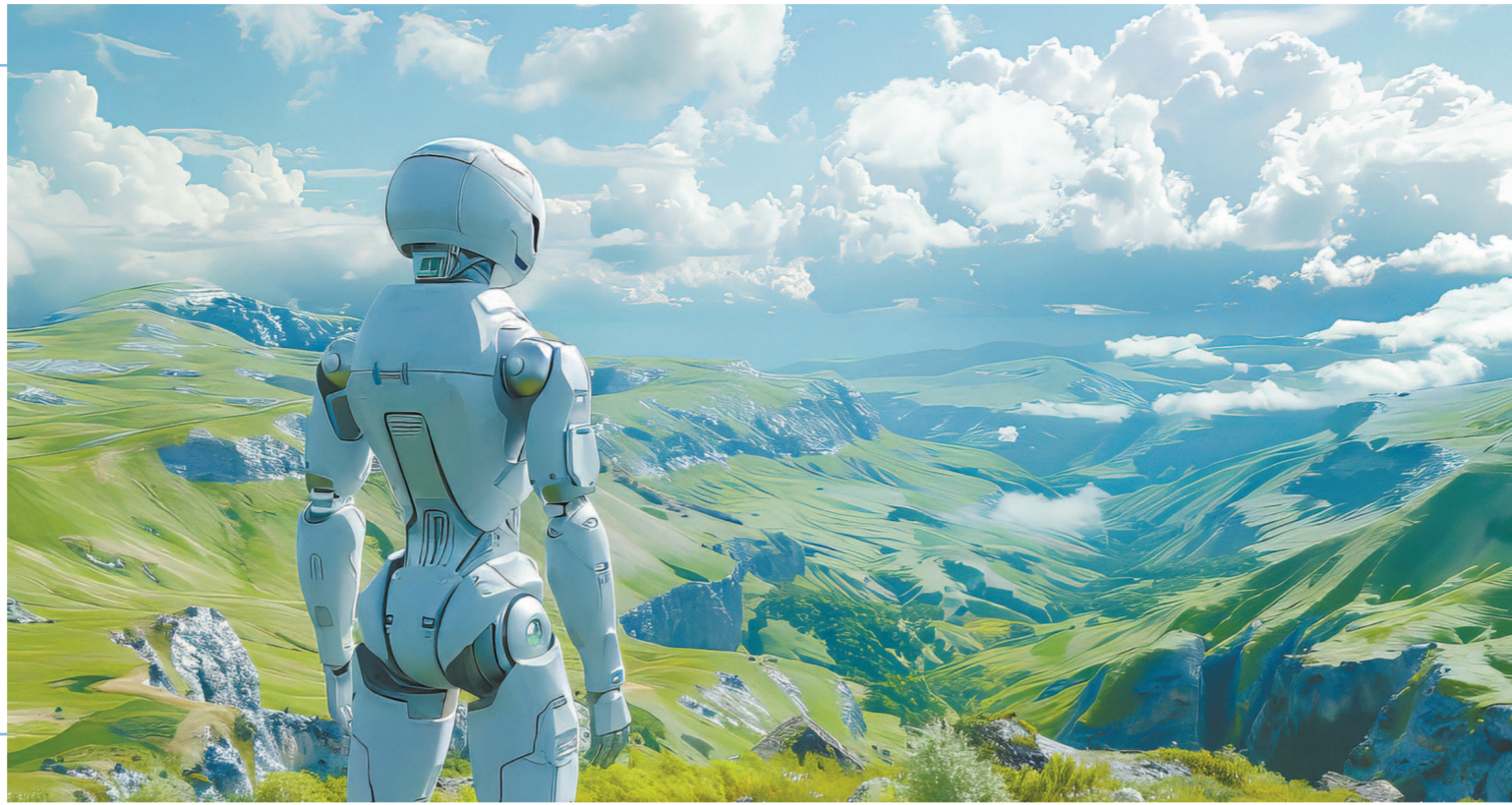


今年的诺贝尔奖像一个风向标,预示科学研究范式的变更——

AI+科学：“双螺旋引擎”开启科学智能新纪元



今年的诺贝尔奖,从结果来看,既是预想之外,又在意料之中:人工智能(AI)成了其中的关键词。

如果说诺贝尔物理学奖是物理学基础研究对AI的赋能,那么化学奖则是AI赋能科学研究、给人类世界带来改变的科技之光——如此一看,AI仿佛在今年的诺奖中“画”了一个闭环。

对高校和科研机构的研究者们来说,今年的诺奖更像一个风向标,预示着科学研究范式的变更,继而开启科学智能新纪元。

——编者

■漆远

无论从AI还是从基础科学的角度看,近日颁发的诺贝尔奖都引发了热烈的思想激荡。

10月8日,2024年诺贝尔物理学奖授予美国科学家约翰·霍普菲尔德(John Hopfield)和英国裔加拿大科学家杰弗里·欣顿(Geoffrey Hinton),以表彰他们在使用人工神经网络的机器学习方面的奠基性发现和发明。次日的诺贝尔化学奖授予三位科学家。其中,来自华盛顿大学西雅图分校的大卫·贝克(David Baker)成功完成了构建全新蛋白质这一几乎不可能完成的任务;而来自谷歌DeepMind公司的德米斯·哈萨比斯(Demis Hassabis)和约翰·江珀(John Jumper)开发了一种名为AlphaFold2的人工智能模型,这种模型解决了一个已有50年历史的难题:预测蛋白质的复杂结构。

2024年两个诺奖的颁布,标志着一个新时代的来临。AI与基础科学的深度融合,将开启AI+科学“双螺旋引擎”共振驱动的科学新范式:一方面,AI在科学研究中逐步走向核心地位,成为科研突破的新引擎,AI for Science(科学智能)将极大加速和扩展科学研究的效率和能力边界;另一方面,底层科学的第一性原理和知识将成为AI向AGI(通用人工智能)进化的根本驱动和扩展引擎,下一个AI领域的里程碑很可能来自Science for AI的启发,而两者作为“双螺旋

引擎”,将共振驱动并定义科学智能新范式“AI+科学”,带领我们进入科学发现涌现和人类文明跃迁的新纪元。

Physics for AI 是物理,也是人工智能

从学科划分的角度看,机器学习与物理学边界井然,物理学是研究真实世界的自然现象,而机器学习则属于计算机科学、数据、代码和算法的范畴。

不过,霍普菲尔德与欣顿的研究早已跨越了学科边界。

20世纪80年代,霍普菲尔德提出的“霍普菲尔德网络”,借鉴了物理学中的自旋相互作用原理,用来模拟大脑中的记忆存储与重构。欣顿则通过引入玻尔兹曼分布的概念,发展出了玻尔兹曼机,这一模型为机器学习中的深度学习生成模型奠定了基础。

从底层逻辑看,物理学的使命是研究从宇宙到粒子尺度复杂系统第一性原理。而从信息论角度看,机器学习研究的范畴也会落在物理学覆盖的领域之内,基于能量的模型(Energy based model)可以将优化问题和物理问题融合起来。

事实上,人工智能的很多关键算法都受到了物理学思维的启发。

物理学家在研究自然界时,往往通过方程和数学模型来描述复杂系统的行为。这种通过变量间的关系来揭示规律的方式,与AI中的建模非常类似。比如在经典物理学中,牛顿方程、热力学等模型都是通过数据推导出来的,而AI模型同样依赖大量数据和数学模型去学习、预测和优化。

另外,人工智能常常需要处理复杂的高维数据,这与物理学家研究多体系统、混沌系统等复杂现象时的挑战相似。比如,神经网络的复杂性和自适应性,与物理学中的自组织现象类似。在此基础上,许多AI模型借鉴了物理学的思维方式,尤其在处理复杂系统时,将物理法则与AI结合以提高模型的解释力和效率。

再举一个例子。想象在水池中滴入一滴高温水,随着时间的推移,热量向四周扩散,直到整个水池温度趋于一致。这是物理学中的热扩散过程,能量逐渐扩散并分布得更加均匀。AI的扩散模型(Diffusion model)则通过类似的过程进行数据生成。它从噪声开始,将数据逐步扩散为无序状态,类似一滴高温水的扩散过程。然后,模型通过逆向过程将无序状态逐步还原为有序数据,最终生成高质量的样本。这个“从无序到有序”的生成过程,完美模拟了热扩散过程,即通过逐步降低热量差异,使系统恢复到稳定状态。

在机器学习领域,很多学者早期都是

物理学背景。麦克斯·韦林(Max Welling)是机器学习领域的顶尖学者之一,专注于概率图模型、深度学习和贝叶斯方法。他曾是一位理论物理学家,研究量子场论,并在物理学领域积累了丰富的数学和计算技巧。这些技巧后来被应用到人工智能中,特别是在变分推断(Variational inference)和深度生成模型的研究上,由此他提出了著名的VAE算法。

受益于我在麻省理工学院的导师汤姆·雅科拉(Tommi Jaakkola,拥有理论物理学硕士和计算神经科学博士学位,于1998年加入麻省理工学院电子工程与计算机科学系担任教授),我在研究AI时也学习了统计物理学的知识。他和同在麻省理工的宇宙学家麦克斯·泰格马克(Max Tegmark)都拥有物理学背景,他们合作开发的一种全新的受物理启发的生成模型家族,统一了扩散模型和泊松流生成模型(PFGM)。这两个模型都与物理学相关,扩散模型与物理学热扩散原理相似,泊松流生成模型则是受到高维电磁理论启发。

AI for Chemistry 科学突破的新引擎

再来看看今年摘获诺贝尔化学奖的AlphaFold2。在2018年发布的AI-

phaFold基础上,由DeepMind公司开发的这款人工智能程序在2020年的蛋白质结构预测大赛(CASP)中取得了突破性进展,以接近实验水平的精度预测蛋白质的三维结构,该成就被认为是人类在21世纪取得的最重要的科学突破之一。这一成果不仅推动了生命科学的发展,还加速了针对癌症、病毒的抗生素、靶向药物和新效率的蛋白酶的研

究。AlphaFold2的成功展示了AI在基础科学研究上的巨大潜力。同获诺奖的贝克的Rosetta软件的开源性质,使得全球科研人员都能够利用这一工具进行研究。同获诺奖,其实早在科学家的意料之中。

AlphaFold2的核心是使用大量的蛋白质序列数据训练出的一个深度学习模型,该模型能够理解氨基酸序列与蛋白质结构之间的复杂关系。它首先搜索同源序列和模板,然后通过多序列比对和对特征联合嵌入来构建蛋白质结构的预测。AlphaFold2采用了一种新的输出表示和相关损失函数,实现了精确的端到端结构预测。此外,它还使用了一种新的等变注意力体系结构,通过中间损失实现预测的迭代细化,并与结构联合训练。

尽管AlphaFold2取得了巨大成功,但它仍然存在预测方面的局限性。今年5月问世的AlphaFold3结合了神经网络和变分推断方法,并引入了多种生

物物理学知识,形成了强大的结构预测统一框架,涵盖了前所未有的广度和精确度。AlphaFold系列证实了AI作为科学发现新引擎的强大力量。

Science for AI 和 AI for Science 定义一个全新的科学研究范式

在AI与科学研究的交汇点上,“Science for AI”和“AI for Science”构成了驱动科学进步的“双螺旋引擎”。这两个概念不仅代表了科学与技术之间的相互作用,也定义了一个全新的科学研究范式。

Science for AI指的是利用物理学等基础科学的原理和方法来启发和改进AI技术。

如前文所述,物理学中的热扩散原理启发了AI中的Diffusion model。在机器学习领域,很多关键算法都受到了物理学思维的启发,比如利用能量函数、玻尔兹曼分布等概念来构建模型。此外,符号计算也是Science for AI的一个重要方面,它涉及到使用计算机代数系统来执行数学符号的计算和推理。

AI for Science则是将人工智能技术应用于科学研究,以解决科学问题,包括引导科学假设的生成,自动实验和验证,并推进科学发现。

总之,无论是数字世界还是物理世界,要实现发现复杂世界的未知规律这一科学终极目标,需要Science for AI和AI for Science的“双螺旋驱动”,就类似DNA和RNA的双螺旋结构。

在Science for AI和AI for Science的共振中,数据驱动和第一性原理的融合是关键。深度学习是数据的拟合,第一性原理和知识则可以外插到缺少或者没有数据的地方,体现更好的扩展性。

当规则、知识和关键数据矛盾,可以调整知识规则,就像爱因斯坦把有限的物理数据、他自己头脑实验的合成数据(比如坐着光会看到什么样的宇宙)和黎曼几何结合起来,把牛顿力学扩展到了刻画宏观宇宙的相对论。在这个意义上,打造AI爱因斯坦是科学智能(AI+ Science)的终极目标,也是AGI的终极目标——发现未知规律,推动科学和人类的进步。在这个激动人心的科学探索和创造未来的旅程中,让我们一起同行。

(作者系复旦大学清教授、上海科学智能研究院院长)

诺奖的AI年,带来哪些启示?

■余元玺 钟博子 洪亮

今年的诺贝尔奖将人工智能(AI)推到了科学舞台的中央。这不仅是对几位杰出科学家的认可,也是对AI在科学进步中作用的肯定。这些奖项的颁发,标志着AI在科学研究中的重要地位得到了认可,预示着AI正在重塑我们的世界,尤其是在科学探索和创新的范式上掀起了新的浪潮。

人工智能势不可挡,正在改变我们的世界和科学研究的方式:AI for science, Science all in AI(科学智能与人工智能中的科学)。展望未来,AI将继续引领科学的发展,为人类社会创造更大的福祉。我们正处于一个新的时代的开端:拥抱AI,将开启无限的可能。

AI与物理学的交汇 从霍普菲尔德网络到深度学习

先简单回顾一下今年两位新晋诺贝尔物理学奖得主的贡献。

1982年,约翰·霍普菲尔德提出了著名的霍普菲尔德网络,这是一种具有自组织能力的递归神经网络。霍普菲尔德网络模拟了生物神经网络的联想记忆功能,能够通过能量最小化的原理,实现对部分缺失信息的补全和模式识别。

霍普菲尔德网络的理论基础深植于物理学,类似于统计物理学中的伊辛模型(Ising Model)用于描述磁性材料中自旋相互作用的系统,通过能量函数的最小化来确定系统的稳定状态。霍普菲尔德巧妙地将这一概念应用于神经网络,使网络状态的演化可以被视为能量函数的下降过程,最终达到稳定的记忆存储状态。霍普菲尔

德网络的出现,为神经网络的理论研究提供了坚实的物理学基础,也为后来的机器学习和人工智能发展奠定了重要的基石。

杰弗里·欣顿被誉为“深度学习之父”,他在1985年与特里·谢泽诺斯基在霍普菲尔德网络的基础上共同提出了玻尔兹曼(Boltzmann)机。这是一种基于随机性和能量函数的神经网络模型,可以通过模拟退火算法学习复杂的概率分布。玻尔兹曼机的名称来源于物理学中的玻尔兹曼分布(Boltzmann distribution),这是统计物理学中描述粒子能量分布的基本概念。玻尔兹曼机利用这种分布来定义网络中神经元状态的概率,从而实现对数据的生成和特征学习,也为后续生成模型的发展提供了思路。

欣顿在1986年与大卫·罗密尔顿和罗纳德·威廉姆斯共同推广了反向传播算法,解决了神经网络做不深的问题,使得深层网络的训练成为可能。借助受限玻尔兹曼机(Restricted Boltzmann Machine)引入无监督的逐层预训练方法,他在2006年提出深度信念网络(Deep Belief Network),解决训练深层神经网络时梯度消失的问题。这一突破为深度学习的迅猛发展奠定了极为重要的基础。

诺贝尔奖委员会将他俩的成果评价为“为机器学习革命奠基性的工作”。我们由此可以看到,统计物理学为机器学习的早期发展提供了思路,如今的深度学习也在可控核聚变、天文观测等研究方向惠及了物理学研究。

生命科学重新认识AI潜力 从Rosetta软件到AI驱动的创新

人体内拥有数万种蛋白质,已知的



本版图片:视觉中国

蛋白质数量也超过数亿。然而,在浩瀚的蛋白质序列宇宙中,这不过是冰山一角。科学家们既希望优化现有的工具蛋白,也渴望探索蛋白质宇宙中的“暗物质”,即那些尚未发现的功能蛋白。这正是蛋白质设计的核心使命——通过这一技术,科学家们可以创造出全新、自然界中从未存在的蛋白质。这些蛋白质不再受限于传统进化规则,而是完全由人类设计,具备定制化功能特征。

在蛋白质设计领域,华盛顿大学大卫·贝克教授的团队无疑站在了最前沿。过去20年中,该团队致力于开发计算驱动的蛋白质设计方法,从蛋白质结构入手,从头开始设计出形态和功能各异的蛋白质。随着时间的推移,这一领域经历了巨大的进步:从最初基于物理和统计方法的Rosetta软件,到如今依靠深度学习的AI方法,如ProteinMPNN和RFdiffusion,蛋白质设计技术不断革新,但从头设计蛋白的目标始终不变。

如今,贝克团队设计的蛋白质已具备多种功能,包括从头设计的联合疫苗

的RSV/hMPV、能够识别非天然底物的荧光素酶,以及用于药物研发的细胞因子类似物和抗体。蛋白质设计已经完成了概念验证阶段,正逐步拓展在各类生物医学领域的实际应用。

在生命科学领域,乃至整个自然科学中,AI最具影响力的应用莫过于AlphaFold。其诞生故事看似简单:一位横跨数学、物理、化学、生物、计算机的年轻博士约翰·江珀,与谷歌DeepMind CEO德米斯·哈萨比斯共同领导的顶尖跨学科团队,经过三年努力,打造出一个专门用于解决蛋白质结构预测任务的模型。

尽管AlphaFold2当时在蛋白质复合物结构预测以及药物分子、核酸、修饰蛋白等方面还存在局限,但三年后的AlphaFold3解决了这些问题。

这一突破性的成功使整个生物学界重新认识了AI的潜力,并推动了新的算法开发,如前文提到的蛋白质设计方法,也催生了生物技术公司的蓬勃发展,改变了生物学家研究蛋白质的方式。

诺贝尔奖一直以来被视为科学领域的最高荣誉,用于表彰在物理学、化学、生理学或医学等领域作出突出贡献的个人。然而,随着人工智能技术的飞速发展,AI正在深刻改变着各个学科的研究方式和方向,成为数据推断的新范式。此次诺奖将物理学奖和化学奖同时颁发给AI领域的先驱,正是对这一趋势的最好回应。

自AlphaFold问世以来,它已被广泛应用于各式各样的生物学研究中。在蛋白质结构数据库PDB中,通过实验方法解析的蛋白质结构数量已达到20万,这也成为了AlphaFold训练数据的重要来源。然而,借助AlphaFold等结构预测工具,科学家已经预测了近10亿个蛋白质的三维结构,其中大多数都具备很高的精度和质量。

AlphaFold正逐步成为生物学研究中的便捷且精准的AI工具,融入很多生物科学领域的研究当中——曾经耗费大量时间和资金才能获得的蛋白质结构,现在“点击就送”。展望未来,AlphaFold的后续版本有望解决更多复杂问题,带来更多意想不到的应用场景。

AI获得诺奖 AI正深刻改变各科学研究方向

这次诺贝尔化学奖和物理学奖都

给了AI,尤其是物理学奖直接颁发给了欣顿(一个计算机学家)还是令人非常震惊的。除了对欣顿基于物理启发的人工智能算法的开发的认可,更体现了诺奖委员会对科学范式变革的预见。

物理这门学科归根结底是探索理解这个世界的方法论。传统的物理方法论或者说占统治地位的方法论是搞清楚底层机理,通过不断地叠加近似来解读复杂事物。但AI反其道行之,是基于数据推断,端到端给出预测。这也能解决问题,就像诺贝尔化学奖的“蛋白质结构预测”,最开始研究这个问题的统计物理学家,他们基于物理计算来预测蛋白结构,但深度学习在这个问题上获得了完胜传统物理计算的精度。

如果科学问题本身就是如何预测一个给定序列蛋白的三维结构,那么显然我们的物理底层知识和方程是不够的,但基于大数据的AI方法是能解决这个科学问题的。基于数据的推断就是在这个科学问题上更好的方法论。这无疑是对传统物理方法论的一种冲击。这次诺奖的颁发显示了物理学的包容,我们期待看到更多基于数据推断的工具在物理学涌现,帮助我们找到更好的超导材料,帮我们找到更优的聚变控制方法等等。

诺贝尔奖一直以来被视为科学领域的最高荣誉,用于表彰在物理学、化学、生理学或医学等领域作出突出贡献的个人。然而,随着人工智能技术的飞速发展,AI正在深刻改变着各个学科的研究方式和方向,成为数据推断的新范式。此次诺奖将物理学奖和化学奖同时颁发给AI领域的先驱,正是对这一趋势的最好回应。

(作者单位:上海交通大学自然科学研究院)