

文章编号: 1000-7032(2024)08-1219-13

数据驱动研制发光材料的策略与挑战

黄霖, 解荣军*

(厦门大学 材料学院, 福建 厦门 361005)

摘要: 人工智能在高效处理数据、精准预测、自动化执行任务以及个性化服务等方面为人类的生产、生活和科学研究带来了极大的便利。机器学习与高通量计算在材料领域的广泛渗透与成功应用,也为发光材料的研制开辟了新路径——通过算法进行高效挖掘和大规模数据处理,加速新材料的筛选与设计,进而推动材料的创新发现与应用进程。本文综述了近年来基于数据驱动发光材料研究的前沿进展,从相关研究案例出发,对数据驱动材料研究进行全流程梳理,详细阐述发光材料研制场景下数据获取阶段的重要性及实施策略,并对如何提取表征材料性能的核心特征进行分析,同时探讨发光材料领域适用的模型选择与优化方案。最后,就当前数据驱动式发光材料研究面临诸如高质量数据匮乏、复杂结构-性能关联模型构建困难的问题,从发光材料数据库平台的构建、高通量实验的实施以及相应数据生产规范的建立等方向提出突破设想。

关键词: 数据驱动; 发光材料; 机器学习; 高通量计算; 发光材料数据工厂

中图分类号: O482.31

文献标识码: A

DOI: 10.37188/CJL.20240098

Strategies and Challenges of Data-driven Research and
Discovery in Luminescent Materials

HUANG Lin, XIE Rongjun*

(College of Materials, Xiamen University, Xiamen 361005, China)

* Corresponding Author, E-mail: rjxie@xmu.edu.cn

Abstract: Artificial intelligence has been bringing a great of convenience for human in manufacture, life, science and technology, for its abilities of efficient data analyses, accurate prediction, automatic task executing, and personalized service. Machine learning and high-throughput computing have extensively permeated and successfully applied in the field of materials, opening up new horizons for innovative research and design methods in luminescent materials. By employing efficient algorithms for mining and processing large-scale data, the screening and design process of new materials is accelerated, thereby driving the discovery and application progress of novel materials. This article provides an overview of the recent advances in data-driven research on luminescent materials. Based on relevant research cases, it outlines the entire process of data-driven material research and elaborates on the importance and implementation strategies of data acquisition in the development of luminescent materials. It also conducts a thorough analysis of how to accurately extract core features that characterize material performance, while exploring algorithm selection strategies applicable to the field of luminescent materials. Finally, the bottleneck of the current research on data-driven luminescent materials is pointed out, such as a lack of high-quality data and difficulties in constructing complex structure-performance correlation models. It also provides an outlook on future development directions, with a particular emphasis on the construction of luminescent material database platforms, implementation of high-throughput experiments, and establishment of corresponding data production standards.

Key words: data-driven; luminescent materials; machine learning; high-throughput computing; luminescent materials data fab

收稿日期: 2024-04-12; 修订日期: 2024-04-29

基金项目: 国家重点研发计划“稀土新材料”重点专项(2022YFB3503800, 2022YFB3503801)

Supported by the National Key Research and Development Program(2022YFB3503800, 2022YFB3503801)

1 引 言

近年来,荧光粉转换发光二极管(Phosphor converted-light emitting diode, pc-LED)得益于其节能环保、寿命长、体积小等优势,替代白炽灯、卤素灯等传统照明工具成为了新一代照明市场的主流^[1]。作为发光元件主要组成部分之一,大部分荧光粉是由稀土离子或过渡金属离子掺杂到无机材料中受到外部激发后发光,具有稳定性高、发光性能稳定以及价格较低的优点,其中, $\text{CaAlSiN}_3:\text{Eu}^{2+}$ 、 $\text{K}_2\text{SiF}_6:\text{Mn}^{4+}$ 、 $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Ce}^{3+}$ 、 $\beta\text{-SiAlON}:\text{Eu}^{2+}$ 是目前使用最为广泛的几种LED用荧光粉^[2-5]。尽管这些荧光粉的综合性能已使LED照明上升到了更高的水准,但伴随着全球的能源危机,市场在追求更高色彩饱和度、更高亮度的同时,亦追求更低的能耗,这就对固态照明与显示领域中发光材料的能量转换效率、热猝灭以及光谱特性等^[6]也提出了更苛刻的要求。因此,亟需探索新方法、新思路以加速目标发光材料的创制和应用。

新材料的研发思路大致可以分为两类。一类是在不考虑性能的前提下,发现一种未曾报道的材料组分或结构。这类新材料可以通过实验试错法、化学元素替换法、单颗粒检测法^[7]等方法发现。这些方法的运用在一定程度上加快了新材料发现的速度,但其结果存在一定的偶然性。另一类是基于性能导向的新材料研发,设定某个性能目标后,在理论知识、经验的指导下,先缩小筛选范围再进行实验检验。出于过往经验可能存在的局限性,其实验结果时常不达预期。

近些年,人工智能(Artificial intelligence, AI)领域的飞速发展给新材料的研发也带来了新的思路和机遇^[8-11]。机器学习应用于材料领域时,可将新材料研发拆解为数据、特征、算法等问题,通过构建相对应的模型进行学习和训练,最终实现对材料特征特性的预测。尽管在材料科学中,常常不具备数据量十分充足的条件,但通过恰当的数据处理和利用^[12]并结合AI工具,也可加快材料的研制和应用。目前,材料结合机器学习的研究方法已在合金材料、热电材料、高分子材料等材料方向取得了一定的成功^[13-16]。

在发光材料领域也同样出现了与机器学习相结合的研究模式,取得了有趣的实验结果。本文将从数据驱动材料研究的流程出发,结合现有研究案例,分别对数据获取、特征工程、模型的构建

与优化以及材料的预测与研制进行阐述与分析,并基于发光材料领域的特性提出数据驱动式发光材料研究现存的挑战以及未来发展方向。

2 数据驱动式发光材料研究流程

应用数据驱动方法辅助发光材料研发,通常以构建材料的组分、工艺、结构、性能之间的隐性构效关系为策略^[17],进行数据获取、特征构建与筛选、算法模型的搭建与优化以及材料的性能预测与研制。图1为数据驱动式发光材料研究流程示意图,本节从四个主要流程出发,综述了在发光材料领域应用数据驱动方法的研究进展。

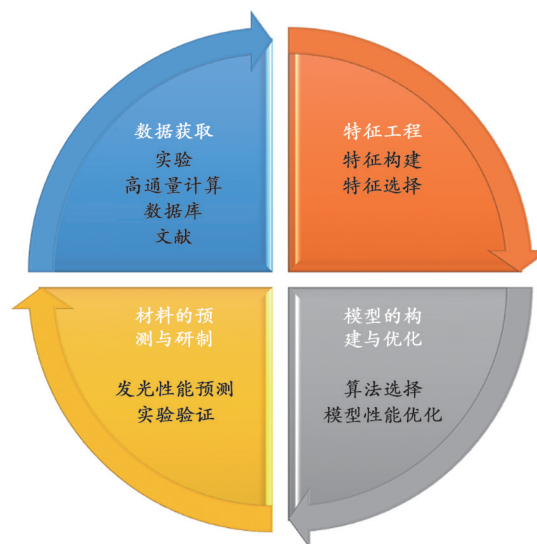


图1 数据驱动式发光材料研究流程

Fig. 1 The workflow of materials machine learning

2.1 数据获取

构建机器学习预测模型所需的数据集内包括输入特征数据与目标性能数据,在目前已报道的相关文献中,两部分数据的来源主要有四种,分别为高通量计算、数据库、文献以及实验。以下将分别对这四种数据获取途径进行阐述分析,并介绍一种尚未应用于发光材料领域的AI-ready数据获取方式,即高通量实验法。

高通量计算为目前构建特征、性能数据集的主流方法,所获数据数量多且具有同源性。其中较成熟且广泛使用的计算方法有基于电子密度的密度泛函理论法(Density functional theory, DFT)^[18-20]、分子动力学法(Molecular dynamics)^[21-22]、蒙特卡洛模拟(Monte Carlo)^[23-24]以及有限元法(Finite elements)^[25-26]等。其中DFT是在多粒子薛定谔方程的基础上,采用绝热近似、单电子近似等方

法,把多粒子问题等效成一个无相互作用的单电子问题,从而使得用量子力学研究实际材料变得切实可行。在各种密度泛函理论程序中,原子尺度材料模拟的计算机程序包(Vienna Ab initio simulation package, VASP)^[27-28]使用最为广泛。美国休斯顿大学 Zhuo 等在 2018 年发表的用机器学习发现热稳定好、高量子效率的无机发光材料一文中^[29],利用禁带宽度、德拜温度作为高量子效率的指征,基于 DFT 计算了 2 071 种材料的禁带宽度,同时参考平均声子速率公式^[30]构建了德拜温度预测模型的数据集,其中块体模量和剪切模量也取自 Materials Project^[31]数据库中基于 DFT 得到的计算结果。

随着高性能计算机的飞速发展以及密度泛函理论的建立,国内外建立了许多大规模材料信息数据库,常用的数据库信息如表 1 所示。其中无机晶体结构数据库(Inorganic crystal structure database, ICSD)^[32]以及开放晶体学数据库(Crystallography open database, COD)^[33]等可以提供的数据信息包括晶体的组成与结构、CIF 文件等,Materials Project^[31]能够提供由计算得到的电子能带结构、相图、散射图谱等相关数据。从数据库中选取数据用于机器学习材料性能模型的搭建是较为便捷的数据获取方式。Amil 等基于 Materials Project^[31]、OQMD^[34]、ICSD^[39]数据库中的原子间势能构建出晶体结构稳定性预测模型 GNoME^[40],成功地在短时间预测出 381 000 种新型稳定晶体结构,进一步证实了数据库提供的大规模数据以及 DFT 计算数据在利用人工智能进行新材料发现时的巨大潜力。但在发光材料领域,目前还没有建立起规模性的发光材料数据库,这也成为机器学习在发光材料领域更进一步应用的一道阻碍。

在发光材料领域,由于发光机理复杂,通过计算手段直接获取相关性能数据较为困难。因此,

科技文献中提供的实验数据也常被用于构建目标性能数据集。例如,2020 年,美国加州大学圣地亚哥分校 Mahdi 等为构建荧光粉热猝灭温度预测模型,从不同文献中提取出 29 个 Eu²⁺或 Ce³⁺掺杂氧化物荧光粉的热猝灭温度组成目标性能数据集^[41]。2022 年,北京科技大学 Jiang 等以高温合成、蓝光激发以及热猝灭测试温度为样本筛选条件,在原有 120 个来自文献的 Ce³⁺掺杂石榴石结构荧光粉样本数据基础上,筛选出 70 个样本组成训练集,用于预测有合适发射波长以及热猝灭温度的材料^[42]。然而,由于目前发光材料的文献数据主要通过人为检索、阅读相关文献获取,数据获取的人力、时间成本较高,所得训练样本数量较少。在金属材料等领域中^[43-44],有研究者采用自然语言处理(Natural language process, NLP)的方法从文献中提取数据作为训练数据构建机器学习预测模型^[45],该方法能够加快提取文献数据的速度。但与此同时,由于文献中的数据其主要作用在于提供事实,不可避免地存在数据的主观性、异构性和团聚性,因此使用文献中的实验数据时需考虑到数据的准确性对模型的最终性能与可信度所产生的影响。

通过实验合成多个材料并测得发光性能数据来构建性能数据集,可以在一定程度上解决数据不同源、异构性的问题。如武汉理工大学 Yuan 等依据不同 x 、 γ 取值合成了 85 个(TDMP)Pb(Br _{x} Cl _{$1-x$})₄: γ Mn 样本,并通过实验测得其光谱、色坐标以及色温,以构建杂化卤化铅荧光粉的显色指数、色温预测模型^[46]。该方法基于同一类材料展开实验获取数据,会使该类型材料性能的预测更加精确,但由于数据涵盖材料范围小,对该体系以外的待预测样本适用性低。

目前,国内外已经出现“无人化学实验室”“自动化实验工厂”^[47]进行大规模的数据生产,进而获取 AI-ready 数据,即专为数据驱动材料研发生产相应数据以构建初始数据集。如图 2 所示,实验数据工厂通过系统化的高通量实验合成以及多参数的性能表征,形成一个具备全面实验设施的科研集成平台,而计算机数据工厂则由配备的科学计算软件及硬件进行规模化的高通量计算。由此,“数据工厂”能够实现快速获取大规模、同源同构和高分散性的材料数据,解决材料基因工程研究中面临的数据供给困境。

表 1 常用数据库

Tab. 1 Databases

数据库	链接	Ref.
ICSD	https://icsd.products.fiz-karlsruhe.de/	[32]
Materials Project	https://next-gen.materialsproject.org/	[31]
COD	https://www.crystallography.net/cod/	[33]
OQMD	https://oqmd.org/	[34]
AFLOW	http://www.aflowlib.org/	[35]
NOMAD	https://nomad-lab.eu/prod/rae/gui/search	[36]
Atom Work	https://crystdb.nims.go.jp/	[37]
MatNavi	https://mits.nims.go.jp/	[38]

图 2 “数据工厂”概念图^[47]Fig. 2 Conceptualization of Data Fab^[47]

2.2 特征工程

机器学习特征工程是指包括特征构建、特征筛选与分析在内的构建能够表征待预测性能的特征数据集的过程,特征也被称为描述符。传统的实验试错法也包含特征构建过程,以经验作为特征构建依据,再对符合经验导向特征的材料进行实验验证。如以“结构刚度较高的材料作为荧光粉基质时通常有更好的热稳定性”为依据^[48],筛选对称性好的结构作为掺杂基质进行实验合成,寻找热稳定性好的材料。这样的经验总结在过去的研究中提供了不少材料研发的新思路,但发光性能的影响因素不唯一,导致经验指导实验的方法考虑因素不够全面。机器学习手段的引入恰能弥补这一不足,通过算法找到多特征与性能之间的潜在关系构建模型,以筛选出更符合预期的材料。但若不具备所预测性能的相关领域知识,盲目选取大量与目标性能不直接相关的特征进行模型构建,算法得到的数学关系由于不具有物理可解释性,模型的泛化能力也存疑。因此,从领域的角度选取具有物理意义的特征量,同时利用机器学习的算力更全面地找到不同结构特征与发光性能间的对应关系,从而实现材料性能的预测,是目前研究人员在以上两大问题中尝试寻找平衡的主要方式。

在掺杂无机发光材料中, $\text{Ce}^{3+}/\text{Eu}^{2+}$ 掺杂材料源于其 5d 电子与基质材料中近邻阴离子配体之间存在强的交互作用,共价键效应、晶体场劈裂和电声子耦合作用可以有效调控其激发光谱和发射光谱,从而赋予 $\text{Ce}^{3+}/\text{Eu}^{2+}$ 掺杂无机固体高效可调的优异性能,为基于组分-结构-性能的构效关系设计新型荧光粉提供了理论依据和可行性^[49]。因此在运用机器学习方法实施新型荧光粉的研发时,通常以相同掺杂离子作为区分点,挖掘 $\text{Ce}^{3+}/\text{Eu}^{2+}$ 掺杂材料中存在的潜在发光规律,实现对未掺杂基质掺杂后发光性能的预测。构建发光性能对应

的特征集也一般以发光机理为切入点,从材料元素组成、结构特性、电子结构等方面选取对发光性能贡献较大的因子作为特征。下文将以两大发光性能为例,综述目前研究中各性能预测模型所选取的特征以及特征构建思路。

2.2.1 发射波长

Ce^{3+} 、 Eu^{2+} 的发光主要来自于其在基质晶体中的 5d-4f 能级跃迁,其发射波长 λ 主要受到激发态 $4f^n-15d^1$ 与基态 $4f^n$ 能级之间的能量差以及 Stokes 位移的影响^[50-51],而如何将其进行量化计算尚未出现一个普适定论。在利用机器学习方法预测发射波长时,合适的特征量将在较大程度上影响模型的适用范围及性能,在不同的模型中对其特征选取方案也有着不同的阐释。

在诸多发射能量 (E_{em}) 的影响因素中,掺杂离子在基质晶体中的局域配位环境对其影响最为直接。图 3(a) 为厦门大学 Li 等构建的基于局域配位的静电参数模型 (Harmonic covalent-electrostatic parameterization, HCEP)^[13],其阐释了除 Madelung potential 以外,质心位移 ϵ_c 、晶体场劈裂 ϵ_{cfs} 以及禁带宽度 E_g 对发射能量 E_{em} 的影响。其中,质心位移表示局域 5d 电子与非局域配位电子之间的相互关系,极化率 α_{sp} 、掺杂离子与配位阴离子之间的配位键长 L_{A-N} 体现了掺杂离子与配位阴离子 (A—N) 之间的键强度,而由于氮化物中的阴离子极化率要大于其他阴离子,这就导致在氮化物中,阳离子与配位阴离子 (A—N) 之间的相互作用以及阳离子与阳离子之间的键长 L_{A-N} 会对质心位移有着比其他体系更为重要的影响。因此,以极化率 α_{sp} 、配位键长 L_{A-N} 、阳离子间键长 L_{A-A} 以及缩合度 κ 作为质心位移的描述符,由阳离子配位多面体的角动量参数 Θ_q^k 来描述晶体场劈裂,最终选取了图 3(b) 所示 7 个参数作为 Eu^{2+} 掺杂氮化物发射波长参数模型的特征量,模型的均方根误差 (Root mean square error, RMSE) 为 0.03 eV,预测

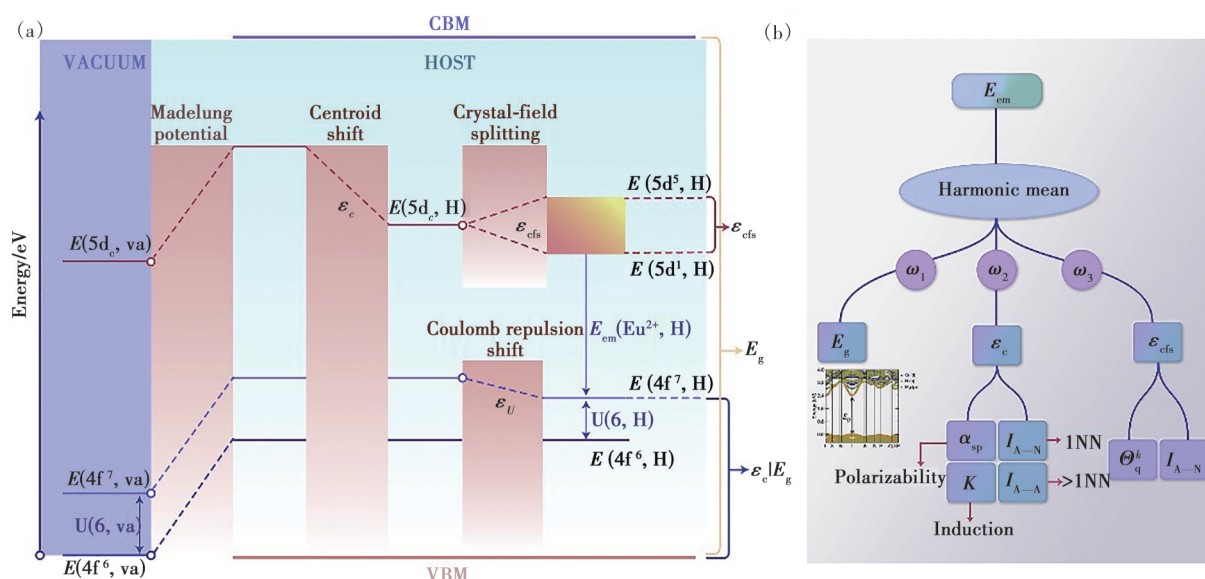


图3 基于局域配位的静电参数模型^[13]。(a) Eu^{2+} 在真空条件下和基质中的能级分布;(b) $\text{Eu}(\text{II})$ 荧光粉发射峰(E_{em})的HCEP模型示意图

Fig. 3 The harmonic covalo-electrostatic parameterization (HCEP) model^[13]. (a) The free (va) and in-crystal (H) energy level diagram of Eu. (b) A schematic of the developed HCEP model for predicting the emission peak (E_{em}) of $\text{Eu}(\text{II})$ phosphors

出具有极长发射波长的 $(\text{Sr}, \text{Ba})_3\text{Li}_4\text{Si}_2\text{N}_6:\text{Eu}^{2+}$ 荧光粉,经实验验证其 $\lambda_{\text{em}}=800\sim 830\text{ nm}$ 且具有 30%~40% 的较高量子效率。

韩国世宗大学 Park 等也构建了 Eu^{2+} 掺杂荧光粉的发射能量预测模型^[52],选择空间群、离子掺杂位点数、掺杂位点对称性、晶胞参数、配位数、平均键长、离子半径以及电负性等 29 个特征构成特征集。其特征选择与上述 HCEP 模型有所不同,这主要是受样本体系的影响,该训练集中有超过 80% 是氧化物,而 HCEP 针对的是氮化物体系。在不同材料体系中,部分特征与目标性能之间的数量关系存在差别,造成不同体系所构建特征集的差异。例如, Dorenbos 综合配体极化、共价性和电子云扩展效应对质心位移的影响,提出了光谱极化率 α_{sp} 的概念^[53],其与平均阳离子电负性 χ_{av} 之间存在以下相关关系:

$$\alpha_{\text{sp}} = \alpha_0 + \frac{b}{\chi_{\text{av}}^2}, \quad (1)$$

其中 α_0 是极大 χ 情况下的极限极化率, b 代表阴离子由于与金属键合而改变其极化率的敏感性。然而,在 Ce^{3+} 掺杂的氮化物、氟化物、氧化物体系中,由于阴离子与金属键合时电子云的极化程度按照由 F 到 O 到 N 的顺序依次增加,使得 α_{sp} 与 χ_{av} 之间的线性关系具有不同的斜率,即 $b^{\text{N}} > b^{\text{O}} > b^{\text{F}}$ ^[54],表明构建多阴离子体系材料的质心位移预测模型时,运用平均阳离子电负性 χ_{av} 作为描述符难以揭示

整体样本的发光构效关系,影响模型的可解释性。

2.2.2 热猝灭温度

热猝灭温度能够反映荧光粉的发光强度随着环境温度升高造成的损耗程度,一般用发光强度降低至初始水平 50% 时的温度 T_{50} 来表示。在 LED 应用中,受芯片激发的影响,器件温度将在使用过程中相应升高,因此一般要求用于 LED 的荧光粉在温度升高至 150 °C 时仍能保持较好的热稳定性。

从数据驱动的角度构建热猝灭温度预测模型,需要从材料热猝灭的相关机理出发,目前得到普遍认可的 $\text{Ce}^{3+}/\text{Eu}^{2+}$ 掺杂无机发光材料的发光猝灭机制或非辐射跃迁过程主要有三种^[48-49,55-56]: 基于位形坐标模型的基态和激发态的交叉弛豫、基于材料电子结构的 5 d 电子到导带的电离化以及杂质和缺陷导致的发光猝灭。以上三种发光猝灭机理被考虑进热猝灭温度预测模型时,分别找到针对性的描述符是实现综合考虑全部机制的关键,但目前材料缺陷难以表征、发光机理复杂使其成为一个难题。美国加州大学圣地亚哥分校 Mahdi 等构建的针对氧化物热猝灭温度的预测模型中^[41],特征集综合考虑了材料的结构特征以及热猝灭现象的物理机制,并提出一个新的描述符——激活离子环境分布 (Activator environment distribution, AED) 来表征掺杂离子局域环境稳定性。通过分子动力学方法分别计算了 300 K 和

500 K 下的 AED 值,发现其差值 ΔAED 与实验 TQ 值之间存在线性关系。在 Dorenbos 提出的热离化势垒模型^[55,57-58]基础上,如公式(2)所示,利用密度泛函理论计算量进行修正,并结合 ΔAED 得到如下热猝灭温度(T_Q)预测公式(3)^[41]:

$$E_a^i(A^{x+}, H) = E_g - \left[E(A^{x+}, \text{free}) - \varepsilon_c(A^{x+}, H) - \frac{\varepsilon_{\text{cfs}}(A^{x+}, H)}{\gamma(H)} + \varepsilon_s(A^{x+}, \text{free}) \right] - E(4f^n - \text{VBM}),$$
$$T_Q = 1 - (1 - K \cdot \Delta\text{AED}) \frac{1 + \Gamma \cdot \frac{-E_a^i}{k_B T_1}}{1 + \Gamma \cdot \frac{-E_a^i}{k_B T_2}},$$

(2)

(3)

其中, E_g 是基质禁带宽度, $E(A^{x+}, \text{free})$ 为修正后的自由离子质心位置, $\varepsilon_c(A^{x+}, H)$ 指 A^{x+} 在基质 H 中的质心位移, $\varepsilon_{\text{cfs}}(A^{x+}, H)$ 为在基质 H 中的晶体场劈裂, $\gamma(H)$ 表示晶体场劈裂和晶体场位移的比率, $\varepsilon_s(A^{x+}, \text{free})$ 表示质心位置与自由离子在最低 5d 能级的能量差, $E(4f^n - \text{VBM})$ 表示离子在 $4f^n$ 与最高价带 VBM 之间的能量差。公式(2)中, k_B 为玻尔兹曼常数, T_1 、 T_2 分别表示初始温度和最终温度, Γ 是热猝灭尝试速率与 5d 态下辐射衰减率的比

值, E_a^i 是 Dorenbos 公式^[55]中的热离化势垒。最终,通过计算获得 29 个样本的禁带宽度、晶体场劈裂、质心位移值作为描述符,构建热猝灭温度预测模型,其均方根误差(RMSE)约为 3.1%~7.6%。

在上述案例中,特征构建从物理模型出发,针对性地筛选出数量较小的特征量,无需再进行特征删减处理。但在一些案例中,为尽可能囊括到更多的特征以发现其中的相关性,会产生数量大于样本个数的特征集。Zhuo 等在构建 Eu^{3+} 掺杂氧化物发光材料的热猝灭温度预测模型时^[59],得到了包含 178 种特征的特征集,但由于特征数量大于材料样本个数时容易造成模型的过拟合,从而降低模型性能,因此采用了递归特征消除法(Recursive feature elimination, RFE)对比不同特征数目下模型的性能。如图 4(a)所示,找到最优特征个数为 51 个,并对各个特征的权重进行排序比较(图 4(b)),剔除掉不重要的特征,最终得到决定系数(R^2)为 0.71 的预测模型,并对预测结果中的五种材料进行实验合成及性能表征,分别是 $\text{Sr}_2\text{ScO}_3\text{F}:\text{Eu}^{3+}$ 、 $\text{Cs}_2\text{MgSi}_5\text{O}_{12}:\text{Eu}^{3+}$ 、 $\text{Ba}_2\text{P}_2\text{O}_7:\text{Eu}^{3+}$ 、 $\text{LiBaB}_9\text{O}_{15}:\text{Eu}^{3+}$ 和 $\text{Y}_3\text{Al}_5\text{O}_{12}:\text{Eu}^{3+}$,发现实验结果与预测值相吻合,具有良好的热稳定性和合适的激发波长,有望应用于 pc-LED。

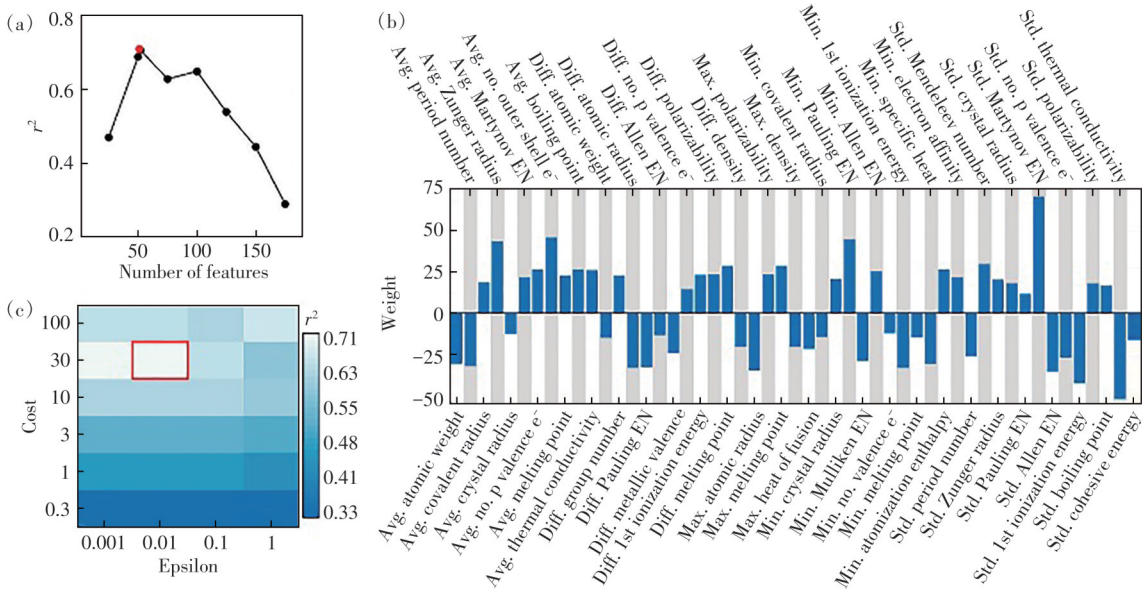


图 4 (a)特征个数与 r^2 的关系;(b)各个特征对应权重;(c)寻找最佳参数组合^[59]

Fig. 4 (a) Coefficient of determination (r^2) as a function of the number of features. The red dot is showing the best r^2 . (b) Weight assigned to each feature. (c) Plot of parameter optimization for cost and epsilon. The optimized cost and epsilon are highlighted by the red box^[59]

2.3 模型构建与优化

机器学习预测任务可以分为回归预测和分类

预测。回归模型用于预测输出具体的、连续的数值,目的是寻找一个最优的模型与数据拟合。分

类模型则根据非连续、离散的数据进行二分类或多分类,以寻找将数据分类的决策边界。在对发光材料进行性能预测时,通常需要得到一个具体的数值,如发射波峰的峰位或者热猝灭温度值等。因此,在多数情况下将用到回归模型进行拟合。回归模型常见的有 Lasso 回归、线性回归(Linear regression)、岭回归(Ridge regression)、逻辑回归(Logistic regression)、随机森林回归(Random forest regression)和支持向量回归(Super vector regression, SVR)等,不同模型核心区别在于处理过拟合的方式不同^[60],通常需要对已有数据集应用于不同算法拟合时进行交叉验证后的性能对比,并对具体模型进行超参数调节。图 4(c)所示为在 SVR 模型中,通过对比超参数 cost 和 epsilon 进行不同组合时模型的评分(r^2),找到最佳参数组合为 cost:30, epsilon:0.01,对其余模型进行超参数调节与之同理。

机器学习的目的在于使习得的模型可迁移至新样本,即具有合适的泛化能力。当模型的学习能力过强时,会将训练样本自身的特殊性质当作一般性质进行学习,导致模型泛化能力较弱,出现“过拟合”的情况,一般表现为训练集的性能明显优于测试集。“欠拟合”即学习器没有学习到训练样本的一般特性,训练出的模型即便在训练样本中表现也不理想。因此,为了判断拟合后的回归模型泛化能力,会利用留出法、交叉验证法、自助法等手段将数据集进行训练集与测试集的划分,分开进行性能评估。常用的回归模型评估指标如表 2 所示,其中除决定系数(R^2)的最优值为 1 以外,其余评估指标值为 0 时代表误差最小。图 5 为韩国世宗大学 Park 等选择模型时对不同类型的机器学习算法进行交叉验证后的均方误差(Mean square error, MSE)(图 5(a))、 R^2 (图 5(b))评分对比^[52]。由图中可以看出训练集的表现平均优于测试集,即出现一定的过拟合现象。通过对比图中不同算法模型的性能评分,综合筛选出 4 个误差处于可接受范围内的算法模型,其均符合 $R^2>$

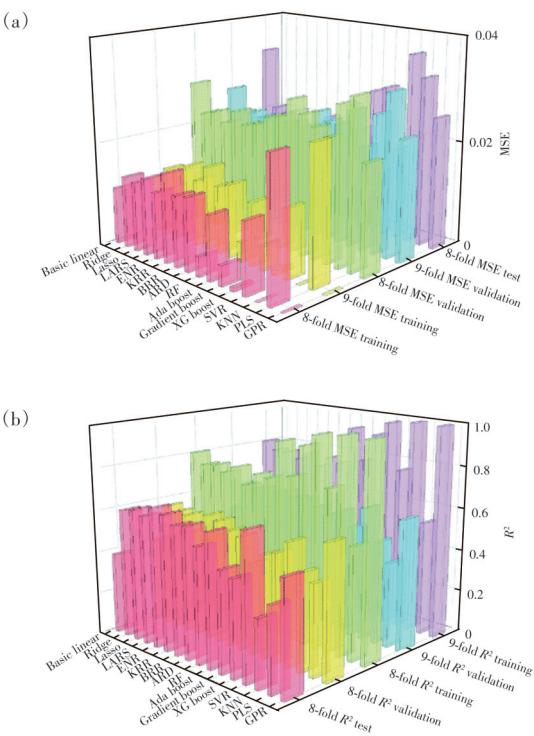


图 5 训练集、验证集和留出法测试数据集分别在 9 折交叉验证和用留出法的 8 折交叉验证后, MSE 分数(a)和 R^2 分数(b)^[52]

Fig. 5 The training, validation and hold-out dataset test results for PEW prediction in terms of MSE (a) and R^2 (b) for 9 cross validation and 8-fold cross validation with a holdout dataset test^[52]

0.6, $MSE<0.02$, 过拟合指数(测试集 R^2 /训练集 R^2) >0.77 , 最终发射峰波长、激发带边缘波长的预测结果选取 4 个模型的平均值。

在材料的实际应用场景中,综合性能是决定其能够投入工程应用的关键,如发光材料运用于显示领域时需要同时满足高量子效率、窄半峰宽、良好热稳定性等。但由于材料性能影响因素众多,性能间的影响因素交错,通过机器学习预测得到具有良好综合性能的材料存在一定困难^[17]。Jiang 等运用多目标优化(Multi-objective optimization, MOO)的手段^[61],尝试在实现长波长、高量子效率双目标的同时优化帕累托前沿(Pareto front),同时运用主动学习的策略,预测具有最大收益的数据点进行实验验证,将验证数据反馈到机器学习模型进行再训练,经 4 次迭代训练及 16 次实验验证(如图 6 所示),得到发射峰位为 763 nm、内/外量子效率分别为 90%/53.1% 的近红外荧光粉材料 $Gd_3Mg_mAl_{2-m}Ga_{3-m}Ge_mO_{12}:0.04Cr^{3+}(m=0.5)$ 。主动学习策略还被应用于光伏材料^[62]、生物材料^[63]、

表 2 回归模型性能评估方法		
Tab. 2 Evaluation method of regression model performance		
评估指标	描述	最优值
Mean absolute error(MAE)	平均绝对值误差	0
Mean square error(MSE)	均方误差	0
Root-mean-square error(RMSE)	均方根误差	0
R -squared(R^2, r^2)	决定系数	1

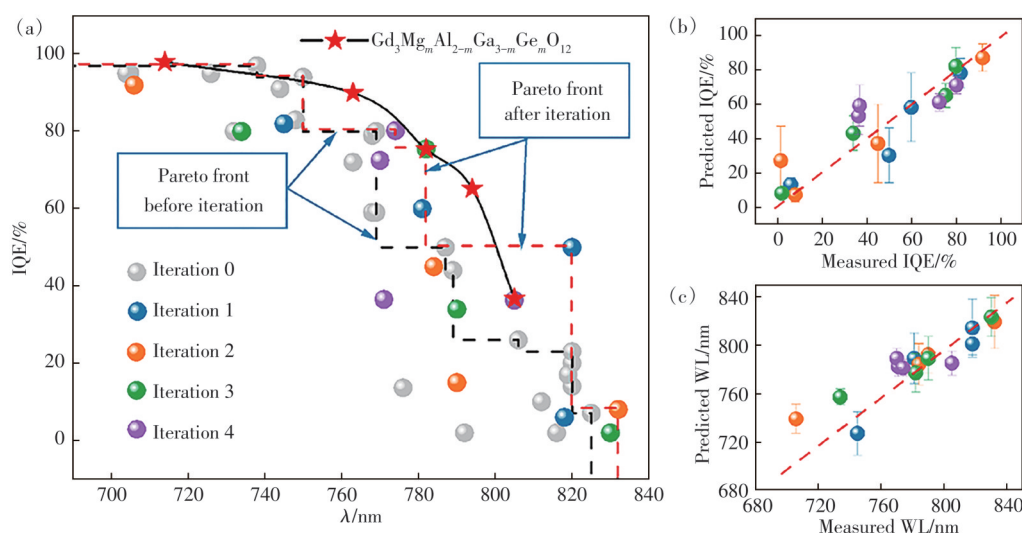


图 6 荧光粉内量子效率(IQE)与发射波长(WL)训练集中的迭代反馈循环。(a)训练集和迭代结果;(b)预测量子效率值与实验值对比;(c)预测发射波长与实验值对比^[61]

Fig. 6 IQE and WL of phosphors in the training set and from the iterative feedback loops. (a) Training set and outcomes from iterations. Predicted IQE(b) and WL(c) versus measured values^[61]

电池材料^[64]和压电材料及储能介质^[65-66]等方面,能够较快提升机器学习模型的准确性。

2.4 材料的预测与研制

机器学习预测模型的构建与优化为筛选潜在高性能发光材料提供了直观依据。如 Zhuo 等通过机器学习构建了德拜温度(Θ_D)的支持向量回

归模型以及基于 DFT 计算的禁带宽度(E_g)预测模型^[29]。有研究表明,结构刚性有助于抑制发光猝灭从而实现较高的发光效率,而德拜温度可用于判定材料的结构刚性^[67],因此从 2 071 个化合物中筛选同时满足德拜温度高、禁带宽度宽的材料作为 Eu^{2+} 掺杂的基质(如图 7(a)所示)。图 7(b)~(e)

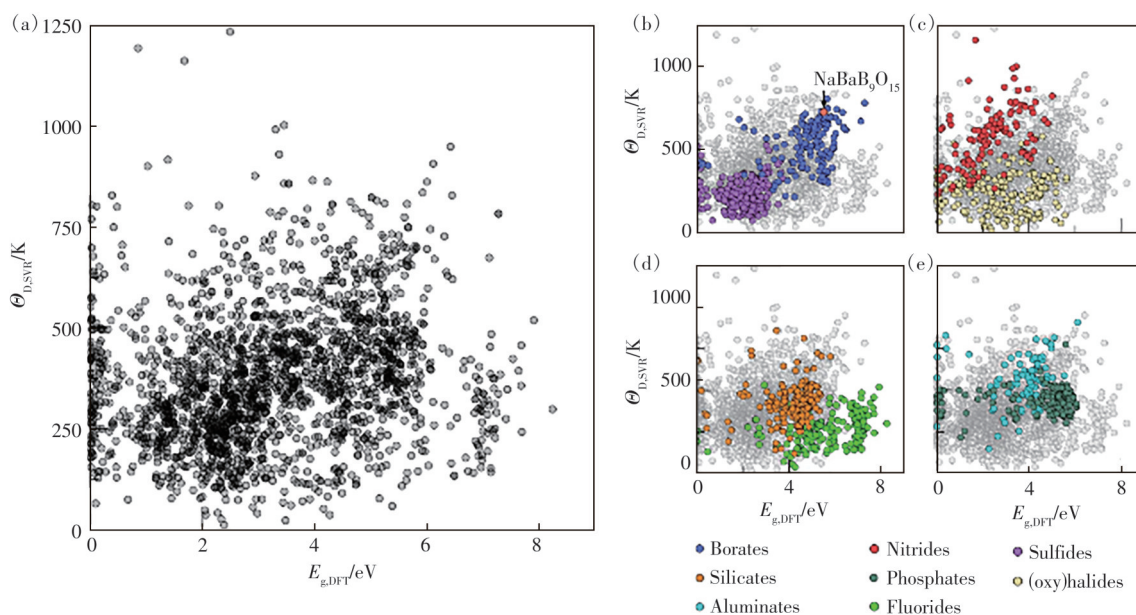


图 7 德拜温度的机器学习预测结果与禁带宽度计算值。(a)机器学习预测的 2 071 种化合物的德拜温度($\Theta_{D,SVR}$)和密度泛函理论计算的禁带宽度,较暗区域表示有数据重叠,同类型的荧光粉基质被标亮,包括硼酸盐和硫化物(b)、氮化物和(氧)卤化物(c)、硅酸盐和氟化物(d)、铝酸盐和磷酸盐(e)^[29]

Fig. 7 Machine learning predicted Debye temperature against calculated bandgap. (a) Machine learning predicted Debye temperature($\Theta_{D,SVR}$) against the density functional theory calculated bandgap($E_{g,DFT}$) for 2 071 compounds predicted. The darker regions occur where there are overlapping data. Classes of common phosphor hosts are highlighted, including borates and sulfides(b), nitrides and (oxy)halides(c), silicates and fluorides(d), aluminates and phosphates(e)^[29]

分别表示不同化合物所属类型的德拜温度与禁带宽度的预测值,由预测结果得知,图7(b)中的 $\text{NaBaB}_9\text{O}_{15}$ 的预测德拜温度达729 K,计算禁带宽度为5.5 eV,为潜在的具有良好发光性能的材料。经实验证明, $\text{NaBaB}_9\text{O}_{15}$ 由于其结构中存在独特的共角 $[\text{B}_3\text{O}_7]^{5-}$ 多阴离子骨架,具有较好的结构刚性。通过掺杂不同浓度的 Eu^{2+} ,对 $\text{NaBa}_{1-x}\text{Eu}_x\text{B}_9\text{O}_{15}$ ($x=0, 0.005, 0.01, 0.02, 0.03, 0.04, 0.05$)进行实验合成及光学表征,得到掺杂浓度为3%时,在紫外光激发下量子效率最高达95%,发射峰位于416 nm,并具有较窄的半高宽34.5 nm。

东南大学Lu等利用后位消除法(Last-place elimination)结合梯度增强回归(Gradient boosted regression)用于新型卤化物有机-无机钙钛矿材料(Hybrid organic inorganic perovskites, HOIPs)的发现^[68],通过第一性原理计算了212个HOIPs的禁带宽度构成目标性能数据集,运用后位特征消除法得到14个重要特征构成特征集,经机器学习算法训练得到HOIPs禁带宽度预测模型。通过替换钙钛矿化学式 ABX_3 ($X=\text{F}, \text{Cl}, \text{Br}, \text{I}$)中的A、B格位阳离子,得到5 504种可能存在的HOIPs结构,并对潜在结构进行稳定性计算,根据应用需求进行禁带宽度范围筛选,排除毒性材料,最终得到6种无毒、稳定的适用于太阳能电池的HOIPs材料。

3 挑战与展望

人工智能时代,数据已成为一种可生产、可获取、可再资源的资源。发光材料领域数据具有数量较少、以实验数据为主的特点,而大部分已报道数据分散于期刊文献或专利中,批量获取困难,测试条件的差异也导致其被再利用时标准程度不一。因此,目前符合科学数据FAIR原则的发光材料领域数据较少,即难以同时满足可发现(Findable)、可获取(Accessible)、可互操作(Interoperable)和可再利用(Reusable)^[69]属性。

建立“发光材料数据工厂”进行数据资源的生产,即在具备全套发光材料合成、分析与测试设施的实验条件下,搭建发光材料实验数据工厂进行规模化的高通量实验,将变革数据作为研究副产品的传统收集方式,实现流水线般以标准化的方式批量产出AI-ready发光材料数据。为保证高通量实验数据满足科学数据FAIR原则,需要建立起发光材料领域针对性的“数据生产规范”,为材料的发光性能测试制定统一标准。包括但不限于规

定所用仪器需提供第三方认证其满足仪器精确度标准的证明、测试热稳定性时需固定统一的温度变化区间、提供统一标准温度下的光谱、量子效率与荧光寿命等发光性能数值。

构建发光材料数据库平台,通过对不同来源的数据及其出处、实验条件与可靠性等信息的集成管理,让已有数据可被更合理、简便地获取与再利用。同时,允许满足“数据生产规范”的数据被自主上传至数据库平台,提供批量获取数据资源并用于机器学习材料研发的端口,实现发光材料在与人工智能相结合时全球AI-ready数据的共通、互信、共享,推动发光材料的智能化研发。

在过去,受碍于有效样本数目小且发光机理复杂,构建单一材料体系的发光性能预测模型准确度更高,且具有可解释性,但很难迁移至不同材料体系,适用范围受限。而难以构建适用于不同体系材料的预测模型的关键在于描述符,一部分发光性能重要描述符在不同材料体系中表现出的性能相关性存在较大差异,如何找到一个能体现并权衡材料体系间差异的发光性能描述符,暂无完美应对之策。

在机器学习算法的选择上,尽管目前计算机专业领域也在不断地开发适用于小样本的算法模型,但从材料领域现已发表的机器学习结合材料研发的相关研究来看,大部分人优先使用传统机器学习算法,侧面反映跨领域进行学科交叉研究时,客观存在的专业壁垒给材料领域与前沿性人工智能的结合带来的困难。但这也启示从事该交叉研究的人员要关注并跟进新兴的技术出现,注重与人工智能领域专业人员的交流与分工合作。

4 结 论

结合人工智能进行材料交叉研究已取得不少令人瞩目的成就,其强大的计算力和学习力将人类的研究能力带到了新的高度。机器学习与高通量计算的运用,在加快发光材料研制、探明发光构效关系等方面体现出了巨大的潜力。本文基于数据驱动材料研制的各个流程,综述了机器学习辅助发光材料研发的重要研究进展及现存的问题与挑战。数据是数据驱动研发的基础,高通量实验的开展与实施将促进大规模数据资源的生产,发光材料数据库的构建将提供极为便捷地获取数据的方式,进而推动机器学习在发光材料领域的广泛应用。发展具有可解释性的预测模型,提升模

型在不同材料体系中的普适性,也将进一步推动材料发光机理和规律的挖掘,成为未来数据驱动发光材料研究的重点发展方向。

本文专家审稿意见及作者回复内容的下载地址:
<http://cjl.lightpublishing.cn/thesisDetails#10.37188/CJL.20240098>.

参 考 文 献:

- [1] PUST P, SCHMIDT P J, SCHNICK W. A revolution in lighting [J]. *Nat. Mater.*, 2015, 14(5): 454-458.
- [2] HIROSAKI N, XIE R J, KIMOTO K, *et al.* Characterization and properties of green-emitting β -SiALON: Eu²⁺ powder phosphors for white light-emitting diodes [J]. *Appl. Phys. Lett.*, 2005, 86(21): 211905.
- [3] UHEDA K, HIROSAKI N, YAMAMOTO Y, *et al.* Luminescence properties of a red phosphor, CaAlSiN₃: Eu²⁺, for white light-emitting diodes [J]. *Electrochem. Solid State Lett.*, 2006, 9(4): H22-H25.
- [4] YOSHIMURA K, FUKUNAGA H, IZUMI M, *et al.* Achieving superwide-color-gamut display by using narrow-band green-emitting γ -AlON: Mn, Mg phosphor [J]. *Jpn. J. Appl. Phys.*, 2017, 56(4): 041701.
- [5] WANG L, WANG X J, KOHSEI T, *et al.* Highly efficient narrow-band green and red phosphors enabling wider color-gamut LED backlight for more brilliant displays [J]. *Opt. Express*, 2015, 23(22): 28707-28717.
- [6] DOE BTO Solid-state Lighting Program. 2022 DOE SSL R&D opportunities [R/OL]. 2022-02-02. <https://www.energy.gov/eere/ssl/articles/doe-publishes-2022-solid-state-lighting-rd-opportunities>.
- [7] HIROSAKI N, TAKEDA T, FUNAHASHI S, *et al.* Discovery of new nitridosilicate phosphors for solid state lighting by the single-particle-diagnosis approach [J]. *Chem. Mater.*, 2014, 26(14): 4280-4288.
- [8] RAMPRASAD R, BATRA R, PILANIA G, *et al.* Machine learning in materials informatics: recent applications and prospects [J]. *Npj Comput. Mater.*, 2017, 3: 54.
- [9] BUTLER K T, DAVIES D W, CARTWRIGHT H, *et al.* Machine learning for molecular and materials science [J]. *Nature*, 2018, 559(7715): 547-555.
- [10] SCHLEDER G R, PADILHA A C M, ACOSTA C M, *et al.* From DFT to machine learning: recent approaches to materials science: a review [J]. *J. Phys. Mater.*, 2019, 2(3): 032001.
- [11] SCHMIDT J, MARQUES M R G, BOTTI S, *et al.* Recent advances and applications of machine learning in solid-state materials science [J]. *Npj Comput. Mater.*, 2019, 5: 83.
- [12] BATRA R, SONG L, RAMPRASAD R. Emerging materials intelligence ecosystems propelled by machine learning [J]. *Nat. Rev. Mater.*, 2020, 6(8): 655-678.
- [13] LI S X, AMACHRAA M, CHEN C, *et al.* Efficient near-infrared phosphors discovered by parametrizing the Eu(II) 5d-to-4f energy gap [J]. *Matter*, 2022, 5(6): 1924-1936.
- [14] WU S, KONDO Y, KAKIMOTO M A, *et al.* Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm [J]. *Npj Comput. Mater.*, 2019, 5: 66.
- [15] WANG T, ZHANG C, SNOUSSI H, *et al.* Machine learning approaches for thermoelectric materials research [J]. *Adv. Funct. Mater.*, 2020, 30(5): 1906041.
- [16] IWASAKI Y, SAWADA R, STANEV V, *et al.* Identification of advanced spin-driven thermoelectric materials *via* interpretable machine learning [J]. *Npj Comput. Mater.*, 2019, 5: 103.
- [17] 谢建新, 宿彦京, 薛德祯, 等. 机器学习在材料研发中的应用 [J]. *金属学报*, 2021, 57(11): 1343-1361.
XIE J X, SU Y J, XUE D Z, *et al.* Machine learning for materials research and development [J]. *Acta Metall. Sinica*, 2021, 57(11): 1343-1361.
- [18] ADAMO C, BARONE V. Toward reliable density functional methods without adjustable parameters: the PBE0 model [J]. *J. Chem. Phys.*, 1999, 110(13): 6158-6170.
- [19] GRIMME S, ANTONY J, EHRLICH S, *et al.* A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu [J]. *J. Chem. Phys.*, 2010, 132(15): 154104.
- [20] CAR R, PARRINELLO M. Unified approach for molecular dynamics and density-functional theory [J]. *Phys. Rev. Lett.*, 1985, 55(22): 2471-2474.
- [21] KRESSE G, FURTHMÜLLER J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set [J]. *Phys. Rev. B*, 1996, 54(16): 11169-11186.

- [22] KRESSE G, FURTHMÜLLER J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set [J]. *Comput. Mater. Sci.*, 1996, 6(1): 15-50.
- [23] JAMES F. Monte carlo theory and practice [J]. *Rep. Prog. Phys.*, 1980, 43(9): 1145-1189.
- [24] ANDRIEU C, DE FREITAS N, DOUCET A, *et al.* An introduction to MCMC for machine learning [J]. *Mach. Learn.*, 2003, 50(1-2): 5-43.
- [25] MELENK J M, BABUSKA I. The partition of unity finite element method: basic theory and applications [J]. *Comput. Methods Appl. Mech. Eng.*, 1996, 139(1-4): 289-314.
- [26] BRENNER S C. Preconditioning complicated finite elements by simple finite elements [J]. *SIAM J. Sci. Comput.*, 1996, 17(5): 1269-1274.
- [27] PERDEW J P, BURKE K, ERNZERHOF M. Generalized gradient approximation made simple [J]. *Phys. Rev. Lett.*, 1997, 78(7): 1396.
- [28] VIRTANEN P, GOMMERS R, OLIPHANT T E, *et al.* Author correction: SciPy 1.0: Fundamental algorithms for scientific computing in python [J]. *Nat. Methods*, 2020, 17(3): 352.
- [29] ZHUO Y, TEHRANI A M, OLIYNYK A O, *et al.* Identifying an efficient, thermally robust inorganic phosphor host *via* machine learning [J]. *Nat. Commun.*, 2018, 9(1): 4377.
- [30] ANDERSON O L. A simplified method for calculating the debye temperature from elastic constants [J]. *J. Phys. Chem. Solids*, 1963, 24(7): 909-917.
- [31] JAIN A, ONG S P, HAUTIER G, *et al.* Commentary: the materials project: a materials genome approach to accelerating materials innovation [J]. *Apl Mater.*, 2013, 1(1): 011002.
- [32] ZAGORAC D, MÜLLER H, RUEHL S, *et al.* Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features [J]. *J. Appl. Crystallogr.*, 2019, 52(5): 918-925.
- [33] GRAŽULIS S, CHATEIGNER D, DOWNS R T, *et al.* Crystallography open database: an open-access collection of crystal structures [J]. *J. Appl. Crystallogr.*, 2009, 42(4): 726-729.
- [34] KIRKLIN S, SAAL J E, MEREDIG B, *et al.* The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies [J]. *Npj Comput. Mater.*, 2015, 1: 15010.
- [35] CURTAROLO S, SETYAWAN W, WANG S D, *et al.* AFLOWLIB.ORG: a distributed materials properties repository from high-throughput *ab initio* calculations [J]. *Comput. Mater. Sci.*, 2012, 58: 227-235.
- [36] DRAXL C, SCHEFFLER M. The NOMAD laboratory: From data sharing to artificial intelligence [J]. *J. Phys. Mater.*, 2019, 2(3): 036001.
- [37] XU Y B, YAMAZAKI M, VILLARS P. Inorganic materials database for exploring the nature of material [J]. *Jpn. J. Appl. Phys.*, 2011, 50(11S): 11RH02.
- [38] YAMAZAKI M, XU Y, MURATA M, *et al.* Nims structural materials databases and cross search engine-MatNavi. [C]. *BALTICA Conference on Life Management And Maintenance for Power Plants, Jyväskylä*, 2007: 193-207.
- [39] 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 Crystallogr. B*, 2002, 58(3): 364-369.
- [40] MERCHANT A, BATZNER S, SCHOENHOLZ S S, *et al.* Scaling deep learning for materials discovery [J]. *Nature*, 2023, 624(7990): 80-85.
- [41] AMACHRAA M, WANG Z B, CHEN C, *et al.* Predicting thermal quenching in inorganic phosphors [J]. *Chem. Mater.*, 2020, 32(14): 6256-6265.
- [42] JIANG L P, JIANG X, ZHANG Y, *et al.* Multiobjective machine learning-assisted discovery of a novel cyan-green garnet: Ce phosphors with excellent thermal stability [J]. *ACS Appl. Mater. Interfaces*, 2022, 14(13): 15426-15436.
- [43] CRUSE K, BAIBAKOVA V, ABDELSAMIE M, *et al.* Text mining the literature to inform experiments and rationalize impurity phase formation for BiFeO₃ [J]. *Chem. Mater.*, 2024, 36(2): 772-785.
- [44] WANG W R, JIANG X, TIAN S H, *et al.* Alloy synthesis and processing by semi-supervised text mining [J]. *Npj Comput. Mater.*, 2023, 9: 183.
- [45] LIU Y, GE X Y, YANG Z W, *et al.* An automatic descriptors recognizer customized for materials science literature [J].

- J. Power Sources*, 2022, 545: 231946.
- [46] YUAN H L, QI L Y, PARIS M, *et al.* Machine learning guided design of single-phase hybrid lead halide white phosphors [J]. *Adv. Sci.*, 2021, 8(19): 2101407.
- [47] WANG H, XIANG X D, ZHANG L T. On the data-driven materials innovation infrastructure [J]. *Engineering*, 2020, 6(6): 609-611.
- [48] DANG P P, WANG W, LIAN H Z, *et al.* How to obtain anti-thermal-quenching inorganic luminescent materials for light-emitting diode applications [J]. *Adv. Opt. Mater.*, 2022, 10(6): 2102287.
- [49] 王淑欣, 宋振, 刘泉林. $\text{Ce}^{3+}/\text{Eu}^{2+}$ 掺杂无机发光材料构效关系与唯象理论 [J]. *发光学报*, 2022, 43(9): 1319-1339.
- WANG S X, SONG Z, LIU Q L. Structure-property correlation and phenomenological theory in $\text{Ce}^{3+}/\text{Eu}^{2+}$ -doped inorganic luminescent materials [J]. *Chin. J. Lumin.*, 2022, 43(9): 1319-1339.
- [50] WANG S X, SONG Z, LIU Q L. Recent progress in $\text{Ce}^{3+}/\text{Eu}^{2+}$ activated LEDs and persistent phosphors: focusing on the local structure and the electronic structure [J]. *J. Mater. Chem. C*, 2023, 11(1): 48-96.
- [51] WANG S X, SONG Z, KONG Y W, *et al.* Relationship of Stokes shift with composition and structure in $\text{Ce}^{3+}/\text{Eu}^{2+}$ -doped inorganic compounds [J]. *J. Lumin.*, 2019, 212: 250-263.
- [52] PARK C, LEE J W, KIM M, *et al.* A data-driven approach to predicting band gap, excitation, and emission energies for Eu^{2+} -activated phosphors [J]. *Inorg. Chem. Front.*, 2021, 8(21): 4610-4624.
- [53] DORENBOS P. Relating the energy of the [Xe] 5d1 configuration of Ce^{3+} in inorganic compounds with anion polarizability and cation electronegativity [J]. *Phys. Rev. B*, 2002, 65(23): 235110.
- [54] WANG T, XIA Z G, XIANG Q C, *et al.* Relationship of 5d-level energies of Ce^{3+} with the structure and composition of nitride hosts [J]. *J. Lumin.*, 2015, 166: 106-110.
- [55] DORENBOS P. Thermal quenching of Eu^{2+} 5d-4f luminescence in inorganic compounds [J]. *J. Phys.: Condens. Matter*, 2005, 17(50): 8103-8111.
- [56] BLASSE G, GRABMAIER B C. *Luminescent Materials* [M]. Berlin: Springer, 1994.
- [57] DORENBOS P. A review on how lanthanide impurity levels change with chemistry and structure of inorganic compounds [J]. *ECS J. Solid State Sci. Technol.*, 2013, 2(2): R3001-R3011.
- [58] DORENBOS P. Energy of the Eu^{2+} 5d state relative to the conduction band in compounds [J]. *J. Lumin.*, 2008, 128(4): 578-582.
- [59] ZHUO Y, HARIYANI S, ARMIJO E, *et al.* Evaluating thermal quenching temperature in Eu^{3+} -substituted oxide phosphors *via* machine learning [J]. *ACS Appl. Mater. Interfaces*, 2020, 12(5): 5244-5250.
- [60] 周志华. 机器学习 [M]. 北京: 清华大学出版社, 2016.
- ZHOU Z H. *Machine Learning* [M]. Beijing: Tsinghua University Press, 2016. (in Chinese)
- [61] JIANG L P, JIANG X, WANG C X, *et al.* Rapid discovery of efficient long-wavelength emission garnet: Cr NIR phosphors *via* multi-objective optimization [J]. *ACS Appl. Mater. Interfaces*, 2022, 14(46): 52124-52133.
- [62] LANGNER S, HÄSE F, PEREA J D, *et al.* Beyond ternary OPV: high-throughput experimentation and self-driving laboratories optimize multicomponent systems [J]. *Adv. Mater.*, 2020, 32(14): 1907801.
- [63] ATTIA P M, GROVER A, JIN N, *et al.* Closed-loop optimization of fast-charging protocols for batteries with machine learning [J]. *Nature*, 2020, 578(7795): 397-402.
- [64] VERDUZCO J C, MARINERO E E, STRACHAN A. An active learning approach for the design of doped LLZO ceramic garnets for battery applications [J]. *Integr. Mater. Manuf. Innov.*, 2021, 10(2): 299-310.
- [65] XUE D Z, BALACHANDRAN P V, YUAN R H, *et al.* Accelerated search for BaTiO_3 based piezoelectrics with vertical morphotropic phase boundary using Bayesian learning [J]. *Proc. Natl. Acad. Sci. USA*, 2016, 113(47): 13301-13306.
- [66] XUE D Z, BALACHANDRAN P V, HOGDEN J, *et al.* Accelerated search for materials with targeted properties by adaptive design [J]. *Nat. Commun.*, 2016, 7: 11241.
- [67] BRGOCH J, DENBAARS S P, SESHADRI R. Proxies from *ab initio* calculations for screening efficient Ce^{3+} phosphor hosts [J]. *J. Phys. Chem. C*, 2013, 117(35): 17955-17959.
- [68] LU S H, ZHOU Q H, OUYANG Y X, *et al.* Accelerated discovery of stable lead-free hybrid organic-inorganic

perovskites *via* machine learning [J]. *Nat. Commun.*, 2018, 9(1): 3405.

- [69] WILKINSON M D, DUMONTIER M, AALBERSBERG I J J, *et al.* The FAIR guiding principles for scientific data management and stewardship [J]. *Sci. Data*, 2016, 3: 160018.



黄霖(1999-),女,湖南岳阳人,博士研究生,2021年于湘潭大学获得学士学位,主要从事人工智能与发光材料的交叉研究。

E-mail: linhuang1800@163.com



解荣军(1969-),男,江苏镇江人,博士,教授,1998年于中国科学院上海硅酸盐研究所获得博士学位,主要从事稀土发光材料、量子点和发光器件的研究。

E-mail: rjxie@xmu.edu.cn