

Nature 速递：利用大模型程序搜索产生数学发现



集智俱乐部 | + 关注

2023-12-20 11:13 江苏 来源：澎湃新闻·湃客

字号▼

原创 集智编辑部 集智俱乐部



关键词：大语言模型，数学发现，组合优化，计算机搜索，可解释性

nature

[Explore content](#) [About the journal](#) [Publish with us](#)[nature](#) > [articles](#) > [article](#)Article | [Published: 14 December 2023](#)

Mathematical discoveries from program search with large language models

[Bernardino Romera-Paredes](#), [Mohammadamin Barekatin](#), [Alexander Novikov](#), [Matej Balog](#), [M. Pawan Kumar](#), [Emilien Dupont](#), [Francisco J. R. Ruiz](#), [Jordan S. Ellenberg](#), [Pengming Wang](#), [Omar Fawzi](#), [Pushmeet Kohli](#) & [Alhussein Fawzi](#) [Nature](#) (2023) | [Cite this article](#)46k Accesses | 841 Altmetric | [Metrics](#)

论文题目：Mathematical discoveries from program search with large language models

论文来源：Nature

论文地址：<https://www.nature.com/articles/s41586-023-06924-6>斑图地址：<https://pattern.swarma.org/paper/21c6c47a-9ad5-11ee-bc79-0242ac17000e>澎湃
中晚报

江苏多地停课调休

美私企月球着陆器“奥月

动画模拟 | 广州沥心沙

[查看更多](#)澎湃
财经

嘉实沪深300ETF成深交

富时中国A50指数期货

[查看更多](#)澎湃
填空

人类对食物加工的第一

“羽之神勇，千古无二”

[开始答题](#)澎湃
热榜

1 外甥正月理发舅舅遇

2 郑州一海洋馆潜水员

3 广州南沙沥心沙大桥

4 直播 | 广州南沙沥心

5 浙大邵逸夫医院：娃

6 浙大邵逸夫医院：宗

7 民政部：全流程

8 广州沥心沙大桥事故

11. 以研究开发的 FunSearch 基于大语言模型，可以超越数学家和计算机科学家的认知，发现针对组合优化问题的全新且性能更好的解决方案。组合优化是一个研究如何计算包含有限多个对象的集合的可能排列的数学领域。

FunSearch 的工作原理是将预先训练好的 LLM 与自动评估器配对，前者的目标是以计算机代码的形式提供创造性的解决方案，后者则负责防止出现幻觉和错误的想法。通过这两个组件之间的来回迭代，初始解决方案“进化”为新知识。该系统搜索用计算机代码编写的“函数”，因此得名 FunSearch (searching in the function space 的简称，即在函数空间中搜索)。

FunSearch 使用由 LLM 支持的进化方法，促进和发展得分最高的创意。这些想法以计算机程序的形式表达，因此可以自动运行和评估。首先，用户以代码的形式对问题进行描述。该描述包括一个用于评估程序的程序和一个用于初始化程序池的种子程序。FunSearch 会自动向经过专门训练的 LLM 发出请求，要求它编写简短的计算机程序，以生成特定数学问题的解决方案。

然后，系统会快速检查这些解决方案是否优于已知解决方案。如果不是，系统就会向 LLM 提供反馈，以便它在下一轮中加以改进。LLM 在此基础上创造性地生成新程序，并自动对其进行评估。最好的程序会被添加回现有程序库，形成一个自我完善的循环。FunSearch 使用的是谷歌的 PaLM 2，但它也兼容其他经过代码训练的 LLM。

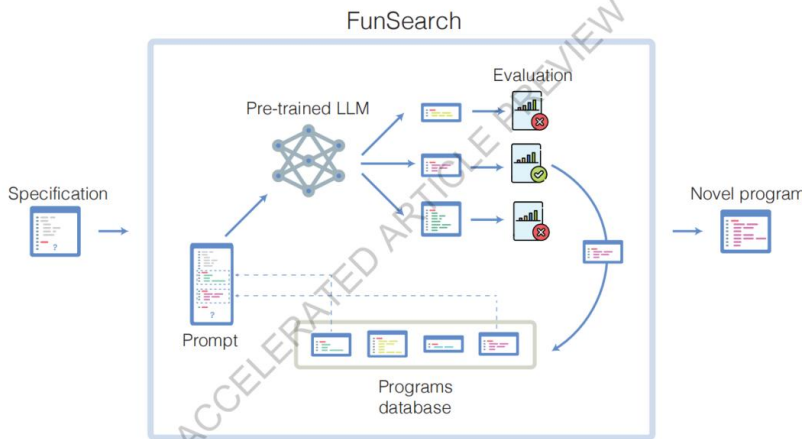


图1：FunSearch的工作模式示意图

用于测试FunSearch的问题由遗传学家Marsha Falco在 20 世纪 70 年代发明的游戏 Set (the cap set problem) 演化而来，知名数学家陶哲轩 (Terence Tao) 曾说这是他最喜欢的开放性问题。Set 卡组包含 81 张牌。每张牌上都有一个、两个或三个在颜色、形状和阴影上完全相同的符号，每个符号都有三种可能的选择。这些可能性加起来就是 $3 \times 3 \times 3 \times 3 = 81$ 。玩家必须翻开扑克牌，找出三张牌的特殊组合，这种组合被称为“套牌”。FunSearch 通过生成满足游戏所有要求的牌组，改进了 $n = 8$ 的下限。为了证明 FunSearch 的实用性，研究中还利用它为“bin packing”问题发现了优于现有算法的程序，该问题的应用无处不在，比如提高数据中心的效率。

	OR1	OR2	OR3	OR4	Weibull 5k	Weibull 10k	Weibull 100k
First Fit	6.42%	6.45%	5.74%	5.23%	4.23%	4.20%	4.00%
Best Fit	5.81%	6.06%	5.37%	4.94%	3.98%	3.90%	3.79%
FunSearch	5.30%	4.19%	3.11%	2.47%	0.68%	0.32%	0.03%

表1：Bin 搜索问题的理论最优解，已知解及 FunSearch 得到的解

光，直升机在现场救援

12 李一舟AI课程小程序暂停服务，个人视频号被禁

13 孩子误吞“水精灵”

14 广州沥心沙大桥事故：撞击桥墩的系5000吨货船

15 被船撞断的广州南沙桥，去年刚完成新型防撞

16 广州沥心沙大桥被撞：交车上有1人，救援正在进

17 3D动画模拟 | 广州南桥被船只撞断

18 沥心沙大桥事故坠桥事：他去接岛上孩子上学，个”

19 娃哈哈声明：宗庆后因病接受治疗，集团业务正

20 电话卡自动续费咪咕服务，八旬老人被“偷偷”



扫码下载
澎湃新闻客户端

关于澎湃 加入澎湃 联系我

法律声明 隐私政策 澎湃帮

沪ICP备14003370号
沪公网安备31010602000299
互联网新闻信息服务许可证：沪
增值电信业务经营许可证：沪
© 2014-2024 上海东方报业



反馈



尽管发现新的数学知识本身意义重大，但与传统的计算机搜索技术相比，FunSearch 方法还提供了额外的优势。这是因为，FunSearch 并不是一个仅仅生成问题解决方案的黑盒子，相反，它生成的程序会描述这些解决方案是如何得出的。这种“展示你的工作”（show-your-working）的方法通常是科学家的工作方式，通过产生新发现或新现象的过程来解释它们。

FunSearch 喜欢用高度紧凑的程序来寻找解决方案，即具有较低 Kolmogorov 复杂性的解决方案。简短的程序可以描述非常庞大的对象，从而使 FunSearch 能够扩展到大型的问题。此外，这也让研究人员更容易理解 FunSearch 的程序输出。

更重要的是，FunSearch 程序的这种可解释性可以为研究人员提供可操作的见解。例如，在使用 FunSearch 的过程中注意到，它的一些高分输出的代码中存在着有趣的对称性。这让研究者对问题有新的认识，利用这种认识，可完善引入 FunSearch 的问题，从而找到更好的解决方案。这是人类与 FunSearch 在数学领域的许多问题上进行合作的典范。

FunSearch 证明，如果我们能防止 LLM 产生幻觉，那么大模型的力量不仅可以用来产生新的数学发现，还可以用来揭示重要现实世界问题的潜在解决方案。对于科学和工业领域的许多问题——无论是长期存在的问题还是新问题——使用 LLM 驱动的方法生成有效和量身定制的算法将成为常见做法。

编译 | 郭瑞东

原标题：《Nature 速递：利用大模型程序搜索产生数学发现》

[阅读原文](#)

特别声明

本文为澎湃号作者或机构在澎湃新闻上传并发布，仅代表该作者或机构观点，不代表澎湃新闻的观点或立场，澎湃新闻仅提供信息发布平台。申请澎湃号请用电脑访问<http://renzheng.thepaper.cn>。

0 收藏

我要举报

相关推荐



民政部通报福利彩票2.2亿大奖情况：全流程严格按照规定程序进行

[9:00am] 起床
[9:30am] 搭乘Waymo前往Mission SF，途中在Tartine买个牛油果吐司
[9:45am] 阅读OpenAI范儿，向优化之神致敬，学习《The Bitter Lesson》（强化学习之父Rich Sutton著）
[10:00am] 在Google Meet上开会，讨论如何在更多数据上训练更大的模型
[11:00am] 跑代码，在更多数据上训练更大的模型，搭档是Hyung Won Chung
[12:00pm] 去食堂吃午饭（纯素且无麸质）
[1:00pm] 真正开始在大量数据上训练大模型
[2:00pm] 处理基础训练问题（读者：由于被铲除了得为晚餐从master分支拉代码？）
[3:00pm] 监督模型训练进展，玩Sora
[4:00pm] 给刚训练的大模型上展示工程
[4:30pm] 短暂休息，坐在牛油果椅子上，欣赏Gemini Ultra的特别定制有多棒

阿里大模型员工也自曝了996作息表



中国参与国际法院巴勒斯坦被占领土咨询意见案口头程序

评论 (0)

反馈



0 / 800 发表

已经到底了

