

基于机器学习的电化学能源电池宏微观设计

李金金, 蔡俊飞, 韩彦强, 汪志龙, 陈 安, 叶思敏

(上海交通大学电子信息与电气工程学院, 薄膜与微细加工教育部重点实验室, 上海 200240)

摘 要: 能源储存系统是电动汽车、电子设备等高新技术的重要基础。近年来基于机器学习的电池设计能够快速连结材料微观结构-材料微观性能-电池宏观性能的复杂关系, 成为了热点研究。本文从能源电池的微观材料设计和宏观状态预测两方面系统性地综述了电池设计中机器学习的应用现状和前景, 概括综述了机器学习电池设计的研究数据来源、算法的优缺点及其在电池领域的应用场景以及近年来的相关创新性工作及展望, 以期机器学习在能源储存系统的宏微观设计提供了参考。

关键词: 机器学习; 数据挖掘; 电池材料; 电池状态

中图分类号: TP181 文献标志码: A 文章编号: 0454-5648(2023)02-0438-14

网络出版时间: 2022-11-30



Macro-/Micro-Design of Electrochemical Energy Battery Based on Machine Learning

LI Jinjin, CAI Junfei, HAN Yanqiang, WANG Zhilong, CHEN An, YE Simin

(Shanghai Jiao Tong University and Electrical Engineering, Key Laboratory of Film and Microfabrication, Ministry of Education, Shanghai 200240, China)

Abstract: The energy storage systems are an important basis for electric vehicles and electronic devices. The existing battery design based on machine learning is able to quickly connect the complex relationship among material microstructure, material properties, and battery macroscopic properties. This review represented the applications and prospects of machine learning in micro-material design and state estimation of batteries. The data sources of machine learning battery design, advantages and disadvantages of algorithms and their application scenarios in the battery field, related innovative work in recent years and their prospects were discussed. This review can provide a reference for machine learning in the macro-/micro-design of energy storage systems.

Keywords: machine learning; data mining; battery materials; battery state

能源的储存系统一直是各种传统产业或者高新技术的核心问题之一^[1-2]。其中, 以锂离子电池为代表的新型储存设备在新能源汽车、计算机、便携式设备等高新技术的发展中都是不可或缺的^[3-5]。可见, 能源的革新正处在科技革新的时代洪流之中。从 19 世纪法国物理学家 Gaston Planté 发明了铅酸电池以来, 能源的发展方向概括而言就是提高循环寿命, 比容量, 能量密度, 功率密度和安全性等电池重要参数^[6-7]。时至今日, 电化学能源电池仍然在不断革新和探索之中。然而, 以往的实验性研究存在着诸多难以解决的问题, 比如: 巨大的试错难度

和成本; 无法揭示能源设备与其材料、工作环境的定性关系; 能源电池充放电过程中的结构-性能演变等。如何解释设备性能与材料之间的关系; 如何提高能源设备设计的效率; 如何解释离子、电子的运输机制, 给电池设备的设计和革新带来了巨大的挑战^[8-11]。

包括以量子力学为基础的密度泛函理论(DFT)、Hartree-Fock 方法; 以统计力学主导的分子动力学方法、Monte-Carlo 方法; 以连续介质力学为主导的有限差分方法和有限元方法等, 能够以相对低的成本以及更快的速度对材料进行性能测试^[12-13]。伴

收稿日期: 2022-08-04。 修订日期: 2022-08-09。

基金项目: 国家科技部重点研发计划(2021YFC2100100); 国家自然科学基金(21901157); 上海市自然科学基金(21JC1403400)。

第一作者: 李金金(1985—), 女, 博士, 研究员。

Received date: 2022-08-04. Revised date: 2022-08-09.

First author: LI Jinjin (1985—), female, Ph.D., Professor.

E-mail: lijijin@sjtu.edu.cn

随着电池材料数据和电池性能实验测试数据的不断积累以及大量数据库的建立。近几年来, 电池设备的发展俨然进入了所谓数据驱动的科学范式, 有希望极大加速这一项技术的发展^[14-15]。

数据驱动的人工智能在电池的设计领域的应用场景包括材料设计和状态估计^[2, 16-17]。其中, 材料设计包括电极^[18]、电解质^[19]的设计。比如, Cai 等^[20]通过机器学习建模结合从头算分子动力学的方法对镁离子电池和锌离子电池的潜在尖晶石正极材料进行电子导电性能、离子导电性能以及稳定性进行了筛选, 最终选出 6 种符合要求的尖晶石正极材料。Wang 等^[21]通过机器学习加速设计, 从 29 008 个石榴石之中筛选出了 12 种具有高电子导电性、锂离子导电性以及高稳定性的优秀石榴石。电池状态估计的研究包括荷电状态(SOC)估计^[22]、健康状态(SOH)估计^[23]以及剩余健康寿命(RUL)估计^[24]。如图 1 所示, 机器学习在电池设计中的流程可以分为 4 步: 1) 通过理论计算或者实验生成用于电池设计的数据; 2) 通过生成的数据建立相关数据库, 包括以 DFT、实验表征计算为基础的材料数据库, 文本数据挖掘为基础的电池表征性能数据库(例如电池能量密度, 功率密度等); 3) 机器学习模型的选择与训练; 4) 通过已训练的机器学习模型对目标性质进行预测。将在第 1 节(数据产生与获取)中重点介绍第 1 步和第 2 步数据相关的内容; 在第 2 节中介绍线性模型、神经网络等具代表性的机器学习算法的原理、优缺点及其应用; 在第 3 节中总结目前人工智能电池设计领域的创新性工作并展望机器学习在

电池设计的后续发展。目前已经有关于电池设计的机器学习综述得到相应的报道。Shi 等^[25]综述了数据驱动的机器学习在电化学储能材料研究中的应用。Xie 等^[26]综述了机器学习方法在电池状态估计中的应用现状。但是目前没有机器学习的相关综述文章针对电池微观材料设计和宏观性能测试研究进行综合评述。分析在不同层次的电池设计研究时有效率的选择和使用机器学习算法对电池设计的发展具有重大意义, 本文综合评述了电池设计研究中以机器学习方法加速微观材料设计以及电池宏观性能测试的应用。分析了不同机器学习算法、数据获取模式对宏微观设计的适用性、优缺点。

1 数据产生与获取

对于人工智能算法的构建, 具有足够“质”和“量”的数据是机器学习模型成功建立的基础, 数据必须保证自身的正确性和能够代表全集对象的典型性。良好的数据资源将会极大降低机器学习算法训练的难度。因此本节将分别介绍电池设计在材料设计和电池状态估计 2 个不同领域研究的数据来源。

首先是材料设计的数据来源。对于材料设计而言, 数据主要来源包括抽样测试的方法和数据库。抽样测试方法即在构建数据全集后从中以某种原则进行子集抽样并对子集中的数据实施性能测试(理论计算或者实验)。例如, Zhang 等^[27]为了抑制锂硫电池充放电过程中硫元素的穿梭效应, 需要从 342 个 AB₂ 型二维材料(材料全集, 研究目的即预测这 342 个

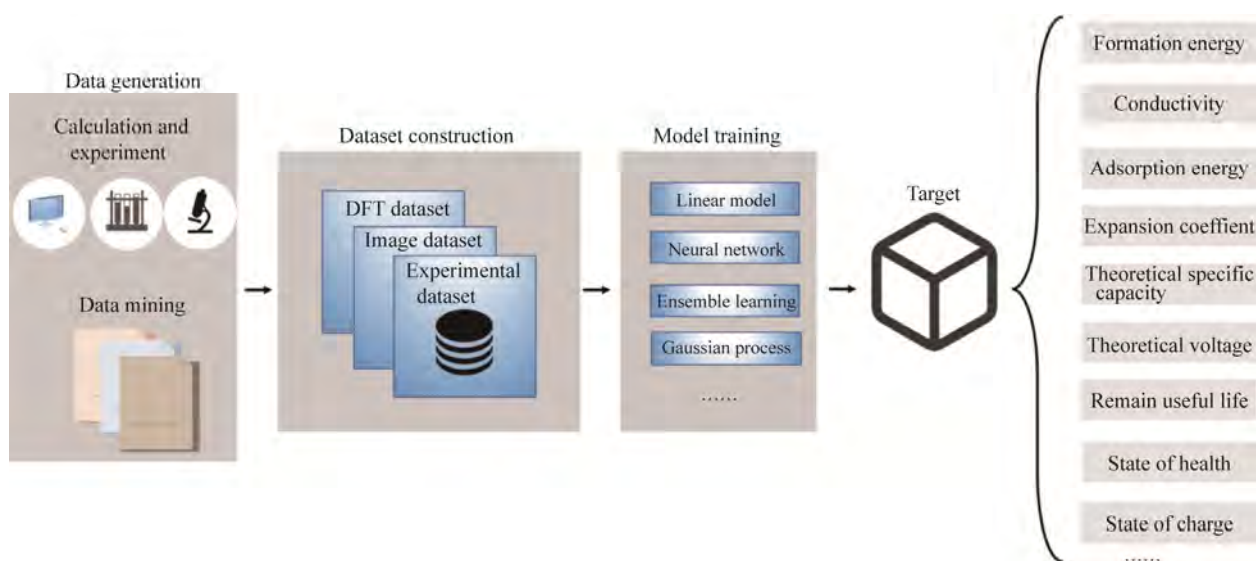


图 1 机器学习在电池设计中的基本流程

Fig. 1 Basic process of machine learning in battery design

材料对多硫化物的吸附能)。Zhang 等从材料全集中选择了 65 个材料(抽样产生的子集)并进行了 DFT 计算(吸附能计算), 然后以子集数据为训练集建立了机器学习模型, 成功预测了全集材料的吸附性能。这种以抽样测试的数据获取方法是直接且有效的, 但采样过程必须保证子集数据对全集数据是具有代表性的。对于大样本容量的全集, 昂贵的时间、金钱成本导致了高通量实验以及理论计算的局限性, 因此采用合适的采样方法获取合适的有训练样本有利于后续的模式训练。目前, 常见采样方法有随机抽样和均匀抽样^[28]。前者指随机地从全集中进行采样, 后者指根据范数等数学定义上的样本距离均匀等距的进行采样。以二维特征空间为例, 假设每个数据是由参数 1 和参数 2 构成的二维向量, 随机抽样和均匀抽样的示意图如图 2a 和图 2b 所示。其中横坐标为特征参数 X_1 , 纵坐标为特征参数 X_2 , 图 2 中取样点向坐标轴的投影可以表示参数的覆盖度。假设从样本空间中取 4 个点(红、蓝、绿、黑 4 个数据点), 将每个点所在坐标向横纵坐标进行投影, 并

用黄色阴影表示数据对参数空间的覆盖程度。比如, 在随机抽样中黑点和绿点重复对相同的 X_1 数值进行了抽样, 在均匀抽样中蓝点和红点选择对相同的 X_2 数值抽样。同时, 图中白色部分表示这部分数据的某些特征在抽样数据集中并不存在, 可以清楚看到随机抽样和均匀抽样方法都导致了一部分的取样的空白。这种抽样难以避免地导致训练数据集中的数据未能填充完整的参数空间, 训练数据对全体数据的代表性较弱, 且重复采样导致的成本问题、随机性导致的复现问题也被研究者们质疑。目前, 有望解决采样问题的方法是拉丁超立方抽样(LHD 技术)^[29]。LHD 技术得名于统计学中的拉丁方阵(每行、每列仅有 1 个样本的方阵), 其基本原理是令 2 个采样点之间的距离最大化并最小化具有相似距离值的点的数量。如图 2c 所示, LHD 的采样点覆盖了参数空间所有参数, 说明采样结果对全体数据的代表性更强。当然, 目前 LHD 在电池宏微观设计中的应用仍然比较有限, 仍然需要进一步的发展和有研究。

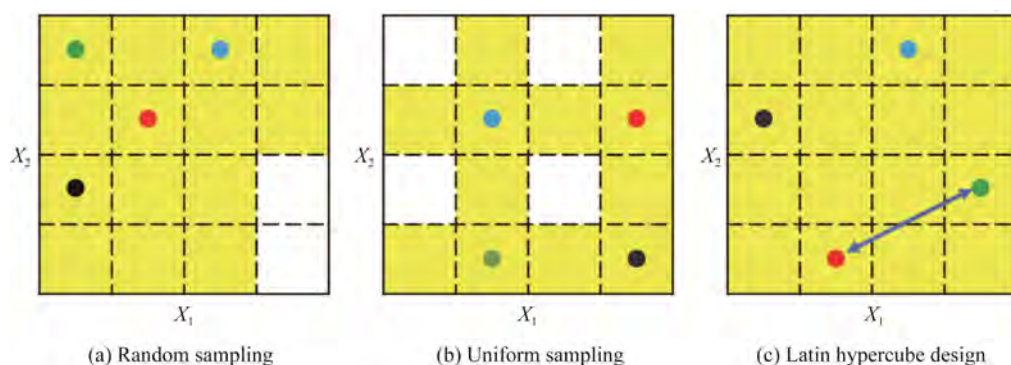


图 2 随机抽样、均匀抽样以及拉丁方超立方抽样的示意图
Fig. 2 Schematic diagram of random sampling, uniform sampling and Latin hypercubic design

除了抽样实验或计算的方法之外, 科学数据库也是机器学习在能源电池设计中的另一重要数据来源。在 2011 年美国的材料基因组计划提出前后, 全世界的材料领域研究者通过理论计算模拟或者是实验测量, 建立起了许多知名的材料数据库以促进材料数据的共享和传播。这些数据库基于大量的实验经验和计算, 涵盖了全面的材料性质。包括结构信息、力学信息和电学信息, 并且跨越了金属、高分子、合金、二维材料等不同的类别。如表 1 所示, 概括了国内外知名的材料数据库, 涵盖了有机/无机, 金属/非金属以及晶体/非晶体材料等不同类型。如表 1 所示, 无机材料的研究可以从国家标准技术研究所无机晶体结构数据库(NISTICSD)、无机晶体

结构数据库(ICSD)、材料工程(MP), 开放量子材料数据库(OQMD)、Pauling File、AFLOW 等知名数据库中获取。其中 NISTICSD 是由美国国家标准技术局从 20 世纪初就开始主导的一个数据库项目, 收集包括键长-角, 粉末衍射数据, 空间群等 21 万条以上的晶体实验信息。由于发展时间比较早, NISTICSD 的数据数量和质量都优于很多现代晶体实验数据库, 包含了约 21 万条实验数据, 但是数据并非开源而且界面较为老旧, 因此用户友好度相对较低。作为另一个知名的订阅制数据库, FIZ Karlsruhe-ICSD 则由于更为现代的管理和用户友好度, 更加受材料研究者的关注。ICSD 是德国的非盈利公司 FIZ Karlsruhe 主导运营的一个无机晶体材料

表 1 知名数据库以及对应的材料类型、数据来源及其网址

Table 1 Well-known databases and corresponding material types, data sources and uniform resource locator (URL)

Name	Type	Data resource	URL	Reference
NISTICSD	Inorganic	Experiment	https://icsd.nist.gov/	[32]
FIZ Karlsruhe ICSD	Inorganic	Experiment	https://icsd.products.fiz-karlsruhe.de/	[33]
Pauling File	Inorganic	Experiment	https://www.paulingfile.com/	[34]
MatWeb	Inorganic/Organic	Experiment	https://www.matweb.com/	
Materials Project	Inorganic	Calculation	https://materialsproject.org/	[35]
OQMD	Inorganic	Calculation	https://oqmd.org/	[15, 30]
AFLOW	Inorganic	Calculation	https://aflowlib.org/	[36]
PolyInfo	Polymer	Experiment	https://polymer.nims.go.jp/en/	
Total Materia	Metal/Nonmetal	Experiment	https://www.totalmateria.com/	
CSD	Organic, MOF	Experiment	https://www.ccdc.cam.ac.uk/solutions/csd-core/components/csd/	[37]
ChemSpider	Organic	Experiment/calculation	http://www.chemspider.com/	[38]
NIST	Normal material	Experiment/calculation	https://materialsdata.nist.gov/	
COD	Organic/Inorganic	Experiment/calculation	http://www.crystallography.net/cod/	[39]

数据库, 收罗了 26 万条以上的晶体结构信息(包括了实验数据, 理论数据以及文献数据), 全面而庞大的数据库特性使得其不仅在作为机器学习数据集来源方面意义深远, 而且催生了其他计算材料数据库的诞生。例如, OQMD 基于量子化学计算模拟和 ICSD 数据的应用产生, 保存了 100 万条以上的晶体热力学性能和结构信息数据^[30]。显然, OQMD 作为以理论计算为基础的开源数据库, 数据量远大于作为实验数据库的 NISTICSD 以及 ICSD。其优势在于不仅能够提供相应的材料数据, 而且能够提供对应化合物的相图计算、结构可视化以及基态化合物确定等功能。知名 DFT 数据库物料工程(MP)与 OQMD 类似, 包含了 14 万条以上的无机晶体性能、6 万条以上的分子性能以及 50 万条以上的纳米空隙材料性能等数据。MP 和知名的 Python 材料分析库 Pymatgen 有独特的应用程序编程接口(API)连接, 使得用户对数据的提取与使用更为方便友好。更具特色的是, MP 数据库提供分块化材料数据分析软件, 囊括了电池电极材料搜索, 界面效应计算, 相图计算、化学反应计算等其他数据库没有的创新性功能。具有全面金属、非金属材料力学性能分析功能的 Total Materia 数据库收罗了应力应变曲线图、金属成形曲线、疲劳实验数据等金属力学实验的重要数据, 因此在金属力学性质的相关研究中表现突出。上述无机材料数据库中 MP, OQMD、Pauling File 和 AFLOW 的数据是开放获取的, 数据获取的难度和成本都更低且有更高的数据使用自由度。对于有机材料的相关研究, 则可以从 MatWeb、PolyInfo、剑桥结构数据库(CSD), ChemSpider 以及 COD 中获取。其中 MatWeb、PolyInfo、CSD 都是基于实验的

材料数据库, 而 Chemspider 和晶体学开放数据库(COD)同时收罗了实验数据和计算数据。上述的几种有机材料数据来源中, 除了 MatWeb 之外, 其他均为开源获取的数据, 为机器学习对能源电池的研究提供了极大的帮助。

相对于实验性数据库而言, 以量子力学计算为基础产生的数据库在拟合材料性质方面时有更大的误差, 这是因为目前理论计算方法的描述精度与实验科学存在一定的偏差。例如求解 Schrödinger 方程过程中密度泛函理论中交换关联函数的局部密度近似等固有误差是无法避免的。但这些基于理论计算的无机晶体材料数据优点在于独特的 API 和大量基于高通量筛选的虚拟材料数据, 允许研究者轻松和快速的访问, 并支持加速新材料发现。另一方面, 实验性产生的数据集在不同研究设备和不同研究环境因素下, 产生的数据误差波动情况较大, 因此作为机器学习的训练数据集也具有较大的噪声, 为模型训练提高了不少难度。相反, 不论应用了那种近似方法, 计算产生的数据本身就经过了统一的数学求解方法(Schrödinger 方程的求解), 使得研究者们更容易在数学上找到从数据到目标性能的映射。对于实验性数据或者理论计算性的数据的取舍是明显的: 以实验数据为训练数据集的模型能更好与实际耦合, 但是训练难度也比较大。以计算数据为训练集的模型更容易训练, 但和实际物理过程相比不可避免地存在量子力学计算和机器学习预测地误差叠加问题。

作为电池设计的另一重要研究体系, 状态估计(SOC、SOH、RUL), 其数据的来源主要来源于实验^[16, 22-31]。Álvarez Antón 等^[17]根据他们的实验数

据, 建立了一个支持向量机模型用于大容量磷酸铁锂电池的 SOC。他们的模型训练主要基于实验得到的电池电流、电池电压、电池温度以及从电池充放电测试周期中提取模型参数。Wang 等^[16]通过实验获取了锂离子电池工作过程的微分热伏安曲线(DTV 曲线)并使用滤波器的方法将其平滑化。然后他们将 DTV 曲线中的不同维度, 包括峰值位置、峰值和谷值等数据提取为影响锂离子电池健康因素的特征并通过相关性分析的方法进行优化, 输入高斯回归过程算法之后得到了一个精确的电池性能预测模型。可见, 在电池的状态预测领域, 其数据特征来源一般是实验过程中的条件或者是从实验数据中提取的特征。

综上所述, 材料理论计算方法、实验数据特征提取、数据库等技术为数据驱动的电池设计研究提供了丰富的数据来源, 支持了人工智能在能源电池领域的发展。但是另一方面, 不论计算还是实验, 数据的扰动都是不可避免的, 对于理论计算方法, 即便是目前最具先进性的理论模拟计算方法仍然在模型、方法中有固有误差存在。而实验数据一般来源于不同的实验环境、设备, 且电池本身存在的缺陷、污染等问题均无法被数据库所记录。对于数据工作者而言, 最初数据的质量决定了后续工作的进行, 如果收集数据时每个数据的误差不断叠加, 将导致训练数据集的噪声以指数型增加, 极大提高了模型训练的困难程度。因此, 数据库的构建急需统一的测试标准、全面的参数管理方法以及更全面的数据描述, 以进一步帮助人工智能工作者们进行数据筛选和收集。

2 算法及其应用

机器学习模型的选择在人工智能模型训练预测结果中起着关键作用。一个合适的机器学习模型不仅可以保证最终输出结果的可靠性, 也能够大量减少训练需要花费的时间。在各大会议或期刊中, 已经有上千种不同的机器学习算法被报道, 总体而言分为监督学习、无监督学习 2 大类^[40–42]。监督学习在电池设计研究中是应用最广泛的一种算法, 它需要的数据包括输入数据(Input)和相应的目标数据(Target)。监督学习可以从训练数据中建立学习模型, 然后根据模型推断出新的样本, 目的是通过机器学习建立 1 个从输入数据到目标数据的映射(Mapping)^[43]。根据目标数据的不同, 监督学习又可以分成 2 类: 目标数据是一个连续值时, 对应的问

题为回归问题(Regression); 目标数据是离散的预测标签时, 对应的问题是分类问题(Classification)。无监督学习则是对输入的数据进行自动分类, 没有目标数据的存在——也就是说, 它不需要配套的输入与输出^[44]。无监督学习的优势在于能够自主解释数据并有可能发现一些新的模式, 挖掘数据中的信息。比如常见的关联分析、聚类问题和降维问题都是无监督学习中的经典问题^[45]。在本节中, 将介绍一些具有代表性的机器学习算法及其优缺点, 并综述每一类算法的优缺点以及近年来以其为基础的电池研究案例。

2.1 线性模型

线性模型数据机器学习中最简单的模型之一。其目的是建立输入数据 X 到输出数据 y 之间的线性映射($\hat{y} = w^T X$)。线性模型的模型简单、计算成本低, 常常无法解决复杂性过高的问题, 但是由于可解释性强, 使得其在机制解释, 性能影响因素的研究中表现极其突出。在这一节中, 将着重介绍在电池材料设计和状态预测研究中更具代表性的 2 种线性模型: 套索回归(LASSO)以及弹性网络。

LASSO 是一种结合普通线性模型和 L1 正则化而提出的线性模型^[45] (在机器学习中, 正则化的目的是为了控制模型复杂度, 进而减小过拟合的风险)。L1 正则化通过在目标函数中添加惩罚项, 通常表示为 $\|w\|_1$ (权重系数 w 的绝对值之和), 达到稀疏化权重参数的目的(即不重要的数据权重降低, 重要的数据权重上升)。正则化之后, 只有少数更重要的参数决定了最终的学习结果, 因此抗噪能力比较强。同时由于正则化之后可以分析出哪些参数的权重更加多, 因此 LASSO 有着更优秀的可解释性, 即解释输入向量中的不同维度数据的重要性^[47]。Yoshida 等^[48]以优化锂离子电池的循环性能为出口, 通过第一性原理计算研究最佳的掺杂 LiNiO_2 (锂离子电池正极材料)并最终确定铌元素的掺杂是最佳方法。此外, 他们通过 LASSO 进行数据的分析, 并且成功指出掺杂剂的电荷密度是影响电池循环性能最重要的因素。Min 等^[49]同样通过建立 LASSO 模型, 从模型解释中得出了一个全新的描述符, 可以用于描述异质结构的本征性质和氧还原反应的线性关系。近年来在工程领域, 尤其在催化、材料、能源的开发中, 避免黑盒属性的机器学习导致的模糊性结论是很有必要的^[50]。如何探索特定材料或者特定电池优异性能的影响因素和机制始终是能源数据分析领域的重要课题, 因此诸如 LASSO 回归此类的

高解释性算法在电池设计研究中仍然有巨大的潜力。

与 L1 正则化类似, L2 正则化向目标函数添加一个基于权重系数 w 各元素的平方和, 但不同的是 L2 正则化几乎涉及到了全部的变量(但每个数据的权重值都会近似于 0, 因此被称作是权重衰减), 比如经典的岭回归模型就是以 L2 正则化为基础。弹性网络则是同时应用上述 2 种正则化模型进行训练的线性模型。由于 2 种正则化方法的结合, 弹性网络能够学习到一个类似套索模型的稀疏化模型, 但不会出现 LASSO 中的高权重值, 保持了类似岭回归的正则化性质^[51]。相对于 LASSO, 弹性网络最大的优点是能够更有效地处理多特征相互关联的问题, 而这个问题在电池的宏观设计中是十分常见的。Severson 等^[52]测试了 124 个商业磷酸铁锂-石墨电池在快速充放电循环下寿命的变化, 建立了一个全面的数据集。基于早期循环的放电电压曲线, 他们建立了弹性网络等模型, 使用 100 个循环数据进行回归预测(测试误差为 9.1%)和 5 个循环数据进行分类预测(测试误差为 4.9%)。这个结果表明了弹性网络在数据更少的情况下并没有出现误差剧增的问题, 说明弹性网络的可靠性。

2.2 Bayesian 学习

Bayesian 学习是一类基于 Bayesian 学派观念的学习方法, 即可以假设参数服从某一个先验分布, 并基于输入数据计算其后验分布, 是一种生成式的非线性模型, 通常用于监督学习。Bayesian 学习的发展过程中, 逐渐有多种不同的算法相继被提出并在电池设计优化研究中扮演者越来越重要的角色^[53]。由于 Bayesian 学习在小样本数据的分析中表现优异, 因此 Bayesian 学习在电池状态估计方面的应用更为广泛, 而电池材料设计方面则相对较少。

首先是最简单的朴素 Bayesian 学习, 基本原理是通过训练数据集学习从输入数据到目标数据的联合概率 $P(X,Y)$ 。对于待分类的输入数据, 由学习到的 $P(X,Y)$ 结合条件概率公式(Bayesian 原理)计算出每个类别的后验概率分布 $P(Y|X)$ 。换句话说, Bayesian 学习最终的目的是计算出某个特定类别 $Y=y$ 在特定输入数据 x 条件下的概率——最后该条件下最大概率的类别将会被学习器认定为最终的输出。由于在这个过程有一个强的条件独立假设——特征数据之间在确定类情况下是相互独立的, 所以被冠以“朴素”。相对而言, 朴素 Bayesian 学习的优点在于算法简单, 适合于小数据量的分类问题, 但其分类决策的错误率还是比较大。

稀疏 Bayesian 算法(SBL), 也称为相关向量机(RVM), 最早由 Tipping 等在 2001 年提出, 由于其模型的稀疏性, 其泛化能力相对而言更强, 具有高学习能力, 训练过程简单等优点^[54]。由于 SBL 算法的稀疏性和不确定性表达能力, 使得其在电池状态预测领域的应用十分广泛^[55]。例如 Hu 等^[56]从充电曲线中提取了 5 个特征: 初始充电电压、横流充电电压、恒压充电电压、最终充电电压以及最终充电电压, 最终使用 SBL 从充电数据中推断出了电池容量。Zhou 等^[57]提出, SBL 策略十分适合样本量小、预测性能差的锂离子电池 RUL 预测。但由于 SBL 算法在多步非线性问题的预测能力较低, 因此对 RUL 的预测结果不尽人意。通过组合 SBL 子模型, Zhou 等^[57]开发了一种非迭代式的新型组合 SBL 策略, 提高了锂离子电池 RUL 的预测精度。可以注意到, 虽然在电池的状态预测中 SBL 的表现十分强势, 但是它在电池材料筛选设计方面的应用却不多, 根本原因仍然在于 SBL 更加适合样本空间小的、稀疏性强的输入数据。

Bayesian 分层模型(BHM)是一种基于层次建模的 Bayesian 学习方法, 有多篇相关工作报道了分层 Bayesian 模型在电池优化的优势。Jiang 等^[53]开发了一种分层 Bayesian 学习方法用于在很少的周期测量中准确评估不同循环充电协议下的电池寿命。结果表明即使在进行少数的几次测量之后(仅持续 3 个循环周期的单个电池测试), 该方法仍然能够提供精确的预测。应用在锂-磷酸铁/石墨电池中, 整体平均误差仅有 6.5%。同样的, Jafari 等^[58]在分层建模的 Bayesian 模型下, 将外部因素(例如驾驶行为习惯和充电选项、电网服务等)与电池容量衰减联系起来, 使用 85%的数据作为训练集并用其余的 15%进行测试, 其研究成功建立了从生态系统到概率过程的映射关系。

总体而言, Bayesian 学习在电池领域的应用着重表现在以小样本数据以及稀疏性数据的训练上。因此, Bayesian 学习更适合应用于电池的状态估计以及相关的研究当中。

2.3 Gaussian 过程回归(GPR)

GPR 是一种数据驱动的非参数机器学习模型, 具有超参数自适应特性, 易于实现, 不涉及参数的预设^[59]。GPR 的数据驱动属性和非参数性使得它能够有效的适应复杂的输入条件^[60]。与线性模型、非线性模型不同的是, GPR 是非预设性的机器学习算法——即 GPR 在训练之前不预设问题是线性或是

非线性,通常用于监督学习器的训练。在核函数的形式不受限制的情况下,理论上 GPR 能够实现任意不同函数曲线的任意逼近。因此,Ren 等针对超级电容器的状态预测问题曾提出一种以隐性函数为核函数的 GPR 方法,试图通过非预设性的核函数进行高斯过程学习,结果表明以隐函数为核函数的 GPR 相对显性函数 GPR 具有更低的误差^[61]。作为一种相对比较成熟的机器学习算法,GPR 在电池材料设计以及状态预测 2 个不同的领域都表现优异^[62]。

材料设计方面,Ha 等^[62]通过稀疏 GPR 和 DFT 的结合,对高能量密度正极材料(锂过量阳离子无序岩盐材料)的铝掺杂实验进行了分析。他们成功发现在含有 4d 轨道元素的正极中,铝元素的掺杂能够高效地抑制电池的电压/容量衰减。电池状态预测方面,Li 等^[64]通过从电池的容量曲线中进行特征提取,从截距、斜率和峰值等不同尺度提取了总共 4 个训练特征并建立了基于 GPR 的电池 SOH 估计模型。此外,他们将这个 GPR 模型和非线性回归相结合实现了 RUL 的预测。最终结果显示,SOH 估计的平均绝对误差(MAE)和均方根误差(RMSE)均小于 2%,而 RUL 预测误差低于 26 个周期。另外,Lyu 等^[65]提出一种基于部分充电曲线的锂离子电池容量估计数据融合模型方法,通过双 Gaussian 回归模型建立了电池退化模型并使用粒子滤波器抑制训练过程中的噪声。这种粒子滤波器结合双 Gaussian 模型的预测方法与传统的 Gaussian 回归预测模型相比,使用了电池退化的公共数据库进行验证,有效降低异常退化特征(噪声)的影响,估计精度高,鲁棒性强。可见,Gaussian 回归过程在电池领域的材料设计和状态估计都有不俗的表现,尤其在状态估计研究中的算法改良目前仍然是相关领域的一大热点。

2.4 神经网络

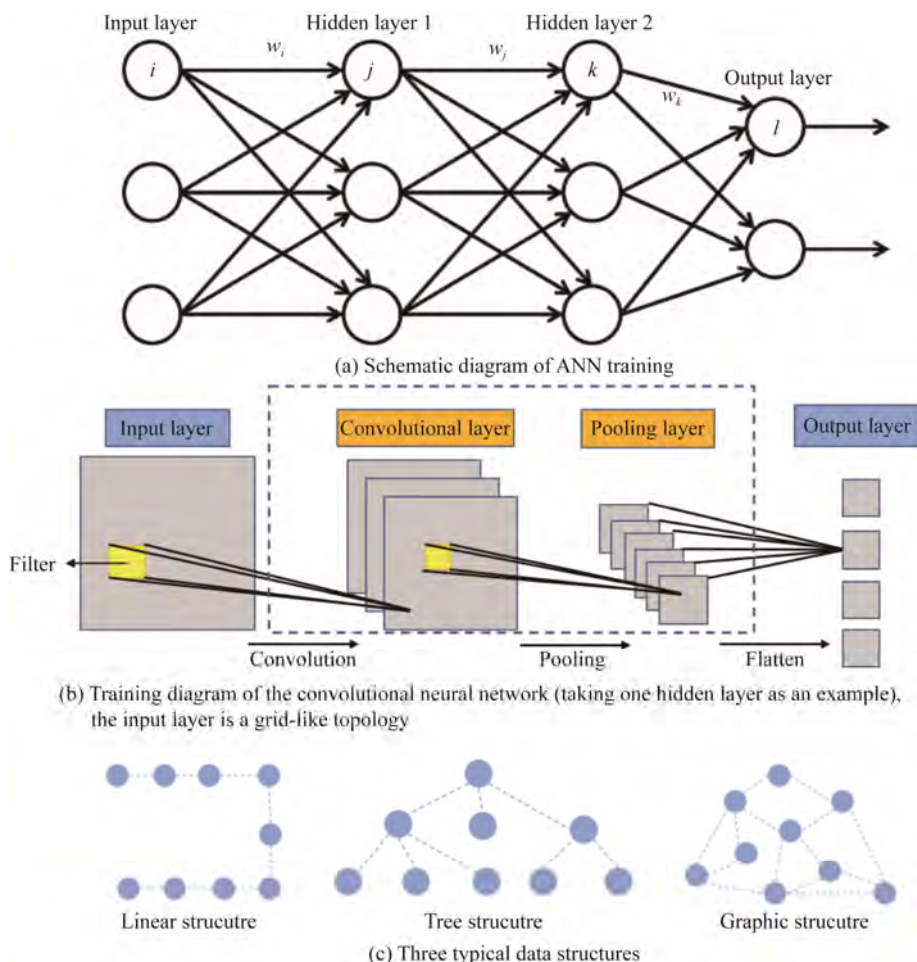
作为一种极受欢迎的非线性模型,神经网络算法能够胜任监督学习、无监督学习等不同的机器学习方式。神经网络最早来源于计算机科学家们对人工智能的初步设想——对人或动物的大脑进行仿生模拟。虽然在几十年的发展后神经网络已经渐渐成为一种数学概念,但还是遗留了最早称谓。实际上在 21 世纪初神经网络在机器学习的研究中一直不受到重视,直到 2012 年 Geoffrey Hinton 教授及其学生 Alex Krizhevsky 设计的 Alexnet^[66]获得了 ImageNet 竞赛的冠军,人工智能研究者们才逐渐重视起层数更多的神经网络(也称作深度学习^[66]),并

在之后的几年中提出了更多更复杂的神经网络(比如 GoogleNet, ResNet 等)^[68–69]。神经网络的基本原理是在每一个网络层中学习每个神经元的权重结合激活函数(Sigmoid 函数、ReLU 函数以及 Tanh 函数等)得到该层的输出,最终得到输入层到输出层的映射(神经网络的权重是通过优化一个损失函数来测量神经网络输出和期望目标之间的差值来学习的)。在包括催化、能源和材料的工科研究中,诸如人工神经网络、卷积神经网络、图神经网络等优秀算法的应用也在近年来渐渐出现。神经网络作为目前最受欢迎的机器学习算法,其优点包括:1) 自学习,自适应能力强,能够处理不确定性较高的问题,适合开发端到端的预测模型;2) 具有很高的鲁棒性和容错性;3) 对复杂问题的解决能力强,并能够充分地逼近这些复杂问题的潜在映射(不论线性或非线性的)^[70]。缺点也很突出,那就是几乎无解释性的黑盒子属性使得神经网络的结果无法像上文提到的 LASSO 回归等算法那样进行机制、特征重要性的解释。因此,单纯神经网络在电池设计的应用中一般是结果导向,追求模型的预测质量而舍弃过程中对特征重要性的分析。由于神经网络具有优秀的函数逼近能力,所以它不论在电池宏观状态估计或是微观材料设计中都有大量的相关工作被报道^[71–73]。这里本文将重点介绍几种在电池设计领域具有代表性的神经网络算法,包括人工神经网络(ANN),卷积神经网络(CNN),图神经网络(GNN)。

如图 3a 所展示的是全连接的人工神经网络结构(即每个层间的神经元全部两两相互连接)。这种模型是通过对损失函数的优化(基于向后传播算法),学习层间任意 2 个神经元之间的权重系数 w 并通过与激活函数结合去逼近从输入到输出的目标映射。ANN 进行全连接结构训练时每个神经元之间的参数都是独立不共享的,适合训练数据之间相互独立的问题解决。因此,ANN 在电池的状态估计研究中应用较为广泛(因为电池的应用场景、温度等数据之间的关系并不大)。例如,He 等^[74]建立了一种基于 ANN 的电池模型,根据测量的电流和电压来估计基于磷酸铁锂的锂离子电池的 SOC 并在后续工作中使用 Kalman 滤波器进行模型的进一步优化。He 等不仅针对性地开发了一种建设性的搜索方法,而且建立起了一种非等效电路模型的 SOC 预测方法。Bian 等^[75]结合 ANN 模型、结合开路电压模型以及增量容量分析,提出了一种融合性的 SOH 预测方法。经过检验后,提出的模型平均绝对误差仅有 7.6 至 8.8 mV。

CNN 算法的名字源于在训练过程中的卷积过程(通俗上指卷积核在矩阵型数据遍历的过程)。与人工神经网络相比, CNN 多了网格搜索(卷积)和池化的过程, 如图 3b 所示, 输入层的数据一般是具有网格拓扑的一类数据, 通过卷积核的搜索, 每一个隐藏层的输入都是上一层数据的特征提取(一个包含上一层数据特征的全新网格数据)。之后经过池化层数据被进一步提取特征, 得到了下一层的输入数据。在隐藏层训练结束后, CNN 会将最终的网格数据进行 Flatten (即把网格数据展开为一维的向量数据)得到一个维度更高的特征数据。这个特征数据在后面会作为全连接层的神经网络结构的输入数据, 用以学习最终的目标映射函数。自从 2016 以 CNN 算法为核心^[75]的围棋人工智能 AlphaGo 出现以来, CNN 算法在人工智能及其相关领域的研究热度一直是最高之一, 近年来更是在图像分类、目标检测

等复杂问题上独占鳌头^[76-77]。从 CNN 的输入数据类型可以知道, CNN 擅长处理具有网格状拓扑结构(比如图像, 衍射结构或者晶体结构)的数据^[78-79]。在能源、材料领域的复杂问题分析中, CNN 表现优异^[80-84], 尤其是在电池测试、材料表征图像分析领域, CNN 更是大幅度领先其他算法——比如高分辨率透射电子显微镜图像的特征分割^[85], 基于 X 射线衍射图像的缺陷分析^[86-87], 粉末 X 射线衍射的定量分析^[88]。而且, CNN 在电极材料相关性能的预测(尤其是电学性能)中已经被证明是一种鲁棒性很高的方法, 例如 Saidi 等^[89]使用分层卷积神经网络进行结构和电子输运特性的预测, Dong 等利用从头计算得到的带隙和相应的结构作为训练数据集, 针对石墨烯及其掺杂产物进行了带隙的预测并为 CNN 的执行提出了一种新的描述符, 最终得到了高精度的预测模型^[90]。



w_i , w_j and w_k represent the weight vector from the input layer to the hidden layer 1, the hidden layer 1 to the hidden layer 2, and the hidden layer 2 to the output layer, respectively.

图 3 机器学习算法常见算法与数据结构示意图

Fig. 3 Machine learning algorithms Common algorithms and data structure diagram

与 ANN 和 CNN 相比, GNN 是一种基于图形结构的机器学习算法。对于一般神经网络算法, 数据必须是结构化的数据, 即序列化或者网格化的数据, 如文本、图像等。GNN 擅长解决非结构化的输入数据, 例如社交网络、知识图谱等^[91]。如图 3c 是 3 种典型的数据结构——线形结构、树形结构和图形结构, 其中图形结构是一种典型的非结构化的数据类型, 相对于线形结构和树形结构, 图形结构的数据点之间的关系更加错综复杂。类似的, 晶体、蛋白质等物理、化学概念中, 原子和原子之间, 分子和分子之间的关系一样是错综复杂的图形关系, 使得 GNN 在物理、化学微观分析领域的应用极具潜力。例如, Xie 和 Grossman^[92]建立了一种晶体图卷积神经网络(CGCNN), 用于从晶体中原子间的关系学习材料的性质。这种方法能够从不同类型晶体结构特征和成分特征数据预测晶体的包括形成能、绝对能量、带隙、体积模量、剪切模量、泊松比等性能。Ahmad 等^[93]在对超过 12 000 的样本进行了基于 CGCNN 的材料弹性模量和体积模量预测, 并用于筛选能够有效抑制锂-金属电池负极树枝晶生长的固态电解质; 另外, Zhou 等^[94]针对锌电池的正极材料, 利用 CGNN 从 Materials Project 和 AFLOW 的数据中成功筛选出了 70 种全新的优秀正极材料, 以及 10 种已经被实验证明先进的锌离子电池正极材料。Zhang 等^[27]通过结合 GNN 和迁移学习方法, 从 2D MatPedia 数据库^[95]中挖掘出 14 种(PdN_2 、 TaS_2 、 PtN_2 、 TaSe_2 、 AgCl_2 、 NbSe_2 、 TaTe_2 、 AgF_2 、 NiN_2 、 AuS_2 、 TmI_2 、 NbTe_2 、 NiBi_2 和 AuBr_2)新型的二维 AB_2 型正极主体材料用于解决锂硫电池充放电过程中的硫元素穿梭效应问题。Wu 等^[96]使用 GNN 算法模型从材料数据库种提取了 6 531 个二维材料并从中筛选出了 38 种高效的催化剂用以抑制电池中的析氢反应, 该模型的精确度高达 95.2%。此外在预测过程中, Wu 及其团队通过 GNN 的遍历了二维材料的表面进行了吸附活性位点分析, 保证了吸附预测的合理性和正确性。

整体而言, 神经网络的预测模型在包括电池 SOC 预测, 电极材料图像分析以及活性位点分析等复杂问题中应用广泛。随着电池领域的材料数据和实验测试数据不断积累, 新能源电池的应用场景和性能需求越来越趋于多样性, 进一步佐证了神经网络在电化学能源电池的设计的巨大潜力。

2.5 集成学习方法

集成学习方法是多个机器学习算法的集成, 即

通过多个弱的基学习器(Base learner)的结合来完成最后的学习任务, 因此集成学习方法的类型也取决于不同基学习器的选取, 其主要方法可以分为 3 种不同的类型: 自助法(Bagging), 提升法(Boosting)和堆叠法(Stacking)^[97]。集成学习的构建首先以并行的或者顺序的方法构建基学习器, 然后通过某种组合算法(如多数投票, 权重平均等)将基学习器组合起来。集成学习通过联合不同的模型, 具有更高的预测性能且容易实现, 而且基学习器一般比较简单, 所以能够有效地避免出现过拟合问题^[98]。然而, 由于集成学习是一种多模型混合结构, 所以集成学习模型的预测结果往往也是难以理解的, 即解释性相对较低。这里将主要介绍随机森林(Random forest)和极限梯度提升树(eXtreme Gradient Boosting, XGBoost) 2 种算法的优势与应用。

随机森林是一种以决策树(Decision tree)模型为基学习器的自助集成学习方法。随机森林的模型预测步骤如下: 1) 数据集的数据被随机有放回的选取并随机选出部分特征作为训练的输入, 得到 N 子数据集; 2) 基于子数据集进行训练得到最终得到 N 个基学习器; 3) 预测时每个基学习器都会输出自己的预测产生一个 N 维的向量; 4) 求向量中数据的平均值或者进行多数投票产生最终的预测结果。由于上述的随机划分数据集的做法, 随机森林对类别不平衡的数据适应力很强, 可以处理大量的数据输入, 且学习过程相当迅速。目前随机森林算法已被报道能够胜任在线式的电池容量估计^[99]、SOH 估计^[100]和电池电极材料筛选^[101]。

XGBoost 是梯度提升树(GBDT)机器学习算法的工程实现。其作为提升算法的重要实现方式, 不论是分类问题还是回归问题, 其效果都十分优秀。XGBoost 特别之处在于对缺失值的处理和防止过拟合的方法。对于缺失数据, XGBoost 会将其分配到不同子树并进行信息增益的计算, 最后将信息增益最高的子树作为该缺失值的分类。另外, XGBoost 采用学习率收缩、提前停止优化、选区建树等方法避免来模型的过拟合, 因此在很多问题中能够拥有比其他传统模型更高的精度。Chen 等^[102]针对金属/半导体异质结之间的界面电阻问题, 基于元素周期表建立了一个样本容量为 1 092 的异质结全集并通过第一性原理计算和 XGBoost 算法模型的结合预测这些异质结之间的 Schottky 势垒。最终, Chen 等将筛选出 6 种具有 Ohm 接触以及高隧穿概率的二维 van der Waals 金属-半导体异质结, 为相关器件的发

展提供了可靠的理论指导, 而且在机器学习训练过程中采用了创新性的无监督-监督模型混合预测方法, 为其他类似材料的设计与筛选提供了有力的指导和借鉴。Wang 等^[103]基于 Material Project 数据库的数据, 对 3 880 个尖晶石结构的电学性能进行了预测, 并同时构建了分类模型和回归模型用于预测所有尖晶石材料的带隙并成功筛选出 8 种不仅稳定且具有直接带隙的尖晶石材料(CaAl_2O_4 、 CaGa_2O_4 、 SnGa_2O_4 、 CaAl_2S_4 、 CaGa_2S_4 、 CaAl_2Se_4 、 CaGa_2Se_4 、 CaAl_2Te_4), 加速了尖晶石型材料的结构设计和性能预测研究。

上述介绍的是在电池领域极具代表性的几种人工智能算法。当然, 许多优秀的算法在电池材料设计以及状态预测中不断被探索与发展, 例如符号回归^[104-105]是一种能够从可用数据中提取材料特性显式函数形式的直接方法; 循环神经网络适合用于高分子材料长链结构、与时间序列相关的电池应用场景以及电池文献数据挖掘; 生成式对抗网络在利用化学反应遍历材料合成参数空间的应用中表现卓著^[28]。此外, 在能源电池材料设计领域中, 小样本数据量和高维的数据类型之间的矛盾、模型准确性及模型简单性的矛盾、模型预测结果和材料领域的既有知识之间的矛盾为机器学习目前面对的三大矛盾^[106]。本章节综述的相关研究中总体而言追求大样本的数据量, 更高的准确性, 且较少通过物理、化学等材料基础架构知识对机器学习模型预测结果进行机制性的解释——这是目前研究的缺陷所在, 也是未来研究的一个重要努力方向。机器学习在电池设计领域的应用在近年来发展迅速, 因此在第 3 部分将综述电池设计或者可供电池设计借鉴的创新性研究, 展望未来机器学习在电池设计领域的进一步发展。

3 创新性研究与未来展望

目前在现代能源科学实验室中, 多种人工智能方法已经有极具规模的发展。不论对于理论计算实验科学, 机器学习的应用经验都在相关领域专家的不断实践与创新中变的不同。本节将总结一些创新性的研究并做出一定的展望。

3.1 基于自然语言处理的数据提取

有关新能源电池的计算、实验研究不断地创建各种材料性能(带隙、离子电导率等)、电池性能数据(能量密度、功率密度、库伦效率等), 但这些数据隐藏在各大期刊文本数据之中未被开发, 且未开

发内容正呈指数级增长。作为一种有效的文本数据挖掘技术, 自然语言处理(NLP)已经被应用到电池设计中用于挖掘文献文本、科学数据和表格, 以发现有趣的设计趋势。例如, Swain 等^[107]开发的 ChemDataExtractor 软件是一种基于 NLP 文献科学数据抓取工具, 能够从科学文件中自动提取化学实体及其相关属性、测量和关系软件, 并填充结构化化学数据库。剑桥大学 Cavendish 实验室的 Huang 等^[108]就基于 ChemDataExtractor 软件汇总了若干电池(实验上)的化学信息和相应的电容、导电性、库伦效率、能量密度以及工作电压等信息, 建立了容量高达 292 313 的数据库。同样的, 基于 NLP 得来的大量数据, 研究者们可以从中进行机器学习建模并提取有价值的信息, 比如 He 等^[109]通过 NLP 从文本数据中提取了固体合成的前体数据并利用机器学习从中提取出合成前体的相似性信息。可以预见的是, 在未来 NLP 的极强的数据提取能力将会使其在电池设计的数据准备上越来越重要。

3.2 电池实验逆向设计

如何将机器学习的应用带入实验室一直是相关领域的重要研究课题, 同样的电极引导实验程序中的应用目前仍然处于早期阶段。在上节提到的 NLP 技术发展中, 研究者们提出了一种从文献中抓取合成参数的方法。例如, Kononova 等^[110]利用使用文本挖掘和 NLP 方法从 53 538 个实验合成文献中检索到的 19 488 条合成参数条目。每个条目都包含目标材料、起始化合物、所用操作、合成条件以及合成反应的平衡化学方程式。这种实验参数相关数据的出现催生了机器学习针对合成参数的优化, 实现合成方法、实验流程、实验参数的设计^[111]。Attia 等^[112]开发了一种机器学习方法用于有效地优化电流和电压分布图的实验过程(能够以 6 步的实验获取)。基于早期预测模型和 Bayesian 优化算法他们减少了最大化电池循环寿命所需的实验次数和持续时间, 并最终在实验上验证了他们的方法的有效性和准确性。Liow 等^[9]通过对文献中 11 个领域专家衍生描述符进行聚类, 采用逆设计代理模型建立实验参数-属性关系。该模型无需进行反复试验, 就可以预测设计变量, 从而建立了锂离子电池正极合成方法优化的有效策略。相对于虚拟材料的筛选, 直接的实验参数优化对电池实验过程进行设计能够为电池实验科学家提供更加直接的指导, 也更容易在实验中得到验证。因此, 未来实验参数优化或者材料合成设计将会是机器学习在电池设计研究领域的重

要研究方向。

3.3 图像-性质预测

从图像预测材料的某个性质或者从材料特征预测某个图像是材料领域的新兴研究。Lu 等^[113]通过以下步骤结合了 3 种不同的模型: 1) 使用密集卷积神经网络(DenseNet)从 Raman 光谱预测光致发光图谱; 2) 使用 XGBoost 消除拉 Raman 谱本身对光致发光图谱的影响; 3) 使用支持向量机模型在 Raman 光谱上投影光致发光性能。他们的工作成功揭示了 Raman 光谱和光致放光之间的关系。Kong 等^[114]则是根据材料的特征提取, 建立了一种材料-图谱的预测模型, 成功用材料性能数据预测对应材料的声子态密度以及电子态密度。经过预测后, Kong 等使用从头算进行了验证并发现了候选热电体和透明导体。可见, 材料性能表征图像和性质预测之间的关系连结既有可行性, 也有必要性。对于电池实验而言, 包括电化学阻抗谱, 循环伏安法图像以及横流充放电等图像中均含有大量的信息等待提取。因此, 建立容量大的图像数据库以及针对性的机器学习预测模型建立很可能会成为电池人工智能设计的热点发展方向之一。

3.4 交互式机器学习软件开发

随着机器学习方法的逐渐成熟, 开发一种集成式的机器学习交互软件是促进机器学习在电池设计领域的重要方式。以迁移学习为例, 迁移学习一种在机器学习领域的前沿研究方向, 其基本目的是将相关领域学习到的知识或模型重新应用到不同的但相关的问题中(实现模型举一反三式的迁移)。比如, Zhang 等^[27]通过迁移学习获得了 Li_2S_6 与底物之间吸附位点的任意预测, 且 MAE 低于 0.05 eV, 帮助他们系统地理解了二维材料的吸附性能。这种迁移学习的方式结合无监督学习可以显著减少数据标注所需的工作量, 从而降低计算和工作的成本。但是目前并没有一种软件能够将监督学习、无监督学习以及上述的迁移学习集成起来供研究者们方便地使用。开发一种交互式的机器学习模型训练辅助软件, 将会为有机机器学习训练需求的材料、能源科学家提供指导和帮助, 进一步促进领域的发展和进步。本课题组目前正致力于开发一种基于 Linux 系统的交互式命令行软件(AIMSLab, <http://www.aimslab.cn/>), 围绕监督学习、无监督学习和迁移学习 3 种类型的人工智能范式, 涵盖数据预处理、特征提取、超参数优化, 模型训练等重要功能, 为能源领域研究者们训练相关的机器学习模型提供参考^[115]。随着相关

软件开发的不断完善, 材料、能源领域的人工智能将不用进行很多重复性的数据处理与训练, 并在很大程度上加速机器学习在电池设计领域的发展。

4 结语

相对于试错性的电池表征实验或者电池材料的理论模拟计算, 机器学习可以有效避免昂贵的经济成本和时间损耗。此外, 机器学习以最优化和概率统计的理论为基础, 从数学角度能够有效探索能源电池的微观结构、外部条件、宏观性质的构效关系, 为能源储存系统的机制解析和理论发展提供了合理且高效的指导作用。本综述针对电池领域以机器学习为框架的材料设计(电极材料以及电解质材料)和状态估计(SOC、SOH 以及 RUL 的预测)研究进行了全面的综述, 展示了机器学习在能源电池设计领域的应用现状和应用前景。首先介绍了电池设计的数据获取方式, 包括基于实验或者第一性原理计算的抽样检测方法以及从数据库中获取的方式。然后几种具有代表性的算法(包括线性模型中的 LASSO 模型和弹性网络模型; Bayesian 相关模型的朴素 Bayesian 算法、稀疏 Bayesian 模型以及分层 Bayesian 模型; 高斯过程回归模型; 人工神经网络、卷积神经网络以及图神经网络; 随机森林和极度梯度提升树)在本综述中得到了充分的介绍, 包括他们的优缺点和适用场景, 为电池设计领域的机器学习研究者们提供了模型选择指导。另外, 还强调了近年来机器学习在电池设计领域的创新性研究, 并指出自然语言处理的数据收集、基于实验参数的逆向优化, 实验图像与性质的映射连接以及相关集成软件的发展将会是机器学习在电池设计领域的重点研究方向。本综述旨在为相关工作者们提供可能的研究参考, 并为高速、高精度的电池设计指明方向。

参考文献:

- [1] KOOHI-FAYEGH S, ROSEN M A. A review of energy storage types, applications and recent developments[J]. J Energy Storage, 2020(2): 101047.
- [2] CHEN C, ZUO Y, YE W, et al. A critical review of machine learning of energy materials[J]. Adv Energy Mater, 2020, 10(8): 1903242.
- [3] BACHMAN J C, MUY S, GRIMAUD A, et al. Inorganic solid-state electrolytes for lithium batteries: Mechanisms and properties governing ion conduction[J]. Chem Rev, 2016, 116(1): 140–162.
- [4] FU Z H, CHEN X, ZHANG Q. Review on the lithium transport mechanism in solid-state battery materials[J]. WIREs Comput Molecular Sci, 2022. Doi: 10.1002/wcms.1621.
- [5] WANG Z, ZHENG X, CHEN A, et al. Unraveling the anchoring effect of mxene-supported single atoms as cathodes for aluminum-sulfur

- batteries[J]. ACS Mater Lett, 2022, 4(8): 1436–1445.
- [6] LU C, FANG R, CHEN X. Single-atom catalytic materials for advanced battery systems[J]. Adv Mater, 2020, 32(16): 1906548.
- [7] LI M, LU Jun, JI Xiulei, et al. Design strategies for nonaqueous multivalent-ion and monovalent-ion battery anodes[J]. Nat Rev Mater, 2020, 5(4): 276.
- [8] LIU Y, GUO B, ZOU X, et al. Machine learning assisted materials design and discovery for rechargeable batteries[J]. Energy Storage Mater, 2020, 31: 434–450.
- [9] LIOW C H, KANG Hyeonmuk, KIM Seunggu, et al. Machine learning assisted synthesis of lithium-ion batteries cathode materials[J]. Nano Energy, 2022, 98: 107214.
- [10] KNØSGAARD N R, THYGESEN K S. Representing individual electronic states for machine learning GW band structures of 2D materials[J]. Nat Commun, 2022, 13(1): 468.
- [11] WANG Y, CAI Junfei, HAN Tianli, et al. *In-situ* growing polyaniline nano-spine array on FeVO₄ nanobelts as high-performance rechargeable aluminum-ion battery cathode[J]. Appl Surface Sci, 2022, 591: 153157.
- [12] JENA P, SUN Q. Theory-guided discovery of novel materials[J]. J Phys Chem Lett, 2021, 12(28): 6499–6513.
- [13] WANG Z, CAI Junfei, HAN Yanqiang, et al. Computational screening of spinel structure cathodes for Li-ion battery with low expansion and rapid ion kinetics[J]. Comput Mater Sci, 2022, 204: 111187.
- [14] PATIL M A, TAGADE Piyush, KRISHNAN S H, et al. A novel multistage support vector machine based approach for li ion battery remaining useful life estimation[J]. Appl Energy, 2012, 159: 285–297.
- [15] 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]. J Materiomics, 2013, 65(11): 1501–1509.
- [16] WANG Z, YUAN C, LI X. Lithium battery state-of-health estimation via differential thermal voltammetry with gaussian process regression[J]. IEEE Transact Transport Electric, 2021, 7(1): 16–25.
- [17] ÁLVAREZ ANTÓN J C, GARCÍA NIETO P J, BLANCO VIEJO C, et al. Support vector machines used to estimate the battery state of charge[J]. IEEE Transact Power Electron, 2013, 28(12): 5919–5926.
- [18] JIANG Z, LI J, YANG Y, et al. Machine-learning-revealed statistics of the particle-carbon/binder detachment in lithium-ion battery cathodes[J]. Nat Commun, 2020, 11(1): 2310.
- [19] WANG C, FU K, KAMMAMPATA S P, et al. Garnet-type solid-state electrolytes: Materials, interfaces, and batteries[J]. Chem Rev, 2020, 120(10): 4257.
- [20] CAI J, WANG Z, WU S, et al. A machine learning shortcut for screening the spinel structures of Mg/Zn ion battery cathodes with a high conductivity and rapid ion kinetics[J]. Energy Storage Mater, 2021, 42: 277–285.
- [21] WANG Z, LI X, HAN Y, et al. Harnessing artificial intelligence to holistic design and identification for solid electrolytes[J]. Nano Energy, 2021, 89: 106337.
- [22] ZHAO F, LI Y, WANG X, et al. Lithium-ion batteries state of charge prediction of electric vehicles using rnns-cnns neural networks[J]. IEEE Access, 2020, 8: 98168–98180.
- [23] DAI H, ZHAO G, LIN M, et al. A novel estimation method for the state of health of lithium-ion battery using prior knowledge-based neural network and markov chain[J]. IEEE Transac Ind Electron, 2019, 66(10): 7706–7716.
- [24] LI P, ZHANG Z, XIONG Q, et al. State-of-health estimation and remaining useful life prediction for the lithium-ion battery based on a variant long short term memory neural network[J]. J Power Sources, 2020, 459: 228069.
- [25] SHI S, TU Z, ZOU X, et al. Applying data-driven machine learning to studying electrochemical energy storage materials[J]. Energy Storage Sci Technol, 2022, 11(3): 739.
- [26] XIE Y, CHENG X. Review of state estimation of lithium-ion battery with machine learning[J]. Automotive Eng, 2021, 43(11): 1720–1729.
- [27] ZHANG H, WANG Z, CAI J, et al. Machine-learning-enabled tricks of the trade for rapid host material discovery in Li–S battery[J]. ACS Appl Mater Interfaces, 2021, 13(45): 53388–53397.
- [28] BATRA R, SONG L, RAMPRASAD R. Emerging materials intelligence ecosystems propelled by machine learning[J]. Nat Rev Mater, 2021, 6(8): 655–678.
- [29] MORRIS M D, MITCHELL T J. Exploratory designs for computational experiments[J]. J Statist Plan Inference, 1995, 43(3): 381–402.
- [30] 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(1): 15010.
- [31] EL MEJDOUBI A, CHAOUI H, SABOR J, et al. Remaining useful life prognosis of supercapacitors under temperature and voltage aging conditions[J]. IEEE Transact Ind Electron, 2018, 65(5): 4357–4367.
- [32] BERGERHOFF G, HUNDT R, SIEVERS R, et al. The inorganic crystal structure data base[J]. J Chem Inform Computer Sci, 1983, 23(2): 66–69.
- [33] 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.
- [34] VILLARS P, BERNDT M, BRANDENBURG K, et al. The pauling file, binaries edition[J]. J Alloys Compd, 2004, 367(1): 293–297.
- [35] JAIN A, NG 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.
- [36] CALDERON C E, PLATA J J, TOHER C, et al. The AFLOW standard for high-throughput materials science calculations[J]. Comput Mater Sci, 2015, 108: 233–238.
- [37] GROOM C R, BRUNO I J, LIGHTFOOT M P, et al. The cambridge structural database[J]. Acta Crystallogr Section B, 2016, 72(2): 171–179.
- [38] PENCE H E, WILLIAMS A. ChemSpider: An online chemical information resource[J]. J Chem Educat, 2010, 87(11): 1123–1124.
- [39] QUIRÓS M, GRAŽULIS S, GIRDZIJAUSKAITĖ S, et al. Using SMILES strings for the description of chemical connectivity in the crystallography open database[J]. J Cheminform, 2018, 10(1): 23.
- [40] 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(1): 83.
- [41] SUN S. A survey of multi-view machine learning[J]. Neural Comput Applic, 2013, 23(7): 2031–2038.
- [42] ALOM M Z, TAHAT M, YAKOPCIC C, et al. A state-of-the-art survey on deep learning theory and architectures[J]. Electronics, 2019, 8(3): 292.
- [43] ALZUBAIDI L, ZHANG Jinglan, HUMAIDI A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions[J]. J Big Data, 2021, 8(1): 53.

- [44] LÄNGKVIST M, KARLSSON L, LOUFI A. A review of unsupervised feature learning and deep learning for time-series modeling[J]. *Pattern Recognition Lett*, 2014, 42: 11–24.
- [45] SAXENA A, PRASAD M, GUPTA A, et al. A review of clustering techniques and developments[J]. *Neurocomputing*, 2017, 267: 664–681.
- [46] EMMERT-STREIB F, DEHMER M. High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection[J]. *Mach Learn Knowl Extract*, 2019, 1(1): 359–383.
- [47] LI Z, SILLANPÄÄ M J. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection[J]. *Theoret Appl Genet*, 2012, 125(3): 419–435.
- [48] YOSHIDA T, HONGO K, MAEZONOR. First-principles study of structural transitions in linio_2 and high-throughput screening for long life battery[J]. *J Phys Chem C*, 2019, 123(23): 14126–14131.
- [49] MIN Y, YUAN H, WANG W, et al. Design of heterostructures of mxene/two-dimensional organic frameworks for Na- O_2 batteries with a new mechanism and a new descriptor[J]. *J Phys Chem Lett*, 2021, 12(11): 2742–2748.
- [50] ESTERHUIZEN J A, GOLDSMITH B R, LINIC S. Interpretable machine learning for knowledge generation in heterogeneous catalysis[J]. *Nat Catal*, 2022, 5(3): 175–184.
- [51] PAVLOU M, AMBLER G, SEAMAN S, et al. Review and evaluation of penalised regression methods for risk prediction in low-dimensional data with few events[J]. *Statist Med*, 2016, 35(7): 1159–1177.
- [52] SEVERSON K A, ATTIA P M, JIN Norman, et al. Data-driven prediction of battery cycle life before capacity degradation[J]. *Nat Energy*, 2019, 4(5): 383.
- [53] JIANG B, GENT W E, MOHR F, et al. Bayesian learning for rapid prediction of lithium-ion battery-cycling protocols[J]. *Joule*, 2021, 5(12): 3187–3203.
- [54] ZHOU Y, KANTARCIOGLU M, THURASINGHAM B. Sparse Bayesian Adversarial Learning Using Relevance Vector Machine Ensembles[C]//2012 IEEE 12th International Conference on Data Mining, 2012: 1206–1211. Brussels, Belgium
- [55] AFSHARI S S, CUI S, XU X, et al. Remaining useful life early prediction of batteries based on the differential voltage and differential capacity curves[J]. *IEEE T Instrum Meas*, 2022, 71: 1–9.
- [56] HU C, JAIN G, SCHMIDT C, et al. Online estimation of lithium-ion battery capacity using sparse Bayesian learning[J]. *J Power Sources*, 2015, 289: 105–113.
- [57] ZHOU J, LIU D, PENG Y, et al. Combined Sparse Bayesian Learning Strategy for Remaining Useful Life Forecasting of Lithium-ion Battery[C]//2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control, 2012: 457–461. Harbin, China
- [58] JAFARI M, BROWN L E, GAUCHIA L. Hierarchical bayesian model for probabilistic analysis of electric vehicle battery degradation[J]. *IEEE Transact Transpor Electrifi*, 2019, 5(4): 1254–1267.
- [59] SCHULZ E, SPEEKENBRINK M, KRAUSEA. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions[J]. *J Math Psychol*, 2018, 85: 1–16.
- [60] RICHARDSON R R, OSBORNE M A, HOWEY D A. Gaussian process regression for forecasting battery state of health[J]. *J Power Sources*, 2017, 357: 209–219.
- [61] REN J, CAI J, LI J. High precision implicit function learning for forecasting supercapacitor state of health based on Gaussian process regression[J]. *Scientif Rep*, 2021, 11(1): 12112.
- [62] DERINGER V L, BARTÓK A P, BERNSTEIN N, et al. Gaussian process regression for materials and molecules[J]. *Chem Rev*, 2021, 121(16): 10073–10141.
- [63] HA M, HAJIBABAEI Amir, KIM Dong Yeon, et al. Al-doping driven suppression of capacity and voltage fadings in 4d-element containing Li-ion-battery cathode materials: machine learning and density functional theory[J]. *Adv Energy Mater*, 2022, 12(30): 2201497.
- [64] LI X, WANG Z, YAN J. Prognostic health condition for lithium battery using the partial incremental capacity and gaussian process regression[J]. *J Power Sources*, 2019, 421: 56–67.
- [65] LYU Z, GAO R, LIX. A partial charging curve-based data-fusion-model method for capacity estimation of Li-ion battery[J]. *J Power Sources*, 2021, 483: 229131.
- [66] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks//Advances in Neural Information Processing Systems[C]//2012. <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- [67] SCHMIDHUBER J. Deep learning in neural networks: An overview[J]. *Neural Networks*, 2015, 61: 85–117.
- [68] LU X, XIE Z, WU X, et al. Hydrogen storage metal-organic framework classification models based on crystal graph convolutional neural networks[J]. *Chem Eng Sci*, 2022, 259: 117813.
- [69] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436–444.
- [70] VIDAL C, MALYSZ P, KOLLMAYER P, et al. Machine learning applied to electrified vehicle battery state of charge and state of health estimation: State-of-the-art[J]. *IEEE Access*, 2020, 8: 52796–52814.
- [71] KIM B, LEE S, KIM J. Inverse design of porous materials using artificial neural networks[J]. *Sci Adv*, 2020, 6(1): 9324.
- [72] GOH G B, HODAS N O, VISHNUA. Deep learning for computational chemistry[J]. *J Comput Chem*, 2017, 38(16): 1291–1307.
- [73] SO S, BADLOE T, NOH J, et al. Deep learning enabled inverse design in nanophotonics[J]. *Nanophotonics*, 2020, 9(5): 1041–1057.
- [74] HE W, WILLIARD N, CHEN C, et al. State of charge estimation for Li-ion batteries using neural network modeling and unscented Kalman filter-based error cancellation[J]. *Int J Electr Power Energy Systems*, 2014, 62: 783–791.
- [75] SILVER D, HUANG A, MADDISON C, et al. “Mastering the game of Go with deep neural networks and tree search[J]. *Nature*, 2016, 529(7587): 484–489
- [76] GU J, WANG Zhenhua, KUEN Jason, et al. Recent advances in convolutional neural networks[J]. *Pattern Recognit*, 2018, 77: 354–377.
- [77] ZHOU D-X. Universality of deep convolutional neural networks[J]. *Appl Comput Harmon Anal*, 2020, 48(2): 787–794.
- [78] HU W, HUANG Y, WEI L, et al. Deep convolutional neural networks for hyperspectral image classification[J]. *J Sens*, 2015, 2015: 258619.
- [79] SUN Y, XUE B, ZHANG M, et al. Evolving deep convolutional neural networks for image classification[J]. *IEEE Transact Evol Comput*, 2020, 24(2): 394–407.
- [80] YAGHOUBI M, ALAEI M. Machine learning for compositional disorder: A comparison between different descriptors and machine learning frameworks[J]. *Computat Mater Sci*, 2022, 207: 111284.
- [81] JU J H, SEO Seunghwan, BAEK Sungpyo, et al. Two-dimensional mxene synapse for brain-inspired neuromorphic computing[J]. *Small*,

- 2021, 17(34): 2102595.
- [82] NARAYAN B, NARAYAN A. Machine learning non-Hermitian topological phases[J]. *Phys Rev B*, 2021, 103(3): 035413.
- [83] LYM J, GU G H, JUNG Y, et al. Lattice convolutional neural network modeling of adsorbate coverage effects[J]. *J Phys Chem C*, 2019, 123(31): 18951–18959.
- [84] LEE J-W, PARK W B, LEE J H, et al. A deep-learning technique for phase identification in multiphase inorganic compounds using synthetic XRD powder patterns[J]. *Nat Commun*, 2020, 11(1): 86.
- [85] HORWATH J P, ZAKHAROV D N, MÉGRET R, et al. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images[J]. *npj Comput Mater*, 2020, 6(1): 108.
- [86] LIM B, BELLEC E, DUPRAZ M, et al. A convolutional neural network for defect classification in Bragg coherent X-ray diffraction[J]. *npj Comput Mater*, 2021, 7(1): 115.
- [87] KOPP R, JOSEPH J, NI X, et al. Deep learning unlocks X-ray microtomography segmentation of multiclass microdamage in heterogeneous materials[J]. *Adv Mater*, 2022, 34(11): 2107817.
- [88] DONG H, BUTLER K T, MATRAS D, et al. A deep convolutional neural network for real-time full profile analysis of big powder diffraction data[J]. *npj Comput Mater*, 2021, 7(1): 74.
- [89] SAIDI W A, SHADID W, CASTELLI I E. Machine-learning structural and electronic properties of metal halide perovskites using a hierarchical convolutional neural network[J]. *npj Comput Mater*, 2020, 6(1): 36.
- [90] DONG Y, WU C, ZHANG C, et al. Bandgap prediction by deep learning in configurationally hybridized graphene and boron nitride[J]. *npj Comput Mater*, 2019, 5(1): 26.
- [91] WANG S H, SUN J, PHILLIPS P, et al. Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units[J]. *J Real-Time Image Proc*, 2018, 15(3): 631–642.
- [92] XIE T, GROSSMAN J. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties[J]. *Phys. Rev. Lett.*, 2018, 4(14): 145301.
- [93] AHMAD Z, XIE T, MAHESHWARI C, et al. Machine learning enabled computational screening of inorganic solid electrolytes for suppression of dendrite formation in lithium metal anodes[J]. *ACS Cent Sci*, 2018, 4(8): 996–1006.
- [94] ZHOU L, YAO A M, WU Y, et al. Machine learning assisted prediction of cathode materials for Zn-ion batteries[J]. *Adv Theory Simulat*, 2021, 4(9): 2100196.
- [95] ZHOU J, SHEN LEI, COSTA M D, et al. 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches[J]. *Scientif Data*, 2019, 6(1): 86.
- [96] WU S, WANG Z, ZHANG H, et al. Deep learning accelerates the discovery of two-dimensional catalysts for hydrogen evolution reaction[J]. *Energy Environ Mater*, 2021. Doi: 10.1002/eeem.12259.
- [97] DONG X, YU Z, CAO W, et al. A survey on ensemble learning[J]. *Frontiers Comput Sci*, 2020, 14(2): 241–258.
- [98] WANG Z, HAN Y, LIN X, et al. An Ensemble Learning Platform for the Large-Scale Exploration of New Double Perovskites[J]. *ACS Appl Mater Interfaces*, 2022, 14(1): 717–725.
- [99] LI Y, ZOU C, BERECIBAR M, et al. Random forest regression for online capacity estimation of lithium-ion batteries[J]. *Appl Energy*, 2018, 232: 197–210.
- [100] LIN M, WU D, MENG J, et al. A multi-feature-based multi-model fusion method for state of health estimation of lithium-ion batteries[J]. *J Power Sources*, 2022, 518: 230774.
- [101] SHANDIZ M A, GAUVIN R. Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries[J]. *Comput Mater Sci*, 2016, 117: 270–278.
- [102] CHEN A, WANG Zhilong, ZHANG Xu, et al. Accelerated mining of 2d van der waals heterojunctions by integrating supervised and unsupervised learning[J]. *Chem Mater*, 2022, 34(12): 5571–5583.
- [103] WANG Z, ZHANG H, LI J. Accelerated discovery of stable spinels in energy systems via machine learning[J]. *Nano Energy*, 2021, 81: 105665.
- [104] SUN S, OUYANG R, ZHANG B, et al. Data-driven discovery of formulas by symbolic regression[J]. *MRS Bull*, 2019, 44(7): 559–564.
- [105] WANG Y, WAGNER N, RONDINELLI J M. Symbolic regression in materials science[J]. *MRS Commun*, 2019, 9(3): 793–805.
- [106] LIU Y, ZOU X, YANG Z, et al. Machine learning embedded with materials domain knowledge[J]. *J Chin Ceram Soc*, 2022, 50(3): 863.
- [107] SWAIN M C, COLE J M. ChemDataExtractor: A toolkit for automated extraction of chemical information from the scientific literature[J]. *J Chem Inform Model*, 2016, 56(10): 1894–1904.
- [108] HUANG S, COLE J M. A database of battery materials auto-generated using ChemDataExtractor[J]. *Scientif Data*, 2020, 7(1): 260.
- [109] HE T, SUN W, HUO H, et al. Similarity of precursors in solid-state synthesis as text-mined from scientific literature[J]. *Chem Mater*, 2020, 32(18): 7861–7873.
- [110] KONONOVA O, HUO Haoyan, HE Tanjin, et al. Text-mined dataset of inorganic materials synthesis recipes[J]. *Scientif Data*, 2019, 6(1): 203.
- [111] SEGLER M H S, PREUSS M, WALLER M P. Planning chemical syntheses with deep neural networks and symbolic AI[J]. *Nature*, 2018, 555(7698): 604–610.
- [112] ATTIA P M, GROVER Aditya, JIN Norman, et al. Closed-loop optimization of fast-charging protocols for batteries with machine learning[J]. *Nature*, 2020, 578(7795): 397–402.
- [113] LU A-Y, MARTINS L G P, SHEN Pin-Chun, et al. Unraveling the correlation between raman and photoluminescence in monolayer MoS₂ through machine learning models[J]. *Adv Mater*, 2022, 34(34): 2202911.
- [114] KONG S, RICCI F, GUEVARRA D, et al. Density of states prediction for materials discovery via contrastive learning from probabilistic embeddings[J]. *Nat Commun*, 2022, 13(1): 949.
- [115] WANG Z, HAN Y, CAI J, et al. Vision for energy material design: A roadmap for integrated data-driven modeling[J]. *J Energy Chem*, 2022, 71: 56–62.