

Competitive Programming with Large Reasoning Models (大型推理模型在编程竞赛中的应用)

📅 2025年2月17日 ⌚ 1 分钟阅读

#Competitive Programming

#Large Reasoning Models

#AI

Competitive Programming with Large Reasoning Models (大型推理模型在编程竞赛中的应用)

Competitive Programming with Large Reasoning Models (大型推理模型在编程竞赛中的应用)

一、研究背景与模型介绍

研究背景：竞争性编程是评估推理和编码能力的重要基准，复杂算法问题求解需要高级计算思维与解题技能，且便于客观评分，是测试 AI 系统推理能力的理想平台。近年来，大型语言模型（LLMs）在程序合成方面取得进展，模型规模与性能呈正相关，强化学习（RL）和特定任务微调可提升准确性，推动了大型推理模型（LRMs）的发展。

OpenAI o1：经 RL 训练处理复杂推理任务，通过生成内部思维链并借助 RL 优化，能分解任务、纠错和探索多种解法，还可利用外部工具在安全环境中编写与执行代码以验证输出，在 CodeForces 基准测试中表现优异，超越非推理 LLM，其进一步训练后的版本取得更高评分。

目录

文章信息

字数

阅读时间

发布时间

更新时间

标签

#Competitive Programming

#Large Reasoning Models

#AI

OpenAI o1-ioi

训练与策略：基于 o1 针对 2024 年国际信息学奥林匹克竞赛 (IOI) 进行扩展 RL 训练，聚焦编程任务，强化 C++ 代码生成与运行时检查，并按 IOI 提交格式输出。测试时将问题拆分为子任务，为每个子任务从模型采样 10,000 个解决方案，经聚类 and 重排序筛选提交。

性能表现：在 CodeForces 模拟竞赛中，经多种优化策略后评分显著提升；在 IOI 2024 现场竞赛中，按规则生成和筛选提交方案，取得 49 百分位成绩，放松提交限制后可达金牌水平，证明领域特定优化可提升竞赛编程效果。

OpenAI o3：探索仅依靠 RL 训练的极限，无需人工设计测试时策略。在 CodeForces 基准测试中，相比 o1 和 o1-ioi 有大幅性能提升，能解决更广泛复杂问题，其思维链更具洞察力，可自动验证和改进解决方案；在 IOI 2024 基准测试中，采用不同采样和选择策略，在 50 次提交限制下超越 o1-ioi 并达金牌分数，表明大规模 RL 训练可实现先进编码和推理性能。

二、软件工程评估

HackerRank Astra 数据集：含 65 个项目导向编程挑战，模拟真实开发场景，无公开测试用例。评估显示，思维链推理提升了模型性能，o1 及其改进版本在 pass@1 和平均得分指标上表现优于对比模型，证明其在复杂行业相关软件开发任务中的有效性。

SWE - Bench Verified 数据集：OpenAI 准备的经过人工验证的子集，能可靠评估模型解决实际软件问题能力。结果表明，强化学习增强了模型推理能力，o3 相比 o1 有更大提升，说明推理技能在软件工程等实际任务中具有适用性。

三、研究结论

通过 o 系列大型推理模型实验，证明思维链推理在编程任务中效果显著，从竞赛编程到软件工程挑战均有体现。增加 RL 训练和测试时计算资源可提升模型性能，使其接近人类顶尖水平，有望在多领域解锁新应用场景。

论文10问解读论文

1. 论文的研究背景是什么？

论文研究了大规模推理模型（Large Reasoning Models, LRMs）在复杂代码生成与推理任务中的表现，尤其是在竞争性编程（如IOI和CodeForces）中的应用。作者指出，竞争性编程是评估AI推理与编程能力的理想场景，因为这些任务需要复杂的算法思维，并且可以通过客观评分进行量化。

2. 论文的研究目标是什么？

作者希望比较两种推理方法的优劣：

领域特定的手工设计方法（如o1-ioi模型，专为IOI优化）。

通用的强化学习方法（如o3模型，通过大规模强化学习训练，无需手工设计策略）。

目标是评估是否可以通过通用的强化学习方法超越手工设计的领域特定方法。

3. 有哪些主要研究问题？

手工设计的领域特定推理策略（如o1-ioi）与通用强化学习策略（如o3）的表现差异是什么？

增加模型规模和强化学习训练是否能自然涌现复杂的推理能力？

通用推理模型是否可以在没有人为干预的情况下，在竞争性编程和其他推理任务中达到甚至超越领域专用模型的性能？

4. 提出的模型或方法是什么？

论文提出并比较了以下模型：

o1模型：一个早期的大规模推理模型，使用强化学习提升推理能力。

o1-ioi模型：o1的改进版，专为2024年国际信息学奥林匹克（IOI）设计，结合了手工设计的推理策略（如任务分解和聚类）。

o3模型：一个更通用的强化学习模型，不依赖手工设计的推理策略，而是通过强化学习涌现出复杂的推理能力。

5. 使用了哪些数据集和评估基准？

CodeForces：一个全球性的编程竞赛平台，提供实时竞赛环境。

IOI 2024：国际信息学奥林匹克竞赛，测试模型在高难度算法问题上的表现。

HackerRank Astra 和 **SWE-Bench Verified**：模拟真实世界的软件工程任务，测试模型在复杂、多文件、长上下文场景下的表现。

6. 实验结果如何？

CodeForces：

o1模型达到1673分（89%分位），o1-ioi达到2214分（98%分位），而o3模型达到2724分（99.8%分位），接近顶级人类选手。

IOI 2024：

o1-ioi在正式比赛中得分213分（49%分位），在放宽限制下得分362.14分（达到金牌标准）。

o3模型在相同限制下得分395.64分，超越o1-ioi，并且无需手工设计推理策略。

软件工程任务：

在HackerRank Astra数据集上，o3模型的任务通过率（Pass@1）达到了71.7%，显著高于o1的63.92%。

在SWE-Bench Verified数据集上，o3模型解决了71.7%的任务，比o1模型高出22.8%。

7. 论文的主要发现是什么？

强化学习的优势：与手工设计的推理策略相比，通用强化学习方法（如o3）能够自然涌现复杂的推理能力，且表现更优。

领域专用 vs 通用模型：虽然领域专用模型（如o1-ioi）在特定任务上表现良好，但通用模型（如o3）在不依赖手工设计的情况下，表现更强。

推理能力的扩展性：通过增加模型规模和强化学习训练，o3模型在多个任务中展现了强大的推理能力，接近甚至超越人类顶尖选手。

8. 论文的创新点是什么？

全面比较：首次系统性比较了领域专用推理策略与通用强化学习策略在竞争性编程和实际软件工程任务中的表现。

涌现推理能力：展示了通过强化学习训练，模型可以在没有人干预的情况下自然涌现复杂推理能力。

跨领域应用：研究不仅限于编程竞赛，还扩展到实际的软件工作任务，证明了推理能力的广泛适用性。

9. 论文的局限性是什么？

人类对比：虽然o3模型接近顶级人类选手，但仍未完全超越顶尖人类选手的解决率（CodeForces顶级选手解决率超过85%，o3为70%左右）。

训练成本：o3模型的训练需要大量计算资源，可能难以复制。

特殊场景表现：在某些特定任务中，手工设计的策略仍可能具有优势。

10. 对未来研究的启示或建议是什么？

进一步扩展模型规模：探索更大规模的通用模型，进一步提升推理能力。

优化训练效率：降低强化学习训练的计算成本，使其更具可复制性。

跨领域应用：将o3模型的推理能力扩展到科学研究、数学证明等其他复杂任务领域。

自动化手工策略：研究如何通过强化学习自动生成类似o1-ioi的手工推理策略，以结合两者的优势。

脑洞建议：

“AI编程竞赛联盟”：组织一个由顶尖AI模型（如o3）组成的编程竞赛团队，与人类顶尖选手进行实时对抗，研究AI与人类协作的潜力。

“AI自我进化平台”：开发一个平台，让AI模型通过不断参与实际编程任务（如开源项目）进行自我强化学习，形成真正的“自我进化”能力。

“推理模型+机器人”：将o3模型的推理能力与机器人结合，用于解决实际工程问题，如自动化建模、优化物流等。

论文列表 2025-02-03 [Competitive Programming with Large Reasoning Models](#)

分享这篇文章



相关文章推荐

OpenAI 推理模型最...

本文总结了
OpenAI推理模...

DeepSeek V3 论文解读

本文介绍了深度
求 ...

DeepSeek 微调

本文介绍了如何
使用合成推理...