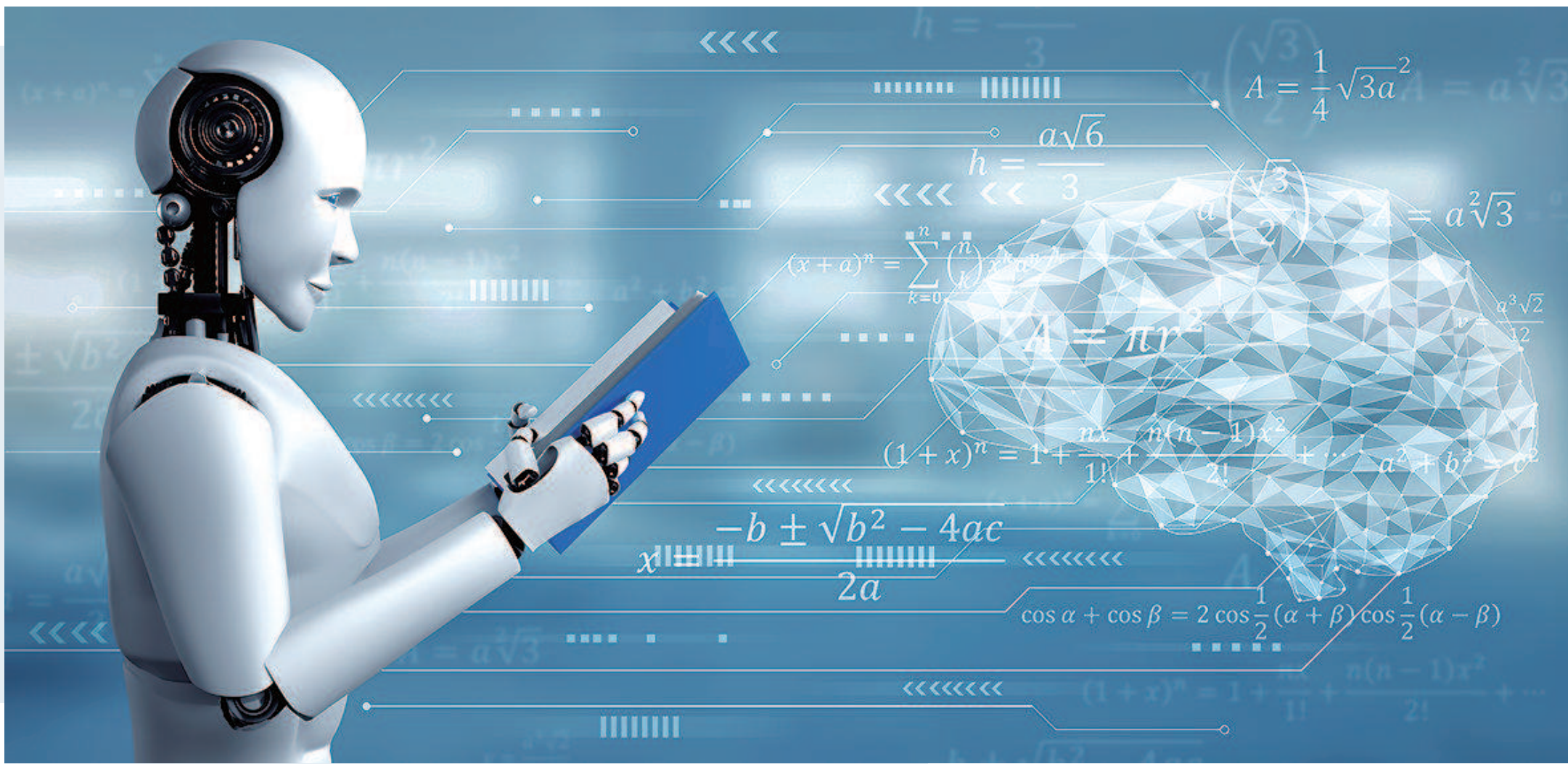


AI正在习得人类水平的数学推理能力,表现直逼国际奥数优秀选手

“征服”数学, AI 是否有能力“回答世界”



最近,专门为人工智能(AI)设立的AI国际奥林匹克数学竞赛(IMO)即将进入尾声,其结果将随今年7月于英国巴斯举行的65届IMO大会同步揭晓。这项赛事的目的是推动发展大语言模型的数学推理能力,训练出更高数学水平的新一代AI模型。

纯数学领域中的重大发现是推理和创造力的灵感结晶,往往意味着人类智慧极限的突破。迄今为止,解决或协助解决高等数学难题,仍是AI领域一项极其复杂的多维度重大挑战。随着AI越来越多出现在数学领域,这是否预示着AI开始具有人类水平的推理能力?它是否会为通用人工智能(AGI)铺平道路?

猜想提出约60年后,英国伦敦数学科学研究所的何杨辉教授和他的同事们想试试AI能否提出新的见解。于是,他们让AI神经网络在一个包含数百万条椭圆曲线的数据库中搜索曲线的某个特征,结果AI没有让人失望。在对其预测进行更详细的分析后,何杨辉和同事们发现了一个惊人模式,并将其绘制在图表上——它很像棕鸟成群结队时产生的流体形状。

这项工作吸引了更多研究人员,他们后来找到描述这种模式的方程,并发现其他重要的数学函数中也会出现相似模式。何杨辉说:“AI能够让我们去尝试一些东西,这些东西如果只凭我们自己的直觉,可能不会去尝试。”

一年后,数学家与DeepMind合作,利用机器学习算法,在结组理论和表示论这两个数学分支的大型数据库中,也有了类似的发现。与此同时,该算法还发现了数据库中一些潜在的有趣关系。进一步分析后,研究人员还找到了一些新的猜想。

戴维斯表示,尽管数学家们对机器学习是否对数学研究有实际帮助抱有极大怀疑,但在发现复杂数据集中的新模式时,AI显然可完成人类数学家无法完成的任务,尽管它的工作原理有时“有些透明”。

威廉森感觉,与这些系统一起工作,就像有一个不能很好沟通的合作者,“在与AI合作时,我总有一种感觉,那就是它在某种程度上知道答案,但却无法告诉我为什么”。

不过,AI并不总是那么不友好。2022年,OpenAI推出了ChatGPT,这是一种基于大语言模型(LLM)的AI聊天机器人。它们通过吸收大量文本,能对各种提示做出流暢的、类似人类的回应。对于ChatGPT的出现,数学家们并不感到惊讶,但更想知道,它的底层架构(一种被称为Transformer的神经网络)是否能被改造成一种更懂数学的工具。他们最大的担心是,它虽然有生成文本的能力,却在过滤错误答案或发现自己的错误方面无能为力。

因此,DeepMind的研究人员在构建FunSearch系统时,创建了一个LLM,以计算机程序的形式编写数学问题的解决方案,并将其与一个按性能对程序进行排名的系统相结合——它会将效果最好的程序反馈给LLM,LLM则据此反复改进版本,直到有新的发现。与DeepMind合作开发FunSearch的艾伦伯格觉得,这个AI系统比之前想象的要好用很多。

此后,另一支DeepMind团队又用同样的方式建立起一个名为“阿尔法几何”(AlphaGeometry)的数学大模型,解决了国际数学奥林匹克(IMO)中的复杂几何问题。解答国际数学奥林匹克的数学题,需要强大的头脑创造力,而AI历来在解答此类问题中的表现不佳。但“阿尔法几何”经过针对性训练后,在逻辑检



黄昏时,棕鸟一起飞翔形成的曲线。

查系统的加持下,其几何学的解题表现几乎与最优秀的人类选手不相上下。

证明“形式化” AI正加速学习数学思维

一旦AI能够阅读人类数学库中的语料,并掌握人类使用这些数学库的方法,或许它们将具备证明一些数学问题的能力

不少数学家认为,在人工智能的帮助下,他们可能会获得更多的新发现。有些人甚至认为,这也可能是数学创造力的最初萌芽。威廉森猜想,可能这就是创造力,“数学家就像诗人、音乐家或小说家,只是拥有非常好的创造力和非常敏锐的评估能力”。

不过,这些最新突破也提出了一个更诱人的可能性:如果能再进一步,用大量研究级别的数学语料库来训练一个类似“阿尔法几何”的系统,那么它就有可能开始为现有的猜想找到证明,并提出全新证明和猜想,而不需要人类输入提示——这可能就相当于人类水平的推理和理解了。

但问题在于,绝大多数尖端数学语料都无法被计算机读取——让计算机读懂数学的过程,既棘手又耗时。许多数学家宁愿把时间花在数学本身的研究上。不过,这个方向正吸引越来越多的关注者,尤其当计算机辅助和计算机检验在现代数学的证明过程中越来越重要。一位日本科学家对数论问题说:“证明是数学的中心,这是数学学科的真谛”。

当安德鲁·怀尔斯在1993年最终证明著名的费马大定理后,他还必须将其与数论和代数几何领域的前沿成果相融合,从而推动进一步的研究。但是,长达数百页的证明也日益成为数学家的“眼中钉”。一位日本科学家对数论问题说:“abc猜想的证明长达500页,面对密密麻麻的符号,很少有人能够读懂”。

这时,AI就可以提供帮助。它们不仅能快速检查证明步骤是否正确,还能帮助数学家编写证明。尽管目前AI能做的还远远落后于当前的数学研究,但这种情况正在改变——如果能将足够多的证明“形式化”,让人工智能可以访问它们,那么AI工程师就能以此为语料训练这些系统,让它们比人类更快,甚至更有效地生成猜想和证明。这样,人工智能就能通过数学思维学会推理。

为证明“凝聚数学”理论中的一个关键结果,2018年菲尔兹奖得主、德国波恩大学的彼得·舒尔茨发起了“液态张量实验”,并用数学软件Lean对结果进行形式化。在整个团队的努力下,他们于2021年5月完成了一个关键定理的形式化,又于2022年7月得到了完整定理的形式化。由于这个项目大幅扩充了Lean的

数学库,使其他的Lean形式化项目效率得到了提升,越来越多的顶尖数学家也将他们的研究纳入到Lean中。去年,他们证明了一个重要猜想,并在不到一个月的时间内将其形式化。这样一来,数学领域的书面证明和形式化证明之间的时间差从数年缩短到数周。

英国伦敦帝国理工学院的凯文·巴萨德认为,这对AI的发展意义重大,“一旦AI能够阅读人类数学库中的语料,并掌握人类是如何使用这些数学库的,或许它们将具备证明一些数学问题的能力”。

离AGI更近 “AI数学家”呼之欲出

如果AI习得人类水平的数学推理,人类将向AGI迈出重要一步。不过,推理仅仅只是真正的AGI所具备的诸多技能之一

不过,当AI可以做到这一步,它们能够发明新的概念吗?换言之,它们是否能够在没有人类输入的情况下提出新的数学见解——那些对人类来说,都十分罕见的灵光一现?

对许多数学家来说,答案是否定的,或者至少还需要几十年时间才可能实现。“数学家所依赖的理解和推理过程是人类独有的,AI迄今所做的一切都没能改变这一点。”巴萨德说,“到目前为止,我们还没有看到这方面的证据”。

即便如此,自2022年菲尔兹奖研讨会以来,许多世界顶尖数学家一直在讨论AI将在多大程度上改变他们的工作,以及他们可能如何适应。

另一方面,能够解决高等数学问题,对整个AI的进步意味着什么,则是另一问题,而这可能对所有人都产生影响。有一部分人认为,“超人AI数学家”的推理,会让AGI离我们更近。

这其中的逻辑非常清楚:如果数学是人类推理的最高形式,如果AI能够像人类最优秀的数学家一样完成数学推理,甚至做得更好,那么这将代表着人类向AGI迈出了重要一步。不过,戴维斯指出,真正的AGI还需要具备更广泛的技术,推理仅仅是其中之一。

克里斯蒂安·塞格迪是一名计算机科学家,曾致力于用AI进行数学运算和自动形式化。他相信,“超人AI数学家”到2026年就会出现,“一旦拥有了推理这种新技能,AI不仅可以拥有人类的直觉,而且还将大大超越”。

如果塞格迪是对的,那么在通往人工智能的道路上,机器数学家可能会带着人类走得比许多人愿意承认的更远。哪怕退一步来说,数学高级别难题所带来的挑战也会将AI的发展推向新高度。

威廉森说:“数学在描述宇宙的许多方面有着非凡的能力。”假设我们有一个系统,它普遍能够回答高难度的数学问题。那么这样一个系统也应该能够回答我们这个世界的难题。

张磊/编译

由于人们担心人工智能(AI)驱动的聊天机器人可能会泄露私人数据,计算机科学家们正努力教机器学习模型如何遗忘。虽然这非常困难,但“机器不学习”的解决方案已经开始出现。这项工作除了能解除人们对隐私和错误信息的担忧外,还能起到一个至关重要的作用——如果我们真的想让AI像人类一样学习和思考,可能需要让它们学会遗忘。

被遗忘权

由大型语言模型(LLM)支撑的新一代AI聊天机器人能根据用户的提示生成文本。这些模型是在大量数据的基础上训练出来的,其中大部分数据抓取自互联网上的公开信息。由此,它们学会了预测句子中最有可能出现的下一个词,从而流利地回答每一个问题。

与此同时,这也意味着当大模型学习到某些东西时,就无法取消学习。澳大利亚最大的国家科研机构联邦科学与工业研究组织的AI研究员兼工程师大卫·说,大语言模型根据汇总数据生成回复,因此它们无法像谷歌等搜索引擎那样轻松地遗忘或“删除”特定信息,甚至个人也无法准确追踪AI应用程序对自己的了解程度。

这在隐私方面造成一个重大问题。2018年生效的《欧盟通用数据保护条例》(GDPR)规定用户享有“被遗忘权”,即当用户要求企业清除和停止传播个人的隐私信息时,企业应当采取合理措施,及时销毁用户的个人数据,否则将面临巨额处罚。

瑞士苏黎世联邦理工学院的计算机科学家弗洛里安·特拉梅尔认为,大模型公司必须想办法解决这个问题,尤其是当这些公司开始用更敏感的信息对大模型进行训练。

更令人担忧的是,由AI驱动的聊天机器人本身也很容易受到攻击。已有研究证明,一些黑客技术可以让聊天机器人在用户设备上远程运行代码,或要求用户提供银行账户信息。

诱导“失忆”

机器学习模型对数据有很强的依赖性,尤其像ChatGPT这样拥有海量训练数据和千亿级参数的大型AI系统,面对用户频繁的删除请求,重新训练整个模型会消耗巨额资源。那么,是否能在不用重新训练模型的情况下,移除或至少屏蔽特定信息呢?

2019年,加拿大多伦多大学的尼古拉斯·帕贝尔诺及其同事提出了一种被称为SISA的方法——它将数据集分割成不同小块,分别对每个小块进行模型训练,并像游戏一样保存节点进度,再合并结果。当遇到取消学习请求时,它可以返回节点,切断相关数据块,再从那里开始重新训练。经测试证明,这种方法大大加快了重新训练速度。

由于删除数据会严重影响机器学习模型的性能,一些团队选择了隐藏或遮蔽相关数据的策略。例如,微软和美国俄亥俄州立大学的研究人员在用于训练模型的数据中引入了噪音,使大模型随后输出的信息由数据中更泛化的模式,而不是特定的、可识别的例子所决定。这在理论上保证了大模型不会透露训练数据中的个人信息。

但是,这种泛化往往会在一定程度上削弱AI的学习能力。为了规避这个问题,韩国科学家尝试了一种“知识解除学习”的方法,其目的是切断某项数据对算法的影响,而不是完全删除它,这样聊天机器人就永远不会引用它。目前,这一方法是该领域最有前途的方法之一,因为它能在更短时间内使用更少的计算资源完成工作。

为了激励研究人员提出更高效的解决方案,谷歌公司还组织了一场竞赛。这不仅表明这项挑战的重要性,也意味着或许我们将找到更多好方法,让新一代大模型拥有“遗忘”的能力。

选择性记忆

英国伦敦政治经济学院研究人工智能的哲学家阿里·博伊尔认为,虽然人类的遗忘倾向通常被视为认知缺陷,但有时也是有益的——通过遗忘,人们可以更有效地检索有用的记忆。

AI系统可能也是如此。2017年,谷歌DeepMind公司的研究人员开发了一种可以玩多种游戏的AI系统。通过存储并回忆自己玩游戏的记忆,它能更有效地归纳出知识。当研究人员改进了模型,使其优先存储和回忆一部分事件而忘记其他数据时,系统性能得到了提高。

这意味着,选择性遗忘可以提高人工智能的性能,而关键是在记忆过多和过少之间找到适当的平衡点。如果AI研究人员的最终目标是建立能像人类一样学习和思考的系统(这也是该领域的最初目标之一),那么他们就必须设计出能选择性遗忘的系统。在博伊尔看来,遗忘不是设计缺陷,而是一个高效、运作良好的记忆系统的必要特征。

如何让「机器不学习」?科学家正寻找解决方案

学会「遗忘」将使新一代AI更出色

章珂/编译

数学一直被视为人类智力的“天花板”,数学家则不断在挑战人类智力的极限。同时,数学难题也一直是大型模型难以“征服”的领域之一。

数学是推理的语言。长期以来,人们一直在为这类难题的复杂推理而苦恼。近年来,越来越多的数学家认为,这个领域正进入一个崭新时代:机器智能的数学能力正不断提升,在一些任务中的表现甚至超过人类数学家。

去年底,谷歌DeepMind公司发布的数学大模型FunSearch,针对人类历史上的诸多数学难题给出了新的解法。这是大模型第一次对数学领域中具有挑战性的开放性问题给出新的发现或解法,也将为解决悬而未决的数学难题开辟新的途径。

DeepMind公司亚历克斯·戴维斯和数学家一起,建立了用于协助数学研究的机器学习框架。他认为数学是推理的语言,“如果大模型能学会流利地使用数学,我们将创造出一个非常值得信赖的智能伙伴”。

要理解AI从事复杂数学的意义,就必须了解人类数学家的工作。与应用数学相比,纯数学是在不考虑实际目的的情况下进行的。美国威斯康星大学麦迪逊分校的乔丹·艾伦伯格认为,从根本上说,数学研究的目的是通过研究抽象的对象和概念,如数字、代数和几何,找到基本的关系和原理。研究纯数学需要复杂的推理、直觉和创造力。澳大利亚悉尼大学的乔治·威廉森说:“推理是数学过程的精髓。”

事实上,很难找到一门比数学更能体现推理的人类思维学科了。这就解释了为什么数学推理长期以来被视为一种即使是最强大的计算机也无法企及的东西。尽管数学家使用计算机已有数十年之久,但仅限于比较简单粗暴的计算。说到人工智能,即使是大获成功的深度学习神经网络,也无法在数学推理方面取得多大进展。不过,最近有迹象表明,最新的AI大模型可能会改变这种状况。

数学大模型

表现直逼人类最优秀选手

尽管AI的工作原理“有些透明”,但它们在发现复杂数据集中的新模式时,可以完成人类数学家难以完成的任务

伯奇和斯温纳顿-戴尔猜想是数论中一个悬而未决的重大问题。这个猜想是美国克雷数学研究所在21世纪之交选出的七个“千年大奖问题”之一。这些问题都被认为是极具挑战性,悬赏奖金高达100万美元。

2019年,在伯奇和斯温纳顿-戴尔



本版图片均视觉中国