanjing¹

(1. Institute for Advanced Materials and Technology, University of Science and Technology Beijing, Beijing 100083, China)

(2. Beijing MatDao Technology Co., Ltd., Beijing 100081, China)

(3. Material Digital R&D Center, China Iron & Steel Research Institute Group, Beijing 100081, China)

Abstract: With the continuous advancement of artificial intelligence (AI) technology, material databases are increasingly playing a pivotal role in materials science research. This paper aims to explore how the integration of AI technology can expand the application scope and enhance the core value of material databases. Through a literature review, the current classifications of material databases were systematically analyzed, including material basic databases, production and processing databases, and application service databases, and outlined the application of supporting technologies such as machine learning, deep learning, and data standardization. Despite the international development of material databases showing trends towards intelligence, networking, assetization, and decentralization, challenges remain in terms of data quality, data sharing, intellectual property rights, and market operation and maintenance. The future development of material databases is expected to benefit from the integration with emerging technologies such as material data factories, blockchain, privacy computing, and AI large models, which will provide innovative means and tools for the research and development and application of new materials.

Key words: material database; big data technology; AI; machine learning; large model

收稿日期: 2024-11-29 修回日期: 2025-03-28

基金项目: 国家重点研发计划项目(2022YFB3505202)

第一作者: 冯建发, 男, 1999 年生, 硕士研究生

通讯作者: 苏航, 男, 1969 年生, 教授级高级工程师,
博士生导师, Email: hangsu@vip.sina.com

DOI: 10.7502/j.issn.1674-3962.202411028

1 前言

材料是制造业的基础, 新材料的研发和应用分布在

制造业的各个环节,因此材料是承载装备全产业链、全生命周期成本和安全最重要的物质载体,是制造业的“骨骼”。材料科学领域的快速发展带来了大量材料科学数据的积累^[1],这些数据涵盖了材料的各个方面,包括成分、结构、性能、制备工艺等,构成了一个庞大且复杂的信息网络。如何高效地管理、整合和利用这些海量数据,成为当前材料科学研究亟需解决的重要问题。

2011年,美国提出的材料基因组计划^[2]将材料数据库确定为其中的3大基础平台之一;随后我国启动了材料基因工程专项^[3],也注重了材料数据库的建设,加速了材料数据库的发展。另外,人工智能、机器学习、大模型等新兴技术的发展为材料数据库的应用提供了深度分析、知识挖掘和验证平台,可以推动材料科学的前沿研究。本文综述了国内外材料数据库的分类、应用以及目前存在的问题,并提出了未来材料数据库可能的发展及技术突破的方向。

2 材料数据库概述

2.1 材料数据库的工程价值

材料数据库对新材料研发的重要性主要体现在以下几个方面。

(1) 服务于工程选材、评估与优化

工程选材过程中存在信息分散、数据格式不一、难以互联等问题。不同材料数据库采用不同格式和标准,使得信息获取和比对困难。利用人工智能和数据挖掘可自动匹配不同数据库信息,解决数据差异问题,实现联通。这使材料数据从“小数据”迈向“大数据”,为工程选材和设计提供了重要支持^[4]。德国的 Keyto Steel 和 Mat-match 等在线数据库专注于提供多国牌号对照和相似材料查询,主要用于产品牌号数据对比。杨丽等^[5]介绍了中国工程材料数据库的发展,并描述了 AI 自动匹配技术在国际材料牌号匹配、跨标准和私有数据库的关联查询、焊接材料自动选配方面的工作,自动匹配牌号技术得出的结果与《世界钢号对照手册》《焊材手册》等基于专家推荐的材料牌号高度一致。

(2) 结合机器学习进行新材料开发与性能预测

在大数据时代,数据科学成为驱动科学发现的关键工具。材料数据正以爆炸式增长的势头不断积累,为材料科学研究提供了丰富的机器学习资源,如 Tehrani 等^[6]利用 Materials Project 数据库训练机器学习模型,成功预测了大量化合物的弹性模量,实验验证误差控制在 10% 以内。Stanev 等^[7]使用 NIMS (The National Incident Management System) 的 Super Con 数据库进行超导临界温度的机器学习建模,建立了高准确率的预测模型。Cheon

等^[8]通过机器学习模型在 Materials Project 数据库中成功识别出大量二维层状和一维链状无机晶体材料,显著提高了对这些材料类型的识别能力。

(3) 服务于材料与工艺计算

材料数据库为计算软件提供了必要的数据库支持,使得工程设计和模拟分析更加高效。尽管现有的计算方法和软件能够满足数据挖掘、相图生成分析等需求^[9],但它们仍然依赖于专业数据库的配合^[10, 11]。面对外部数据获取的挑战,建设具有自主知识产权的数据库对于支持我国的材料计算和设计、推动中国材料领域的创新至关重要。ALKEMIE 平台^[12]拥有超过 18 万条材料结构数据和 1 万条材料性能数据,并且集成了多种计算引擎,可以实现全自动建模、运行和数据分析等功能。CISRI-DLab (<https://dlab.cisri.com/>) 数字化研发平台^[5]是一项成功的企业级材料数字化研发平台,集成了 40 余款材料计算和模拟软件以及众多金属材料国内外标准数据库,提供远程、排队和移动计算,并众筹了 700 多个材料 APP 应用场景。2023 年推出的行业级开放平台 Material-DLab 面向所有材料领域的学者开放,旨在提供便捷高效的材料计算与设计工具,为材料科学的创新与发现提供平台支持。

2.2 材料数据库类别

材料数据库的分类方式繁多,有按材料类型划分,如金属、非金属、复合材料等;也有按使用场景区分,涵盖航空航天、汽车、电子等各个行业。本文基于一种更为系统和全面的分类方法进行综述,即按照材料基础数据库、生产加工数据库和应用服役数据库进行划分,这种分类方法不仅涵盖了材料的基本属性、制备工艺,还延伸到了材料在实际应用中的性能表现方面。

2.2.1 材料基础数据库

材料基础数据库,是指在材料研发过程中通过实验室、文献期刊或材料计算软件等渠道获取的数据资源^[13]。图 1 展示了材料基础数据库发展的历程,这些数据库包含了晶体结构、热力学性质和结构特性等多方面的材料信息,例如,ICSD (Inorganic Crystal Structure Database) 是目前世界上最大的无机晶体结构数据库^[14, 15],收录了自 1913 年以来出版的超过 210 000 条完全表征的无机晶体结构详细信息;MatNavi 是由日本国立材料科学研究所开发的材料数据库系统^[16],集成了多个数据库和应用系统,是全球材料领域最大的数据库之一;AFLOW (Automatic-FLOW for Materials Discovery) 是一个开源的晶体结构数据库^[17],专注于无机、有机和金属有机框架的晶体结构数据;OQMD (The Open Quantum Materials Database) 是一个开放的、基于互联网的数据库^[18],旨在为

材料科学社区提供量子力学计算得到的材料性质数据；MGED (Materials Genome Engineering Databases) 是一个专注于高通量实验数据的数据库，它通过整合来自不同实验条件下的材料性能数据^[19, 20]，为材料的筛选和优化提供了强有力的支持；钢研新材道是一个免费的材料云平台^[5]，提供钢材、焊材、有色金属标准数据、实验数据，并配备智能匹配、选材算法。然而，尽管这些数据库极大地促进了全球范围内的数据共享和交流，它们仍面临

一些挑战。主要问题包括数据整合不足导致的碎片化，缺乏标准化造成的数据不一致性以及更新频率低引起的数据低时效性。

2.2.2 生产加工数据库

生产加工数据库汇集了材料加工领域，如焊接、切削、热处理和锻造等工艺过程中产生的大量数据。利用这些数据可以对材料执行精确的成型、加工、连接和表面处理，从而满足多样化的工程需求。

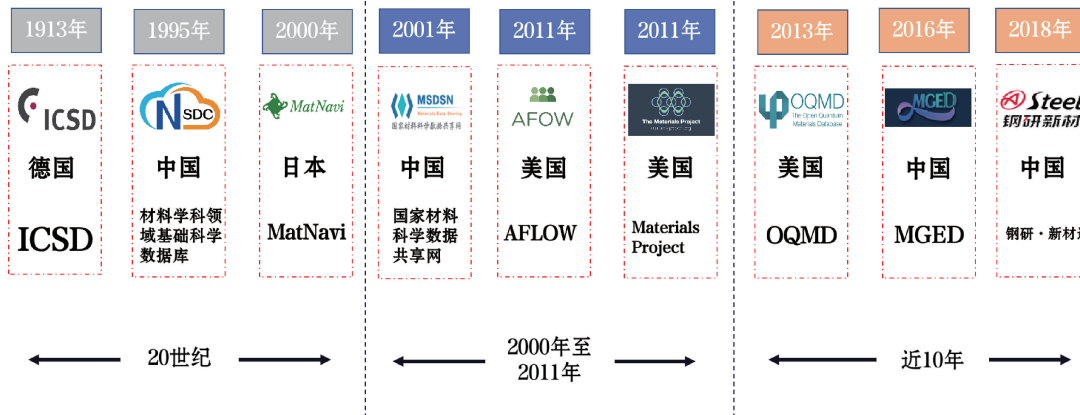


图 1 材料基础数据库发展历程

Fig. 1 Evolution of material fundamental databases

(1) 焊接数据库

焊接数据库主要以焊接材料、焊接工艺及焊接工艺评定数据库为主^[21, 22]，也包括其他功能的焊接数据库，国内外不同类型焊接数据库如表 1 所示^[23-26]。从中可以看出焊接数据库的种类繁多，这些数据库通常为特定的需求而建，如针对某种焊接技术或材料。由于这些数据库多由个人或组织开发用于内部，它们的使用范围相对较窄、专业性强，主要服务于焊接工程师和研究人员。

(2) 切削数据库

除了焊接数据库外，还有一些材料切削数据库^[27]，如表 2 所示^[28-31]。从表 2 中可以观察到，大部分切削数据库都是由高校自行建立的，这些数据库往往服务于特定的课题或者实验室需求。相较于焊接数据库，切削数据库的应用范围通常更为狭窄，主要用于切削工艺的优化、刀具材料的选择以及加工质量的评定等方面。

表 1 国内外不同类型焊接数据库对比
Table 1 Comparison of global welding databases

类型	名称	国别	时间	单位	概述
焊接材料数据库	焊接连续冷却转换图 (CCT 图) 数据库 ^[23]	日本	1986 年	日本国立金属材料技术研究所	可查询材料成分、相变临界点的 CCT 图, 用户输入焊接参数后可计算出指定点的 CCT 和热影响区的组织成分与硬度
焊接材料数据库	焊林院	中国	2017 年	上海孔德信息技术有限公司	提供焊接知识答疑、生成焊接工艺规程、查询金属标准等功能, 还包括焊接词典、焊材查询、设备信息等实用工具
焊接工艺数据库	搅拌摩擦焊工艺变量数据库 ^[24]	印度	2021 年	达纳拉克什米工程学院	创建了航空铝合金搅拌摩擦焊工艺数据库, 发现拉伸强度与工艺参数相关
焊接工艺判定数据库	轨道车车辆架焊接工艺编制软件 ^[25]	中国	2020 年	南京焊接智能科技有限公司	通过工作流集成支持多部门协作编制和审核焊接工艺规程, 确保钢轨焊接系统的高效运行, 并整合焊接接头信息以减少重复评定
其他功能的焊接数据库	激光焊接数值模拟的模型数据库 ^[26]	中国	2020 年	南京航空航天大学	建立了可实现任意结构的激光焊接数值模拟的模型数据库。借助该模型数据库, 可以快速有效地实现建模

表 2 切削数据库对比

Table 2 Comparison of cutting databases

名称	单位	概述
高速切削工艺数据库	同济大学	将工艺参数对切削力的影响结果存入数据库,指导后续实验 ^[28]
细长轴零件车削加工数据库	南昌航空大学	库中录入基本加工信息,实现加工工艺查询,并设计了切削加工参数优化功能 ^[29]
轴类零件车削加工数据库	北京林业大学	提供四大数据库:车削工件材料、车削刀具、加工设备和切削液,根据工件特征输出匹配信息 ^[30]
针对汽车发动机缸盖零件的切削数据库	上海交通大学	基于振动稳定性叶瓣图实现切削参数优化,确定了合适的参数范围 ^[31]

(3) 热处理数据库

热处理数据库是材料科学中用于指导金属热处理工艺设计的重要工具,通过收集和分析材料在不同热处理条件下的性能数据支持热处理工艺的模拟和优化。Xu 等^[32]建立了一个包含 23 种组分的多组分镁合金热力学数据库,用于指导凝固模拟和热处理方案的选择。He 等^[33]通过基于 DEFORM-3D 的二次开发成功集成了 16MnR 材料热处理数据库,显著提升了压力容器热处理过程数值模拟的准确性和实用性。Watanabe 等^[34]通过建立基于有限差分法的热处理模拟程序和包含 Jominy 试验与回火参数的数据库高效预测了热处理后钢材的硬度,显著简化了工艺设计流程。

(4) 铸锻数据库

铸锻数据库为锻造、铸造工艺提供了关键的数据支持,包括材料属性、工艺参数和模具性能等,尽管铸锻数据库的应用范围可能不如焊接数据库那样广泛,但它们对于理解和改进铸造和锻造工艺至关重要。Chen 等^[35]提出了一种面向对象的知识建模方法,并构建了锤锻设计知识库,以支持和引导设计师进行有效的模具设计和决策。Ohashi^[36]开发了一款模具寿命分析仪,结合互联网上的冷锻模具寿命数据库,利用风险树网络和模糊语言方法预测模具寿命,并通过分析寿命缩短的原因来指导锻造工艺设计。Giese^[37]开发了一个铸铁物理特性数据库,用于预测铸件中的金属渗透缺陷,并通过算法模型和接触角数据库推导出渗透指数,指导铸造厂有效预防缺陷。

2.2.3 应用服役数据库

应用服役数据库涵盖了材料在实际使用过程中的关键性能数据,如磨损、疲劳、耐腐蚀性和电学性能等,对评估材料在工程应用中的表现至关重要。

(1) 腐蚀数据库

腐蚀数据库为研究人员和工程师提供了详尽的腐蚀数据,这些数据可以支持材料耐久性的设计、性能评估及维护策略的制定。美国 NACE (National Association of

Corrosion Engineers) 和 NIST (National Institute of Standards and Technology) 合作开发了腐蚀数据管理系统^[38],旨在整合和传播腐蚀数据。欧盟、日本国立金属研究所和美国的 MPD (Materials Property Database) 数据库也提供了丰富的材料腐蚀信息^[39]。我国的国家材料环境腐蚀科学数据库是广泛使用的资源之一,累积了超过 40 万个数据点,并建立了多个子库^[40]。API (American Petroleum Institute) 581^[41]标准则为工业设施的腐蚀和磨损评估提供了基于风险的检验方法,考虑了多种环境和操作条件,提供了关于材料选择和兼容性的建议,以确保在特定环境下选用合适的材料以抵御腐蚀及其他损伤。

(2) 疲劳数据库

材料疲劳数据库整合了材料在反复加载下的性能数据,这些数据库不仅是预测材料疲劳寿命的重要依据,同时也对优化材料设计、提升结构安全性能具有关键作用。Total Materia 数据库 (<https://www.totalmateria.com>) 提供了全面的材料疲劳数据,包括应变寿命和应力寿命参数等。NIMS (National Institute for Materials Science) 疲劳数据库是一个包含结构材料疲劳特性的详尽数据库^[42],记录了多种材料在不同疲劳试验下的性能。英国的 BS (British Standard) 7608 规范^[43]基于大量焊接结构疲劳数据为船舶结构的疲劳评估和设计提供了指导,确保船舶结构的长期安全和稳定。

(3) 磨损数据库

磨损数据库汇集了在各种工作条件下材料的磨损特性数据,对于深入剖析材料磨损机制、精准预测实际应用性能以及指导耐磨设计具有不可替代的作用。为了有效地整理和比较 DLC (diamond-like carbon) 涂层的摩擦学性能,Sedlacek 等^[44]开发了一个磨损数据库,它不仅便于新数据的添加,还能够快速检索和分析不同实验条件下的测试结果。德国钢铁研究所 VDEh (Verein Deutscher Eisenhüttenleute) 及其会员企业建立了一个基于互联网的钢材材料信息和营销平台 Stahl-Datenbank,该平台提供了包括 Stahl-Eisen-Liste 内容、6000 次单独磨损试验数

据、工具损伤数据、450 种钢级的物理特性、超过 3200 个不同的时温转变图谱数据及来源标注以及根据 SEP (Stahl-Eisen-Werkstoffblatt) 1664 计算硬化性的方程式和在线硬度变化计算课程在内的综合磨损数据库^[45]，为工程师、科学家以及钢材加工企业的采购商提供了宝贵的知识资源。

(4) 蠕变数据库

蠕变数据库汇集了材料的蠕变曲线、蠕变速率、蠕变寿命等关键数据，为深入研究材料的蠕变机理、预测长期性能以及指导材料选择和设计提供了有力支持。Kimura 等^[46]基于扩展的蠕变数据库对 Grade 91 钢焊接接头的蠕变断裂因子和焊缝强度降低因子进行了综述和重新评估，将蠕变断裂因子和焊缝强度降低因子定义为横焊缝试样的平均蠕变断裂强度与基材平均蠕变断裂强度的比值。Koizumi 等^[47]通过在相同的制造工艺和测试条件下对商业合金进行蠕变性能测试，建立了一个包含多种 Ni 基单晶超合金蠕变性能的数据库，为涡轮设计和虚拟喷气发动机计算机模拟提供了重要数据支持。

(5) 断裂数据库

断裂数据库汇集了各种材料在不同条件下的断裂特性、断裂韧性、裂纹扩展速率等重要数据，有助于深入了解材料的断裂机理，预测材料在实际应用中的断裂行为，指导材料的设计和选择。Lamborn 等^[48]提出了建立管道材料数据库的概念，该数据库将集成断裂韧性和疲劳裂纹扩展数据，以提高管道结构完整性预测的准确性，并支持管道行业的性能优化。Moskovic^[49]对 22NiMoCr37 铁素体钢锻件在不同厚度和温度下的断裂韧性数据库进行了统计分析，采用基于竞争风险的统计方法评估断裂韧性，并量化了解理断裂的概率。Shi 等^[50]开发了一个裂纹扩展数据库，用于训练神经网络预测 304 不锈钢的裂纹扩展行为，并揭示了其与环境因素的电化学关联。

2.3 存在的问题与挑战

材料数据库发展目前仍有一些亟待解决的问题，主要包括以下几个方面。

(1) 数据质量问题

在当代科学研究的版图中，AI for Science^[51, 52] (人工智能用于科学) 正逐渐显露其重要性，成为一个具有划时代意义的新兴领域。AI for Science 的核心在于借助人工智能的强大计算力，对科学数据进行深度分析，构建精确的预测模型，从而推动科学知识的边界。该领域已经渗透至自然科学的各个分支，包括生物学、物理学、化学和天文学等，通过提供先进的数据分析工具和模拟技术，AI for Science 正在塑造科学研究的新方向。

不同于 AI for Science 在天文学、分子生物学等自然

科学领域的成功，材料领域传统研发、生产数据存在的一些固有问题，阻碍了 AI 技术在材料领域更为广泛的应用，如表 3 所示，这些问题^[53]需要改变传统材料研发以数据为副产品的陈旧观念，建立专业化、标准化的材料数据生产设施比如材料数据工厂，才能从根本上得以解决。

表 3 数据质量问题一览表

Table 3 List of data quality issues

数据主观性	实验设计或计算过程中由于个体性，数据可能受到设计者个人观点的影响，导致数据的客观因果关系受到干扰
数据团聚性	大量生产数据集中在很小的材料成分、工艺范围内，导致出现大量相似或重复的信息，使 AI 学习的变量范围过窄
数据异构性	不同来源的数据可能在格式、标准等方面存在显著差异，没有统一的标准，使得数据的整合和分析变得复杂困难
数据碎片性	存在大量孤立、零散的实验数据，高质量数据的稀缺与低质量数据的过剩共存，导致数据收集的平衡，严重影响机器学习建模的质量
数据噪声性	材料检测数据可能存在错误或者误差，这些噪声可能会在后续的机器学习和应用中放大，影响数据的准确性和可靠性

(2) 知识产权和共享问题

在大数据时代，数据不仅是科技创新的核心资源，更是推动经济发展的关键要素。然而数据的离散性和非实体性特征在法律层面上造成了产权界定和保护的难题，尤其在涉及企业生产、研发和应用数据时，这些数据往往包含敏感的商业机密。这一困境在全球范围内对数据共享构成了挑战^[54]。为了推动材料迭代升级，打通产业链之间的壁垒，实现数据共享成为必不可少的路径。FAIR 原则^[55]是国际科学数据共享的重要指导方针，它代表了可查找 (findable)、可访问 (accessible)、可互操作 (interoperable) 和可重用 (reusable) 4 个关键特性。这些原则旨在提高科学数据的可用性和价值，以便研究人员和机器都能够方便地发现、访问、整合和利用数据。国内 CSTM (Chinese Society for Testing & Materials) 等相关机构已经构建了 CSTM/FC91 材料产业区块链标准化领域委员会，积极推动数据共享领域相关技术^[56]和规范的建立。通过区块链技术，CSTM 旨在建立一个安全、高效的数据共享平台，保障数据的完整性和可追溯性，同时保护企业的知识产权。

(3) 数据库的商业运维问题

材料数据库的商业化运营是确保其可持续发展的关键，但目前面临以下挑战：① 缺乏可持续的商业模式：在 21 世纪初期，国家科技部、工信部、发改委及地方政府部门纷纷投资支持了大量数据库平台的建设，然而令人遗憾的是，这些数据平台中实现长期稳定运营的并不

多见。主要原因在于这些平台往往以项目形式支持建立，一旦项目执行期结束，团队自然而然解散，缺乏持续的运营和维护，导致许多平台陷入了无人监管的境地。② 用户需求多样化：不同用户（如企业、科研机构、个人研究者）对数据的需求差异较大，难以提供个性化服务。③ 数据价值变现困难：材料数据的价值难以量化，导致商业化路径不清晰。④ 参考国外成功运营的材料数据库平台如 MatWeb、Matmatch 等，可以发现一个共同的成功经验——建立一个以数据为生的运营团队。这样的团队可以使数据库走向商业化、市场化，从而实现可持续发展。而这些平台通过订阅服务、数据交易和增值服务等手段，吸引了用户和资金，实现了商业化和市场化。例如，MatWeb 通过订阅模式为用户提供高质量的材料性能数据和选材工具，Matmatch 则利用机器学习算法为用户提供个性化的材料推荐服务。我国材料数据库的发展应借鉴这些经验，建立和完善商业运维机制，包括构建可持续的商业模式、形成稳定的资金来源以及培养专业的运营团队。同时，政府需出台政策支持材料数据库的商业化运营，推动行业合作，建立数据共享联盟，从而在全球科技竞争中占据有利地位。

3 材料数据库的支撑性技术

我国材料数据库虽然取得了进步，但与发达国家相比仍有差距，需要在技术、应用和管理等方面持续创新。目前材料数据库的发展呈现出以下主流趋势：在线化和网络化，提供便捷的在线访问和数据共享；标准化和智能化，实现数据的一致性和智能分析；去中心化和共享化，推动知识的开放和数据的安全分布；商业化和资产化，满足个性化需求并视数据为可交易的资产。

材料数据库的建设和优化涉及多种技术的综合应用，近期机器学习、深度学习以及数据标准化等关键技术受到广泛关注。

3.1 机器学习技术

机器学习^[57]是一门跨学科领域，涵盖计算机科学、概率论、统计学等学科知识。其核心是通过数据和算法让计算机模拟人类学习过程，从而实现智能决策。传统计算方法依赖人工输入的程序，而机器学习则通过数据学习规律，无需人工干预。机器学习分为监督学习、无监督学习和强化学习 3 类。监督学习依赖标记的输入和输出进行学习，无监督学习尝试在数据中寻找模式，而强化学习通过决策接收反馈。根据任务不同，机器学习可分为分类、回归、聚类和降维等 4 类任务，如表 4 所示^[58-70]。在材料科学领域，机器学习技术已被证明是推动材料设计和性能优化的强大工具。通过分析大量的实验和计算数据，机器学习模型能够揭示材料成分、结构与性能之间的复杂关系，并预测新材料的性质。如 Zhang 等^[71]利用机器学习对高熵合金的相结构进行预测，提高了材料设计的准确性和效率。Ward 等^[72]利用机器学习技术对材料的弹性模量进行预测，通过分析大量实验数据，成功构建了高精度的预测模型，能够准确预测不同成分和结构下的材料弹性性能，为材料设计和筛选提供重要参考。此外，机器学习在优化镁合金的热处理工艺方面也展现了其潜力，如 Chen 等^[73]通过机器学习辅助的方法优化了 ZE62 铸态镁合金的性能，实现了屈服强度和塑性的显著提升。AI for Science 正在成为材料研究的新范式。例如在钙钛矿材料领域，数据与模型驱动的智能计算框架显著提升了预测精度^[74]。针对高性能金属材料的定制化研发，张闫等^[75]与张思雨等^[76]分别系统论述了机器学习在钛合金成分优化及轻质合金组织预测中的应用进展，通过构建数据驱动的映射模型，为实现材料减重与性能平衡提供了新路径。对于镍基单晶高温合金等复杂体系，崔壮等^[77]指出，利用数智化技术驱动的合金化设计已成为应对极端服役环境、提升材料环境适应性的关键趋势。

表 4 机器学习技术分类
Table 4 Machine learning technology classification

学习类型	监督学习		无监督学习	无监督学习/预处理
学习任务	回归	分类	聚类	降维
主要功能	通过学习输入和输出之间的关系，用于预测连续数值	通过学习样本特征与标签之间的映射关系，将样本分到不同离散类别	通过发现数据内在的相似性，将样本划分为组	通过保留数据关键信息的同时减少特征维度，提高计算效率并防止过拟合
常用算法	线性回归 ^[58]	Logistic 回归 ^[59] 人工神经网络 ^[66] 深度学习 ^[67] K 近邻 ^[68] 决策树 ^[69] 支持向量机 ^[70]	K 均值聚类 ^[60] AP 聚类 ^[61] 层次聚类 ^[62] DBSCAN ^[63]	主成分分析(PCA) ^[64] 判别分析(LDA) ^[65]

3.2 深度学习技术

深度学习^[78]是机器学习的一个子领域，利用多层神经网络模拟和学习复杂数据表征，通过逐层提取和抽象数据特征实现对输入数据的高阶表示和模式学习^[79]。目前深度学习领域发展应用最广泛的为卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)。CNN^[80]是一种多层可训练结构的深度学习算法，通过卷积层和池化层进行特征提取，在图像^[81]、文本^[82]等领域得到广泛应用。在材料科学中，Jung 等^[83]利用 CNN 对材料的 SEM 照片进行分析，成功识别了材料的晶粒尺寸和缺陷分布，为材料性能优化提供了重要依据。Azimi 等^[84]利用深度学习技术对钢的微观结构进行分类，通过 CNN 成功识别了不同钢种的微观组织特征，如马氏体、贝氏体和铁素体等，为钢材的性能优化和质量控制提供了重要依据。RNN^[85]是处理序列数据的强大工具，能保持数据间的依赖关系，适用于处理长度不固定的输入序列。在材料科学中，RNN 可以用于预测材料在动态条件下的性能变化，如疲劳寿命^[86]、流变行为^[87]和微观结构演变^[88]。在 RNN 基础上发展起来的 Transformer 架构是大语言模型(large language model, LLM)成功的基础。大语言模型近年来发展迅速，以 ChatGPT4.0 为代表，它基于人工智能内容生成技术(artificial intelligence generated content, AIGC)，打破了之前基于数据场景建模的局限性，可以在特定领域形成知识体系化、闭环化、可自学习的人工智能模型，在知识高度分散的材料领域应用前景广阔。

3.3 数据标准化技术

在材料科学研究中，数据预处理对于保障数据质量至关重要。预处理流程涵盖数据清洗、变换、特征工程以及规范化和标准化等步骤，旨在提升模型的预测性能并确保数据集间的可比性。尽管预处理技术为数据分析奠定了基础，但它们并不直接解决数据互操作性问题。在多种数据基础设施共存的背景下，统一的数据标准化技术对于简化数据收录、存储和使用过程至关重要。

标准化^[89]是为了在一定范围内实现最佳秩序和共同利益，通过制定共同使用的条款和发布应用文件来解决现实或潜在问题。材料数据的特点是数量庞大、种类繁多、形式多样、产出单位不一和知识产权复杂。在多种数据基础设施共存的当下，缺乏统一标准会导致数据收录和存储变得复杂^[90]，也不便使用。因此，建立统一的数据标准对于规范化管理材料数据至关重要，它是实现数据驱动研究范式和材料领域广泛应用人工智能方法的基础。欧美国家通过建立材料科学本体和数据联盟组织如 FAIRmat^[91]和 OPTIMADE(The Open Databases Integration for Materials Design)^[92]，致力于改善材料数据的标准化和互操作性，以便更有效地管理和利用多源异构数据，尽管这需要开发数据转换器和共享数据模式来实现。中国在材料基因工程领域早期就重视标准化工作^[93]，成立了 CSTM 材料基因领域标准化委员会，致力于建立一个面向未来、全面覆盖材料数据生命周期的标准化体系，以支持数据驱动的研究和人工智能应用。

4 材料数据平台的技术前景

未来以人工智能应用、数据资产入表为代表的场景需求，将会为材料数据库的发展和引入全新的技术支撑。

4.1 材料数据工厂

材料数据的碎片化、主观化、噪音化以及标准化程度不足导致了优质试验数据的稀缺，而这正是材料机器学习所面临的主要挑战，在这一背景下，国外学者率先提出了“Data Fab”^[94]的概念，强调数据与人工智能相结合是推动材料基因工程发展的未来。随后，我国学者也提出了“材料数据工厂”的概念^[95]，进一步加强了这一理念。“数据工厂”可以是集成原位制备和多参数表征手段为一体的实验设施，流水线般标准化地批量产生数据^[94]，提高数据的准确性和一致性，也可以是各种高通量计算软件及硬件平台，通过批量计算产生大量系统的综合的材料数据，如图 2 所示。利用数据标识码技术，

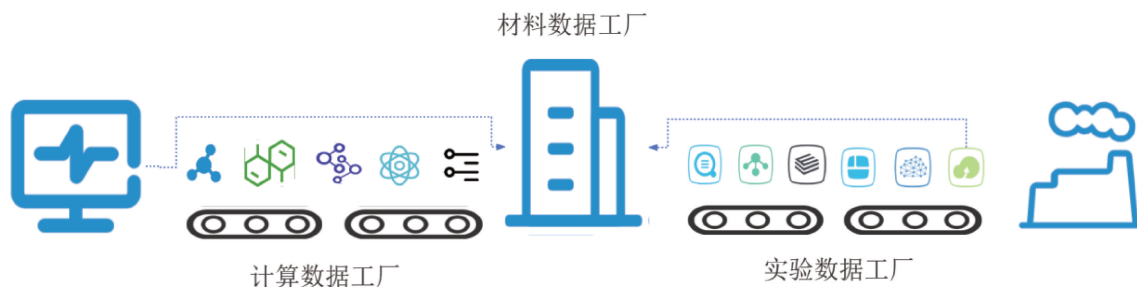


图 2 材料数据工厂概念图

Fig. 2 Material data factory concept diagram

结合高通量实验(或高通量计算)数据格式标准,就可以从实验线站上导出记录样品信息、实验条件和实验源数据(或计算条件和计算源数据)的具有唯一标识的、符合 FAIR 原则^[55]的数据,供社会使用。这些数据将为材料学者提供丰富的实验和计算结果,支持快速、高效的材料设计、优化和发现过程。结合人工智能和机器学习算法,可以从这些大规模数据中挖掘出隐藏的模式和规律,加速材料研究的进展。

通过建立材料数据工厂,可以构建一个完整的材料数据生态系统,促进数据的共享和交流,推动材料领域的协同创新和跨学科研究。未来的材料数据工厂将为材料学者提供丰富而大量的材料数据,数据与人工智能的结合将成为国际上推动材料基因工程向工程化发展的共识路径。

4.2 材料区块链

区块链技术^[96]通过其去中心化的分布式账本确保数据交易的透明性和不可篡改性^[97],同时强大的加密技术保障了信息安全。智能合约的机制使得合同条款能自动执行^[98],而交易记录的公开性进一步增强了透明度。区块链技术与材料数据库的结合代表了一种跨学科的前沿合作,Wang 等^[99]提出了一种基于区块链的材料数据发

现和共享基础设施——InterMat,这一创新性框架模型融合了云平台 and 区块链技术构建一个可信赖的材料大数据管理和共享平台,通过区块链技术确保了材料数据的安全共享和知识产权保护。InterMat 的提出标志着材料科学领域对于区块链技术的应用探索取得了初步进展。虽然目前正处于开发与应用的初期阶段,但在未来将在以下方面展现出广泛的应用场景:实现材料的全面质量溯源与认证;构建去中心化的智能合约驱动的材料交易市场;确保材料数据的知识产权保护;为国家重大工程提供坚实的数据支持,保障工程的安全和持续发展。

传统的中心化数据库^[100]将数据集中存储于单一服务器,适用于规模较小的企业和应用,但存在安全和数据管理的挑战。去中心化或分布式数据库^[101]将数据存储于多个节点上,提高了系统的容错性和扩展性,但需要解决数据一致性和网络通信等问题。图 3 是数据库架构图,其中图 3a 是中心化数据库,图 3b 是去中心化数据库。目前,材料领域普遍使用中心化数据库,如 Materials Project、NIST 和 OQMD。杨丽等^[5]指出,北京新材道数智科技有限公司正在开发去中心化材料数据库,通过旗下“新材道”平台推动数据透明度和可访问性的提升,展示去中心化技术在材料科学中的潜力。

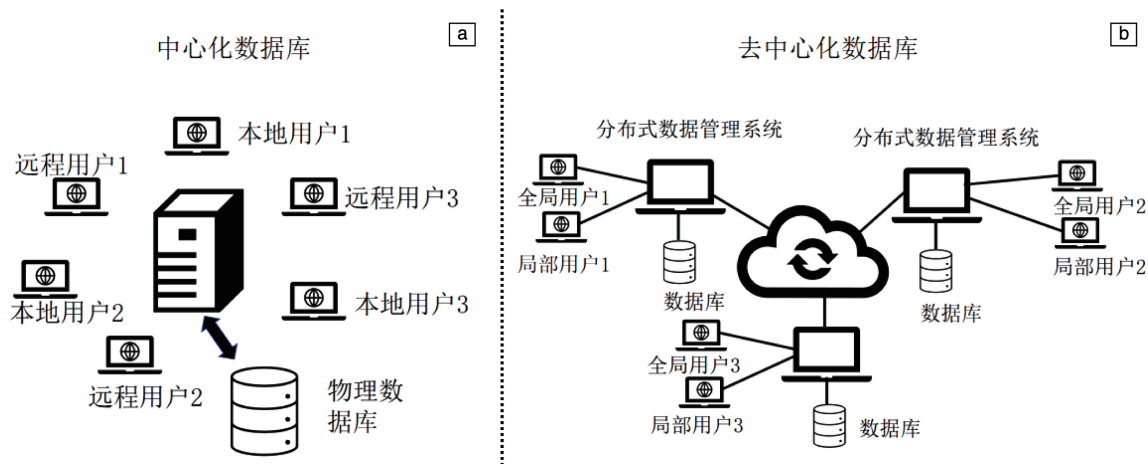


图 3 数据库架构图: (a) 中心化数据库, (b) 去中心化数据库

Fig. 3 Database architecture diagram: (a) centralized database, (b) decentralized database

4.3 隐私计算

隐私计算(privacy computing)^[102]是指在保证数据不对外泄露的前提下,由 2 个或多个参与方联合完成数据分析计算的技术,其主要技术有安全多方计算、联邦学习、可信执行环境等。当材料数据库与隐私计算技术相融合,将为材料科学领域带来革命性的变革。通过横向联邦学习^[103](图 4a),不同材料研发机构的数据可以在保护数据隐私的前提下进行建模和应用,从而解决数据孤岛问题^[104]。这意味着各个机构可以共享数据并

共同建立材料模型,加快材料研发的速度和效率。同样地,通过纵向联邦学习^[103](图 4b),材料生产、制造和应用机构的数据可以在不暴露关键信息的情况下进行聚合建模。该方法与材料区块链技术相结合,可以使材料工艺、性能、成本优化等材料大数据典型应用在保护数据隐私的前提下进行,促进整个材料价值链的协同优化。

4.4 材料大模型

大模型,也称大规模预训练模型^[105],是指拥有大量

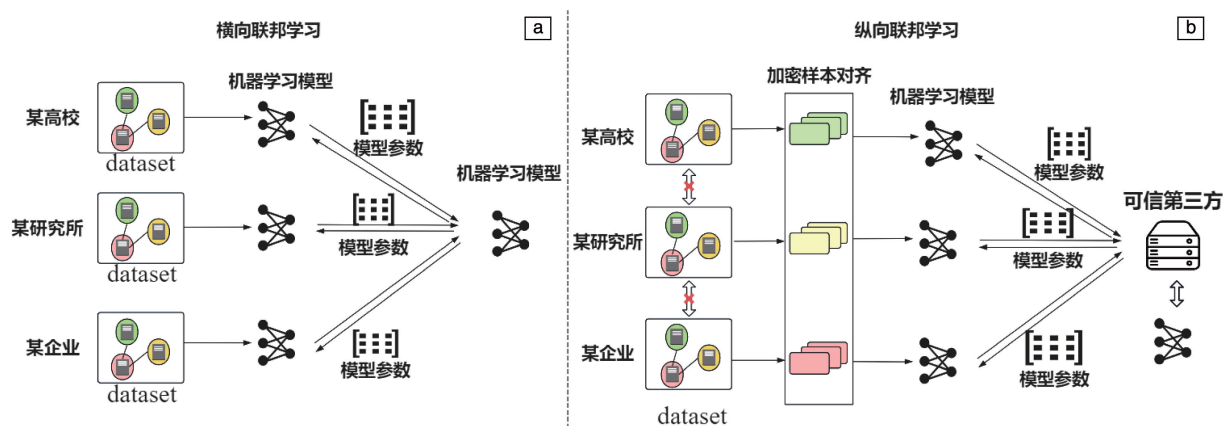


图 4 联邦学习场景图：(a) 横向联邦学习，(b) 纵向联邦学习

Fig. 4 Federated learning scenario diagram; (a) horizontal federated learning, (b) vertical federated learning

参数和复杂结构的深度学习模型，其训练需要大量数据和计算资源。ChatGPT 是一种专注于自然语言处理任务的大规模预训练模型，也叫做大语言模型^[106]，这类模型在文本处理和语言理解方面表现出色。

材料领域已经出现大模型的影子^[107]，2023 年 11 月，DeepMind 推出了 GNOME^[108]，这是一个针对材料科学的 AI 强化学习模型，结合高通量第一性原理计算，成功识别出超过 380 000 种热力学稳定的晶体材料，显著提高了新材料发现的速度。同年 12 月，微软发布了 Matter-Gen^[109]，一个能够根据特定性质需求预测新材料结构的 AI 模型，极大地提升了材料研发的效率。2024 年 1 月，微软联合西北太平洋国家实验室，通过 AI 和高性能计算技术，从数百万种无机材料中筛选并实验验证了一种全

固态电解质材料，实现了从理论预测到实验验证的研究闭环。Tian 等^[110]利用大型语言模型开发了一种钢材设计方法，通过构建 SteelBERT 材料语言编码器和多模态深度学习框架，实现了对钢材力学性能的高精度预测。这些进展表明 AI 大模型在加速材料科学领域的创新和发现中扮演着越来越关键的角色。

在未来，可以基于材料知识库、产品库、标准库、研发资源数据库等数据，训练一个大规模预训练模型，即一个人工智能内容生成技术架构下的全面的、专业化的、闭环的材料科学大模型^[111]。如图 5 所示，这个大模型可以被视为一个超级材料专家，为材料行业提供全生命周期的知识服务、数据服务、计算服务、试验服务和咨询服务。

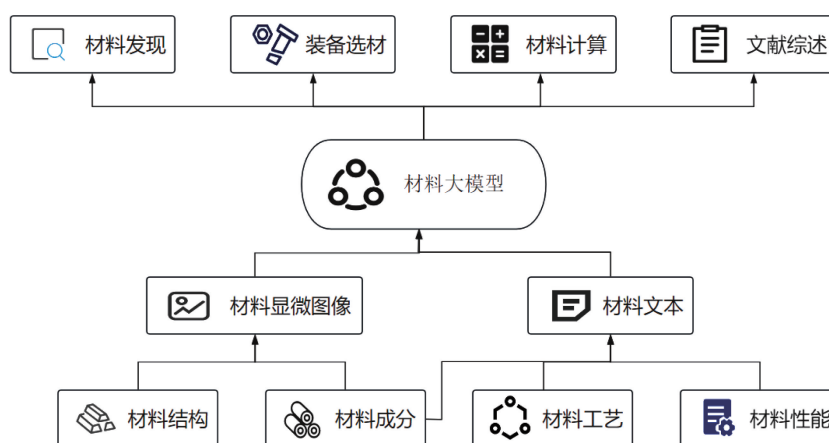


图 5 材料大模型的简易架构图

Fig. 5 Simple architecture diagram of the materials large model

材料科学大模型在未来将扮演重要角色。它可以通过图文和语义交互的方式，提供丰富的专业化应用，如材料发现、装备选材、实验设计、材料计算、图谱解析、

文献综述、解析化建模等。它将整合和利用大量的材料科学数据和知识，提供高效的材料筛选和设计方案，加速材料研发和应用的过程。由于其专业化和闭环化的特

点,它可以在不同领域的材料应用中发挥重要作用,从而推动材料科学的发展和革新。

5 结 语

(1)通过集成先进的 AI 技术,如机器学习和深度学习,材料数据库不仅能够为工程选材、性能评估与优化提供坚实的数据支持,还能在材料与工艺计算方面发挥关键作用,极大地加速新材料的开发与性能预测过程。

(2)目前材料数据库面临的核心挑战在于数据质量与数据共享问题,具体表现为数据的碎片化、主观化、噪音化以及标准化程度不足,难以支撑 AI for Science 对高质量数据的需求。此外,数据的知识产权保护不足、商业运营模式的欠缺严重制约了材料数据的共享与融通,限制了材料数据的有效利用和材料数据库的可持续发展。

(3)未来的材料数据库的发展应积极探索并融合先进技术,例如使用材料数据工厂批量化生产符合 FAIR 原则的材料数据,应用区块链技术增强数据安全性和透明度,通过隐私计算保护敏感数据,利用 AI 大模型为材料行业提供全生命周期服务。这些技术的融合将推动材料数据库向网络化、智能化、资产化和去中心化的方向发展,为材料科学研究和工业应用带来前所未有的创新机遇。

参考文献 References

- [1] HUNT W H. JOM[J], 2006, 58(7): 88.
- [2] WARD C H, WARD C H, WARREN J A. Materials Genome Initiative; Materials Data[M]. Washington, DC: US Department of Commerce, National Institute of Standards and Technology, 2015.
- [3] 宿彦京,付华栋,白洋,等.金属学报[J],2020,56(10):1313-1323.
SU Y J, FU H D, BAI Y, et al. Acta Metallurgica Sinica[J], 2020, 56(10): 1313-1323.
- [4] 杨丽,苏航,柴锋,等.中国材料进展[J],2019,38(7):672-681+650.
YANG L, SU H, CHAI F, et al. Materials China[J], 2019, 38(7): 672-681+650.
- [5] 杨丽,苏航,刘和平,等.第十四届中国钢铁年会论文集[C].北京:冶金工业出版社,2023:58-59.
YANG L, SU H, LIU H P, et al. Proceedings of the 14th CSM Steel Congress[C]. Beijing: Metallurgical Industry Press, 2023: 58-59.
- [6] TEHRANI A M, OLIYNYK A O, PARRY M, et al. Journal of the American Chemical Society[J], 2018, 140(31): 9844-9853.
- [7] STANEV V, OSES C, KUSNE A G, et al. npj Computational Materials[J], 2018, 4(1): 29.
- [8] CHEON G, DUERLOO K A N, SENDEK A D, et al. Nano Letters[J], 2017, 17(3): 1915-1923.
- [9] ONG S P, CHOLIA S, JAIN A, et al. Computational Materials Science[J], 2015, 97: 209-215.
- [10] ONG S P, RICHARDS W D, JAIN A, et al. Computational Materials Science[J], 2013, 68: 314-319.
- [11] JAIN A, ONG S P, CHEN W, et al. Concurrency and Computation: Practice and Experience[J], 2015, 27(17): 5037-5059.
- [12] WANG G, PENG L, LI K, et al. Computational Materials Science[J], 2021, 186: 110064.
- [13] 李霞,苏航,陈晓玲,等.中国冶金[J],2007(6):4-8.
LI X, SU H, CHEN X L, et al. China Metallurgy [J], 2007 (6): 4-8.
- [14] BELSKY A, HELLENBRANDT M, KAREN V L, et al. Acta Crystallographica Section B: Structural Science[J], 2002, 58(3): 364-369.
- [15] HELLENBRANDT M. Crystallography Reviews[J], 2004, 10(1): 17-22.
- [16] YAMAZAKI M, XU Y, MURATA M, et al. Proceedings of the BAL-TICA VII International Conference on Life Management and Maintenance for Power Plants[C]. Helsinki: [s. n.], 2007: 193-207.
- [17] CURTAROLO S, SETYAWAN W, WANG S, et al. Computational Materials Science[J], 2012, 58: 227-235.
- [18] SAAL J E, KIRKLIN S, AYKOL M, et al. JOM[J], 2013, 65(11): 1501-1509.
- [19] 尹海清,刘国权,姜雪,等.科技导报[J],2015,33(10):50-59.
YIN H Q, LIU G Q, JIANG X, et al. Science & Technology Review [J], 2015, 33(10): 50-59.
- [20] 尹海清,姜雪,张瑞杰,等.中国科技资源导刊[J],2016,48(3):58-65.
YIN H Q, JIANG X, ZHANG R J, et al. China Science & Technology Resources Review[J], 2016, 48(3): 58-65.
- [21] 朱昱颖,孙宏伟,朱加雷,等.电焊机[J],2022,52:36-45.
ZHU Y Y, SUN H W, ZHU J L, et al. Welding Machine[J], 2022, 52: 36-45.
- [22] 路勇超,苏航,尹士科.金属功能材料[J],2019(5):14-18.
LU Y C, SU H, YIN S K. Metallic Functional Materials[J], 2019 (5): 14-18.
- [23] FUJITA M, KINUGAWA J, OKADA A, et al. Data Science Journal [J], 2003, 2: 136-145.
- [24] HARIHARAN V, ELUMALAI P, NAMBIRAJ M, et al. Materials Today: Proceedings[J], 2021, 47: 6830-6834.
- [25] ZHANG Y, WEI Y, PU J, et al. The International Journal of Advanced Manufacturing Technology[J], 2021, 112: 273-284.
- [26] HE S, CHEN S, ZHAO Y, et al. Journal of Manufacturing Processes [J], 2021, 63: 121-129.
- [27] 曹拯.基于多元特征的切削数据库系统的研究与开发[D].上海:华东理工大学,2022.
CAI Z. Research and Development of a Multi-Features Based Cutting Database System[D]. Shanghai: East China University of Science and Technology, 2022.
- [28] 胡绕脖.高速切削工艺数据库的研究开发[D].上海:同济大

- 学, 2007.
- HU R B. Research and Development of a Technology Database for High Speed Machining[D]. Shanghai: Tongji University, 2007.
- [29] 曾江. 细长轴切削数据库的构建[D]. 南昌: 南昌航空航天大学, 2016.
- ZENG J. Construction of Cutting Database for Slender Shaft[D]. Nanchang: Nanchang Hangkong University, 2016.
- [30] 赵彦钊. 轴类零件车削工艺数据库系统设计与开发[D]. 北京: 北京林业大学, 2018.
- ZHAO Y Z. Design and Development of Turning Process Database System for Shaft Parts[D]. Beijing: Beijing Forestry University, 2018.
- [31] 陈耀峰. 汽车发动机缸盖切削数据库系统的开发与应用[D]. 上海: 上海交通大学, 2020.
- CHEN Y F. Development and Application of Cutting Database System for Automobile Engine Cylinder Head[D]. Shanghai: Shanghai Jiao Tong University, 2020.
- [32] XU G, ZHANG L, LIU L, *et al.* Journal of Magnesium and Alloys [J], 2016, 4(4): 249-264.
- [33] HE B L, BIN W. Proceedings of the International Conference on Information Engineering for Mechanics and Materials (ICIMM 2012)[C]. Hangzhou: Trans Tech Publications, 2012.
- [34] WATANABE S, KAKEHI K, TAKENO H, *et al.* Tetsu to Hagane—Journal of the Iron and Steel Institute of Japan[J], 2009, 95(2): 124-130.
- [35] CHEN X, CHEN J, RUAN X. Forging & Stamping Technology[J], 2005, 30(4): 92-97.
- [36] OHASHI T. JSME International Journal Series A Solid Mechanics and Material Engineering[J], 2003, 46(3): 432-436.
- [37] GIESE S R. An Investigation of Cast Iron Metal Penetration Defects [M]. Tuscaloosa: The University of Alabama, 1996.
- [38] 李娅娟. 船舶腐蚀防护计算机辅助设计系统[D]. 大连: 大连海事大学, 2007.
- LI Y J. Computer Aided Design System for Ship Corrosion Protection [D]. Dalian: Dalian Maritime University, 2007.
- [39] 刘建容, 张万灵, 蔡捷, 等. 武汉工程职业技术学院学报[J], 2012, 24: 10-13.
- LIU J R, ZHANG W L, CAI J, *et al.* Journal of Wuhan Engineering Institute[J], 2012, 24: 10-13.
- [40] 李海涛, 袁森. 海洋科学[J], 2020, 44: 33-38.
- LI H T, YUAN S. Marine Sciences[J], 2020, 44: 33-38.
- [41] NAUBNOME V, HARYADI G D, ISMAIL R, *et al.* Proceedings of the AIP Conference Proceedings[C]. Melville, New York: AIP Publishing, 2016: 020055.
- [42] 黄云, 李少川, 肖贵坚, 等. 航空材料学报[J], 2021, 41(4): 19.
- HUANG Y, LI S C, XIAO G J, *et al.* Journal of Aeronautical Materials[J], 2021, 41(4): 19.
- [43] CICERO S, GARCÍA T, ÁLVAREZ J A, *et al.* Journal of Constructional Steel Research[J], 2016, 120: 221-231.
- [44] SEDLACEK M, PODORNIK B, VIZINTIN J. Materials Characterization[J], 2008, 59(2): 151-161.
- [45] STEINBECK G, SCHMIDT O. Stahl und Eisen[J], 2006, 126(8): 55-60.
- [46] KIMURA K, YAGUCHI M. Proceedings of the ASME Pressure Vessels and Piping Conference[C]. San Antonio: TX, 2019.
- [47] KOIZUMI Y, YOKOKAWA T, HARADA H, *et al.* Journal of the Japan Institute of Metals[J], 2006, 70(2): 176-179.
- [48] LAMBORN L, ZHANG S, LIMÓN S, *et al.* Proceedings of the ASME 13th International Pipeline Conference (IPC)[C]. New York: ASME, 2020.
- [49] MOSKOVIC R. Engineering Fracture Mechanics [J], 2002, 69(4): 511-530.
- [50] SHI J B, WANG J H, MACDONALD D D. Corrosion Science [J], 2014, 89: 69-80.
- [51] ZHANG X, WANG L, HELWIG J, *et al.* Foundations and Trends[®] in Machine Learning[J], 2025, 18(4): 385-849.
- [52] WANG H, FU T, DU Y, *et al.* Nature [J], 2023, 620(7972): 47-60.
- [53] 李姿昕, 张能, 熊斌, 等. 数据与计算发展前沿[J], 2020, 2: 78-90.
- LI Z X, ZHANG N, XIONG B, *et al.* Frontiers of Data & Computing [J], 2020, 2: 78-90.
- [54] 于志强. 国家行政学院学报[J], 2016(6): 99-103+28.
- YU Z Q. Journal of Chinese Academy of Governance[J], 2016(6): 99-103+28.
- [55] WILKINSON M D, DUMONTIER M, AALBERSBERG I J, *et al.* Scientific Data[J], 2016, 3(1): 160018.
- [56] 中国钢研科技集团有限公司. 材料区块链通则: T/CSTM 01057—2023[S]. 北京: 中关村材料试验技术联盟, 2023.
- China Iron & Steel Research Institute Group. General Principles for Material Blockchain: T/CSTM 01057—2023[S]. Beijing: Zhongguancun Materials Testing Consortium, 2023.
- [57] 王珏, 周志华, 周傲英. 机器学习及其应用[M]. 北京: 清华大学出版社有限公司, 2006.
- WANG Y, ZHOU Z H, ZHOU A Y. Machine Learning and Its Applications[M]. Beijing: Tsinghua University Press Co., Ltd., 2006.
- [58] MONTGOMERY D C, PECK E A, VINING G G. Introduction to Linear Regression Analysis [M]. Hoboken, NJ: John Wiley & Sons, 2021.
- [59] FAN R E, CHANG K W, HSIEH C J, *et al.* The Journal of Machine Learning Research[J], 2008, 91: 871-874.
- [60] POLLARD D. The Annals of Statistics[J], 1981, 9(1): 135-140.
- [61] FREY B J, DUECK D. Science[J], 2007, 315(5814): 972-976.
- [62] De SA J M. Pattern Recognition: Concepts, Methods and Applications [M]. Heidelberg: Springer Science & Business Media, 2012.
- [63] WANG Y, YU C, XING L, *et al.* Metallurgical and Materials Transactions A[J], 2020, 51: 940-950.
- [64] HOTELLING H. Journal of Educational Psychology[J], 1933, 24(6): 417.

- [65] BLEI D M, NG A Y, JORDAN M I. *Journal of Machine Learning Research*[J], 2003, 3: 993–1022.
- [66] GOH G B, HODAS N O, VISHNU A. *Journal of Computational Chemistry*[J], 2017, 38(16): 1291–1307.
- [67] LECUN Y, BENGIO Y, HINTON G. *Nature*[J], 2015, 521(7553): 436–444.
- [68] SHAKHAROVICH G, DARRELL T, INDYK P. *IEEE Transactions on Neural Networks and Learning Systems*[J], 2008, 19(2): 377.
- [69] SAFAVIAN S R, LANDGREBE D. *IEEE Transactions on Systems, Man and Cybernetics*[J], 1991, 21(3): 660–674.
- [70] BURGESS C J. *Data Mining and Knowledge Discovery*[J], 1998, 2(2): 121–167.
- [71] ZHANG Y, WEN C, WANG C, *et al.* *Acta Materialia*[J], 2020, 185: 528–539.
- [72] WARD L, AGRAWAL A, CHOUDHARY A, *et al.* *npj Computational Materials*[J], 2016, 2(1): 1–7.
- [73] CHEN Y, TIAN Y, ZHOU Y, *et al.* *Journal of Alloys and Compounds*[J], 2020, 844: 156159.
- [74] 程晋荣, 何鹏飞, 李艺欣, 等. *中国材料进展*[J], 2025, 44(4): 309–318.
CHENG J R, HE P F, LI Y X, *et al.* *Materials China*[J], 2025, 44(4): 309–318.
- [75] 张闫, 薛德祯, 辛社伟, 等. *中国材料进展*[J], 2025, 44(4): 319–329+318.
ZHANG Y, XUE D Z, XIN S W, *et al.* *Materials China*[J], 2025, 44(4): 319–329+318.
- [76] 张思雨, 李日. *中国材料进展*[J], 2025, 44(11): 1009–1017.
ZHANG S Y, LI R. *Materials China*[J], 2025, 44(11): 1009–1017.
- [77] 崔壮, 刘满平, 曾迎, 等. *稀有金属材料与工程*[J], 2024, 53(8): 2375–2389.
CUI Z, LIU M P, ZENG Y, *et al.* *Rare Metal Materials and Engineering*[J], 2024, 53(8): 2375–2389.
- [78] 孙志军, 薛磊, 许阳明, 等. *计算机应用研究*[J], 2012, 29: 2806–2810.
SUN Z J, XUE L, XU Y M, *et al.* *Application Research of Computers*[J], 2012, 29: 2806–2810.
- [79] 康靛, 米晓希, 王海莲, 等. *材料导报*[J], 2020, 34: 21172–21179.
KANG L, MI X X, WANG H L, *et al.* *Materials Reports*[J], 2020, 34: 21172–21179.
- [80] LECUN Y, BOTTOU L, BENGIO Y, *et al.* *Proceedings of the IEEE*[J], 1998, 86(11): 2278–2324.
- [81] 李勇, 林小竹, 蒋梦莹. *自动化学报*[J], 2018, 44(1): 176–182.
LI Y, LIN X Z, JIANG M Y. *Acta Automatica Sinica*[J], 2018, 44(1): 176–182.
- [82] 邓长银, 张杰. *信息通信*[J], 2018(1): 9–12.
DENG C Y, ZHANG J. *Information and Communications*[J], 2018(1): 9–12.
- [83] JUNG J H, LEE S J, KIM H S. *Materials*[J], 2022, 15(19): 6954.
- [84] AZIMI S M, BRITZ D, ENGSTLER M, *et al.* *Scientific Reports*[J], 2018, 8(1): 2128.
- [85] BENGIO Y, SIMARD P, FRASCONI P. *IEEE Transactions on Neural Networks*[J], 1994, 5(2): 157–166.
- [86] HENG F, GAO J, XU R, *et al.* *Fatigue & Fracture of Engineering Materials & Structures*[J], 2023, 46(5): 1979–1996.
- [87] MEI H, LANG L, YANG X, *et al.* *Metals*[J], 2020, 10(12): 1588.
- [88] FARIZHANDI A A K, MAMIVAND M. *Computational Materials Science*[J], 2023, 223: 112110.
- [89] 中华人民共和国国家质量监督检验检疫总局、中国国家标准化管理委员会. *标准化工作指南第 1 部分: 标准化和相关活动的通用术语: GB/T20000.1—2014*[S]. 北京: 中国标准出版社, 2014.
State General Administration of the People's Republic of China for Quality Supervision and Inspection and Quarantine, Standardization Administration of China. *Guidelines for Standardization—Part 1: Standardization and Related Activities—General Vocabulary: GB/T 20000.1—2014*[S]. Beijing: Standards Press of China, 2014.
- [90] HIMANEN L, GEURTS A, FOSTER A S, *et al.* *Advanced Science*[J], 2019, 6(21): 1900808.
- [91] SCHEFFLER M, AESCHLIMANN M, ALBRECHT M, *et al.* *Nature*[J], 2022, 604(7907): 635–642.
- [92] ANDERSEN C W, ARMIENTO R, BLOKHIN E, *et al.* *Scientific Data*[J], 2021, 8(1): 217.
- [93] 路勇超, 汪洪, 张澜庭, 等. *Ai-ready 材料数据、标准及基础设施*[J/OL]. (2022–11–17)[2024–11–29]. <https://chinaxiv.org/abs/202211.00177>.
LU Y C, WANG H, ZHANG L T, *et al.* *AI-Ready Material Data, Standards and Infrastructure*[J]. (2022–11–17)[2024–11–29]. <https://chinaxiv.org/abs/202211.00177>.
- [94] WANG H, WANG X D, ZHANG L T. *Engineering*[J], 2020, 6: 56–61.
- [95] 汪洪, 项晓东, 张澜庭. *科技导报*[J], 2018, 36(20): 15–21.
WANG H, XIANG X D, ZHANG L T. *Science & Technology Review*[J], 2018, 36(20): 15–21.
- [96] ZHENG Z B, XIE S A, DAI H N, *et al.* *International Journal of Web and Grid Services*[J], 2018, 14(4): 352–375.
- [97] JIAO T, SHEN D R, NIE T Z, *et al.* *Journal of Software*[J], 2019, 30(10): 3110–3127.
- [98] 贺海武, 延安, 陈泽华. *计算机研究与发展*[J], 2018, 55(11): 2452–2566.
HE H W, YAN A, CHEN Z H. *Journal of Computer Research and Development*[J], 2018, 55(11): 2452–2566.
- [99] WANG C, SU H, DUAN L, *et al.* *Processes*[J], 2023, 11(11): 3168.
- [100] JARKE M, KOCH J. *A Survey of Query Optimization in Centralized Database Systems*[J], 1984, 16(2): 111–133.
- [101] XU X A. *Advanced Materials Research*[J], 2014(846): 1435–1438.
- [102] LI F, LI H, NIU B, *et al.* *Engineering*[J], 2019, 5(6): 1179–1192.

- [103] YANG Q, LIU Y, CHEN T, *et al.* ACM Transactions on Intelligent Systems and Technology (TIST)[J], 2019, 10(2): 1-19.
- [104] LI L, FAN Y, TSE M, *et al.* Computers & Industrial Engineering [J], 2020, 149(5): 9106854.
- [105] LIU X, HU B, CHEN K, *et al.* Bulletin of National Natural Science Foundation of China[J], 2023, 37(5): 758-766.
- [106] KALLENS P C, KRISTENSEN-MCLACHLAN R D, CHRISTIANSEN M H. Cognitive Science[J], 2023, 47(3): 13256.
- [107] 闫姿霓, 饶梓元, 曾小勤. 中国材料进展[J], 2025, 44(5): 409-423+450.
- YAN Z N, RAO Z Y, ZENG X Q. Materials China[J], 2025, 44(5): 409-423+450.
- [108] ERCHANT A, BENDELEV S, FUCHS R, *et al.* Nature[J], 2023, 624(7990): 80-85.
- [109] ZENI C, PINSLER R, ZÜGNER D, *et al.* Nature[J], 2025, 639(8055): 624-632.
- [110] TIAN S, JIANG X, WANG W, *et al.* Acta Materialia[J], 2025, 285: 120663.
- [111] 李长泰, 韩旭, 蒋若辉, 等. 工程科学学报[J], 2024, 46(2): 290-305.
- LI C T, HAN X, JIANG R H, *et al.* Chinese Journal of Engineering [J], 2024, 46(2): 290-305.