

机器学习在固体氧化物燃料电池、锂电池、CO₂ 电还原催化剂中的研究进展

许建兵^{1,2}, 李翰实³, 谭济民⁴, 韩敏芳⁵, 陈 迪^{1,6}

(1. 清华大学未来实验室, 中国 北京 100084; 2. 清华大学美术学院, 中国 北京 100084; 3. 耶鲁大学机械工程与材料科学系, 美国 康涅狄格州 纽黑文市 06520; 4. 纽约大学数据科学中心, 美国 纽约州 纽约 10012; 5. 清华大学能源与动力工程系, 中国 北京 100084; 6. 清华大学材料学院, 中国 北京 100084)

摘 要: 随着清洁能源系统的推广应用, 锂离子电池、固体氧化物燃料电池作为清洁能源器件受到广泛的关注。然而, 作为复杂的电力动力系统, 电池的商用化一直面临长时间、多维度、高精度的性能预测需求, 一些新型的电池性能预测方法仍处于起步探索阶段。近年来, 随着人工智能的普及与推广, 机器学习这项基于传统人工神经网络的技术被国内外研究者所重视。机器学习等数据科学的最新进展为科学和工程界提供了灵活而快速的预测框架, 在材料研发等方面显示出巨大的应用前景。本工作总结了采用机器学习方法用于固态氧化物燃料电池、锂电池、CO₂ 电化学还原催化剂的最新进展, 并对未来的发展方向提出了若干建议。

关键词: 数据科学; 机器学习; 绿色能源; 固态氧化物燃料电池; 锂电池; 二氧化碳电化学还原催化剂

中图分类号: TQ174.75 文献标志码: A 文章编号: 0454-5648(2022)11-3021-19

网络出版时间: 2022-09-29



Research Progress on Data Science of Solid Oxide Fuel Cells, Lithium Batteries, CO₂ Electroreduction Catalysts

XU Jianbing^{1,2}, LI Hanshi³, TAN Jimin⁴, HAN Minfang⁵, CHEN Di^{1,6}

(1. The Future Laboratory, Tsinghua University, Beijing 100084, China; 2. Academy of Arts & Design, Tsinghua University, Beijing 100084, China; 3. Department of Mechanical Engineering and Materials Science, Yale University, New Haven 06520, CT, USA.; 4. Center for Data Science, New York University, New York 10012, NY, USA.; 5. Department of Energy and Power Engineering, Tsinghua University, Beijing 100084, China; 6. The School of Materials Science and Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Lithium ion batteries and solid oxide fuel cells have become popular as clean energy devices. However, the commercial application of the batteries as complex electric power systems requires the batteries with long-term, multi-dimensional and high-precision performance. Some battery performance prediction methods are still in the initial stage of exploration. With the popularization and promotion of artificial intelligence, machine learning based on the artificial neural network technology has attracted recent attention. Recent advances in data science, such as machine learning, provide the scientific and engineering communities with flexible and rapid prediction frameworks, indicating great application prospects in materials research and

收稿日期: 2022-05-28。 修订日期: 2022-06-30。

基金项目: 国家重点研发计划(2021YFB3502000); 清华大学国强研究院人工智能与机器人领域通用类自由探索研究项目(2020GQG1003); 微软亚洲研究院基金资助(PO99386669); 中国博士后科学基金面上资助(2020M680537); 国家自然科学基金(52102107, 52102137); 新型陶瓷与精细工艺国家重点实验室开放课题(KF202101)。

第一作者: 许建兵(1991—), 男, 博士。

通信作者: 陈 迪(1987—), 男, 博士, 副研究员。

Received date: 2022-05-28. Revised date: 2022-06-30.

First author: XU Jianbing (1991—), male, Ph.D.

E-mail: xujianbing@mail.tsinghua.edu.cn

Correspondent author: CHEN Di (1987—), male, Ph.D., Associate Professor.

E-mail: dichen@tsinghua.edu.cn

development. This review summarized the latest advances on machine learning for the development of renewable energy technologies (*i.e.*, solid oxide fuel cells, lithium batteries) and carbon reduction technologies, and gave some comments for the future development directions.

Keywords: data science; machine learning; renewable energy; solid oxide fuel cells; lithium batteries; carbon dioxide electrochemical reduction catalysts

随着现代工业社会的发展,能源危机和环境污染已成为困扰人类的两大社会问题,开发新型可持续发展的环保能源显得尤为重要^[1–3]。1) 作为清洁的能源转换系统,固体氧化物燃料电池(SOFC)能够在不需要燃烧的情况下,直接将燃料化学能高效转化为电能^[4],因此受到广泛关注。锂离子电池由于能量密度高,循环次数多,成本相对较低等优点被广泛应用于移动设备和电动汽车领域中^[5]。但是在应用过程中,SOFC 与锂离子电池作为动态的能源系统,反应过程中涉及离子的转移等复杂的电化学反应^[6–7]并伴随着电、热、力等耦合过程^[8],电池的寿命无法被简单准确的预测或测量,需要花费大量时间成本来做破坏性试验来得到某种电池的寿命数据。目前对电池性能退化机制尚无清晰的理论认知,更缺乏相应的电池寿命延长策略。2) 化石能源的过度使用造成了严重的碳排放,减碳已成为当前社会的共识。通过电化学方法实现 CO₂ 循环和能量再生被认为是重要的固碳手段,其关键挑战之一是高性能 CO₂ 电化学还原催化剂的研发。当前基于经验指导多轮实验迭代寻找高效催化剂的模式难以实现催化剂的高通量筛选与快速研发。

近年来,研究者通过将理论物理化学、材料科学与人工智能、机器学习、深度学习等数据科学相结合,加快新能源材料的研发进度^[9]。这种数据驱动的方法降低了传统方法的试错实验成本,减少了计算–实验周期时间,并且有望将技术的迭代更新时间和新材料发现速率提高 10 倍^[10]。目前,基于机器学习的数据驱动方法已经在许多方面产生了巨大影响,计算能力的提高和大数据的出现使得机器学习在线上零售购物^[11–12]、医学诊断^[13]、图像和语音识别^[14]、实时配音等方面的取得了巨大成功。

机器学习与材料学等自然科学相结合,将有助于新的方法体系的建立。由于可以自动建立数据之间的内在联系,机器学习技术在处理复杂电池系统的性能预测方面显示出独特的优势。本工作主要概叙述了包括各种机器学习技术在 SOFC、锂电池新能源体系以及 CO₂ 电化学还原催化剂研发方面的应用情况,并对机器学习在绿色能源中的发展进行了

相应的展望。

1 机器学习在电池技术中的应用

1.1 机器学习方法概述

近年来,随着机器学习的应用的增加,数据科学家以及计算机科学家开发了多种机器学习的模型和算法。每一种模型和算法都有各自的优点与不足。机器学习模型依照结构可以分为较为复杂模型比如多层感知器^[15](又称人工神经网络或深度学习)和相对简单的模型,例如决策树和支持向量机等等。其中,深度学习不仅可以用于复杂数据结构中,也可适用于简单的数据分析。但一般来说,深度学习类型的模型需要的计算资源多并且编写调试时间长。

在大多数的数据分析中,数据的结构和逻辑相对简单,通常可以使用一个“数据点—自变量”的表格表示出来所有的数据,并且自变量和因变量之间没有复杂的关系。在这类问题上一些适用的简单机器学习模型往往能在更少的时间内也能给出和深度学习相似的结果。

当输入数据类型为图片、序列、或者类似这种需要从数据原始形态中学习复杂表征的数据,深度学习最为适用。由于这类模型通常有着很高纵向的深度和层级,使得它们能够从最原始的数据形式开始学习并且在多个层级中自己学习出复杂的变换以拟合原始数据内复杂的规律。深度学习虽然强大但也有不足^[16]。由于深度学习是黑箱算法,研究人员很难解释模型的内部状态,所以模型的稳定性不可估量^[17]。而且如果采用过于复杂的深度学习模型,但是只有小规模的数据集,会很容易出现过拟合的情况从而影响模型在测试集上的性能。

总的来说,当输入数据类型为人为整理过的数据或者数据本身和需要预测的变量之间逻辑非常清楚的情况下,线性回归类方法或者其他结构相对简单的机器学习模型最为高效。当输入数据类型为图像,需要使用结构复杂的原始数据并且不适合转换成多个因变量的时候,深度学习更为适合。但是,由于大多数问题的数据结构简单,运用高效的回归类方法,或许依然可以得到类似结果。当研究的问

题不清楚是否需要使用深度学习这一类模型的时候, 可以先使用简单的机器学习模型。这样的模型比深度学习具有更好的可解释性、可预测性和稳定性。如果简单的机器学习模型拟合能力不足, 说明数据过于复杂, 可以换用深度学习来加强模型的学习能力。

目前, 常用的机器学习方法主要有分类、回归、聚类等几种类型, 具体的常用方法包括线性回归、决策树、支持向量机(SVM)、k 邻近算法(kNN)、随机森林(RF)、人工神经网络(ANN)等^[18-19]。线性回归是利用数理统计中回归分析, 来确定 2 种或 2 种以上变量间相互依赖的定量关系的一种统计分析方法; 决策树是 1 个树结构, 其每个非叶节点表示 1 个特征属性上的测试, 每个分支代表这个特征属性在某个值域上的输出, 使用决策树进行决策的过程就是从根节点开始, 测试待分类项中相应的特征属性, 并按照其值选择输出分支, 直到到达叶子节点, 将叶子节点存放的类别作为决策结果; SVM 是在分类与回归分析中分析数据的监督式学习模型与相关的学习算法, 可以分析数据, 识别模式, 用于分类和回归分析; kNN 是一种基本分类与回归方法, 它通过搜索整个训练集内 k 个最相似的实例, 并对这些 k 个实例的输出变量进行汇总, 来预测新的数据点; RF 是对决策树的一种调整, 相对于选择最佳分割点, RF 通过引入随机性来实现次优分割; ANN 是由大量处理单元互联组成的非线性、自适应信息处理系统, 它能自动建立起输入与输出之间的关系, 具有自适应、自动学习的能力。后面根据文献总结情况进一步归纳每种机器学习方法的特点。

1.2 SOFC

近年来, SOFC 吸引了燃料能源领域的关注。首先, 燃料电池拥有固态电解质夹在阳极和阴极之间的简单三明治结构, 并且固态燃料电池不需要昂贵的催化剂。这些特性使得燃料电池的制造成本可以被控制在可接受范围内。其次, 固态燃料电池能够在不需要燃烧的情况下, 直接将燃料化学能高效转化为电能。在一般运行条件下, 固态燃料电池可以提供大约 45% 的能量转化率; 在加压的固态燃料电池-汽轮机系统中, 效率可达 55%^[20]。这样的能量转化率远高于一般燃烧化石能源发电的效率。以上 2 点使得固态燃料电池通常被视为人类完全转移到可再生能源之前的良好的过渡产品。然而, SOFC 系统的复杂程度远超传统的发电汽轮机, 导致我们对于其工作条件原理状态的认识相对不足, 这使得

SOFC 系统的平稳高效持久运行成为了能源科学界的焦点。目前, 为了优化固态燃料电池的输出功率, 运转效率以及延长固态燃料电池的寿命, 研究方向主要集中于通过系统状态监测以及改变运行参数来优化固态燃料电池的运行性能和出现问题时即时故障诊断及修复。

固态燃料电池系统是 1 个集合了众多物理变量的复杂非线性系统。这些变量包括进料出料流量和温度、进料出料口的形状、电压、电流、电池内部温度等等^[19]。这些物理变量相互关联影响, 错综复杂, 导致对于固态燃料电池系统的控制和运行状况的判断成为了难题。此外, 在固态燃料电池的工业运用中, 一些关键运行参数难以直接准确获取, 通常只能依靠经验判断, 这也提高了掌握固态燃料电池系统的运行情况难度。在这种情况下, 快速模拟未知参数并且根据测量和模拟出的参数进行系统运行状态优化成为了固态燃料电池领域的难题, 并且现有技术很难解决这些棘手的问题。在这种情况下, 基于统计和大数据的深度学习因其快速准确的模拟和预测能力吸引了很多学者的关注。下面就机器学习在固态燃料电池领域的应用做了介绍, 汇总情况如表 1 所示。

1.2.1 系统状态与性能监测 SOFC 在工作状态时会发生多个电化学/化学反应, 从而导致复杂的电/热产生过程^[20], 是一个典型的非线性系统。监测总体的系统状态需要知道 SOFC 在所有位置的多物理场数据。虽然可以在电池中放置多个探针, 但是由于每个探针只能测量 1 个点的数据, 仍然无法得到电池内所有位置的数据。一个可行的解决方案是使用多物理场建模(MPS)的程序进行模拟。MPS 可以生成 SOFC 内部运行的所有情况, 可以提取每个位置的信息来得到所需监测的全部数据。常用的 MPS 软件有 Comsol, Ansys, CFD 等。此类建模基于物理和化学反应, 模拟结果相对准确, 但是通常模拟运行时间较长。

在系统监测中, 一个常见的应用场景是通过获得电池的基本信息和的情况, 实时预测电池内的一些关键运行数据例如系统运行的最高温度。通过多物理场建模的方式预测通常会运行一整个系统, 耗费巨大的计算量和时间^[21], 如果用 MPS 预测一个已经运行 1 个月的电池, 通常我们需要模拟过去 1 个月的历史状态才能够得到结果。所以 MPS 不适用于需要通过当前数据进行快速单个或者多个关键数据预测的任务。基于机器学习的方法, 可以建立

表 1 最近关于机器学习在 SOFC 中性能预测、故障诊断和性能优化方面的研究总结

Table 1 Summary of recent research on machine learning for performance prediction, fault diagnosis, and performance optimization in Solid Oxide Fuel Cell (SOFC)

Research contents	Methods	Model classification	Method selection basis	Virtue	Weakness	Data Set	Performance indicators	References
SOFC system status and performance monitoring	Artificial Neural Network (ANN)	Regression	Multifactor input conditions	Small error and high fitting degree	Easy to overfit	Data for 80 different operating conditions	Mean error< 1% maximum error<4%	[24]
	Least square support vector machines (LS-SVMs)	Regression	Establish mathematical relation of dynamic system based on input and output data	Faster convergence (26s) and higher global search precision	Long total calculation time	16 input and output parameters	Root mean square error in series-parallel mode<3%	[25,26]
	Recurrent neural network (RNN)	Sequence, regression	Use long-term data effectively	Implement parallel and sequential computation	Not suitable for long sequences, easy gradient disappearance	1 000 data samples	Maximum error: 5.29%, mean error: 0.25%	[27]
	Bayesian sparse polynomial chaos expansion (BSPCE)	Regression	Handle multiple types of problems for sensitivity analysis	Handle multiple tasks	Accuracy depends on large samples	Data set: $2^{11} \times 27$	Accuracy> 92%	[28]
	Back Propagation Neural Networks (BPNN), Random Forest (RF), Support Vector Machine (SVM)	Classification, rank	BPNN for multifactor input conditions, RF for feature sorting, SVM for classification problems			Data set: 858set, training set: 650set	BPNN(1%), SVM(4%), RF(6%)	[29]
SOFC performance optimization	Deep Neural Networks (DNN), Genetic Algorithm (GA)	Search	Multi-factor input conditions, decision optimization	Find the optimal solution	Overfitting	Data set: 17 384×11	Mean error 1% fuel utilization efficiency >65%	[30]
	SVM, GA	Regression	Solve the constrained nonlinear optimization problem	Find the optimal solution	High sensitivity to parameters	Data set: 1000	Mean error <10%	[31]
	ANN, Radial Mobility Optimizer (RMO)	Regression	RMO handles optimization problems quickly and requires small computational memory	Find the optimal solution	Easy to overfit	Data set: 538, training number: 100	Accuracy: >99%	[32]
	k-Nearest Neighbor (kNN), ANN, Bayes regression, Logistic regression	Regression	Compare the prediction results of different methods			Data set: 538, training number: 250	ANN is the best, accuracy >98%	[33]
	DNN, GA	Search	Multi-factor input conditions, decision optimization	Find the optimal solution	Parameters are complex and difficult to interpret	Data set: 1 882×7, training set: 80%	Achieve durability and high performance	[34]

(to be continued)

(Table 1 continued)

Research contents	Methods	Model classification	Method selection basis	Virtue	Weakness	Data Set	Performance indicators	References
SOFC fault diagnosis	SVM	Classification	Efficient classification technology for multi-class classification problems	Dealing with nonlinear problems	Parameter sensitivity	Data set: 30 000×197	The diagnostic accuracy of single fault is more than 95%	[35]
	RF, SVM	Classification, rank	The valid features are automatically extracted and the weights of the whole neural network are initialized using the raw data in SOFC	Deal with nonlinear, multi-factor problems	Parameter sensitivity	3 fault types and 25 influencing factors	The fault diagnosis rate is more than 90%	[36]
	Stacked Sparse Autoencoder (SSAE)	Representative learning, unsupervised learning	The valid features are automatically extracted and the weights of the whole neural network are initialized using the raw data in SOFC	Multifactor problem	Depend on the original factor	10 fault types and 16 input variables	The accuracy of fault diagnosis is close to 80%	[37]
	Principal Component Analysis (PCA), SVM	Classification	Multivariate statistical analysis combined with supervised machine learning methods	/	/	70 000 data set	The false detection rate is less than 7%	[38]

数据点与数据点之间非线性的映射，不需要遵循物理法则。所以对于同样预测运行 1 个月的电池，可以忽略时间，通过观察到的数据直接预测需要监测的数据。在现在的监测应用当中，我们通常使用 MPS 来生成模拟数据作为机器学习模型的训练数据。

Arriagada^[24]在 2002 年提出使用深度学习模型来预测电池运行状态，如图 1a 所示。Arriagada 首先使用 MPS 生成了训练数据，然后用 1 个 2 层的神经网络当作预测模型在生成的数据集上训练。输入模型的数据为燃料利用率、气体温度、电压和燃料流量。输出的数据为空气流量、电流密度、空气出口温度、燃料出口温度、平均固体温度和可逆电压。训练过的模型在测试集上的预测误差最大不超过 4%。Ghorbani^[33]也使用了一些可观测数据作为输入数据来预测燃料供给。Ghorbani 采用了 4 种不同的机器学习模型，包括 Naive Bayes、Logistic Regression、kNN 和神经网络。其中神经网络和 kNN 表现最好，误差率都低于 4%。

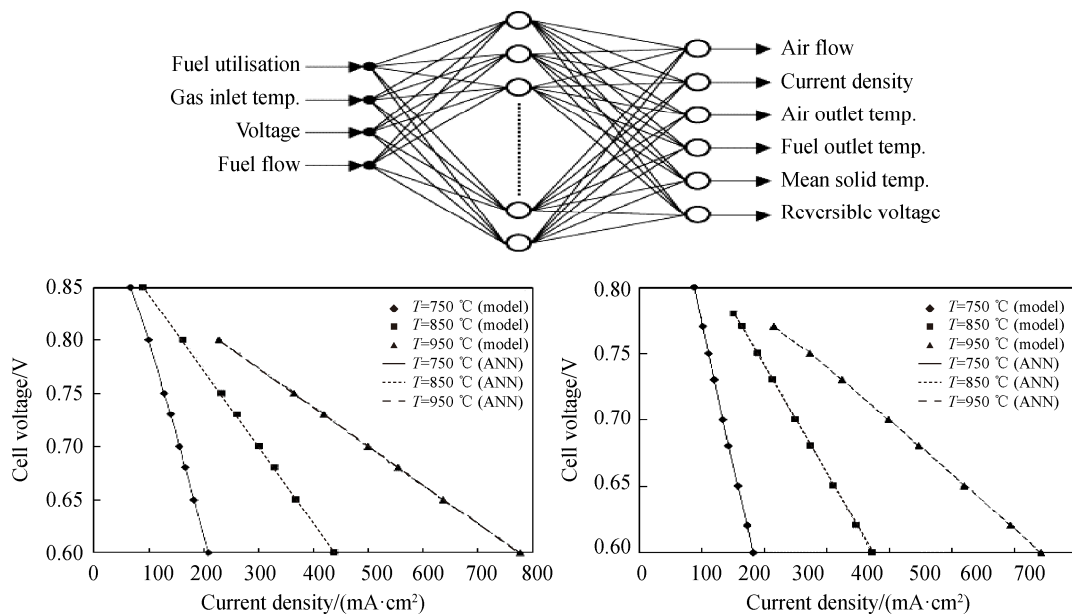
Arriagada 和 Ghorbani 都没有包括真实的实验数据。一般来说，实验数据由于收集过程过于昂贵，

在目前经常作为对 MPS 的检验或者作为测试集对机器学习模型进行检验。Haoran Xu^[30]的研究如图 1b 所示。Haoran Xu 先用 MPS 模型生成了训练数据，然后用真实的实验数据校验了 MPS 模型的准确性。收集了系统运行过程中可以直接观测的数据，包括电流密度、阳极流量、阴极流量、和温度。需要监测的关键数据是最高温度梯度和系统产热，这 2 个监测数据反应了电池运行的状况和效率，但是很难在运行过程中直接测量。研究者首先用 MPS 模型生成了 17 384 个包含可观测数据和监测数据的数据点，然后用 3 层的深度神经网络(DNN)作为预测模型，在 80%的生成数据上进行训练，然后用剩下的 20%进行测试。用可观测的数据预测监测数据，在测试集上的准确率高达 99.9%。

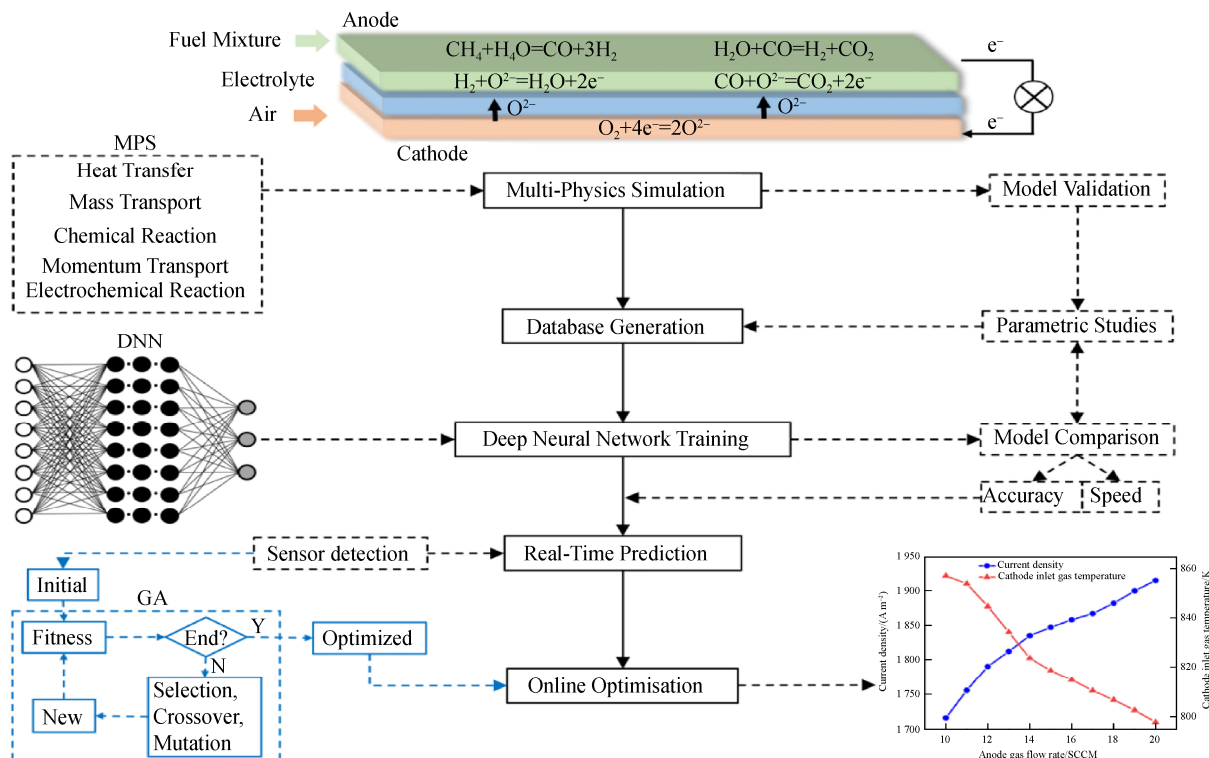
1.2.2 性能优化 预测和分析电池数据，其主要目的是通过得到的数据来调整电池参数，从而优化电池运行效率和延长电池寿命。想要得到最优的参数组合有 2 大计算上的难点，第一在于常规的实验甚至 MPS 模拟也需要耗费大量时间，这个问题可以通过以上训练过的机器学习模型来解决。由于机器学习模型不需要物理建模，可以直接一步生成需要

预测的结果, 速度很快。第二在于如何找到最优的参数组合, 燃料电池中有着大量的可调参数, 并且需要搜索的参数组合数量随着参数的增加以指数级上涨。即使单独的性能预测时间很短, 使用穷举

法暴力求得全局最优解依然需要花费大量时间。所以, 为了解决寻找最优解的问题, 可以使用计算机科学中的优化算法, 其中包括遗传算法, 模拟退火, Bayes 优化, 随机优化等。



(a) Using Artificial Neural Network (ANN) to predict SOFC performance results to get SOFC voltage and current density relationship with H_2 and CH_4 fuels



(b) SOFC training set database was established by multi-physical field simulation, and then the Deep Neural Networks (DNN) model was used to train the data set, and the SOFC DNN prediction model was established. The SOFC parameters were optimized by Genetic Algorithm (GA), and the optimal solution was obtained to predict the maximum temperature gradient and heat production of SOFC

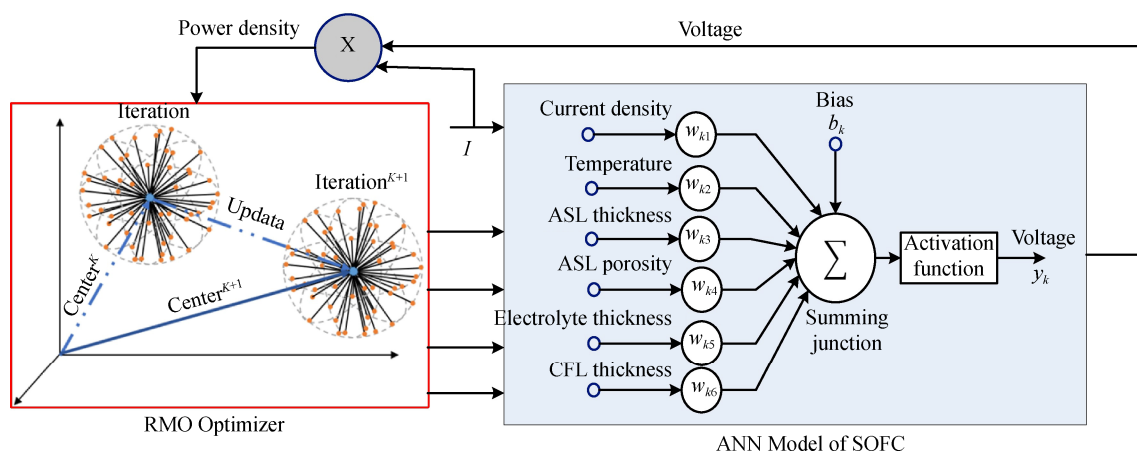
图1 机器学习对 SOFC 性能预测研究^[24, 30]

Fig. 1 Machine learning for SOFC performance prediction^[24, 30]

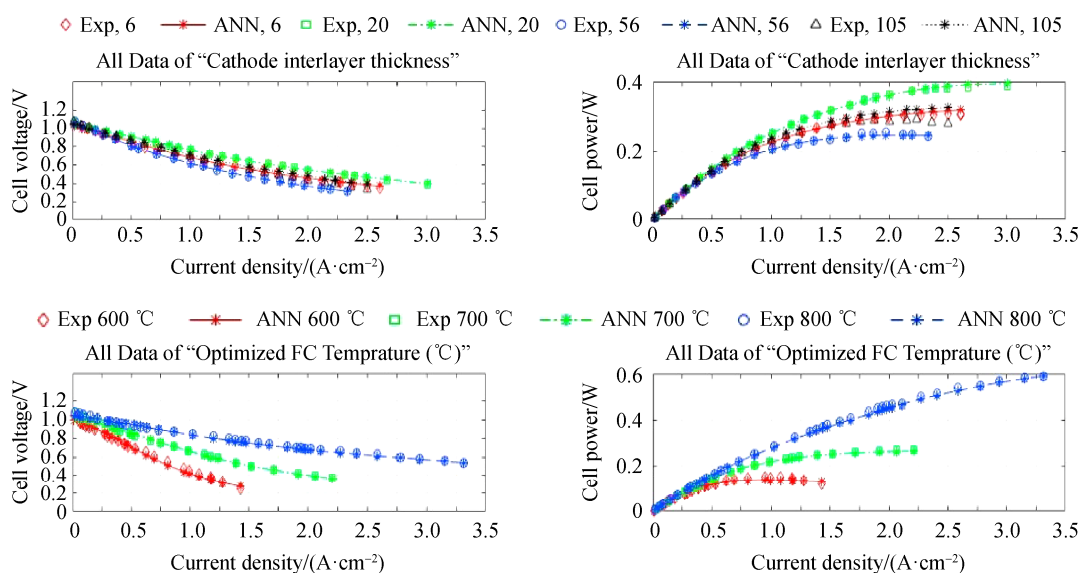
Haoran Xu^[30]采用了遗传算法在 7 种不同的参数中搜索最优组合, 参数包含阴极阳极气体流速、气体流量、阳极气体组成、阴极气体成分的空气、阳极入口温度、阴极进气温度 and 外加电压。遗传算法是 1 种进化优化算法, 以 1 种群中的所有个体为对象, 并利用随机化技术指导对 1 个被编码的参数空间进行高效搜索方法。Haoran Xu 使用训练过的神经网络, 用电流密度作为优化指标, 用遗传算法优化, 进化 20 次后就稳定在最优解模型。筛选出的模型在给定的工作环境中最低效率也高于 65%。

相同地, Nassef^[32]使用了径向移动优化器(RMO)的优化算法来寻找 SOFC 参数的最佳来达到最大的功率如图 2 所示。Nassef 首先将数据处理后,

得到 6 个特征, 分别为电流密度、温度、电解质厚度、阳极支撑层厚度、阳极孔隙率和阴极厚度。然后这些数据给到一个神经网络来预测 SOFC 的电压。此处的预测和 Xu^[30]中的神经网络一样, 都是用来快速得到所需的目标值, 不同的是这里预测的目标值是电压。完成训练的神经网络和作者提出的 RMO 优化算法结合在一起, 能够快速筛选出实现 SOFC 功率最大化的最优参数组合。RMO 利用了加权的局部最优解和当前全局最优解一同寻找优化参数的方向。由于 RMO 只需要当前信息便可以做出下一步优化的方向, 不需要储存所有历史数据, 所以计算速度快而且对存储要求低。与遗传算法对比, RMO 找到的最优解 SOFC 的功率要大 20%。



(a) Relationship between input parameters and output parameters (voltage) was established by ANN for SOFC performance prediction model, and Radial Mobility Optimizer (RMO) was used to quickly optimize the performance of the model



(b) Output parameter of voltage, current density by the neural network and experimental data of SOFC at 600 °C, 700 °C and 800 °C under the inter cathode layer thickness of 6 μm, 20 μm, 56 μm and 105 μm

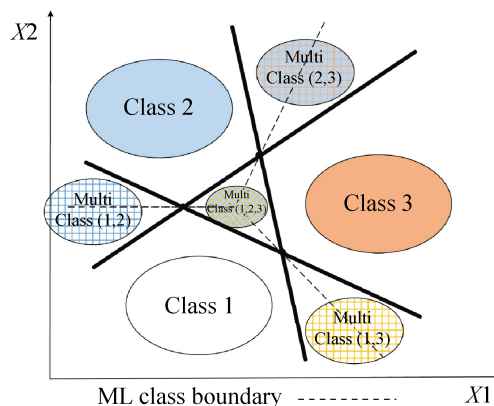
图 2 机器学习对 SOFC 性能优化研究^[32]
Fig. 2 Machine learning for SOFC performance optimization^[32]

除了遗传算法和 RMO, 之前提到的例如 Bayes 优化和模拟退火也可被用在此类优化问题中, 因为我们可以使用神经网络快速地预测电池性能, 所有的优化算法都适用于此类问题。

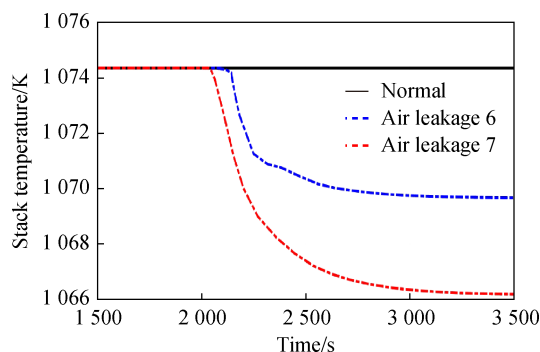
1.2.3 故障诊断 因固态燃料电池通常需要高温高压的运行环境, 系统稳定性和可靠性表现并不令人满意, 经常会出现泄漏气体或者燃料的故障。这类故障会导致整个系统运行功率效率下降或者停止, 极端情况下甚至会导致系统的损坏和高昂的维修费用。在这种情况下, 及时准确的故障辨别和诊断就变得非常重要, 因为这种诊断可以为系统控制提供信息来避开出现问题的部分。对于故障的诊断一般可分为 2 种, 一种是单个类型故障的逐一筛查, 另一种是故障的综合诊断。在固态燃料电池系统的实际使用中, 很多情况下是多种类型故障一起发生, 所以故障的综合分析诊断就显得尤为重要。综合故障诊断有 2 个主要的难点: 1) 电池的多个故障可以叠加, 从而在外表现出不同的形态, 增加了故障诊断的难度; 2) 故障的类型很多, 加上其可以叠加的特性, 使得故障组合的数量呈指数级增加。即使使用计算机生成数据, 涵盖所有可能出现的故障也很困难。由于传统的机器学习模型例如逻辑回归对于样本外的预测能力不强, 如果用在在故障诊断里需要看过所有类型的数据, 否则性能会大打折扣。Zhang 等^[37]提出了一种基于堆叠稀疏自编码器 (SSAE) 和神经网络的模型来解决这个问题, 如图 3 所示。由于穷举所有故障和运行状况不可行, Zhang 等在随机的运行状况下随机取样了 100 000 个故障组合, 分为训练和测试数据。作者横向对比了 SVM, 神经网络和带 SSAE 的神经网络。结果显示神经网络和带 SSAE 的神经网络预测性能更好。主要原因在于神经网络, 尤其是 SSAE 能够解耦复杂故障的相互间作用, 分离出不同单独的故障, 故预测性能更好。深度学习在故障诊断方面, 由于变量解耦的需要, 会持续地高于普通机器学习性能。也推荐在相似的问题上使用深度神经网络。

从机器学习在 SOFC 中的发展历程可以看出, 目前的研究还是基于模拟仿真的数据作为数据集, 对 SOFC 的性能进行预测、优化以及进行故障诊断, 缺乏真实的实验数据。由于 SOFC 收集数据的难度比较大, 暂时还没有研究者采用试验数据进行此方面的研究, 因此 SOFC 中机器学习方法的正确性缺乏实验上的验证。此外, 如果建立的模型过于简单, 采用简单的拟合方法就可以实现, 机器学习方法反

而容易出现过度拟合的现象, 这便是模型复杂性与易用性的矛盾与统一^[18]。未来关于机器学习在 SOFC 领域的应用, 需要采用真实的实验数据作为机器学习的数据集, 这有助于增强模型的准确度和实用性。



(a) SVM is used to construct hyperplane in high dimensional space to classify the samples. This method is used for SOFC fault classification



(b) Typical type of fault diagnosis with infection fuel gas leakage induced by reactor temperature

图 3 机器学习对 SOFC 故障分析研究^[35]

Fig. 3 Machine learning for SOFC fault analysis^[35]

1.3 锂电池

进入 20 世纪以来, 人类社会对于电能储能介质的要求与日俱增。一方面, 伴随着手机平板电脑等移动计算设备不断更新迭代, 处理器性能不断提升, 这些设备的耗电量也不断增大, 有限的机内空间与增长的耗电量之间的矛盾越来越显著。另一方面, 因为化石能源即将枯竭, 温室效应不断加剧, 汽油价格及其税费居高不下, 对电动汽车的需求日益增加。随着技术的成熟, 锂离子电池显示出了其作为能源储存介质的优点: 能量密度高, 循环次数多, 成本相对较低。由于这些优点, 其被广泛应用于移动设备和电动汽车领域中, 一定程度上解决了过去电能储能介质(例如镍氢电池, 铅酸电池)性能较差

的问题。但是在应用过程中, 锂离子电池的一些问题也暴露出来。首先, 锂离子电池的寿命无法被简单准确的预测或测量, 这导致在我们对电池的测试中必须要花费大量时间成本来做破坏性试验得到某种电池的寿命数据。在实际应用中, 由于不知道锂电池的剩余寿命, 对于其更换的计划也无法准确制定。其次, 由于无法精确观测到电池内离子分布状况, 电池剩余电量的准确检测也是 1 个难题。尽管电池的一些运行数据, 例如电压电流, 可以提供一些线索, 因为变量和影响因素众多, 依靠传统模拟方法很难准确估计电池剩余电量。随着人工智能技术的发展和普及, 科学家们发现利用人工智能方法(例如深度学习)可以更好地解决这 2 个问题。在使用某电池一段较短的时间后, 就可以根据已有电池数据, 应用机器学习的方法比较精确的预测其寿命和剩余电量。机器学习的应用大大提高了预测的效率并减少了成本。此外, 对于电池材料的探索也受益于机器学习。每一种材料都有很多复杂的性质, 1 个电池内不同材料的性质叠加在一起决定了电池的性

能。由于不同材料的性质繁杂且互不相关, 人工利用实验穷尽所有材料组合并找出最佳组合几乎是不可能完成的任务。机器学习利用了计算机高计算能力高存储能力的优势, 可以通过对已有实验结果的学习来预测位置材料的性质。这就给人工实验提供了明确的方向以及很强的指导作用, 表 2 总结了相关的研究情况。

1.3.1 寿命预测 在锂电池研究中, 预测电池的使用寿命是一项重要的任务, 尤其在给电动汽车装配的锂电池上。锂电池的运行是一个动态的过程, 所以电池寿命会被许多因素影响, 包括运行电量、电流、温度以及一些人为因素, 例如电动汽车中驾驶员的驾驶习惯。由于锂电池寿命预测涉及到的变量总体包括电压和电流曲线, 相对简单, 通常来说使用普通的机器学习模型即可拟合损耗曲线作出预测。常用的方法有支持向量回归(SVR)、线性回归以及 DNN。Nuhic^[40]使用了 SVR 搭配一些类似标准化数据预处理的步骤, 使用 2/3 的数据进行训练, 便能在剩下 1/3 的数据中达到约为 0.000 6

表 2 近年来机器学习在锂电池寿命预测、电池健康诊断以及电极材料筛选等方面的研究

Table 2 Machine learning research in Li battery life prediction, battery health diagnosis and motor material screening in recent years

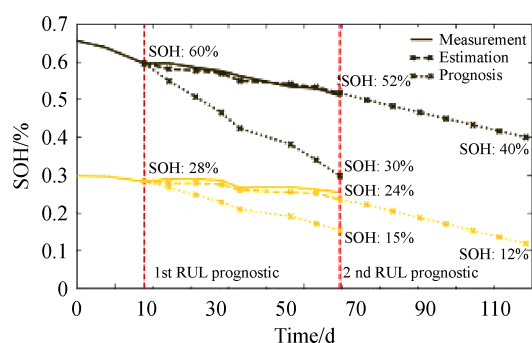
Research contents	Methods	Specific content	Model classification	Method selection basis	Virtue	Weakness	Data Set	Performance indicators	References
Life prediction	Elastic net	Battery life	Regression	Converted into linear model calculation	Simple, efficient and easy to explain	Many data samples are required	Capacity decay curves of 124 lithium batteries in different states, first 100 cycle data	Forecast error: 9.1%	[39]
	Support Vector Regression (SVR)	State of health, battery life	Regression	SVM regression and prediction	Solve the classification and regression problems of high dimensional features	Sensitive to missing data	The dataset contains 396 samples and 2/3 training sets	Forecast error: 0.11%	[40]
	SVM	Battery life	Regression, classification	Input and output relational mapping, classification of battery life	Deal with nonlinear regression problems	Sensitive to kernel functions and missing data	591 sets of cyclic discharge data, 70% of the training set	Forecast error: 6%	[41]
	Gaussian process regression (GPR)	Battery life	Regression	Early data predict later battery behavior	Simple and easy to explain	Higher dimensional space loses its validity	20 000 impedance spectroscopy	Predict the remaining life of the battery without knowing its past performance	[42]

(to be continued)

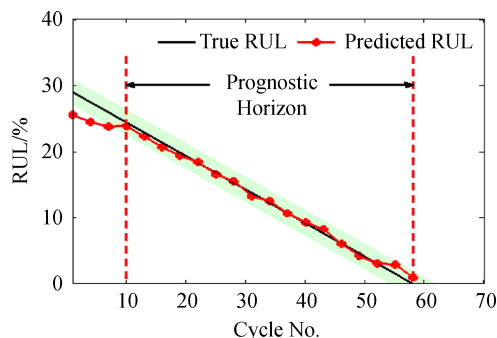
(Table 2 continued)

Research contents	Methods	Specific content	Model classification	Method selection basis	Virtue	Weakness	Data Set	Performance indicators	References
	Linear regression, kNN, RF, Decision tree	Battery temperature and battery performance prediction	Regression, classification	Compare the accuracy of different methods			A total of 18 600 sample sites were recorded at three operating temperatures, 70% of the training set and 30% of the test set	Decision tree (0.99), linear regression (0.5), kNN(0.75), RF(0.94)	[43]
State of charge prediction	Neural Network (NN)	State of charge	Regression	Fuzzy neural network simulation	Small error and high fitting degree	Easy to overfit	40 hidden neurons	Forecast error <5%	[44,45]
	DNN	State of charge	Regression	The weights of input parameters are obtained by self-learning	Multilayer perception	The model is complex and the process is unexplainable	Battery discharge time data of 5 different external temperatures	Forecast error <1.1% (25 °C) forecast error <2.17% (−20 °C)	[46]
	Adaptive Network-based Fuzzy Inference System (ANFIS)	State of charge	Regression	The fuzzy system is realized by combining fuzzy network with neural network	Simple and efficient,	Sensitive to abnormal data	Battery data for 10 different driving environments	Forecast error <0.1%	[47]
	Relevance Vector Machine (RVM), Convolutional Neural Network (CNN)	Battery capacity	Regression	Hybrid machine learning model for higher prediction accuracy			Data of 41 batteries at different charging and discharging times	Compared with RVM and CNN alone, the error is smaller	[48]
Electrode material	BPNN	Battery cathode material optimization	Regression	Supervised learning algorithms	Strong ability of self-adaptation, self-learning and nonlinear mapping	Slow convergence	Data set: DFT calculation results	The predicted voltage distribution is in good agreement with experimental results	[49]
	CNN	Battery microstructure detection	Classification	Hierarchical and multi-scale characteristics of CNN	Higher classification accuracy	Need to adjust parameters, rely on large amount of data	X-ray phase contrast nano tomographic images of 650 electrode material particles	Accurately quantify the microstructure evolution of battery electrode	[50]
	DNN	Screening of battery electrode materials	Regression	Multifactor input conditions	Extract richer data features	Difficult for training	Data set: 4 250	Accuracy >75%	[51]

的平均方差,如图 4a 所示,在加入负载集合之后,平均方差能达到更低的 0.000 4。类似地,Patil^[41]也使用了 SVR 作为回归模型,如图 4b 所示。但是不一样的是这篇文章囊括了 SVM 和 SVR 两个模型并且分阶段执行。作者选择分阶段执行模型的动机在于种类繁多的输入数据会包括很多杂音,如果使用 SVM 这种分类算法先预测并找出能决定电池寿命的关键参数,我们可以挑出这些参数放入 SVR 中来预测准确的电池寿命。在 2 个阶段的建模过程中,作者先采用 SVM 预测电池运行循环的次数,然后通过主成分分析法分析和筛选出最适合预测的变量,然后使用这些变量来训练基于神经网络的 SVR,最后预测出电池寿命。此项研究中,由于原始电压和电流的数据是时间序列,SVR 等模型无法直接处理,所以作者放入模型的变量都是经过预先提取和处理的变量,例如电池或电压曲线的弧度和曲率等等。Wu^[45]也是采用了优化特征选择加上机器学习模型的方法,方式和模型略有不同。作者采用了重要性抽样进行特征筛选然后将选出的特征放入 3 层的神经网络中进行预测,误差低于 5%。



(a) Estimated remaining life of new batteries (brown curve) and aged batteries (orange curve) over the next 50 day



(b) Predicted remaining life of batteries compared to the actual situation

图 4 SVM 用于对 Li 电池剩余寿命的预测^[40-41]

Fig. 4 Residual life of Li battery was predicted by SVM method^[40-41]

与 SOFC 不同,锂电池技术发展更加成熟,获取实验数据较为容易,所以大多数研究都是基于现

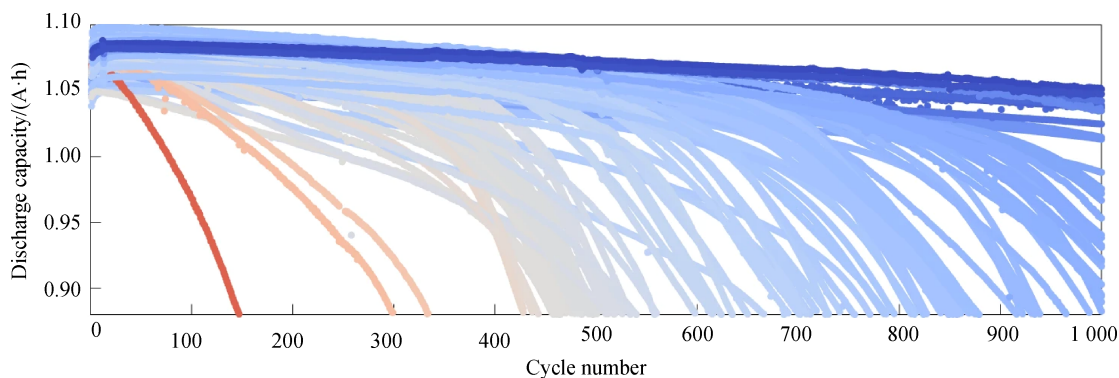
实实验数据,而不是计算机模拟结果。Serverson^[39]通过观察锂电池前 100 个循环的数据预测电池最后的寿命。使用弹性网络,也就是带有 L1 和 L2 正则的线性回归作为模型,准确度可达到 90% (图 5)。

以上所说的研究其寿命预测都是基于电压和电流曲线,如果单单从阻抗谱的角度来说,对电流或者电压进行改动来影响另一个变量,可以通过测量电压或者电流变化的程度来得到材料的当前的物理性质以及电化学反应状况,这些性质都可以通过测量阻抗谱来得到。最大的问题是需要测量的阻抗谱范围很大,很难界定材料不同性质具体对应哪一块光谱区域,例如 Zhang^[42]就提取了从 0.02 Hz–20 kHz 的范围并收集了 20 000 条电化学阻抗谱,采取高斯过程回归的模型也得到了很好的效果,预测的电池衰减曲线和实际曲线的 R^2 值大于 0.75。

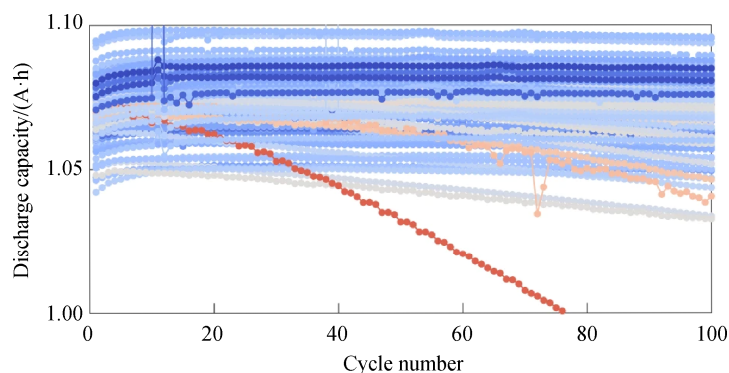
1.3.2 剩余电量预测 同样地,在电量预测任务中,目前大部分的研究都使用电动汽车作为模型依据来提取数据,因为电动汽车的剩余电量估计 (SOC) 是一个重要的议题。电动汽车相对于燃油汽车,续航里程相对更低,从而使得用户对电池剩余电量较为敏感。电动汽车也需要准确的 SOC 估计用来优化和平衡电动汽车的驾驶性能。与电池寿命估计的问题类型相似,电量预测也是取锂电池中可直接观测和获得的数据,然后预测剩余的电量。此类问题一般采用回归类型的机器学习模型便可解决。Chemali^[46]采用当前的电压,温度,平均电流和平均电压来预测当前电池的剩余电量。使用高斯噪声对数据进行模糊处理防止过拟合,得到的平均绝对误差小于 1%。相同地, Zahid^[47]使用了自适应神经模糊系统 (ANFIS), 一种神经网络的变种对 SOC 进行预测。输入变量包括实际电流、温度、功率、功率损耗和热移除。循环的数据是基于计算机软件 NREL's ADvanced VehIcle SimulatOR 的模拟结果。Zahid 横向对比了基础神经网络、Elman 神经网络和 ANFIS, ANFIS 得到了最好的结果,最大绝对误差小于 0.1%。以上的预测模型都是基于普通的全链接神经网络,如果想要包含一定数量的历史数据来做预测难度较大,因为全链接神经网络的隐藏节点个数都提前设置好。但是用于图像和自然语言处理的卷积神经网络 (CNN) 没有这样的限制。Shen^[52]就使用了 CNN 作为预测模型,并且输入数据是 3 个参数 (电压、电流、容量) 的 25 个历史数据,所以输入数据的尺寸是 25 乘 3。卷积神经网络中的卷积核会考虑数据周边其他数据的联系,然后输出结果。在此

应用中,卷积神经网络能够总结所有包含在输入数据中的历史记录信息,从而得到更准确的预测。在此研

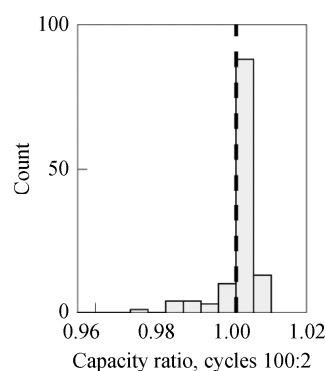
究中,作者对比了 CNN, DNN 和相关向量机的性能, CNN 有着最好的性能,均方根误差低于 1.5%。



(a) Relationship between the capacity of lithium battery and the charging cycle time under the first 1 000 cycles of charge and discharge in different ways. The color of each curve is determined by the cycle life of the battery



(b) Relationship between lithium battery capacity and charging and discharging times under the first 100 cycles of different charging and discharging modes, used in the training set of machine learning model



(c) Prediction results of lithium battery life calculated by machine learning algorithm

图 5 Severson 采用机器学习对 Li 电池寿命预测研究^[39]

Fig. 5 Severson used machine learning to study Li battery life prediction^[39]

1.3.3 故障诊断 随着以锂离子电池为储能物质的电动车的普及,由于锂离子电池故障所引发的火灾和事故引起了人们的注意。事实上,只要能及时的诊断出锂离子电池的运行异常状态,很多事故都是可以避免的。在实验室条件下,经过多年的总结和积累,科学家们总结出了很多可用于锂电池故障诊断的算法。但是这些算法往往不能应用于实际的复杂情况中。

Yang Zhao^[44]运用了电动车运行的大数据,做出了利用机器学习来对锂离子电池进行诊断的模型。作者使用 3σ-MSS 系统建立了判断电池电压的标准,并利用此标准来区分正常与非正常电压。以此为基础,作者运用了机器学习模型来模拟电压错误的发生区域,并发现位于电动汽车前部和底部的电池更容易发生故障。

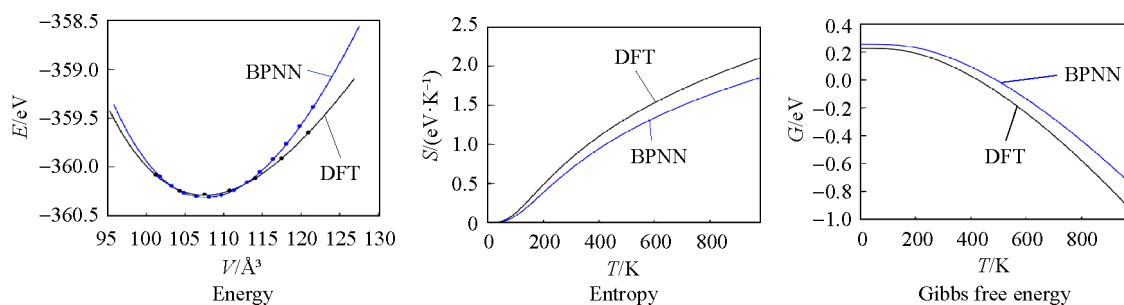
1.3.4 电极材料 锂离子电池在全世界范围内的成功让人们开始探索运用新材料来提高电池性能的

可能性。尽管金属离子电池很早就被人类发明并应用,由于可用的电极电解质材料以及金属离子众多,依靠传统试验方法逐一测试各种不同材料离子组合的电池性能耗时长、成本高,至今只有很少一部分材料的性能被测试并公布。为了大面积测试电池材料性能,筛选出最好的电极电解质材料和金属离子,机器学习被引入到这一领域,来为实验提供指导性意见。

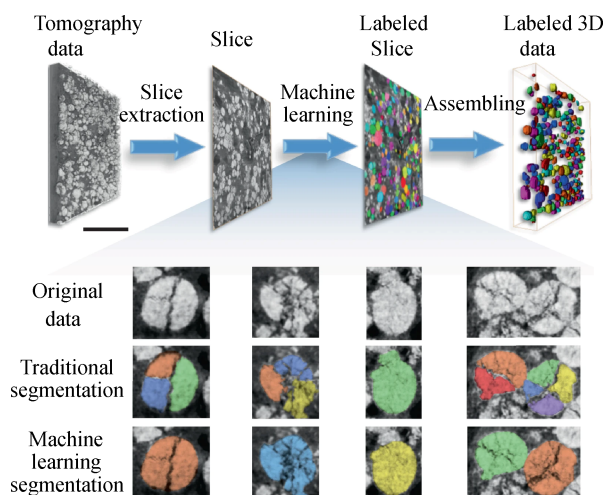
Gregory Houchins 等^[49]基于原子中心对称函数的指纹识别,采用反向神经网络模型,训练密度泛函理论(DFT)的计算结果,来优化锂离子电池阴极材料,如图 6a 所示。复合电极的微观结构决定了单个电池颗粒在锂离子电池中的充电和放电方式,Zhisen Jiang 等^[50]X 射线相衬纳米断层扫描结果图像的基础上,采用遮罩区域卷积神经网络机器学习模型,揭示了锂离子电池阴极微粒-碳/粘结剂分离的统计分布。结果表明,颗粒的脱离程度与带电率呈正相关,较小的颗粒在与碳/粘合剂基体的脱离中

表现出较高的不确定性, 如图 6b 所示。Rajendra Joshi^[49]在 2019 年提出了运用机器学习来预测未知电极材料所能提供的电池电压如图 6c 所示, Rajendra Joshi 运用了 Material Project 中 3 977 组数据, 这些数据中包含 6 种离子电池的参数与性质, 分别是 Li, Mg, Ca, Al, Zn, Y。这些特定的电极材料参数包括电池中的工作离子种类, 有效金属离子在电极中的浓度(例如 Li_{0.5}C₆ 中 Li 的浓度是 LiC₆

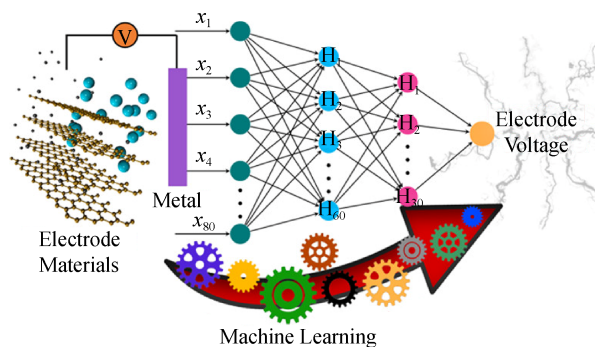
中的二分之一)、晶胞种类、空间群数, 并且添加了从 Ward 等^[53]的实验结果中截取的 237 项元素性质作为机器学习的输入变量。随后 Rajendra Joshi 使用主成分分析法筛选出了 80 余个对结果有影响的变量。利用这些变量和机器学习中的支持向量回归、内核岭回归、深度神经网络模型, 作者较为准确的预测了未知电极材料所能提供的电压, 可以作为 DFT 的一个粗略的替代方案。



(a) Comparison of Back Propagation Neural Networks (BPNN) and Density Functional Theory (DFT) results for thermodynamics predictions of LiNiO₂



(b) Convolutional neural network (CNN) simulating lithium ion battery cathode particles, carbon/binder separation statistical distribution



(c) Using deep neural network, support vector machine, and kernel ridge regression to predict the voltage of metal ion battery electrode materials

图 6 机器学习用于电池材料的研究^[49-51]

Fig. 6 Machine learning used in battery materials^[49-51]

机器学习除了应用于锂电池的电极材料研究, 还可以应用于其他诸如钠电池、锌电池等体系的研究。Isaiah A Moses^[54]建立了电极材料筛选的深度神经网络回归机器学习模型, 根据从材料数据库获得的数据进行训练, 用于预测金属离子电池电极材料充放电时的平均电压和体积变化。通过用钠离子系统地替换原始数据库中的锂离子来生产钠离子电极, 筛选出 22 种在能量密度上表现良好的钠电池电极材料。Linming Zhou^[55]采用晶体图卷积神经网络的机器学习方法从 13 000 多种无机材料中筛选出十几种具有具有高容量(大于 100 mA·h·g⁻¹)和高

电压(大于 0.5 V)的锌电池电极材料。这种方法有望广泛推广到其他电池系统, 包括锂、钠、钾、铝和镁电池。

由于锂电池已经商用化了很长时间, 机器学习在锂电池方面的研究的实用性更强, 从其发展历程可以看出, 大部分研究者采用电池的实验数据作为研究的数据集, 研究锂电池的寿命预测等问题。此外, 对于电极材料方面的研究, 目前也只能基于材料数据库以及结合 DFT 等材料计算数据开展研究, 进行电极材料的高通量筛选。对于电极反应过程中原子尺度上的反应机理, 还是依靠传统的实验方法,

机器学习缺乏对电极材料的本质上的解释。如何结合大数据分析方法, 结合现有的实验表征来加速材料科学创新, 既需要多学科的融合, 也需要进一步基础理论上的突破。

2 机器学习在 CO₂ 电化学还原催化剂

研发方面中的应用

通过使用更加清洁能源取代化石燃料的办法可以在源头上减少 CO₂ 的排放。但是现在全球 90% 的能源仍由化石燃料提供, 这种能源结构在短期内仍然很难改变。尤其是中国, 作为以煤炭为主的能源结构在很长一段时间内不会有根本性的变化^[56]。因此, 在未来很长一段时间内, 人类将继续向大气排放 CO₂。但是现在可以采取一些方法收集排放的

CO₂ 并将其长期储存在开采过的油气井、煤层和深海里^[57–58], 即碳封存。另一种方法是将收集到的 CO₂ 采用化学方法进行还原, 制备成为燃料和有附加值的化学品^[59–61], 如碳氢化合物和酒精^[61], 实现碳循环利用, 从而减少 CO₂ 在大气中的浓度, 减缓全球变暖趋势, 这个过程叫做人工固碳。

CO₂ 的电还原是实现 CO₂ 循环和能量再生的最可能方法之一。推广该技术的关键是开发高性能电催化剂。通常, 高通量计算筛选有助于材料创新, 但仍会浪费大量时间和资源。为了有效地探索用于还原 CO₂ 的电催化剂, 近几年来, 随着数据驱动材料科学的发展, 研究者采用大数据的方法辅助新型催化剂的设计开发, 极大降低了实验成本和时间成本, 加速了 CO₂ 还原电催化剂的开发。表 3 总结了相关的研究成果。

表 3 机器学习 CO₂ 还原电催化剂的开发的总结

Table 3 Research summary of machine learning for the development of CO₂ reduction electrocatalysts

Methods	Research contents	Model classification	Virtue	Weakness	Data Set	Performance indicators	References
Extreme Gradient Boosting Regression (XGBR)	Prediction of catalytic performance of SAC	Regression	Categorical and continuous predictive variables	Sensitive to unusual value	171 designing SAC DFT calculation data, 80% training set	The change of Gibbs free energy of CO adsorption in 1060 SAC was predicted quickly. A total of 94 potential SAC compounds were screened for CO ₂ reduction	[62]
ANN, Density Functional Theory (DFT)	The complex nonlinear interaction between adsorbed and adsorbent	Regression	Small error and high fitting degree	Easy to over-fit, complex model, difficult to explain	About 250 bimetallic systems were randomly selected for the interaction model	Terminated (100) have a low overpotential and are more selective for the electrochemical reduction of CO ₂ to C ₂ species	[63]
NN, DFT	Screening CO ₂ reduction catalysts for transition metals	Regression	Small error and high fitting degree	Easy to over-fit, complex model, difficult to explain	2~8 hidden layer, 75% training set	Search for {100} -terminated polymetallic copper catalysts for electrochemical reduction of CO ₂	[64]
NN	Search exhaustively for active bimetallic surfaces	Regression	Small error and high fitting degree	Easy to over-fit, complex model, difficult to explain	7 000 DFT single point calculation, 583 CO adsorption sites	Reveal the active site of CO ₂ reduction reaction	[65]
NN	Accurate prediction of CO adsorption energy and HOCO formation energy on highly irregular Au surface	Regression	Small error and high fitting degree	Easy to over-fit, complex model, difficult to explain	1 384 DFT calculation data sets and 1 104 training sets	Applied to Au nanoparticles and Au ₃ FeAu ₃ dealloyed surfaces, the active sites of these disordered and irregular surfaces were identified	[66]

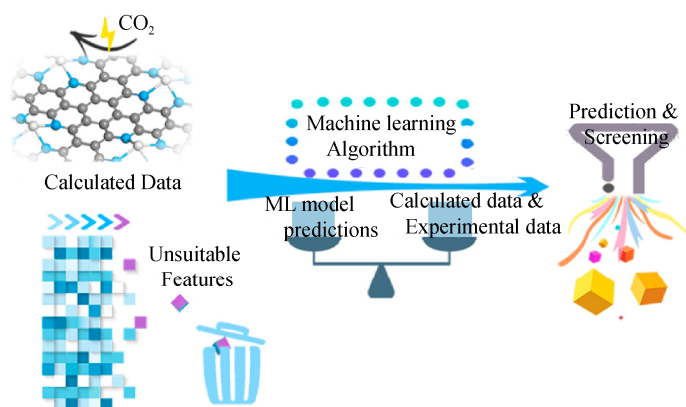
An Chen^[62]基于极限梯度增强回归算法创建了 CO₂ 还原电催化剂机器学习模型。该模型可以准确地描述具有分散的金属-非金属共掺杂石墨烯结构

的单原子催化剂(SAC)的催化性能, 快速地预测出 1 060 种 SAC 中 CO 吸附的吉布斯自由能变化以及它们对电化学析氢反应催化活性。结合预测的催化

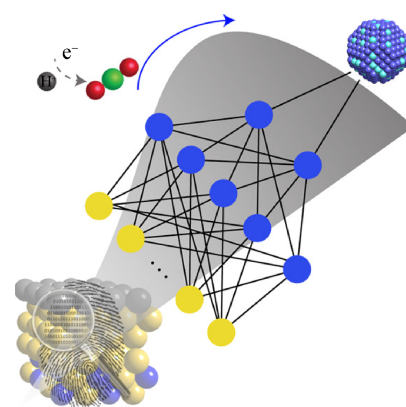
性能, 研究者通过重叠结果快速筛选出 94 种潜在的用于 CO₂ 的电还原的 SAC 催化剂(图 7a)。Xianfeng Ma^[63]结合人工神经网络与第一性原理计算方法建立了 1 种机器学习增强化学吸附模型, 能够对金属合金的表面反应性实现快速、准确的预测, 得到了被吸附物-吸附体之间复杂的非线性相互作用关系。预测结果表明暴露(100)表面的 Cu 合金具有较低的过电位, 对 CO₂ 电还原为 C₂ 物种具有更高选择性(图 7b)。Zheng Li^[64]采用相同的机器学习方法筛选用于 CO₂ 还原的过渡金属催化剂, 得出了类似的结论(图 7c)。Sergey Gusarov^[67]研究了 1 种基于 Fukui 函数描述符的机器学习方法获得催化剂的催化性

能, 用于 CO₂ 的还原(图 7d)。

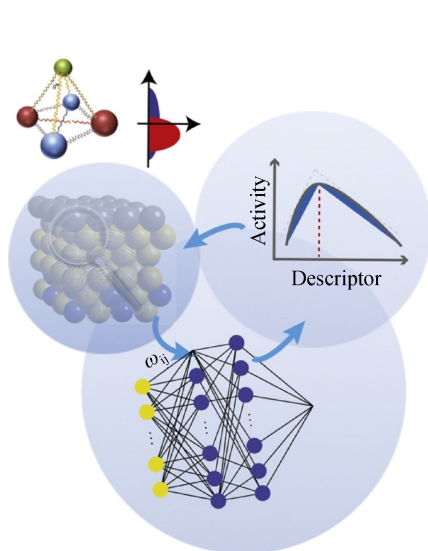
机器学习技术除了应用于 CO₂ 还原催化剂的筛选, 还可以用来探究 CO₂ 电还原的反应机理。Zachary W. Ulissi^[65]采用机器学习方法可以穷举搜索活性双金属面并揭示 CO₂ 还原反应活性位点(图 8a)。Yalu Chen^[66]通过结合机器学习和多尺度模拟来识别合金表面 CO₂ 还原的活性位点。建立了 2 种基于神经网络的机器学习模型, 用于准确预测极不规则的 Au 表面上 CO 吸附能和 HOCO 形成能。将该模型应用于 Au 纳米粒子表面和 Au₃Fe 合金表面, 可以识别出对这种无序和不规则表面的活性位点, 从而能够提高对 CO₂ 还原性能(图 8b)。



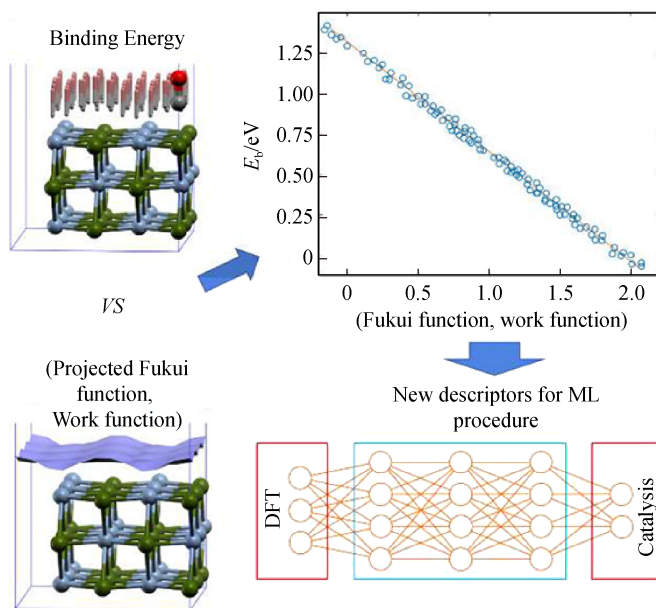
(a) Extreme gradient boosting regression (XGBR) was used to simulate the catalytic performance of a SAC with metal and non-metal co-doped graphene structure, and successfully and rapidly predicted the change of Gibbs free energy of CO adsorption in 1 060 atom-dispersed metal and no-metal co-doped graphene system, greatly reducing the research cost



(b) ANN can quickly and accurately predict the surface reactivity of metal alloys, and obtain the complex nonlinear interaction between adsorbents and adsorbents



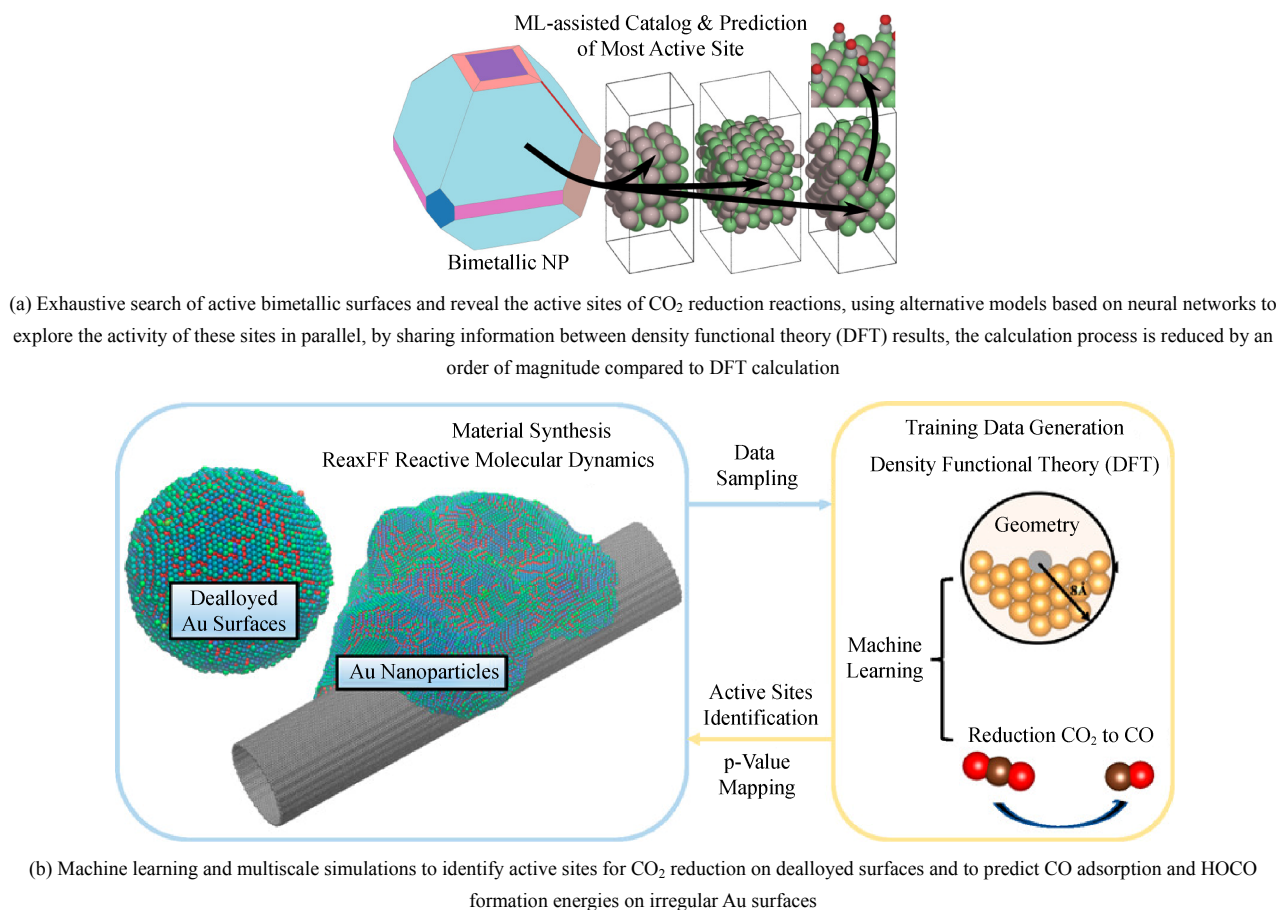
(c) ANN is used to screen transition metal catalysts for CO₂ reduction



(d) Machine learning method based on Fukui function descriptor is used to screen CO₂ reduction catalysts

图 7 机器学习用于 CO₂ 还原电催化剂的开发^[62-64, 67]

Fig. 7 Machine learning for the exploitation of CO₂ reduction electrocatalysts^[62-64, 67]

图 8 机器学习用于揭示 CO_2 还原反应活性位点^[65–66]Fig. 8 Machine learning used to reveal active sites of CO_2 reduction^[65–66]

机器学习可以加快 CO_2 还原电催化剂的开发进度, 实现催化剂的快速筛选。然而对于筛选的过程缺乏物理层面上的解释, 机器学习与实验解释存在明显的逻辑鸿沟, 这便是模型学习结果与专家经验的矛盾与融合^[18]。如何增加机器学习方法的可读性, 赋予其物理学的解释, 也是今后需要考虑的问题之一。

3 展望

过量的碳排放造成了严重的温室效应, 将数据科学应用于清洁能源技术和减碳技术, 可以助力实现人类社会的碳中和。在本文中, 我们对机器学习方法用于开发可再生能源材料系统(SOFC、锂电池新能源体系以及 CO_2 电化学还原催化剂)的研究进展进行了回顾。机器学习与材料学等自然科学相结合, 涉及跨学科交叉, 需要自然科学家与数据科学家之间的紧密协作, 如图 9 所示, 未来相关研究的可以关注以下几个方向:

1) 首先, 自然科学家需要根据自己的研究提炼出其中的科学问题, 并根据实验结果建立相应的数

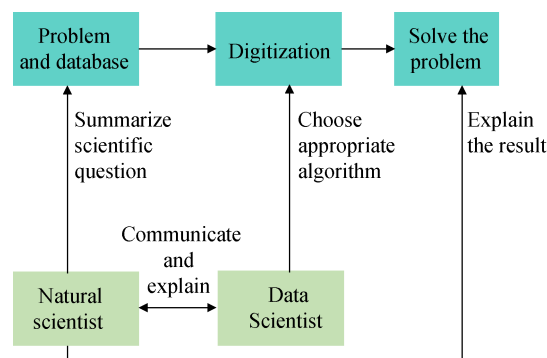


图 9 自然科学与数据科学相结合解决问题流程

Fig. 9 Problem solving process combining natural science and data science

据库。当前很多机器学习训练的数据集依赖于计算仿真模拟数据, 数据数量和质量决定了机器学习的质量, 如何获取高质量的数据库是此类研究的重要问题。自然科学与数据科学的结合, 大多数情况下, 使用的数据库是基于物理化学模型通过软件计算的出来的数据作为训练的数据集。数据库的建立需要自然科学研究者对所研究的问题有深刻地理解, 针对具体的问题建立合适的数据库类型。部分文献中

机器学习的模型计算结果精度非常高,在 90%以上,原因可能是建立的模型简单,有一些是线性的模型,从而易于拟合;还有可能是建立的模型有一些限制条件,有使用的前提,在某些区间精度很高。

2) 其次,在使用机器学习方法求解自然科学问题时,数据科学家需要根据研究的问题选择合适的机器学习算法,这一过程是将物理化学过程转换成

数学处理过程,数据科学家需要与自然科学研究者之间充分对接,理解需要解决的问题,从而能够根据自己的经验选择合适的算法。表 4 总结了本工作列举的常用机器学习方法及其在绿色能源科学中使用情况。因此在当前学科交叉的背景下,理工科研究者可以学习相关的数据科学知识,拓展自己的研究方法体系。

表 4 常用机器学习方法及其在绿色能源科学中使用情况总结

Table 4 Summary of commonly used machine learning methods in this paper and their application in green energy science

Algorithm	Feature	Model type
<i>k</i> Nearest Neighbors	The sample type is determined by <i>k</i> nearby samples in the available data set.	Classification (fault classification, influence factor analysis).
Decision tree	Based on the probability of the occurrence of various situations, the probability that the expected value of net present value is greater than or equal to zero is obtained by constructing decision tree.	Optimization problems (optimal conditions, optimal operating conditions)
Support vector machine	A hyperplane is constructed to classify samples in high dimensional space. It solves the regression problem.	Performance prediction, classification
Artificial neural network	Simulating the neural network processes of the organic brain to process information is the basis of modern neuroscience. Its structure is a multilayer perceptron consisting of many connected neurons.	Multi-factor input variable processing problem (performance prediction, regression analysis)
Principal component analysis	Transform a set of possibly correlated variables into a set of linearly unrelated variables by orthogonal transformation.	Multivariate analysis
Random forest	By combining several weak classifiers, the results are voted or averaged, which makes the results of the whole model have high accuracy and generalization performance.	Classification and regression

3) 最后,由于机器学习采用的是数据驱动的统计方法,在处理能源材料科学问题时,由于完全以数据驱动,计算出的结果缺乏相应的物理意义,各个变量之间的联系是基于多重映射的关系,缺乏明确物理关系,在应用新的技术时,也需要材料学家、物理学家对结果进行归纳解释。

参考文献:

- [1] MAC DOWELL N, FENNELL P S, SHAH N, et al. The role of CO₂ capture and utilization in mitigating climate change[J]. Nat Clim Chang, 2017, 7(4): 243–249.
- [2] CAI Y, SAM C Y, CHANG T. Nexus between clean energy consumption, economic growth and CO₂ emissions[J]. J Clean Prod, 2018, 182: 1001–1011.
- [3] QIAO W, LU H, ZHOU G, et al. A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimizer[J]. J Clean Prod, 2020, 244: 118612.
- [4] ANDERSON T R, HAWKINS E, JONES P D. CO₂, the greenhouse effect and global warming: from the pioneering work of Arrhenius and Callendar to today's Earth System Models[J]. Endeavour, 2016, 40(3): 178–187.
- [5] LOVELOCK C E, CAHOON D R, FRIESS D A, et al. The vulnerability of Indo-Pacific mangrove forests to sea-level rise[J]. Nature, 2015, 526(7574): 559–563.
- [6] HIMANEN L, GEURTS A, FOSTER A S, et al. Data-driven materials science: status, challenges, and perspectives[J]. Adv Sci, 2019, 6(21): 1900808.
- [7] RACCUGLIA P, ELBERT K C, ADLER P D F, et al. Machine-learning-assisted materials discovery using failed experiments[J]. Nature, 2016, 533(7601): 73–76.
- [8] LI B, DUAN Y, LUEBKE D, et al. Advances in CO₂ capture technology: A patent review[J]. Appl Energy, 2013, 102: 1439–1447.
- [9] CHU S, CUI Y, LIU N. The path towards sustainable energy[J]. Nat Mater, 2016, 16(1): 16–22.
- [10] BOTU V, RAMPRASAD R. Learning scheme to predict atomic forces and accelerate materials simulations[J]. Phys Rev B, 2015, 92(9): 94306.
- [11] LI Z, LI H, SHAO L. Improving online customer shopping experience with computer vision and machine learning methods[J]. Lect Notes Comput Sci, 2016, 9751: 427–436.
- [12] KIM K jae, AHN H. A recommender system using GA K-means clustering in an online shopping market[J]. Expert Syst Appl, 2008, 34(2): 1200–1209.
- [13] SAMANT P, AGARWAL R. Machine learning techniques for medical diagnosis of diabetes using iris images[J]. Comput Methods Programs Biomed, 2018, 157: 121–128.
- [14] WU M, CHEN L. Image recognition based on deep learning[J]. Proc - 2015 Chinese Autom Congr CAC, 2015, 2016: 542–546.
- [15] LIPPMANN R. An introduction to computing with neural nets[J].

- IEEE Assp Mag, 1987, 4(2): 4–22.
- [16] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436–444.
- [17] ZHANG C, BENGIO S, HARDT M, et al. Understanding deep learning (still) requires rethinking generalization[J]. Communications of the ACM, 2021, 64(3): 107–115.
- [18] 施思齐, 涂章伟, 邹欣欣, 等. 数据驱动的机器学习在电化学储能材料研究中的应用[J]. 储能科学与技术, 2022, 11(3): 739.
- SHI Siqu, TU Zhangwei, ZOU Xinxin, et al. Energy Storage Sci Technol(in Chinese). 2022, 11(3): 739.
- [19] 刘悦, 邹欣欣, 杨正伟, 等. 材料领域知识嵌入的机器学习[J]. 硅酸盐学报, 2022, 50(3): 863–876.
- LIU Yue, ZOU Xinxin, YANG Zhengwei, et al. J Chin Ceram Soc, 2022, 50(3): 863–876.
- [20] DEBE M K. Electrocatalyst approaches and challenges for automotive fuel cells[J]. Nature, 2012, 486(7401): 43–51.
- [21] GHIABI C, GHAFARINEJAD A, KAZEMI H, et al. *In situ*, one-step and co-electrodeposition of graphene supported dendritic and spherical nano-palladium-silver bimetallic catalyst on carbon cloth for electrooxidation of methanol in alkaline media[J]. Renew Energy, 2018, 126: 1085–1092.
- [22] MEHMETI A, MCPHAIL S J, PUMIGLIA D, et al. Life cycle sustainability of solid oxide fuel cells: From methodological aspects to system implications[J]. J Power Sources, 2016, 325: 772–785.
- [23] CASCOS V, FERNÁNDEZ-DÍAZ M T, ALONSO J A. Structural and electrical characterization of the novel $\text{SrCo}_{1-x}\text{Ti}_x\text{O}_{3-x}$ ($x = 0.05, 0.1$ and 0.15) perovskites: Evaluation as cathode materials in solid oxide fuel cells[J]. Renew Energy, 2019, 133: 205–215.
- [24] ARRIAGADA J, OLAUSSON P, SELIMOVIC A. Artificial neural network simulator for SOFC performance prediction[J]. J Power Sources, 2002, 112(1): 54–60.
- [25] KANG Y W, LI J, CAO G Y, et al. Dynamic temperature modeling of an SOFC using least squares support vector machines[J]. J Power Sources, 2008, 179(2): 683–692.
- [26] WU X J, HUANG Q, ZHU X J. Thermal modeling of a solid oxide fuel cell and micro gas turbine hybrid power system based on modified LS-SVM[J]. Int J Hydrogen Energy, 2011, 36(1): 885–892.
- [27] CHEN J, CHEN Y, ZHANG H. Study on Fuel Utilization Dynamic model of a SOFC-GT Hybrid System Based on Deep Learning Technique[C]//E3S Web of Conferences. EDP Sciences, 2019, 113: 1–6.
- [28] SHAO Q, GAO E, MARA T, et al. Global sensitivity analysis of solid oxide fuel cells with Bayesian sparse polynomial chaos expansions[J]. Appl Energy, 2020, 260: 114318.
- [29] SONG S, XIONG X, WU X, et al. Modeling the SOFC by BP neural network algorithm[J]. Int J Hydrogen Energy, 2021, 46(38): 20065–20077.
- [30] XU H, MA J, TAN P, et al. Towards online optimisation of solid oxide fuel cell performance: Combining deep learning with multi-physics simulation[J]. Energy AI, 2020, 1: 100003.
- [31] LI Y, SHEN J, LU J. Constrained model predictive control of a solid oxide fuel cell based on genetic optimization[J]. J Power Sources, 2011, 196(14): 5873–5880.
- [32] NASSEF A M, FATHY A, SAYED E T, et al. Maximizing SOFC performance through optimal parameters identification by modern optimization algorithms[J]. Renew Energy, 2019, 138: 458–464.
- [33] GHORBANI B, VIJAYARAGHAVAN K. Developing a virtual hydrogen sensor for detecting fuel starvation in solid oxide fuel cells using different machine learning algorithms[J]. Int J Hydrogen Energy, 2020, 45(51): 27730–27744.
- [34] WANG Y, WU C, ZHAO S, et al. Coupling deep learning and multi-objective genetic algorithms to achieve high performance and durability of direct internal reforming solid oxide fuel cell[J]. Appl Energy, 2022, 315: 119046.
- [35] LI S, CAO H, YANG Y. Data-driven simultaneous fault diagnosis for solid oxide fuel cell system using multi-label pattern identification[J]. J Power Sources, 2018, 378(1): 646–659.
- [36] MENG-TING C, XIAO-WEI F, ZHONG-HUA D, et al. Data-Driven Fault Detection for SOFC system based on Random Forest and SVM[C]//2019 Chinese Automation Congress (CAC). IEEE, 2019: 2829–2834.
- [37] ZHANG Z, LI S, XIAO Y, et al. Intelligent simultaneous fault diagnosis for solid oxide fuel cell system based on deep learning[J]. Appl Energy, 2019, 233–234: 930–942.
- [38] ZHENG Y, WU X long, ZHAO D, et al. Data-driven fault diagnosis method for the safe and stable operation of solid oxide fuel cells system[J]. J Power Sources, 2021, 490: 229561.
- [39] SEVERSON K A, ATTIA P M, JIN N, et al. Data-driven prediction of battery cycle life before capacity degradation[J]. Nat Energy, 2019, 4(5): 383–391.
- [40] NUHIC A, TERZIMEHIC T, SOCZKA-GUTH T, et al. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods[J]. J Power Sources, 2013, 239: 680–688.
- [41] PATIL M A, TAGADE P, HARIHARAN K S, et al. A novel multistage Support Vector Machine based approach for Li ion battery remaining useful life estimation[J]. Appl Energy, 2015, 159: 285–297.
- [42] ZHANG Y Y, TANG Q, ZHANG Y Y, et al. Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning[J]. Nat Commun, 2020, 11(1): 1–6.
- [43] TRAN M K, PANCHAL S, CHAUHAN V, et al. Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery[J]. Int J Energy Res, 2022, 46(2): 786–794.
- [44] ZHAO Y, LIU P, WANG Z, et al. Fault and defect diagnosis of battery for electric vehicles based on big data analysis methods[J]. Appl Energy, 2017, 207: 354–362.
- [45] WU J, ZHANG C, CHEN Z. An online method for lithium-ion battery remaining useful life estimation using importance sampling and neural networks[J]. Appl Energy, 2016, 173: 134–140.
- [46] CHEMALI E, KOLLMAYER P J, PREINDL M, et al. State-of-charge estimation of Li-ion batteries using deep neural networks: A machine learning approach[J]. J Power Sources, 2018, 400: 242–255.
- [47] ZAHID T, XU K, LI W, et al. State of charge estimation for electric vehicle power battery using advanced machine learning algorithm under diversified drive cycles[J]. Energy, 2018, 162: 871–882.
- [48] SHEN S, NEMANI V, LIU J, et al. A hybrid machine learning model for battery cycle life prediction with early cycle data[C]//2020 IEEE

- Transportation Electrification Conference & Expo (ITEC). IEEE, 2020: 181–184.
- [49] HOUCHINS G, VISWANATHAN V. An accurate machine-learning calculator for optimization of Li-ion battery cathodes[J]. *J Chem Phys*, 2020, 153(5): 054124.
- [50] 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): 1–9.
- [51] JOSHI R P, EICKHOLT J, LI L, et al. Machine learning the voltage of electrode materials in metal-ion batteries[J]. *ACS Appl Mater Interfaces*, 2019, 11(20): 18494–18503.
- [52] SHEN S, SADOUGHI M, CHEN X, et al. A deep learning method for online capacity estimation of lithium-ion batteries[J]. *J Energy Storage*, 2019, 25: 100817.
- [53] WARD L, AGRAWAL A, CHOUDHARY A, et al. A general-purpose machine learning framework for predicting properties of inorganic materials[J]. *npj Comput Mater*, 2016, 2(1): 1–7.
- [54] MOSES I A, JOSHI R P, OZDEMIR B, et al. Machine learning screening of metal-ion battery electrode materials[J]. *ACS Appl Mater Interfaces*, 2021, 13(45): 53355–53362.
- [55] ZHOU L, YAO A M, WU Y, et al. Machine Learning Assisted Prediction of Cathode Materials for Zn-Ion Batteries[J]. *Adv Theory Simulations*, 2021, 4(9): 1–6.
- [56] HONG C, ZHANG Q, HE K, et al. Variations of China's emission estimates: Response to uncertainties in energy statistics[J]. *Atmos Chem Phys*, 2017, 17(2): 1227–1239.
- [57] MICHAEL K, GOLAB A, SHULAKOVA V, et al. Geological storage of CO₂ in saline aquifers-A review of the experience from existing storage operations[J]. *Int J Greenh Gas Control*, 2010, 4(4): 659–667.
- [58] AMINU M D, NABAVI S A, ROCHELLE C A, et al. A review of developments in carbon dioxide storage[J]. *Appl Energy*, 2017, 208: 1389–1419.
- [59] FU H Q, ZHANG L, ZHENG L R, et al. Enhanced CO₂ electroreduction performance over Cl-modified metal catalysts[J]. *J Mater Chem A*, 2019, 7(20): 12420–12425.
- [60] QIAO J, LIU Y, HONG F, et al. A review of catalysts for the electroreduction of carbon dioxide to produce low-carbon fuels[J]. *Chem Soc Rev*, 2014, 43(2): 631–675.
- [61] HOANG T T H, VERMA S, MA S, et al. Nanoporous Copper-Silver Alloys by Additive-Controlled Electrodeposition for the Selective Electroreduction of CO₂ to Ethylene and Ethanol[J]. *J Am Chem Soc*, 2018, 140(17): 5791–5797.
- [62] CHEN A, ZHANG X, CHEN L, et al. A machine learning model on simple features for CO₂ reduction electrocatalysts[J]. *J Phys Chem C*, 2020, 124(18): 22471–22478.
- [63] MA X, LI Z, ACHENIE L E K, et al. Machine-learning-augmented chemisorption model for CO₂ electroreduction catalyst screening[J]. *J Phys Chem Lett*, 2015, 6(18): 3528–3533.
- [64] LI Z, MA X, XIN H. Feature engineering of machine-learning chemisorption models for catalyst design[J]. *Catal today*, 2017, 280(2): 232–238.
- [65] ULISSI Z W, TANG M T, XIAO J, et al. Machine-learning methods enable exhaustive searches for active Bimetallic facets and reveal active site motifs for CO₂ reduction[J]. *ACS Catal*, 2017, 7(10): 6600–6608.
- [66] CHEN Y, HUANG Y, CHENG T, et al. Identifying active sites for CO₂ reduction on dealloyed gold surfaces by combining machine learning with multiscale simulations[J]. *J Am Chem Soc*, 2019, 141(29): 11651–11657.
- [67] GUSAROV S, STOYANOV S R, SIAHROSTAMI S. Development of Fukui function based descriptors for a machine learning study of CO₂ reduction[J]. *J Phys Chem C*, 2020, 124(18): 10079–10084.