

# 人工智能助力材料学未来已来

■本报记者 袁一雪

近日,《NPJ—计算材料》杂志上刊发了一项研究论文。该论文由美国桑迪亚国家实验室一个实验小组撰写,记录了他们开发的一种机器学习算法。这种算法能够以比正常速度快近4万倍的速度为材料科学家进行模拟计算,相当于人们可以在播种三分钟后吃下一个新鲜的西红柿,或者在半秒钟内从纽约飞到洛杉矶。

近几年,人工智能助力材料学发展的研究比比皆是。

1月8日,牛津大学教授Volker L. Deringer团队开发了一种基于第一性原理计算数据集的原子尺度精确机器学习方法。使用更普适的、基于机器学习模型的、具有第一性原理计算精度的高斯近似势能(Gaussian Approximation Potentials,简称GAP)分子动力学模拟(GAP-MD),研究者对包含10万个硅原子(十纳米尺度)系统的液体—非晶态和非晶态—非晶态转变过程进行了研究,同时预测了其结构、稳定性和电子性质。该方法成功地描述和解释了与实验观察一致的非晶态的全部相变过程。这项里程碑式的成果以封面文章的形式刊登在《自然》。

## 为什么是材料科学

中国科学院物理研究所特聘研究员刘淼更看重后一项研究成果。他在接受《中国科学报》采访时介绍道,GAP方法自10年前诞生后,不断提升。“GAP是用量子力学的方式产生大量的数据,再用这些数据构建机器学习模型,从而提升计算的效率与精度,将计算难度简化,甚至完成之前无法完成的事情。”刘淼解释说,“GAP从大量密度泛函理论数据出发,提取原子间相互作用的信息,将材料计算研究对象的空间尺度放大 $10^3\sim 10^4$ 倍,时间尺度扩展 $10^3$ 倍,且精度接近密度泛函理论计算精度。”

在该项研究中,研究人员通过仿真揭示了更广泛范围内硅的液态和非晶态转变过程,为极富挑战性实验条件下的材料预测建模开辟了新途径。此外,除了硅这一特定材料,原子尺度机器学习方法还具备引领新的科学发现的能力。由于该方法能够获得有关原子稳定结构、原子间相互作用等物理性的精确预测,精度接近量子力学方法模拟,可被用于揭示迄今未知的诸多



人工智能与材料学发展相互影响。

图片来源:unsplash

现象,探索介观系统、液态系统、非晶多晶、生物等复杂体系的动力学演化过程。

而前面提到的美国桑迪亚国家实验室的研究则建立在一个拥有128个处理核的高性能计算集群上。研究人员在该计算集群上进行了一次单独的、无辅助的模拟。机器学习的情况下,同样的模拟在使用36个核的情况下花费了60毫秒——相当于在同等的计算机上快了42000倍。这意味着研究人员现在可以在15分钟内完成通常需要一年时间的计算过程。

而且,一直以来,机器学习被用于快速模拟,计算原子和分子之间的相互作用如何随时间变化。但此次桑迪亚国家实验室公布的结果首次展示了机器学习在相对较大的微观尺度上加速材料模拟。通过这种模拟,科学家可以快速模拟熔化金属的微小液滴在冷却和凝固时是如何聚集在一起的,或者当混合物熔化时是如何分解的。

不仅如此,新算法得出的答案正确率也十分高,与标准模拟的结果相差5%。“我们的机器学习框架达到了与高保真模型基本相同的精度,但计算成本很低。”参与该项目的桑迪亚材料科学家雷米·丁格维尔表示。

中国科学院深圳先进技术研究院研究员欧勇盛在解释这项研究时谈道,这是应用计算机仿真加快计

算过程的尝试,而实现这一目标的原因是材料学本身更注重实验,换言之,实验过程的数据计算是人工智能所擅长的领域。“因此,材料学与人工智能之间擦出了火花。”

## 两者相互影响

其实与材料学一样注重实验的学科很多,因此人工智能与其他学科结合的交叉学科也很多。“近期人工智能飞速发展的动能在于算力进步,算力进步使生产、处理海量数据成为可能,进而使人工智能渗透进了各行各业。”刘淼说,“刚好材料科学进入了数据积累快速发展期,高通量实验和计算等一些新技术正在为材料科学领域生产大量数据,人工智能使大数据分析处理成为可能。”

“通过实验,研究人员可以总结和归纳其中的规律,计算机算法就是计算机实验的过程,之后计算机才能够总结并归纳其中的规律,然后完成某项工作。例如,计算机写诗。”欧勇盛介绍道。

2019年2月,美国国家科学院发布了针对材料研究的第三次十年调查——《材料研究前沿:十年调查》,评估了过去十年中材料研究领域的进展,确定了未来十年材料研究的机遇、挑战和新方向。其中,在纳米材料、高熵合金等前沿材料研究领域,人工智能被寄予厚望。机器学习方法在材料设计和材

料筛选方面表现出巨大潜力,有望极大推动新型材料的发现和传统材料的更新。

反过来,人工智能的发展也离不开材料科学的助力,智能机器人、可穿戴医疗设备、虚拟现实(VR)成像、物联网城市系统……未来的智能传感器需要极高的灵敏度、柔韧度、透明度和稳定性,这对材料提出了新的要求。

“人工智能与材料科学,两者应加深交流,相互促进。”中国科学院院士赵忠贤在2019年粤港澳大湾区科技创新论坛上提出,未来,解决人工智能使用的敏感元器件问题要靠材料科学,同时也有必要根据材料科学的需求去发展人工智能技术和理论。

## 先有数据再有算法

“人工智能是个雏形,可以推动很多领域进步,还需要专业人士去改造。”欧勇盛说,“现在,交叉学科是创新的驱动力,人工智能作为一种工具,更需要与其他学科合作才能发挥其最大的作用。”

作为人工智能深度学习的基础,数据的质量和数量是这一学科的关键。目前,不少国家拥有自己的材料学数据库。“有了数据库,人工智能方法如深度学习等才能提取数据之间的关联和规律。”刘淼表示,“有些学科与人工智能交叉得少,就是因为数据量不够大。”

2018年,由刘淼等人联手正式开始创建我国材料科学数据库Atomly,该数据库已经在2020年8月上线。

“材料学数据库上线后,主要影响有两方面。”刘淼解释说。首先,目前大多数研究都是基于经验指导,即研究人员会根据自己的经验做判断,通过实验合成表征对材料进行试错和验证,而当大数据和人工智能介入后,通过数据驱动的方式进行预测,让实验更加有的放矢。其次,人类个体的认知能力有限,即便是积累毕生所学,依然会遇到瓶颈,计算机则可不断横向扩展,无上限可言。“2016年AlphaGo打败了人类棋手,未来人工智能和材料科学数据库也有望不断进步,成为人类材料科研的好帮手。”刘淼说。

相关论文信息: <https://doi.org/10.1038/s41524-020-00471-8>  
<https://doi.org/10.1038/s41586-020-03072-z>

# 新建数据库填补我国材料学大数据空白

■原鸣

自2020年8月正式上线以来,由中国科学院物理研究所、松山湖材料实验室、怀柔材料基因研究平台共同打造的材料科学数据库Atomly,注册人数已突破1000人。“该数据库填补了我国材料科学领域无大型数据库的空白。”中国科学院物理研究所特聘研究员、Atomly材料数据库创始人刘淼说。

此前,世界其他国家,例如美国、德国、瑞士、日本等都建立了属于自己的材料科学数据库。材料科学数据库把材料的基本性能,如原子结构、电子结构、稳定性、力学性能等信息,带到了研究人员触手可及之处,由此引发了材料科学研究方式的革命性进步。例如,近期兴起的材料基因组方法的核心思想就是借助材料大数据方法提升材料科学的研究效率,降低研发成本。

由此基础出发,Atomly作为材料数据库中的“后起之秀”,不仅集各个前辈之大成,还在某些方面超越了其前辈们,甚至实现了诸多创新功能。

## 像钢铁侠一样制作材料

美国漫威经典人物钢铁侠在全息投影中点击几个化学元素,就能得到由其合成的材料的相关数据,并由电脑匹配出最符合要求的复合材料。

这样的电影场景作为刘淼讲解Atomly系统PPT的开头,引发了不少听众的围观兴趣。事实上,Atomly系统确实不是全息界面,但是在电脑屏幕上的操作界面确实由元素周期表组成。“如果用户想要查询某种化合物的性能,只需要点击化合物的元素组成,系统会自动弹出所有与之相

关的化合物。”刘淼讲解到。

到目前为止,Atomly已经收录了17万余种材料的相关数据,这些材料包含了经过数据库比对去重后的无机晶体结构数据库(ICSD)中的大部分结构。刘淼解释说,ICSD在实验合成及晶体研究领域久负盛名,也包含了一大批以往DFT计算研究中提出的假想结构。因此,Atomly内含的材料数据不仅全面,而且和材料实验的联系十分紧密。

材料库不仅提供已知化合物查询,也为材料创新提供更多可能。“面对尝试新材料的研究人员,我们开发了Run4U这一功能。这一功能支持用户在线自主上传新的结构,我们的后台会对这些结构进行初步的筛选,如果数据库中真的没有包含,就会自动进行第一性原理计算,两三天后用户便可在列表中看到想要的计算结果。”刘淼表示。这一功能也同样适用于不熟悉第一性原理计算软件的用户。使用Run4U功能时,用户可以选择“零学习成本”获得指定材料的DFT的计算数据;同时,计算的结果可以自动被后台分析入库,能复查、复用。

## 为材料学不断突破提供可能

爱迪生尝试了成百上千次才成功找到适合做灯丝的材料。从那时起到LED灯的出现,又过去了120年。由此可见,材料的研发过程十分缓慢。但是通过高通量计算,材料大数据让人工智能寻找新材料,却能让新材料研发过程不断缩短。过去70年人类平均每年发现3.3个氮化物材料,但是美国加州大学伯克利分校的Ceder组通过高通量计算等材料大数据方法,一年内发现了92种新材料,并用实验合成7种。



Atomly作为材料数据库中的“后起之秀”,不仅集各个前辈之大成,还在某些方面超越了其前辈们,甚至实现了诸多创新功能。图片来源:unsplash

“我国研究人员一直使用国外材料数据库,国外数据库不但对国内研究人员在材料信息显示上有所保留,而且对我国的数据安全也有威胁。对方能通过监测我国研究人员访问行为获得我们材料研发的信息,不利于我国材料学的发展。”刘淼坦言。Atomly的上线彻底改变了这一现状。

为了更好地让人工智能学习势函数,刘淼与其他研究人员一起开发了一套精准的机器学习势函数工具包(HAAIFF),可以精确拟合分子动力学中所需的体系能量、原子受力等参量。

此外,在保证精准的前提下,研究人员对程序包进行了优化,使其可在GPU上进行训练以及预测,极大提高了该机器学习势函数的速度,为运行分子动力学提供了便利。用户可以同时获得密度泛函理论计算的精度和经典分子动力学的速度。该工具包可供用户自行使用。同时,为了节省用户收集DFT计算数据带来的成

本,研究人员还提供了由该工具包训练的机器学习势函数,用户可在这些函数库的基础上,进行二度训练,这样既可以节省收集数据时间,又可以扩增机器学习势函数适用范围。

## 上线仅仅是一个开始

材料是人类社会的物质基础,实现材料的按需设计是一直以来人类的终极梦想。如今,材料计算已经成为指导新材料研发的常规方法。

刘淼表示,材料基因工程是物理所近期布局的重要发展领域,Atomly的上线仅仅是一个开始。目前,仍有数以万计的新结构正在计算,各材料的介电函数、声子谱等重要且独特的物理信息也正在上线的路上。

在拥有材料大数据积累的前提下,机器学习等更多新型人工智能方法将使材料数据库的整体性和优越性不断完善和提高,为新材料的研发提供更加智能的捷径。

## 一所一人一事



范峰滔指导学生调试仪器。

## 范峰滔

中科院大连化学物理研究所研究员、催化基础国家重点实验室副主任。主要从事(光)催化剂及(光)催化反应过程的原位、动态先进成像技术的表征研究,在国际上开创了紫外拉曼高温高压条件下研究分子筛合成机理的先河;发展了空间分辨的表面光电电压成像方法并在国际上最早将其应用到微纳尺度光催化材料电荷分离的成像研究中。面向国家重大需求,从事深海资源探测的现场光谱仪研发工作,曾在马里亚纳海沟创造7449米的紫外拉曼探测世界纪录。

中科院大连化学物理研究所(以下简称大连化物所)能源楼有一间地下实验室,被科研人员亲切地称为“小黑屋”。这里摆放着先进的光谱表征仪器和装置,用来进行光催化等与“光”密切相关的前沿科学研究。其中,大部分仪器设备都是由“小黑屋”的家长——研究员范峰滔,带领大家一手搭建起来的。这里,也是他追光之旅的起点。

## 踏上追光新征程

范峰滔于2010年博士毕业,师从中科院院士李灿。博士期间,范峰滔便与“光”结下了不解之缘。那时他利用拉曼成像光谱研究了分子筛合成过程机理,并用原位紫外共振拉曼揭示了分子筛催化活性中心的结构和演化机理。毕业后,范峰滔留所工作。李灿对他寄予厚望,希望他结合组里多年光谱仪器开发和研究的积累,在光催化领域拓展新方向。

光催化领域涉及深奥的半导体物理学知识,这对于一直以来都在从事传统催化表征的范峰滔来说跨度非常大。

2010年,李灿因病在北京医治,受当时基于血管成像的介入手术启发,与范峰滔几次打电话交流,认为光谱成像方向值得考虑。这就像一道光,为迷茫中的范峰滔指明了前进的方向。

追光之路并非坦途。研究团队发现半导体的晶面对光催化反应似乎有很大的影响。经过一段时间的摸索和调研,范峰滔计划以原子力显微镜、光学和半导体晶面性质为基础开展工作。由于缺少合适的仪器设备,工作在前期准备阶段就卡住了。他们不得不亲自设计、自主研发。

范峰滔带领团队制备了大量模型样品和装置,多次往返于北京和大连两地,寻求大仪器公司进行测试,打造合适的配件和仪器。在此期间,范峰滔前往荷兰乌得勒支大学进行访问,进一步感受到了仪器创新对于揭示重大科学问题的重要性,也进一步坚定了要发展新的成像表征方法的想法。

2012年,范峰滔在能源楼地下室找了一间不到20平方米的房间,作为攻坚“战场”。而成型后的仪器设备也陆续搬进去,“小黑屋”就这样诞生了。范峰滔带领团队在此开始了表面光电电压及成像光谱表征方面的研究。

## 时光不负追光人

研究初期,范峰滔团队每天总能收到很多奇怪信号。每每回忆起来,他都感慨颇多:“这些没有规律又无法解释的信号就像梦魇一样每天出现在梦中,我半夜常常会被惊醒。”

同时,李灿对实验结果提出的质疑、对仪器性能和谱图解析能力提出的高标准要求,也激励着范峰滔务实笃行,不懈奋斗。他们一边做实验,一边继续学习专业书籍,逐步改进实验仪器和方法。团队互相扶持,每天坚持近14个小时的高强度工作。终于,在经历了漫长而艰辛的摸索后,他们逐渐获得有规律的

信号。历经四年的努力,相关工作终于取得阶段性成果,2015年7月第一篇表面光电电压及成像光谱文章终于诞生了。

万事开头难,在范峰滔的带领下,相关研究慢慢步入正轨。目前,“小黑屋”小组不断发展壮大,越来越多的年轻人主动选择这一有挑战性的方向。他们面向国际前沿领域,取得了一系列丰硕的成果,受到国际太阳能转化研究领域科学家的广泛关注,已在国内外核心期刊上发表论文70余篇,并多次应邀在国际重要学术会议上作报告。

除进行基础研究外,范峰滔还承担了多项面向国家重大需求的重要科研项目。他将博士期间使用的紫外拉曼光谱方法引入到深海资源探测中,研发的7000米级深海原位探测紫外激光拉曼光谱仪在马里亚纳海沟成功通过7000米海试验证,创造了拉曼光谱仪最高深海探测纪录。

## 追光逐梦的领路人

涉及仪器搭建和测试问题时,范峰滔总是亲自上手,和学生们团队协作。作为导师,范峰滔甘做学生的领路人,为年轻的“追光者们”指引方向。学生们都说他就像大哥哥一样亲切,稳重而又幽默风趣,平时常常和大家聊天谈心,给枯燥的科研生活带来温暖和乐趣。

范峰滔为学生创造了一个开放包容的科研氛围。对于学生的科研灵感和想法,他支持和鼓励他们大胆尝试,勇于挑战新方向。得益于他的信任,学生们大胆探索,不因害怕犯错而畏缩。范峰滔还常常组织“讨论会”,讲述一些身边科学家的故事,鼓励大家不辱时代使命,心系祖国发展,为科技强国建设贡献自己的力量。

不论是为人处世,还是科研态度,范峰滔感染着研究组的每一个人,他是“小黑屋”的追光者,也是领路人。未来,他将继续带领“小黑屋”一众追光学子,一同探索光谱的奥秘,在光催化的前沿科学研究领域深耕细作,再创辉煌。

(作者单位:中科院大连化学物理研究所)

## 进展

# 最新宇宙“天图”发布 归功于数据处理技术

本报讯 近日,中国科学院国家天文台北京—亚利桑那巡天(BASS)团队和暗能量光谱巡天(DES)国际合作项目团队联合发布了最新巨幅宇宙二维天图。据介绍,这幅天图由全球200名科研人员历时6年联合观测和数据分析,覆盖两万平方米的天空(约为全球总面积的一半),容纳10亿亿数码像素,包含20亿天体,是目前人类测量获得的最大宇宙天图。

据了解,由国家天文台联合美国亚利桑那大学建立的BASS巡天团队负责了其中北半球的星空,观测期为2015年1月至2019年3月。为了尽快构建图像数据,中国虚拟天文台团队依托国家天文科学数据中心,为BASS巡天项目提供了专业的数据管理与发布服务。

“我们利用阿里云的云计算资源进行大规模的数据处理和分析,独立开发了数据处理软件,发布了三个国内版本的数据产品。”据国家天文科学数据中心副主任、中国虚拟天文台负责人崔辰州介绍,BASS巡天团队目前已获取2.5亿张天体的图片。“以前完整处

理这些图片需要半个月的时间,但是借助阿里云集群式的计算能力,把它缩短到了不到一天的时间。”

据悉,国家天文台从2017年开始与阿里云合作,并成立大数据联合研究中心,帮助中国虚拟天文台和LAMOST望远镜、FAST射电望远镜数据上云。

“没有这些数据,DES巡天项目就无法开展。”法国原子能和替代能源委员会的宇宙学家、DESI新闻发言人Nathalie Palanque-Deslabrouille评价说,“在收集和发布数据方面,我们取得了重大的成就。”

目前,这些巡天数据已在阿里云的平台向全世界开放,以最大程度地发挥科学价值。据了解,基于DES巡天数据发表的论文已超过200篇,其中单篇引用BASS数据的论文超过80篇。阿里云方面介绍,公众也可以通过架设在阿里云上的在线星图浏览平台,浏览这些天文数据。

另据了解,国家天文台正基于阿里云建设开放协同的天文科研云平台,促进国际天文数据和成果的共享。(赵广立)