

2024 年
第 05 期

先进制造 与新材料

ADVANCED MANUFACTURING
& NEW MATERIALS BRIEFING



上海科学技术情报研究所
上海市前沿技术发展研究中心
技术与创新支持中心(TISC)

人工智能引发材料科学变革

编者按

作为人工智能（AI）发展的一大趋势，国际学术界已对“人工智能驱动的科学研究”形成共识：人工智能将带来科研范式的变革和新的产业业态。

在材料科学领域，人工智能具有许多优势。首先，AI 可以处理大量的复杂数据，并从中提取有价值的信息。这对于材料科学非常重要，因为材料的性能和行为通常受到许多不同因素的影响。其次，AI 可以加速材料开发过程。借助机器学习和数据分析，科学家可以更快地发现新材料和优化现有材料。这有助于缩短研发周期，并提高材料的性能和可靠性。最后，AI 可以提供定制化的解决方案。由于材料的多样性和复杂性，每种材料都有其独特的特性和需求，AI 可以根据特定的要求和约束条件，生成定制化的材料设计方案，满足不同行业的需求。因此，人工智能在材料科学中具有广阔的应用

前景。

本期《先进制造与新材料》简报对国内外在人工智能驱动的材料科学研究方面取得的重要进展予以介绍。

目 录

AI for Science : 科学研究新范式.....	1
人工智能助力新材料的合成.....	5
微软和 PNNL 利用 AI+HPC 加速材料筛选进程.....	6
优美科与微软签订 AI 平台协议，加速电池材料技术开发.....	8
利用 AI 辅助制备高性能分层多孔海绵体钾电负极.....	9
北大在 AI4S-基于 AI 的物质结构解析方面取得重要突破.....	10

AI for Science : 科学研究新范式

随着数字化时代的到来，科技创新转化为直接生产力的速度越来越快。站在科学的分水岭上，人工智能（AI）若想发展为更加通用的“智慧”，则无法绕开“科学”这一命题，用 AI 学习科学原理，解决科学问题成为必经之路。

为贯彻落实《新一代人工智能发展规划》，科技部会同自然科学基金委近期联合启动“人工智能驱动的科学研究”（AI for Science）专项部署工作。围绕基础学科关键问题和重点领域科研需求布局“人工智能驱动的科学研究”前沿科技研发体系。

当前，AI for Science 已成为全球人工智能发展的新前沿，加速了科学研究的范式变革，带来科研模式的重构和新一轮科技革命。

AI 的下一个主战场

既有的科学研究存在两种基本范式，分别是基于大量数据归纳的开普勒范式和基于物理建模演绎的牛顿范式。但随着维数的增加，计数量呈现指数增长的趋势。一方面，数据驱动的办法面临着缺乏数据及数据分析工具的困境，另一方面，从模型驱动角度看，解决实际问题的过程总是陷入精度和速度难以两全的矛盾。中国科学院院士鄂维南认为，“无论从数据驱动的角度，还是从模型驱动的角度，人工智能都有巨大的发展空间。”

近年来，以机器学习，尤其是深度神经网络为代表的 AI 技术发展提供了破局的新思路。对于大量复杂的高维函数而言，作为特殊函数的深度神经网络可

以进行有效拟合逼近，不仅提供了行之有效的数据分析方法，也能对海量数据的高效模拟实现尺度和精度的平衡。

2018 年，鄂维南在全球首次提出“AI for Science”概念，强调利用 AI 学习科学原理、创造科学模型来解决实际问题，使之成为科学研究的新范式。中国科学技术大学教授胡素磊认为，AI 在科学研究中的应用主要包括计算、设计和理论突破这三大步。

目前，AI for Science 已得到国内外学界和业界的普遍认可，2020 年初，美国能源部发布《AI for Science 报告》以促进人工智能在科学上的应用，涵盖高能物理和材料科学到计算技术等领域；2022 年 7 月，微软研究院在全球成立 AI for Science 研究院——微软研究院科学智能中心。

AI for Science 结合机器学习拟合高维函数的强大能力，推动科学研究从单打独斗的“小农作坊”模式走向“安卓模式”的平台科研，直面产业需求，通过规模化和去中心化的测试加速科研和产业的对接，大大提升科研效率和生产力。“要实现安卓模式，需要把科学计算流程抽象化和标准化。”鄂维南说，大家共同建设大平台，共享基本的模型、算法、数据库和知识库等基础设施，在此基础上开发各自团队感兴趣的应用，通过“滚雪球效应”加速科研创新和成果应用。

AI for Science 极大地拓展了科学和人工智能的边界，将成为 AI 的下一个主战场，其快速发展离不开“数据-模型-算法-算力-人才”的共同进步。在中国工程院院士、中科院计算所研究员孙凝晖看来，尽管我们过去已经做了大量研究，但不同板块间的信息流还没有全部打通，引入人工智能的方法最关键是要打通人、机、物之间的信息流，“通过这三者的高速迭代，拓展原本科学研究的边界，求解更复杂的问题。”

AI 真正落地的重要途径

“AI for Science 涉及的应用场景，包括工业制造、材料、药物等都是实体经济不可或缺的，所以它是 AI 真正落地的一个重要途径。”鄂维南说。

作为一种长度单位，一纳米只有一米的十亿分之一（一根头发丝的直径约等于六万纳米），令人震惊的是，正是如此小到“令人咋舌”的纳米却能释放出巨大的能量。世界纳米能源研究领域奠基人、中科院外籍院士王中林团队发明的摩擦纳米发电机，可以收集环境中低频次、低振幅的微小的震动机械，比如人体的心跳、呼吸，空气、水乃至血液的流动等，并将其转化为电能提供给纳米器件。摩擦纳米发电机还可以加以应用，作为无需外部电源的自驱动传感器，而且其对接触物体的材料没有要求，只要人有微小的动作，就能产生并传递电信号，进而影响控制一系列相关操作。王中林举了个例子，在医疗方面，将传感装置放在脖子、指尖或手腕的任何位置，利用纤维之间的触动来监测身体状况，不仅能监测脉搏的跳动，也能通过脉搏提取更多生理参数，分析有没有高血压等潜在隐患，从而对人体进行系统性监测。

在大数据时代，语言、动作、形态等都是非常重要的数据，王中林说，“自驱动就是收集数据，这些数据是 AI 时代分析一切的根据，相信自驱动系统在物联网时代能有重要应用。”

清华大学化工系教授张强以高比能电池研究为例，阐释能源化学与数据科学融合发展的新进展。在能源行业，国家碳中和目标的提出对储能技术提出更高要求。传统的锂离子电池储能技术虽仍占据主导，但其能量密度难以满足社会高速发展对能源的需求，电解液作为锂离子电池中离子传输的载体发挥着关键作用。张强表示，人工智能技术能够加速先进电解液的设计开发与实践，传

统的实验员一次往往只能做一个实验，但现在通过多次高通量的实验可以使几十乃至上百个实验通过一个样品实现。另外，传统的电解液溶剂开发效率不高，在张强团队建立了机器学习的模型后，通过训练机器学习模型来预测溶剂分子和相应电解液性质，并在实验中验证反馈机器学习模型的结果，不仅对电解的属性有了深入理解，也能利用模型预测电解液性质，为选择合适的电解液提供有效方法。

“数据科学会带来新的研究范式，对于构筑高效的储能器件，实现碳达峰碳中和的宏伟目标提供了一个重要的助力手段。”张强说，“机器学习方法能帮助知识发现，在碳中和的时代使命中，能源化学和数据科学逐渐融合在一起，实验、理论和数据正在成为能源化学研究的三驾马车。”

在国外，人工智能在赋能化学、生命科学等领域也取得较大进展。近期，瑞士洛桑联邦理工学院和英国罗切斯特大学的研究人员设计出由大语言模型驱动的化学引擎 ChemCrow，能够简化有机合成、药物发现和材料设计等化学任务的推理过程，且经测试其给出的事实信息准确有效。

此外，在 2022 年英国《自然》杂志（Nature）的一篇论文中，谷歌旗下的 AI 公司 DeepMind 发布的 AlphaTensor 人工智能系统，能显著提高矩阵乘法的运算效率，解决了数学领域一个已困扰多年的问题；在 2021 年美国《科学》杂志公布的十大年度科学突破中，预测蛋白质结构的 AI 模型 Alpha Fold 位列其中，研究人员可以使用人工智能模型来设计应用于疫苗、建筑材料或纳米机器上的全新蛋白质，能基本预测此前人类尚未被解析的蛋白质组结构，中国科学院院士施一公认为，“这是人工智能对科学领域最大的一次贡献，也是人类在 21 世纪取得的最重要的科学突破之一。”

认识发展路径螺旋式上升的规律

任何一场旷日持久的科技革命都并非一蹴而就，而是遵循着一条曲折前进、螺旋上升的发展路径。近年来，随着人工智能的快速发展，AI for Science 也逐渐从星星之火发展成燎原之势。但在鄂维南院士看来，“我们还有很多事情要做。”在科研层面，一方面要持续加大投入，中国科学技术大学教授胡素磊认为，对于构效关系复杂的材料体系，只要投入足够多的研究人员和科研资源，AI 计算模型也能实现突破，也要破除“失败实验不报道”的局限，无论成功与否都要对实验进行记录，在错误中寻找潜在的新机会。另一方面，必须面向产业定义真实的问题。AI for Science 不仅是开发科学研究的平台和工具，也要探索在产业界的垂直落地，总结问题并提出解决方案。在药物研发、基因研究、生物育种、新材料研发等人工智能与科学研究紧密结合的重点领域取得系列突破，包括改变工业研发模式，如利用高精度模拟实现材料定向设计；突破研发瓶颈，在生物医药领域探索难成药靶标的药物发现；等等。

在工具层面，勇于革新。鄂维南认为，DeepModeling 开源平台、物理模型都可以视作工具层面，部分人认为 AI for Science 只是一个由数据驱动的模式，非常有局限性。尽管基本原理层面的革新更具本质性，但“怎么使用这些基本原理真正帮助我们解决实际问题”也不容忽视。深势科技创始人兼首席科学家张林峰表示，“开源是软件的一种去中心化的分布式的评审方式。”一个软件如何作为基础设施应用于各行各业，让更多的人使用，倒逼开发者对开源方案进行调整升级，进而在协同开发的方式下推动“技术的迭代、场景的试错、人才体系的建设和整个生态的打造”。

在创新层面，调整优化科研资源配置，把期望真正寄托在年轻人身上，为其打造成长发展的平台，鼓励更多科研工作者为了初心去奋斗，在原始创新的道路上躬耕不辍。

在文化层面，培育创新文化。鄂维南说，“一个事情没火之前没有人关注，不闻不问，一旦火了很多人都冲过去了，这个现象很普遍。”人们科研的精力和注意力是有限的，发展人工智能，需要用挑战性的场景和问题推动底层系统性创新，而不是盲目跟风炒热点，重复造轮子，“需要开始树立‘既然别人做了，我就不再去做’的思维模式，培育‘不扎堆、不聚集’的创新文化。”

站在科技革命的时代转角，AI for Science 呼唤各行各业的人们打破壁垒，凝聚共识，拥抱人工智能时代的新变化和新机遇。“尽管范式这个词在国内常被庸俗化，但 AI for Science 毋庸置疑是科学研究的一个新范式。”鄂维南说，面向未来，AI for Science 还有巨大的想象空间待开发，帮助我们加快走完从科学研究到产业创新的“最后一公里”。

相关链接：

AI for Science：科学研究新范式[EB/OL]. (2023-05-10)[2024-04-23].
http://www.news.cn/globe/2023-05/10/c_1310715935.htm.

人工智能助力新材料的合成

技术进步已经改进了计算机程序识别新材料的能力，但这个过程面临的主要阻碍，是学习算法如何适应与其所学相反的结果，因为新发现本质上是用新的、创造性的方式去理解数据的能力。

2023 年 11 月 29 日，*Nature* 期刊发表两篇新论文，表明人工智能（AI）驱动的平台可以提高发现和合成新无机化合物的速度和精确性。

第一篇论文来自谷歌旗下人工智能公司 DeepMind，论文题为“Scaling deep learning for materials discovery”。该研究提出“材料探索图形网络”（GNoME）的概念，GNoME 模型能够通过大规模主动学习改进材料发现的效率。这个程序使用现有文献训练，生成多样的潜在化合物候选结构，然后通过一系列学习不断改进这些结构。GNoME 模型发现了超过 220 万个稳定结构，将结构稳定预测的精确性提高到 80% 以上，在预测成分时，每 100 次试验的精确度提高到 33%。相比之下，此前工作中该数字仅为 1%。

第二篇论文来自加州大学伯克利分校，论文题为“An autonomous laboratory for the accelerated synthesis of novel materials”。该研究开发一种自动实验室（A-Lab）系统。这种 A-Lab 系统根据现存科学文献训练，随后结合主动学习，可对拟定化合物创造最多 5 个初始合成配方。随后它可以用机器臂执行实验，合成粉末形态的化合物。如果一个配方产量低于 50%，A-Lab 会调整配方继续实验，在成功达到目标或穷尽所有可能配方后结束。经过 17 天的连续实验，A-Lab 进行了 355 次实验，产生 58 个拟定化合物中的 41 个（71%）。这项研究表明，对决策算法做一些小改动，这一成功率还可提高到 74%；如果计算技术能得到同样改进，还能进一步提高到 78%。相比之下，研究人员需要花费数月去猜测和实验。

上述两项研究所展示的对 AI 的训练，结合计算力的飞速发展和现有文献，其证明使用学习算法辅助发现和合成无机化合物有着极其广阔的前景，未来的自主实验室将能够以最少的人力、最快的速度去发掘新材料。

相关链接：

[1] 同期两篇 Nature：人工智能助力新材料的合成[EB/OL]. (2023-11-30)[2024-04-22]. <https://www.163.com/dy/article/IKPPB4D6053296CT.html>.

[2] AI 仅用 17 天独自创建 41 种新材料[EB/OL]. (2023-12-04)[2024-04-22].

<http://www.news.cn/science/20231204/4f2926747d84466abfa6371d8e8c311e/c.html>.

[3] Amil Merchant, Simon Batzner, Samuel S. Schoenholz et al. Scaling deep learning for materials discovery [J/OL]. Nature. (2023-11-29)[2024-04-22].

<https://www.nature.com/articles/s41586-023-06735-9>.

[4] Nathan J. Szymanski, Bernardus Rendy, Yuxing Fei et al. An autonomous laboratory for the accelerated synthesis of novel materials [J/OL]. Nature. (2023-11-29)[2024-04-22].

<https://www.nature.com/articles/s41586-023-06734-w>.

微软和 PNNL 利用 AI+HPC 加速材料筛选进程

“我们正处于科学发现新时代的黎明，借助最新的人工智能（AI）技术和高性能计算（HPC）能力，我们可以加快科学发现速度，包括发现可以解决时代问题的新分子、新材料。”微软战略使命和技术执行副总裁 Jason Zander 在一次公开报告中说道。

近期，微软与美国能源部太平洋西北国家实验室（PNNL）携手进行一项引人注目化学实验，通过人工智能（AI）与高性能计算（HPC）的协作，他们成功地在绿色电池新材料方面取得突破性成果。

传统筛选约需 20 年，AI+HPC 助缩短至 80 小时

这一合作充分展示将 AI 与 HPC 相结合所能带来的加速科学发现的潜能。在短短 9 个月的时间里，这个团队成功地发现并合成一种适用于资源节约型电池的新型材料。

项目的大部分计算和筛选工作要归功于微软的 Azure Quantum Elements 平台（以下简称 AQE 平台），它大幅加速相关科学研究的进程。据介绍，AQE 平

台集成高性能计算、人工智能和量子计算功能，可实现多方面的计算速度、规模和准确性的提升。同时，通过将人工智能融合至平台中，AQE 平台在寻找复杂问题答案的速度进一步提升。

得益于 AQE 平台的支持，PNNL 团队在 80 小时内针对绿色清洁能源领域取得了突破性进展。AQE 平台负责人 Nathan Baker 表示：“相较于传统计算方法所需的约 20 年时间，我们不到一周的时间就取得了相应成果。”

这项成果的取得，大致经历以下过程。

首先，在研究人员向 AQE 平台寻求较少锂用量的电池材料时，平台迅速地为他们推荐了 3200 万种不同的候选材料。这些材料是根据选定元素替换已知晶体结构中的原子生成的。然后，AQE 平台借助数百万个材料模拟数据点训练的 AI 模型进行必要的计算。与基于 DFT（密度泛函理论）的传统方法相比，该计算过程显著提高材料属性预测的速度，实现超过千倍的速度提升。

接着，科研人员利用 AI 工具根据分子的反应性和导电潜力进行第一轮筛选工作，将 3200 万个候选方案缩减至约 50 万个新型稳定材料。随后，他们将 AI 模型用于预测材料的氧化还原电位和带隙等物理特性，进一步将候选材料数量减少至约 800 种。之后，通过分子动力学模型将这些材料减少到 150 种。

尽管 AI 筛选速度快，但精确度有限。因此，在第二轮筛选工作中，研究人员还采用高性能计算机评估每种材料的实用性。最后，通过结合必要起始材料的可用性和成本等实际考虑因素，系统产生 23 个候选者。在这些材料中，有 5 个为已知材料，从而得到 18 个有潜力的材料。

最后，PNNL 团队在固态反应中合成最佳候选物——一种氯化锂钠钇材料。这种材料相较于其他业界领先的固态电解质，锂含量减少约 70%，并展示出清晰的结构特征，这也解释了该材料卓越的导电率水平。

AI 和 HPC 加速科学发现的潜能

过去，传统的材料发现过程可以分为以下阶段：明确研究问题、收集现有数据、提出假设以及实验验证假设。这种传统的科学研究范式主要依赖于提出假设并通过人工验证。虽然这个过程在概念上很简单，但实际执行中可能面临许多瓶颈，如消耗大量人力资源、极易造成误差等等，都会阻碍项目的成功推进。

随着 AI、HPC 等技术进步，人们有望构建一个协作型的研究闭环。该流程涵盖了大规模的知识提取、整合和推理，深度生成模型自动生成的新假设，以及使用自动化实验测试，旨在提高科学研究的效率和精确性。多元技术协同将助力科学家更好处理科学研究挑战，促进科学发现和科学创新。

PNNL 的科学家感叹，我们正在迈入一个新的范式。该范式的关键是异构能力的协同工作，即不再受单一技术主导，而是趋向多元技术的协同，“使总体结果大于各部分之和”。

回顾科学发展史，核心基础技术的突破推动科学经历多次重大的范式转变。当前，在实证研究的推动下，研究数据的收集和共享使人们更加全面地看待科学问题，不断促进关键基础理论的突破发展。计算系统的出现和成熟让更复杂的计算成为可能，计算技术规模不断扩张。而人工智能和高性能计算技术各具优势，可提高解决科学发现的效率和质量。

PNNL 科学技术副主任 Tony Peurrung 表示：“我们相信，人工智能、高性能计算与人类科学家的结合是加速取得有意义的科学成果的关键。”他还透露，未来，微软和 PNNL 签署了新协议，确定了在下一阶段将继续共享前沿计算技术，深化双方合作关系。

相关链接：

[1] “筛选 3200 万种材料至 18 种，只需 80 小时”[EB/OL]. (2024-02-07)[2024-04-23].

<https://news.sciencenet.cn/htmlnews/2024/2/517359.shtm>.

[2] Chi Chen, Dan Thien Nguyen, Shannon J. Lee et al. Accelerating computational materials discovery with artificial intelligence and cloud high-performance computing: from large-scale screening to experimental validation [J/OL]. Materials Science. (2024-01-08)[2024-04-23].

<https://arxiv.org/abs/2401.04070>.

[3] Discoveries in weeks, not years: How AI and high-performance computing are speeding up scientific discovery [EB/OL]. (2024-01-09)[2024-04-23].

<https://news.microsoft.com/source/features/sustainability/how-ai-and-hpc-are-speeding-up-scientific-discovery/>.

优美科与微软签订 AI 平台协议，加速电池材料技术开发

优美科与微软正式签订协议，利用人工智能（AI）作为促进和加速其电动汽车电池材料技术研究的手段。这是集团的一个关键增长领域。这将使优美科成为应用人工智能方法支持电池科学家开发新型电池材料的先行者，使他们能够更快地将产品推向市场，并提高开发过程的成本效益。

电池材料人工智能平台将利用微软的 Azure OpenAI 服务，并通过大量未发布的特定科学人工智能模型进行扩展。该平台将在优美科专有环境中运行，并完全保障知识产权。优美科将创建一个量身定制的人工智能环境，该环境将分析、综合并汇集优美科专有电池材料研发中数十年的大量复杂数据。此外，它

还将把优美科的数据与外部历史数据以及各种来源的最新技术信息（包括仿真模型、实验或图像）结合起来。

“优美科希望最大限度地利用自 20 世纪 90 年代中期研究电池材料以来所获得的大量电池技术知识。在微软的支持下，优美科将率先应用人工智能作为我们电池科学家的工具，为我们的创新赢得时间、效率和规模，同时保护我们在这一重要研发领域的知识产权。它将帮助我们充分利用所有令人信服的内部和外部数据，更智能地配制新材料，并在了解和抢先满足这一快速增长市场的客户需求方面保持领先地位。”优美科首席执行官 Mathias Miedreich 表示。

多年来，优美科一直将材料、工艺和数据建模作为产品开发的重要工具。在过去几年中，人工智能和机器学习为优美科增添了新的维度，并促成了优美科在电池材料领域申请首批人工智能专利。

优美科与微软在达沃斯世界经济论坛（WEF）上签署协议。“以人工智能为经济社会驱动力”正是此次活动的关键主题之一。

优美科的电动汽车电池材料创新技术为入门级小型汽车、大众级中型汽车以及高端精品汽车提供服务，技术范围涵盖镍、钴、锰（NMC）、富锂锰基（HLM）以及未来的固态和钠离子电池技术。

相关链接：

优美科与微软签订人工智能平台协议，加速和扩大其电池材料技术的开发[EB/OL]. (2024-01-18) [2024-04-22]. <https://www.umicore.cn/zh-s/media/news/umicore-enters-ai-platform-agreement-with-microsoft-to-accelerate-and-scale-its-battery-materials-technologies-development/>.

利用 AI 辅助制备高性能分层多孔海绵体钾电负极

中国科学院大连化学物理研究所周光远研究员团队提出基于人工智能 (Artificial intelligence, AI) 辅助，制备大规模、高性能独特分层多孔海绵体状 N/S 异质元素掺杂的硬碳材料 (NS-C-1100)。

电池储能材料研发是一个复杂的多变量过程。目前的研究流程很大程度上依赖于前向试错方法，并且主要以材料为中心，包括合成材料、制造电解质和电极、组装电池，以及最后评估性能等。即使只考虑这些方面，也有大于 10¹⁰⁰ 种合成活性物质和制备电解质的可能性，几乎有无限种选择电极制造参数的可能性和几十种可能的电池形式，这些远远超过人类大脑所能处理的范围。因此，为了提高电池研究的效率及稳定性，人工智能和机器学习 (Machine learning, ML) 有望能够帮助人们克服处理大量变量和量化数据的方法。

该工作中，研究团队通过人工神经网络模型 (ANN)，筛选得到最优权重文件 (d_{002} , L_a , L_c , SSA , I_D/I_G)；进一步依据权重文件对待测数值进行 ICE、Capacity 准确预测；并在 Github 国际开源网站开源结构模型中建立钾离子电池负极硬碳材料权重参数性能数据库。团队实现了通过 AI 辅助合成大规模高性能硬碳材料，在 1A g⁻¹ 下可逆比容量高达 313.5mAh g⁻¹；即使在 1A g⁻¹ 循环 9000 圈后仍然保持 250mAh g⁻¹ 的高可逆比容量。研究团队还对人工智能在可充放电钾离子电池中进一步发展进行展望，并提出所面临的挑战和可行的解决方案。

相关研究成果以“Hierarchical Porous N/S-doped Carbon with Machine Learning to Predict Advanced Potassium-Ion Batteries”为题，发表在 *Journal of Materials Chemistry A*。

相关链接：

- [1] 我所利用 AI 人工智能辅助制备高性能分层多孔海绵体钾电负极[EB/OL].(2023-04-07) [2024-04-22]. http://www.dicp.cas.cn/xwdt/kyjz/202304/t20230407_6728263.html.
- [2] Ke Bi, Yue Wang, Guangyuan Zhou. Hierarchical porous N/S-doped carbon with machine learning to predict advanced potassium-ion batteries [J/OL]. *Journal of Materials Chemistry A*. [2024-04-22]. <https://pubs.rsc.org/en/content/articlelanding/2023/ta/d3ta00247k>.

北大在 AI4S-基于 AI 的物质结构解析方面取得重要突破

AI for Science (AI4S) 是科学研究的新范式。北京大学深圳研究生院新材料学院潘锋教授团队将图论数学与结构化学相融合创建基于图论的结构化学理论方法，运用图论把晶体结构中的原子及其间的化学键抽象成图论中的点和线，构建出原子间的连接关系图。由此发展以结构基元及其连接关系与相互作用为要素的材料基因挖掘方法，并构建包含 60 万余种独立晶体结构数据库，通过解构晶体成为结构基元与连接关系并结合 AI 方法，自主发展设计新材料、预测结构演化以及自动分析 X 射线/中子衍射图谱解析材料结构的软件系统。使用人工智能技术实现材料结构解析的自动化是物质结构和新材料研究的新范式，它是开发材料领域自驱动实验室的一个关键步骤，对无机化合物、X 射线衍射 (XRD) 分析在材料的结构解析过程起到重要作用。但由于该过程往往需要大量的专家知识，因而实现其自动化仍是一个巨大的挑战。

近日，北京大学深圳研究生院新材料学院潘锋/李舜宁团队设计一个基于残差神经网络的深度学习模型 CrySTINet。该模型能够从 XRD 数据中准确识别出未知材料的结构类型，为自动化 XRD 分析提供新的途径。相关研究成果以“Crystal Structure Assignment for Unknown Compounds from X-ray Diffraction Patterns with Deep Learning”为题，发表于《美国化学会志》（*Journal of the American Chemical Society*）。

在传统的 XRD 分析过程中，研究人员在面对未知材料时如果无法从数据库中找到其物相的晶体结构，则需要借用相近的结构模型，通过对该模型进行调整以获得实测物质的晶体结构。该过程中所借用的结构模型对应未知材料的一种可能的结构类型。对未知材料结构类型的判别一般依赖于专家知识，因而实现该过程的自动化十分困难。由于无机材料的结构类型种类繁多，导致训练得到的深度学习分类模型往往难以获得较高精度。对此，潘锋/李舜宁团队设计一个由多个子模型组合而成的模型框架，每个子模型通过残差神经网络对特定数量的结构类型进行判别，通过联合多个子模型的判别结果可以给出未知材料的最可能结构类型。在该框架下，CrySTINet 可以扩展至新的结构类型而无需对已有子模型进行重新训练，从而使模型能够广泛应用到各类无机材料的研究之中。

研究团队选用 100 种最常见的结构类型的模拟 XRD 数据来训练 CrySTINet 的初始子模型，一共包含 63,963 种无机化合物，覆盖元素周期表中的几乎所有元素。模型在模拟数据集上的准确率达到 80.0%，并且在实验数据集中也拥有同样高的准确率。研究团队进一步使用梯度加权类激活映射（Grad-CAM）来解释 CrySTINet 的分类决策。其结果表明，在每个子模型中，神经网络会将注意力集中至特定衍射角区间以提升子模型中相应结构类型的分类准确率，但这会导致

子模型在面对某些分布外数据时容易给出过高的置信度值。因此，在只依靠神经网络输出的置信度值作为结构类型判定依据时，CrySTINet 的准确度较低，只有 65.7%。而在判定依据中引入与相应结构类型平均 XRD 图谱对比得到的余弦相似度值后，则可以补充 XRD 数据的全局特征信息，从而避免模型陷入对特征峰的过度依赖而导致的误判。以该置信度值与余弦相似度值组合构造的参数作为判定依据，可使 CrySTINet 的准确度最终提升至 80.0%。

相关链接：

[1] 新材料学院潘锋/李舜宁团队在 AI4S-基于人工智能的物质结构解析方面取得重要突破 [EB/OL]. (2024-04-10)[2024-04-23]. <https://news.pku.edu.cn/jxky/2b0b5b39d32e44c98598d4785282ee6e.htm>.

[2] Litao Chen, Bingxu Wang, Wentao Zhang et al. Crystal Structure Assignment for Unknown Compounds from X-ray Diffraction Patterns with Deep Learning [J/OL]. Journal of the American Chemical Society. (2024-03-13)[2024-04-23]. <https://pubs.acs.org/doi/10.1021/jacs.3c11852>.



地址：上海市永福路 265 号

邮编：200031

编辑：吴春莹

责编：崔晓文

编审：林鹤

电话：021-64455555

邮件：istis@libnet.sh.cn

网址：www.istis.sh.cn