



如何拥抱智能时代——以化学学科为例

林京龙, 刘谦益, 莫凡洋*

北京大学材料科学与工程学院, 北京 100871

*通讯作者, E-mail: fmo@pku.edu.cn

收稿日期: 2022-02-13; 接受日期: 2022-03-14; 网络版发表日期: 2022-06-22

国家自然科学基金(编号: 22071004, 22150013, 21933001)资助项目

摘要 本文从人类社会发展的视角论述了拥抱智能时代的重要性, 对第四次工业革命的核心技术——人工智能及其将如何影响人类社会做了简要描述; 接着以化学学科的研究为例, 通过介绍一些有代表性的研究工作, 从基础、工具和应用三个层次, 重点论述了机器智能如何改变学科研究范式以及如何帮助人类加快科学发现; 最后, 对人工智能与其他领域的结合提出了一些方法和路径。

关键词 人工智能, 机器学习, 自动化系统, 化学学科, 研究范式

1 引言

人类发展史就是不断升级手中改造自然的工具使用史, 如石器时代、青铜时代、铁器时代等(图1)。18世纪中叶以后, 伴随着蒸汽机的使用、电的发明以及计算机技术的发展, 三次工业革命将人类文明带到前所未有的高度。当前, 以人工智能为代表性技术的第四次工业革命正如火如荼地进行, 并且带来比前三次更大的生产力飞跃以及人类社会的深刻变革。

蒸汽机、电力、信息技术和人工智能是历次工业革命的代表性技术。每次工业革命都用代表性技术命名, 是因为这些技术都是通用技术, 而且可以对其他行业、学科、领域进行赋能, 是所谓的“工具技术”。搭上“工具技术”这趟快车, 让其赋能, 是又好又快地发展所深耕领域的一个重要思路。

因此, 抓住当前的“工具技术”——以人工智能为代表的第四次工业革命的核心技术, 去加快科学研究,

提高科学知识的生产效率, 用科技的力量去发展科技本身, 然后作用在生产上, 必能带来知识积累和经济发展的强劲增长。

基于此, 本文旨在综述人工智能的发展态势、对科学研究的影响, 特别是在化学学科的发展情况等, 并对其交叉融合发展提出了建议。

2 机器(人工)智能

在人工智能的萌芽阶段, 机器智能一词先被提出。1950年, 艾伦·图灵(Turing A)在《思维》(*Mind*)杂志上发表了其著名论文“计算机与智能”(Computing machinery and intelligence), 并提出了如今广为人知的图灵测试^[1]: 如果一台机器能够与人类展开对话(通过电传设备)而不被辨别出其机器身份, 那么称这台机器具有智能。在1956年的达特茅斯会议上, 人工智能一词由约翰·麦卡锡(McCarthy J)正式提出。

引用格式: Lin J, Liu Q, Mo F. How to embrace the age of intelligence——taking chemistry as an example. *Sci Sin Chim*, 2023, 53: 39–47, doi: 10.1360/SSC-2022-0028

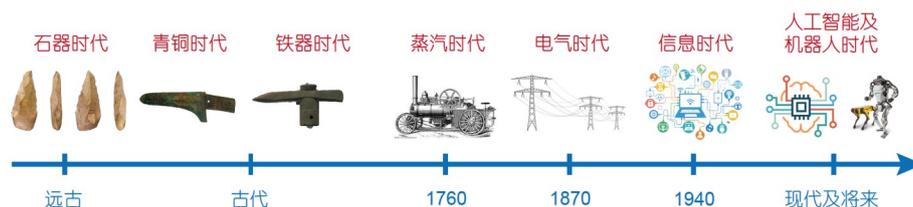


图1 人类改造自然的工具发展史(网络版彩图)

Figure 1 The history of man's tools to transform nature (color online).

这里使用“机器智能”,旨在突出所讨论的主题,不仅在智能上是“非人的”,即“人工的”,还包括机器一词所具有的普遍的、印象化的含义——“机械的”、“自动化的”和“机器人的”等所涉及的物理改造层面。无论有多么强大的信息处理技术,人类社会的生存和发展最终要落在对物质世界的改造上。通常人工智能(狭义)是指通过计算机程序来呈现人类智能的技术。但广义人工智能也包含智能机器系统用于物理改造,如感知与智能机器人等。

人工智能的分类、本文提到的机器学习以及深度学习的关系如图2所示。人工智能是一个包含众多研究领域的集合概念,机器学习只是其中一个分支;而基于神经网络算法架构的深度学习又是机器学习的一种。

2.1 人工智能近期发展态势

人工智能在整个70年的发展历程中有过几次潮起潮落。其原因在于人的主观愿景是不受拘束的,而技术的发展是需要日积月累的。近20年来,随着摩尔定律驱动着计算机计算能力的提高,互联网和移动互联网的崛起带来海量数据的积累,以及蛰伏几十年的神经网络算法的崛起,人工智能又迎来了新一轮的研究和应用热潮。日常生活中常见的推荐系统、人脸识别、语音助手以及智能辅助驾驶等就是人工智能技术的具体应用。此外,人工智能还被应用在能源系统调度、智慧城市、教育和医疗大健康等领域。可以说,人工智能正在以前所未有的方式,深刻地改变着人类社会的生产和生活方式。很多工作被重新定义,甚至已经被取代,但同时人工智能也会创造出一些新的工作。从这个角度说,拥抱人工智能技术就变得特别重要。总的来说,人工智能技术对每一个生产单元的加强以及对整个社会资源调配的优化两方面同时起作用,极大地提高社会生产力,创造更多价值。

2.2 人工智能促进科学研究

人工智能已经、正在并且越来越会成为众多学科科学研究的重要组成部分^[2]。机器学习作为一种研究方法分析高维度、高通量数据用于模式识别、分类和预测,已在广泛的学科中产生重要影响,包括信息科学、数学、物理、化学、材料、生物学、地球科学、管理科学、经济学、心理学以及其他数据密集型经验科学。

与传统的基于if...then...逻辑的硬编码程序不同,机器学习从数据中基于人为设定的目标进行优化计算,寻找隐藏在高维数据中的关联性和规律性,这个过程称之为“学习”。学习是人工智能研究时代的关键词。大数据技术和计算能力的提高使得从海量数据样本中获取特征和信息变得更加高效,从而提供了一种前所未有的处理复杂性的方法。这种能力正是人类所欠缺的。这是要大力发展人工智能的一个重要原因。

当前,我们还处于弱人工智能时代,虽然机器在搜索、计算、存储和优化领域比人类更高效,但它的高级认知功能,如感知和推理等方面,目前还远远比不上人脑。郑南宁院士指出,人机协同的混合增强智能是新一代人工智能的典型特征^[3]。人工智能与人的融合体现在相互之间的功能增强和信息增强,并借助于计算机的高速度,大数据等优势 and 人的强视觉、高灵感、原创力等优势实现优化增强和极限增强,从而开辟通往强人工智能之路。

3 人工智能赋能化学学科的科学研究的科学研究

结构决定性质,是化学学科的主旨。人工神经网络算法理论上可以拟合任何复杂的函数^[4],对于“性质= f (结构)”这样一个映射,只要有足够多的数据,足够“深”和“广”的网络结构,以及可以实施的算力,也应该

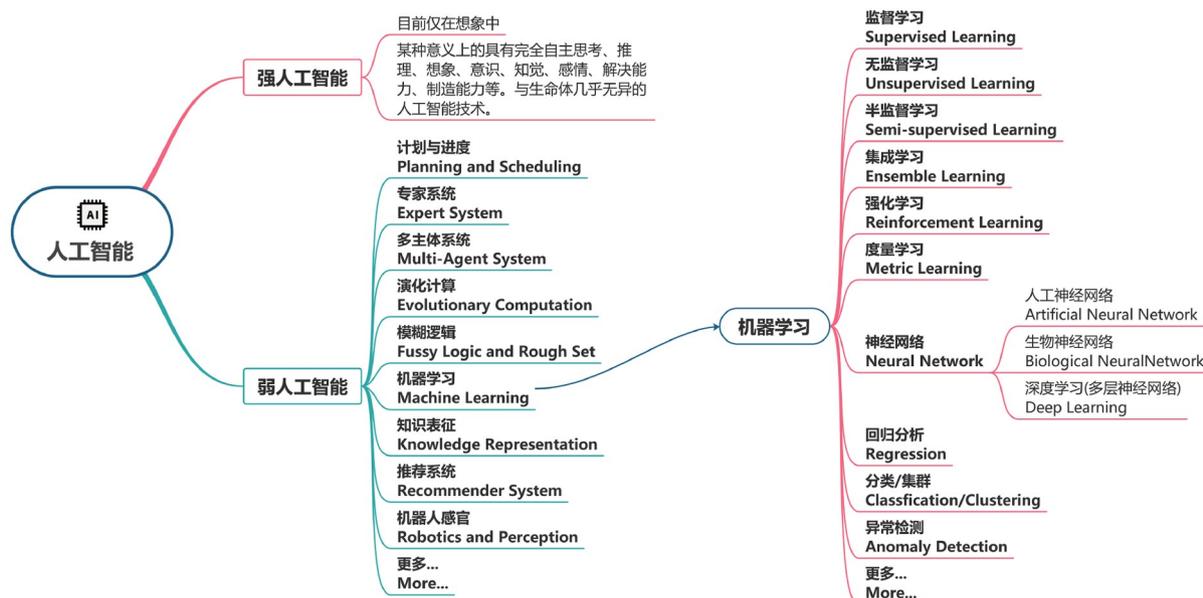


图 2 人工智能及机器学习的分类(网络版彩图)

Figure 2 The classification of artificial intelligence and machine learning (color online).

是一个可求解的问题。事实上, 化学空间的高维度、复杂性以及非线性, 尤其适合通过数据驱动的人工智能算法来求解(预测)。当前, 在化学领域中, 借助人工智能技术的主要研究有应用自然语言处理技术的化学领域文献文摘的自动生成^[5]、化学数据库中的智能检索方法^[6]、化学实验室的自动化与机器人技术^[7]、人工智能辅助合成设计^[8]、人工智能辅助药物筛选^[9]等(图 3)。这些方法无疑为化学家开辟了新的广阔的研究空间。

3.1 基础和工具

3.1.1 化学信息学是人工智能与化学研究结合的基础

化学信息学(Cheminformatics)是使用计算机和信息技术对化学信息进行表示、管理、分析、模拟和传播, 以实现化学信息的提取、转化与共享, 揭示化学信息的实质与内在联系, 促进化学学科的知识创新。

对于实际物理空间和虚拟设计空间中都呈指数级增长的化合物数量来说, 有必要建立一种结构和性质的数学映射关系, 来研究和评估化合物的性质, 减少昂贵且耗时的实验测试, 这就是定量构效关系研究(quantitative structure activity relationship, QSAR), 它是化学信息学的一个重要子领域。研究定量构效关系

首先需要解决的是分子的计算机输入问题, 这就引出了另外一个重要概念——分子描述符。简单地理解, 分子描述符就是分子在某一方面性质的度量, 既可以是分子的物理化学性质, 也可以是根据分子结构通过各种算法推导出来的数值指标。专业一点说, 就是将分子结构所呈现的各种信息进行结构化编码^[10]。分子描述符有上千余种, 而且用户可以根据需要自己定义分子描述符来适合特定问题的研究。

分子指纹一般认为是分子描述符的一种, 是专门针对分子结构的信息化呈现。它将分子的结构、性质、片段或子结构信息用某种编码来表示。常用的分子指纹包括Daylight fingerprints, MACCS keys, MDL 以及public keys等。以下将详细阐述获得这些分子描述符的方法。

3.1.2 计算框架作为研究工具

计算框架是已经开发并封装好的编程环境, 以方便使用人员在此环境下能够快速搭建模型和实现应用, 而不必在算法实现的计算细节上花费时间和精力。这为非计算机编程专业人员提供了便利, 以及为能够使用这些强大工具提供了可能性。

正是基于此, Python以其语法简洁, 容易上手的特点, 已经成为与人工智能应用端结合的最紧密的一种

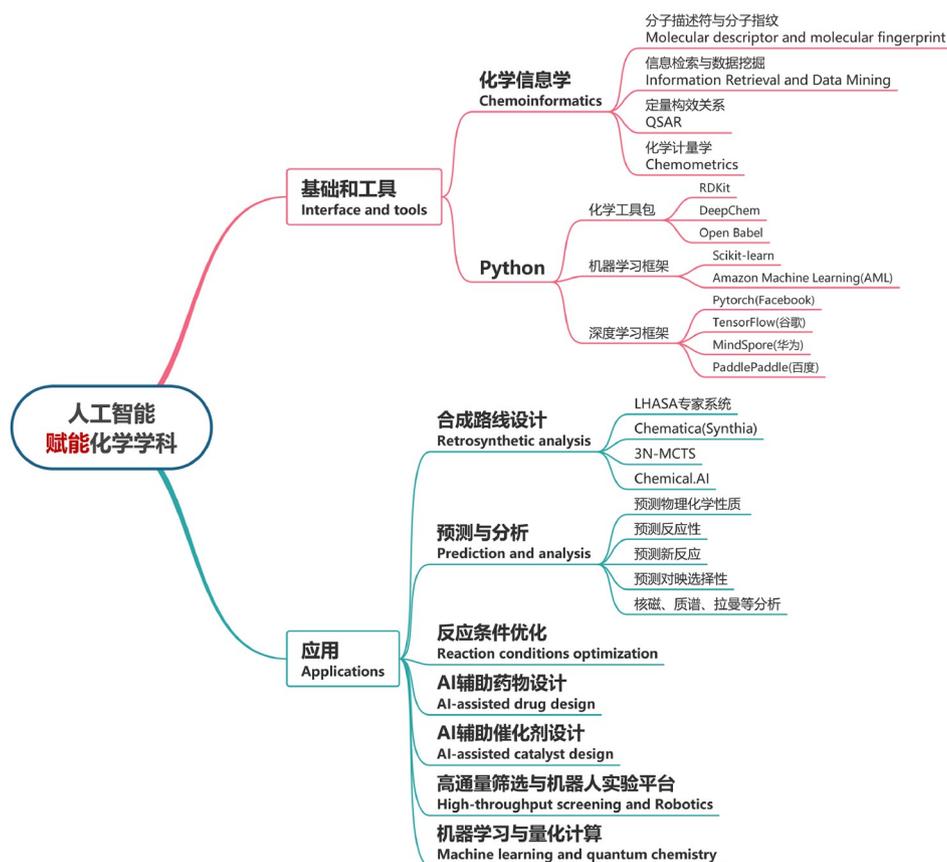


图3 人工智能赋能化学学科(网络版彩图)

Figure 3 Artificial intelligence expedites chemical research (color online).

编程语言,它具有出色的库生态系统,这两者互为因果。近些年,世界顶尖科技公司相继开发并开源了自己的基于Python(或优先支持Python)的AI计算框架,如谷歌的TensorFlow、Facebook的Pytorch以及华为的MindSpore和百度的PaddlePaddle等,Python在人工智能领域更加奠定了其第一语言的地位。这里要指出的是,人工智能核心算法的底层实际上是用C/C++编写的。这与本文关系不大,不再展开。

如前述提及,如何获得分子描述符数据?这里做一简要叙述。计算分子描述符可以使用很多商业和免费的计算软件,如BlueDesc、CDK Descriptor、DRAGON和Multiwfn等。这些软件独立使用起来很方便,但如果做机器学习项目,需要一次输入成百上千个分子,软件必须要提供API(应用程序接口)才行。一些直接以编程语言(如Python)程序包的形式软件,则使用起来更为方便。比如RDKit和DeepChem等,它们作为第三

方库,可以直接导入Python中使用。RDKit是一个开源的化学信息工具包,提供了大量对化学分子的计算操作^[11]。DeepChem则将化学计算和AI算法集成在一起,成为专门为化学和材料领域打造的深度学习计算框架^[12]。

总之,计算框架在科学研究中发挥了基础设施作用。换言之,想要应用AI与领域结合,必须掌握这些计算框架的使用。值得指出的是,近年来AutoML(Automated Machine Learning)技术发展得十分迅速^[13],但这只是在操作层面上简化了许多工作,对于机器学习的理解要求实际上并没有降低。

3.2 应用举例

因文献已详细分类和讨论该交叉领域里的众多研究成果^[14-17],本文不再赘述。我们重点叙述几个有代表性的领域,以点带面,向读者展示机器智能是如何帮

助人类加速科学发现的。

化学是一门以实验为基础的学科。我们将这种以实验为主的学科的科学研究范式分为彼此相对独立又充分地相互作用的两个部分: 前人知识的总结、归纳以及创新性设计, 实验发现和验证进而提出修正以及创造新的知识。前一个是信息的加工整理和创新性思考, 后一个是物理世界的改造。以下在这两个分类的框架下, 结合几个具体的文献实例来详细论述这部分内容。

3.2.1 计算机辅助合成路线设计

化学合成的逆合成分析, 应该是第一个分类中, 最具代表性的研究方向。早在20世纪60年代, Corey等^[18]就提出并研究过这一问题, 并将其系统化、理论化。后来又开发出一些计算机程序LHASA (Logic and Heuristics Applied to Synthetic Analysis), 来进行辅助合成路线设计。然而, 正如前所述, 当时的计算机性能、算法以及合成反应数据库, 都还处于非常低的水平, 因此没有得到广泛应用。近20年来, 技术进步带来了新的契机, 这方面的研究有了长足的发展。

经过多年的努力, Grzybowski教授及其团队^[19,20]开发了一款名为Chematica的逆合成分析软件。这款软件构建了包含700多万有机分子的超大数据库, 并通过相似数量的有机反应将它们彼此连接形成网络, 并且他们手动录入超过5万个有机反应规则来告诉Chematica任何小分子在反应中可能会发生的变化(以上数字可能会有更新)。据文章报道, 在一些测试中, Chematica创建的路线不仅与此前化学家报道的合成路线明显不同, 而且步骤更少或是产率更高, 耗时更短, 成本更低。2017年, Chematica被默克收购, 更名为Synthia, 并提供商业化服务。

如果之前的Chematica还需要人来教它化学反应规则, 那么下面这个成果甚至可以自己学习这些规则。Segler等^[21]报道了基于神经网络与蒙特卡罗树搜索相结合的AI算法, 依靠自动提取的规则数据进行训练和深度学习。当设计一条合成路线时, 其会像人类一样进行选择 and 判断, 根据其学到的设计规则找出最佳的合成路线。这意味着, 通过自我学习和进化, 其可以完全不依靠人类已有的经验和策略, 自行创造新的策略去寻找合成目标分子的最佳路线。

诚然, 在一些具体的问题中, 化学家设计的路线有

可能在科学的优雅性方面, 比计算机所设计出来的要好(但其实这也非常主观), 但这只能是零星出现, 而且保持不了多长时间。未来的设计就像AlphaGo下棋, 完全超出人类顶尖棋手的想象空间。从效率方面看, 人类智能在这方面止步不前, 机器只会越来越快。

未来这方面的发展方向, 应该结合自动化和机器人技术: 机器不仅设计, 还负责实施, 整个过程都不需要人的干预。不过这只是系统整合的事情, 关键的问题是要结合应用需求。

3.2.2 自动化与机器人技术在化学研究中的应用

化学技术帮助人类获得生存和发展的物质基础。化工生产的自动化早已解决, 一方面这种大规模生产非自然人力所能为, 另一方面针对特定产物的自动化系统容易开发和实现, 并带来可观的经济效益。然而, 化学实验室的工作流程基本都是非标准化的, 而且异常复杂。科学研究本是一个发现过程, 在大多数情况下, 无法预见结果, 同时也时刻需要根据实验现象做出判断, 以进行下一项操作和处理。另外, 一般而言, 很难同时具备两个及以上学科的系统且精深的专业知识背景, 因此要通过自动化水平来提升化学实验室的工作效率, 需要多方面人才的共同努力。

虽然有以上很多困难, 但化学实验室的自动化, 一直以来备受关注。剑桥大学的Ley教授在这方面做了许多工作。这篇综述详列了他们课题组以及其他课题组在相关领域内的工作^[7]。比如一种萃取操作的机器视觉自动化实现^[22]、在线溶剂闪蒸装置^[23]、优化色谱分离^[24]以及自动化过滤^[25]等。这些工作很有启发性。

在化学实验室自动化并结合机器学习和深度学习方面, 格拉斯哥大学的Cronin课题组非常活跃。他们课题组的Logo是digital chemistry/programming matter, 可见他们的研究兴趣和致力方向。他们发表了很多工作, 其中一些是有关自动化合成机器^[26,27]。系统的实现是基于几台受计算机控制的泵, 注入反应物到反应瓶中, 以及反应产物的后处理, 如萃取、柱层析、旋转蒸发等操作, 都是靠泵在复杂管线系统中转移液体反应物来实现的。同时还将一些波谱检测加入进来, 如红外光谱和核磁共振波谱等。他们用机器学习算法来读取这些谱图, 获得反应信息, 然后将其反馈给系统, 达到闭环优化的目标。此外, 值得借鉴的是他们同时开发了一套可编程的模块化合成系统软件。这些工作很有开

创性。

利物浦大学Cooper教授与自动化公司Labman^[28]一同研制开发可移动机器人的自动化系统,用于催化剂材料筛选。这项工作给人以很强烈的视觉和心理冲击。机器人完全模仿人类化学家进行科学实验,它会称量固体催化剂,分配液体,启动实验装置。此外它还能操纵分析仪器对实验结果进行检测。在一周的时间里,这个机器人除了充电以外不眠不休,完成了近700个实验。如果人来操作,可能要几个月的时间。更大的亮点在于,除了初始的一些实验条件以外,其余的实验条件是根据实验结果,采用贝叶斯优化算法,不断调整和优化实验参数得到的。这种基于机械臂、可移动、可编程并且带有AI算法的机器人系统,能够在更大程度和更大范围内帮助人们进行化学发现。

应用机器智能,包括物理操作的自动化和信息处理的人工智能算法,从体力和脑力两个层面对人进行解放,使研究人员能够把更多的时间和精力放在真正有价值的科学思考上面,这是一种研究范式的转变。

当然,在实验室开发机器智能系统并不容易,因为其中涉及很多学科和专业技能,需要团队,同时经费投入也很大。机器智能系统的开发一定要以科学问题为导向,以解决实际问题为目标。如果没有这两点,容易给人一种工程化的印象。当然工程化很重要,但通常不一定是一个科学研究所擅长的。

已有的一些报道,给出了一些解决方案,会带来一些启发,但远非标准,更不可迷信和盲从。完全基于复杂管线的系统,如何保证不交叉污染;在自动化系统中,可移动机器人的优势和劣势怎样比较;系统如何解决复杂性带来的不确定性;是否有更简单且可靠的方式解决众多模块物理空间连接的问题;自动化专机和机械臂对比,都有各自的哪些特点;一套复杂系统的搭建,能在多大程度上解决什么规模的问题;将每一个新生成的实验结果放到之前的数据中进行训练,如何平衡模型的拟合能力和泛化能力;在没有人监督的情况下,机器寻找的优化方向一定是对的吗?在实验室自动化和智能化方面,有太多的方案和路径以及细节需要讨论。关键在于,要找到适合自己实际应用场景需求,并紧密结合科学问题。

3.2.3 机器学习促进反应机理探究

在有机合成中,反应物和产物众多,反应条件复

杂,在这样高维空间中理解化学相互作用的机理具有挑战性。因此,有研究者通过机器学习模型关联反应条件和产率,基于模型的可解释性找到对产率影响较大的因素,并从化学的角度进行解释,从而帮助对机理进行探究。

普林斯顿大学Doyle课题组^[29]联合默克研究人员,利用随机森林算法预测Buchwald-Hartwig偶联反应的产率。首先,他们利用高通量筛选平台,在短时间内完成了4608个Buchwald-Hartwig偶联反应,其中包括15种底物、23种异噻唑添加剂、4种钯配体和3种碱。而后,结合核磁数据和DFT计算等,对反应空间进行特征抽提,组成了一个120维的特征向量,反应产率作为标签,形成样本集。采用常见的几种机器学习算法进行回归训练,发现随机森林算法表现最好,以 R^2 为0.92的精度预测产率。

这里,模型以较高精度预测产率的意义在于,可以在秒级时间内,做大规模的“计算机思想实验”。我们可以任意组合反应条件,然后知道它的结果,而不用耗时费力地去进行实验。反过来,模型也能够帮我们迅速精准地“圈定”极可能出现最优反应条件的范围。这对于提升实验室工作效率,加速科学发现,具有重要意义。

在这篇文章中,模型还对特征之于产率预测的重要性进行了研究。发现对应于异噻唑C3位的核磁碳谱位移是重要程度最高的描述符。这启发了研究人员。他们据此进行实验,对不同C3核磁位移的异噻唑添加剂的反应进行研究,发现当异噻唑的C3核磁位移较大时,会得到Pd与N-O键发生氧化加成的产物,因此得出结论: Buchwald-Hartwig偶联反应的产率与异噻唑的竞争反应有关,并从机理上进行了解释。这即是受机器学习寻找到的规律所启发的机理探究过程。试想,若不是借助于机器智能,人类很难发现数据中所隐藏的模式。事实再清楚不过:机器对大量数据具有永恒记忆,能够识别模式并进行大多数人类自身永远无法发现的推理。

3.2.4 计算化学进入机器学习时代

随着计算机技术日新月异的发展,计算化学(Computational Chemistry)成为研究化学反应和分子性质必不可少的工具。针对不同类型的问题,计算化学家不断提出新的方法,以求在计算能力允许的范围内尽可能获得可靠的结果。由于计算化学和机器学习

都是基于数学和计算机技术实现的,其相互结合的门槛较低,因此利用机器学习对计算化学进行优化是彰明较著的发展方向。

密度泛函理论(density functional theory, DFT)是应用最广泛的量子化学关联计算方法之一。然而,由于精确的密度泛函特别是交换泛函是未知的,实际应用的泛函都是一定程度上近似的模型,限制了计算的精度,特别是描述分数电荷或分数自旋时会有很大的系统误差。基于此,DeepMind团队的Kirkpatrick等^[30]报道了一个名为DM21的基于神经网络模型训练建立的密度泛函。在数据集中,精确的能量(输出向量)来自于大量的已知文献报道或非常精确的CCSD(T)计算结果。训练完成之后,用训练得出的密度泛函DM21计算不同的体系和反应,结果表明DM21准确模拟了一系列复杂系统,如压缩的氢原子链(该体系无实际意义,仅作为验证)和带电荷碱基对等。此外,对多个分子数据集的能量计算测试中,DM21的平均误差显著低于传统泛函。由于DM21在训练中遵从了对分数电荷和分数自旋两类系统的已知约束条件,因此克服了传统泛函对这两类问题计算的系统误差。当然,DM21也仍然具有很大的局限性——由于DM21的输入项仍需基于大量对波函数的计算,实际应用中计算代价还是会显著地受到体系大小的影响。

国内的深势科技团队核心成员与合作者创新性地多尺度建模、机器学习和高性能计算融合,突破性地实现了多尺度分子模拟中精度与效率的统一。代表性成果DeePMD-kit开源软件可以实现在保持量子力学精度的基础上,将分子动力学的计算速度大幅提升^[31]。

这方面的研究还有很多,读者可参考该交叉领域最近的一些综述^[32-35]。毋庸置疑,机器学习与计算化学的结合正成为计算化学发展中的下一个重大飞跃。

4 结论和建议

4.1 结论

科技工作者需要拥抱人工智能技术。正如在计算机普及的年代,我们学习Word、Excel以提高工作效率,使用专业软件如ChemDraw、Origin等进行学术研究,我们并不觉得是在跨领域学习计算机科学。其原因有二:一是这些软件容易操作,使我们感受不到学科壁垒;二是这些技能是必备的,容不得我们去选择要不

要学。这些是发生在第三次技术革命,即信息时代的事。但是现在第四次工业革命已经开始,这次工业革命的核心技术我们也必须要掌握。诚然,我们也要承认,这项技术的学习门槛很高,即使在人工智能自己的领域里,仍有大量的科学问题需要解决,人工智能自身也在快速发展。但看问题要看本质:门槛再高,对于我们“外行”,它就是工具技术,本质上和Word没有区别。我们学习Word并不是编写一个能实现Word功能的程序,同样我们学习人工智能技术也不是去创造一个新的优化算法,那是计算机领域专家的工作。我们只是把这些已经开发好的算法应用到自己领域的具体问题中。

本文重点论述了拥抱智能时代的重要性,通过实例说明人工智能技术如何改变科学研究的范式。对于科学教研组来说,始终应该明确的是,应用机器智能不是目的,而是手段,目标应该始终是面向科学问题。实验室的科研工作从劳动密集型过渡到自动化,再从自动化迈向智能化,是发展趋势。在未来,我们还将面临更复杂的问题,而可用的数据量也将难以想象。因此,我们将越来越需要机器来帮助我们了解并理解它。但最终,人是科学活动的创新主体。我们建立新的结构,使机器的智能更强大,并应用机器强大的计算能力来推动新的发现。

4.2 建议

我们对想进入化学和人工智能交叉领域的研究者提出一些建议。

(1) 学习机器学习和深度学习的相关知识。机器学习和深度学习算法背后有较深的数学和统计思维,以及推导过程,对线性代数、概率论和微积分相关知识都有较高的要求。对于已毕业多年的非数学专业的研究者来说,可能不得不重新学习。对于简单应用来说,不需要太在意数学原理;但反过来,对算法了解得越深,就越能在具体问题中将其发挥得更好。诚然,AutoML技术可以解决一部分问题,但对于科学研究,我们总希望能把握住每一个细节。总之,在两个或多个领域都有深刻理解的学者往往能够更具灵感,也更具创造力。这也是交叉科学独有的美感和激动人心之处。

(2) 如果希望在实验室自己动手搭建一套自动化系统,需要掌握以下知识和技能:电气知识、设计简

单的机械结构并能使用CAD软件绘图、机器人编程和操控技术、可编程控制器(programmable logic controller, PLC)、通讯技术等. 最后还要有很强的动手能力. 如果经费允许, 也可以找专业公司, 但价格不菲, 投入产出比是一方面, 后期维护和改造依然是问题.

(3) 鉴于科学研究组力量有限, 而部署人工智能和自动化系统需要很多资源. 一个可行的解决办法是联合业界, 就共同感兴趣的科学问题展开合作. 一方面业界可能具有某一领域里的大量数据和应用场景, 另一方面学界有提出科学问题和解决科学问题的智力资

源. 两方合作, 推动领域进步, 共享科研成果.

(4) 从领域内出发, 解决实际问题. 前不久《科学》杂志将AI预测蛋白质结构评为2021年度突破^[36], 文章称该项技术解决了生命科学领域中50年来一直困扰人们的难题. 很多领域都存在一些长久以来没有得到很好解决的难题, 已有的一些方法和技术不适合这类问题是一个重要原因. 鄂维南院士提出AI for Science概念, 指出AI在传统科学领域具有巨大的发展空间^[37]. 我们期待AI与领域深度交叉, 带来新的科学研究增长点 and 研究范式的革新.

致谢 感谢无锡药明康德新药开发股份有限公司徐艳博士有关化学实验室自动化方面的有益讨论以及北京大学材料科学与工程学院许审镇特聘研究员关于计算化学方面的有益讨论.

参考文献

- 1 Turing AM. *Mind*, 1950, LIX: 433–460
- 2 Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, Liu X, Wu Y, Dong F, Qiu CW, Qiu J, Hua K, Su W, Wu J, Xu H, Han Y, Fu C, Yin Z, Liu M, Roepman R, Dietmann S, Virta M, Kengara F, Zhang Z, Zhang L, Zhao T, Dai J, Yang J, Lan L, Luo M, Liu Z, An T, Zhang B, He X, Cong S, Liu X, Zhang W, Lewis JP, Tiedje JM, Wang Q, An Z, Wang F, Zhang L, Huang T, Lu C, Cai Z, Wang F, Zhang J. *Innovation*, 2021, 2: 100179
- 3 Zheng N. *Chin J Intel Sci Technol*, 2019, 1: 1–3 (in Chinese) [郑南宁. 智能科学与技术学报, 2019, 1: 1–3]
- 4 Hornik K, Stinchcombe M, White H. *Neural Networks*, 1989, 2: 359–366
- 5 Townsend JA, Adams SE, Waudby CA, de Souza VK, Goodman JM, Murray-Rust P. *Org Biomol Chem*, 2004, 2: 3294
- 6 Murray-Rust P. *Nature*, 2008, 451: 648–651
- 7 Ley SV, Fitzpatrick DE, Ingham RJ, Myers RM. *Angew Chem Int Ed*, 2015, 54: 3449–3464
- 8 Szymkuć S, Gajewska EP, Klucznik T, Molga K, Dittwald P, Startek M, Bajczyk M, Grzybowski BA. *Angew Chem Int Ed*, 2016, 55: 5904–5937
- 9 Schneider P, Walters WP, Plowright AT, Sieroka N, Listgarten J, Goodnow Jr. RA, Fisher J, Jansen JM, Duca JS, Rush TS, Zentgraf M, Hill JE, Krutoholow E, Kohler M, Blaney J, Funatsu K, Luebkeermann C, Schneider G. *Nat Rev Drug Discov*, 2020, 19: 353–364
- 10 Mauri A, Consonni V, Todeschini R. *Handbook of Computational Chemistry*. Cham: Springer, 2017. 2065–2093
- 11 Landrum G. *RDKit: A Software Suite for Cheminformatics, Computational Chemistry, and Predictive Modeling*. New York: Academic Press, 2013
- 12 Ramsundar B. *Molecular Machine Learning with DeepChem*. Stanford: Stanford University, 2018
- 13 Yao Q, Wang M, Chen Y, Dai W, Li Y-F, Tu W-W, Yang Q, Yu Y. arXiv: [1810.13306](https://arxiv.org/abs/1810.13306)
- 14 Liu Y, Yang Q, Li Y, Zhang L, Luo S. *Chin J Organ Chem*, 2020, 40: 3812–3827 (in Chinese) [刘伊迪, 杨骐, 李遥, 张龙, 罗三中. 有机化学, 2020, 40: 3812–3827]
- 15 de Almeida AF, Moreira R, Rodrigues T. *Nat Rev Chem*, 2019, 3: 589–604
- 16 Pflüger PM, Glorius F. *Angew Chem Int Ed*, 2020, 59: 18860–18865
- 17 Baum ZJ, Yu X, Ayala PY, Zhao Y, Watkins SP, Zhou Q. *J Chem Inf Model*, 2021, 61: 3197–3212
- 18 Corey EJ, Wipke WT, Cramer Iii RD, Howe WJ. *J Am Chem Soc*, 1972, 94: 421–430
- 19 Klucznik T, Mikulak-Klucznik B, McCormack MP, Lima H, Szymkuć S, Bhowmick M, Molga K, Zhou Y, Rickershauser L, Gajewska EP, Touthkine A, Dittwald P, Startek MP, Kirkovits GJ, Roszak R, Adamski A, Sieredzińska B, Mrksich M, Trice SLJ, Grzybowski BA. *Chem*, 2018, 4: 522–532
- 20 Kowalik M, Gothard CM, Drews AM, Gothard NA, Weckiewicz A, Fuller PE, Grzybowski BA, Bishop KJM. *Angew Chem Int Ed*, 2012, 51: 7928–7932
- 21 Segler MHS, Preuss M, Waller MP. *Nature*, 2018, 555: 604–610

- 22 O'Brien M, Koos P, Browne DL, Ley SV. *Org Biomol Chem*, 2012, 10: 7031
- 23 Deadman BJ, Battilocchio C, Sliwinski E, Ley SV. *Green Chem*, 2013, 15: 2050–2055
- 24 O'Brien AG, Horváth Z, Lévesque F, Lee JW, Seidel-Morgenstern A, Seeberger PH. *Angew Chem Int Ed*, 2012, 51: 7028–7030
- 25 Ingham RJ, Battilocchio C, Fitzpatrick DE, Sliwinski E, Hawkins JM, Ley SV. *Angew Chem Int Ed*, 2015, 54: 144–148
- 26 Granda JM, Donina L, Dragone V, Long DL, Cronin L. *Nature*, 2018, 559: 377–381
- 27 Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, Hinkley T, Aragon-Camarasa G, Kitson PJ, Angelone D, Cronin L. *Science*, 2019, 363:
- 28 Burger B, Maffettone PM, Gusev VV, Aitchison CM, Bai Y, Wang X, Li X, Alston BM, Li B, Clowes R, Rankin N, Harris B, Sprick RS, Cooper AI. *Nature*, 2020, 583: 237–241
- 29 Ahneman DT, Estrada JG, Lin S, Dreher SD, Doyle AG. *Science*, 2018, 360: 186–190
- 30 Kirkpatrick J, McMorro B, Turban DHP, Gaunt AL, Spencer JS, Matthews AGDG, Obika A, Thiry L, Fortunato M, Pfau D, Castellanos LR, Petersen S, Nelson AWR, Kohli P, Mori-Sánchez P, Hassabis D, Cohen AJ. *Science*, 2021, 374: 1385–1389
- 31 Wang H, Zhang L, Han J, E W. *Comput Phys Commun*, 2018, 228: 178–184
- 32 Han Y, Ali I, Wang Z, Cai J, Wu S, Tang J, Zhang L, Ren J, Xiao R, Lu Q, Hang L, Luo H, Li J. *Phys Rep*, 2021, 934: 1–71
- 33 Dral PO. *J Phys Chem Lett*, 2020, 11: 2336–2347
- 34 Goh GB, Hodas NO, Vishnu A. *J Comput Chem*, 2017, 38: 1291–1307
- 35 Ramakrishnan R, von Lilienfeld OA. Machine learning, quantum chemistry, and chemical space. In: *Reviews in Computational Chemistry*. Weinheim: Wiley, 2017. 225–256
- 36 Service RF. *Science*, 2021, 374: 1426–1427
- 37 E W. *Notices AMS*, 2021, 68: 565–571

How to embrace the age of intelligence——taking chemistry as an example

Jinglong Lin, Qianyi Liu, Fanyang Mo*

School of Materials Science and Engineering, Peking University, Beijing 100871, China

*Corresponding author (email: fmo@pku.edu.cn)

Abstract: This perspective provides some helpful reflections on how cutting-edge achievements in artificial intelligence can be applied to academic research. We first discuss the importance of embracing the age of intelligence from a social development perspective, and describe how AI technology, as the core foundation of the fourth industrial revolution, would impact society. Taking chemical research as an example, using representative research work, we then elaborate on how machine intelligence has changed the research paradigm and accelerated scientific discoveries at various levels—basics, tools, and applications. Finally, we put forward some suggestions to the researchers who are willing to conduct interdisciplinary studies by engaging artificial intelligence in their fields of expertise.

Keywords: artificial intelligence, machine learning, automation system, chemistry, research paradigm

doi: [10.1360/SSC-2022-0028](https://doi.org/10.1360/SSC-2022-0028)