进化搜索策略，提高LLM推理深度；学术搜索智能体，RL+PPO；LLM回答置信度

AI研究前瞻 发布于 2025-1-21 12:19

研究背景与意义

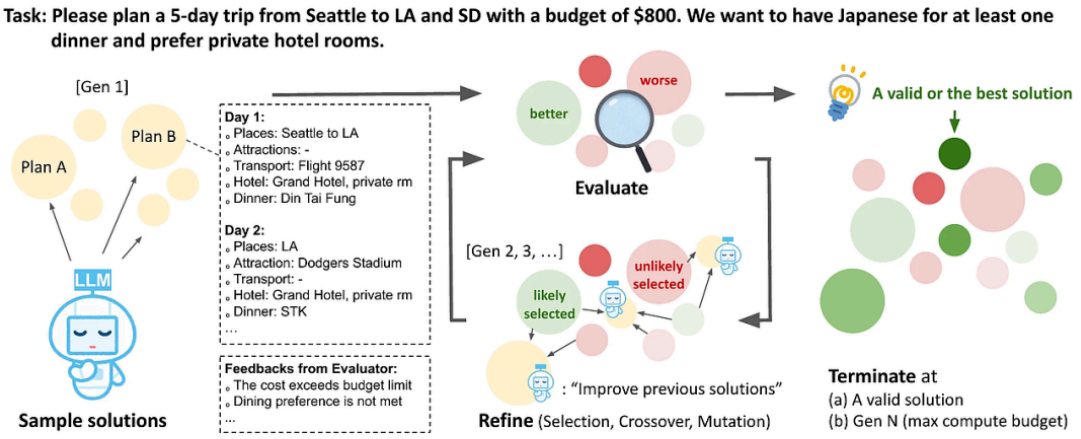

Figure 1 | Mind Evolution is a genetic-based evolutionary search strategy that operates in natural language space. The figure illustrates how Mind Evolution evolves a population of solution candidates toward higher quality candidates for a travel planning task. The candidate population is improved through an iterative process, where an LLM is used to recombine and refine candidates in each iteration.

在当前的人工智能研究中，如何提升大型语言模型（LLM）的推理能力是一个重要的课题。传统的推理方法往往依赖于明确的推理步骤和形式化的问题定义，但这在处理复杂的自然语言任务时显得力不从心。本文提出了一种名为“Mind Evolution”的进化搜索策略，旨在通过利用语言模型生成、重组和优化候选答案，以应对自然语言规划任务中的推理挑战。研究的意义在于，它不仅提供了一种新的思路来提升LLM的推理深度，还展示了如何在不依赖于形式化求解器的情况下，利用评估器进行高效的解决方案生成。

研究方法与创新

“Mind Evolution”方法的核心在于结合自由流动的随机探索与大规模的迭代优化。具体而言，研究者们采用了基因算法的思想，通过语言模型生成多样化的候选解决方案，并利用反馈机制进行逐步优化。与现有的推理策略相比，该方法的创新点在于：

1. **进化搜索策略**：通过生成多样化的候选解并进行迭代优化，Mind Evolution能够在更广泛的解空间中寻找解决方案。
2. **全局评估机制**：与传统的逐步推理不同，该方法仅依赖于全局解决方案评估器，从而简化了推理过程。
3. **适应性强**：该方法能够处理未形式化的问题，尤其适合自然语言规划任务。

通过对比实验，Mind Evolution在多个基准测试中表现出色，成功率显著高于传统方法，如Best-of-N和Sequential Revision等。

实验设计与结果分析

目录

社区精华内容

每日任务

Best-of-N	val	100/180 = 55.6%	472	4.44M	0.47M	US$0.47
Sequential-Revision+	val	149/180 = 82.8%	280	35.53M	0.29M	US$2.75
Mind Evolution	val	172/180 = **95.6%**	174	3.10M	0.18M	US$0.29
(+pro)	val	180/180 = **100%**	(257)	(3.25M)	(0.19M)	(US$0.54)
Mind Evolution	test	952/1000 = **95.2%**	167	3.02M	0.18M	US$0.28
(+pro)	test	999/1000 = **99.9%**	(67)	(3.05M)	(0.18M)	(US$0.33)
Natural Plan [47] Trip Planning						
1-Pass	val	66/320 = 20.6%	1	0.002M	0.001M	<US$0.001
(o1-preview 1-Pass)	val	116/320 = 36.2%	1	0.002M	0.008M	US$0.53
Best-of-N	val	247/320 = 77.2%	274	0.61M	0.18M	US$0.10
Sequential-Revision+	val	238/320 = 74.4%	391	41.57M	0.38M	US$3.23
Mind Evolution	val	308/320 = **96.2%**	168	1.48M	0.19M	US$0.17
(+pro)	val	320/320 = **100%**	(111)	(1.51M)	(0.19M)	(US$0.22)
Mind Evolution	test	1204/1280 = **94.1%**	196	1.78M	0.22M	US$0.20
(+pro)	test	1275/1280 = **99.6%**	(211)	(1.86M)	(0.24M)	(US$0.37)
Natural Plan [47] Meeting Planning						
1-Pass	val	104/500 = 20.8%	1	0.007M	0.001M	US$0.001
(o1-preview 1-Pass)	val	221/500 = 44.2%	1	0.006M	0.006M	US$0.47
Best-of-N	val	347/500 = 69.4%	444	3.99M	0.31M	US$0.39
Sequential-Revision+	val	310/500 = 62.0%	484	32.16M	0.40M	US$2.53
Mind Evolution	val	425/500 = **85.0%**	406	5.35M	0.41M	US$0.52
(+pro)	val	492/500 = **98.4%**	(890)	(13.36M)	(0.91M)	(US$2.55)
Mind Evolution	test	419/500 = **83.8%**	394	5.24M	0.40M	US$0.51
(+pro)	test	491/500 = **98.2%**	(828)	(12.25M)	(0.83M)	(US$2.34)

Table 2 | Experimental results on benchmark natural language planning tasks. "(+pro)" denotes the two-stage results, where we use Gemini 1.5 Pro to solve the problems that were not solved in experiments using Gemini 1.5 Flash. Number of LLM calls, token counts, and API cost are averaged across the validation or test problem set, and they are calculated only on the remaining problems for the "(+pro)" experiments. Here, we also show OpenAI o1-preview results as a reference.

Task	Results (Cipher and Poem)
Message To Encode (M): 10,20,30,40, 50,60,70,80, 90,100,10,20 Style: Poem Inspiration: Shel Silverstein Words Between (B): 4 Topic: "The Fun of Walking"	10 : rooster; 20 : flowers; 30 : bright; 40 : flames; 50 : cherry; 60 : crimson; 70 : sunset; 80 : ruby; 90 : scarlet; 100 : burning;
	I like to walk, I like to stride, With ROOSTER crows and FLOWERS by my side. I like to walk, I like to roam, Past BRIGHT green fields and FLAMES at home. I like to walk, I like to stroll, To see the CHERRY trees and a CRIMSON hole. I like to walk, I like to ramble, To watch the SUNSET and hold a RUBY's gamble. I like to walk, I like to wander, To watch a SCARLET leaf, and a fire, BURNING, under. I like to walk, I like to stride, With ROOSTER crows and FLOWERS by my side.

Figure 10 | **StegPoet example.** Example of the encoding of a StegPoet problem instance (left) and a correct solution (right) that includes the number-to-word cipher and a poem in the style of a children's poetry author. Note that $|M| = 12$ in this instance. We added capitalization to the code words to highlight them.

在实验设计上，研究者们选择了多个自然语言规划任务，包括"Travel Planner"和"Natural Plan"。通过对比Mind Evolution与其他基线方法的表现，结果显示：

1. **成功率**：Mind Evolution在"Travel Planner"任务中达到了95.6%的成功率，而其他方法的成功率普遍较低。
2. **效率**：Mind Evolution在计算成本方面也表现优异，生成的候选解数量和API调用次数相对较少，表明其在资源利用上的高效性。
3. **多场景表现**：无论是在简单还是复杂的任务背景下，Mind Evolution均展现了良好的适应性和稳定性。

结论与展望

本文的研究表明，Mind Evolution为提升LLM的推理能力提供了一种有效的策略。尽管该方法在多个任务中表现优异，但仍存在一些局限，例如在处理极其复杂的任务时可能需要更多的计算资源。未来的研究可以进一步探索如何优化进化策略的参数设置，以及如何将该方法应用于更广泛的自然语言处理任务中。总之，Mind Evolution不仅为LLM的推理深度提供了新的视角，也为相关领域的研究提供了宝贵的借鉴。

目录

per Search

25-01-17 | ByteDance, PKU| ▲ 18

p://arxiv.org/abs/2501.10120v1

ps://huggingface.co/papers/2501.10120

ps://pasa-agent.ai

研究背景与意义

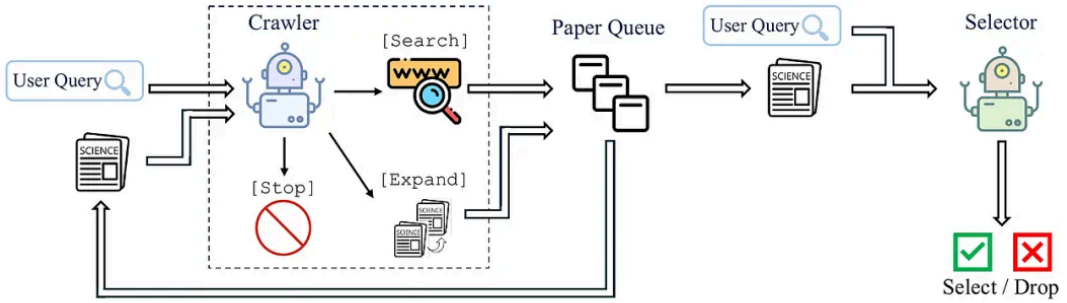

Figure 1: Architecture of PaSa. The system consists of two LLM agents, Crawler and Selector. The Crawler processes the user query and can access papers from the paper queue. It can autonomously invoke the search tool, expand citations, or stop processing of the current paper. All papers collected by the Crawler are appended to the paper queue. The Selector reads each paper in the paper queue to determine whether it meets the criteria specified in the user query.

在现代学术研究中，信息检索的效率直接影响到研究的进展和成果的质量。然而，现有的学术搜索系统（如Google Scholar）在处理复杂的学术查询时，往往无法满足研究者的需求。这种局限性促使研究者花费大量时间进行文献综述，降低了研究效率。因此，开发一种能够自动化、全面且准确地进行学术文献搜索的工具显得尤为重要。本文提出的PaSa（Paper Search Agent）正是为了解决这一问题而设计。

PaSa的设计目标是通过模拟人类研究者的行为，提升学术搜索的准确性和全面性。通过对现有文献检索工具的分析，本文指出了当前系统在处理长尾特定知识、细粒度查询等方面的不足，并阐明了PaSa在优化学术搜索中的潜在价值。

研究方法与创新

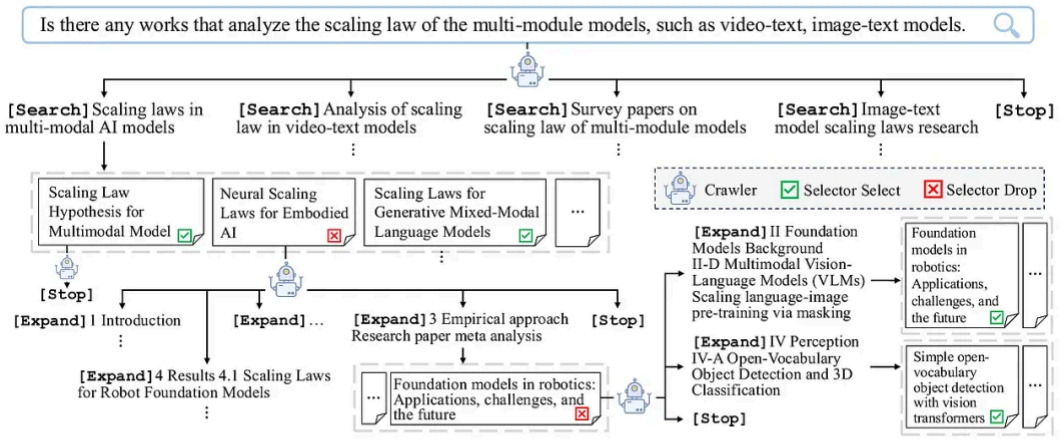

Figure 2: An example of the PaSa workflow. The Crawler runs multiple [Search] using diverse and complementary queries. In addition, the Crawler can evaluate the long-term value of its actions. Notably, it discovers many relevant papers as it explores deeper on the citation network, even when intermediate papers along the path do not align with the user query.

目录

每日任务

在技术实现上，PaSa结合了强化学习（RL）与Proximal Policy Optimization（PPO）算法，针对文献搜索任务的独特挑战进行了优化。具体而言，PaSa通过设计新的奖励机制来应对稀疏奖励和长轨迹问题，从而提升了模型的学习率。此外，PaSa还开发了两个高质量的数据集（AutoScholarQuery和RealScholarQuery），用于训练和评估其性能。

实验设计与结果分析

Method	Crawler Recall	Precision	Recall	Recall@100	Recall@50	Recall@20
Google	-	-	-	0.2015	0.1891	0.1568
Google Scholar	-	-	-	0.1130	0.0970	0.0609
Google with GPT-4o	-	-	-	0.2683	0.2450	0.1921
ChatGPT	-	0.0507	0.3046	-	-	-
GPT-o1	-	0.0413	0.1925	-	-	-
PaSa-GPT-4o	0.7565	0.1457	0.3873	-	-	-
PaSa-7b	0.7931	0.1448	0.4834	0.6947	0.6334	0.5301
PaSa-7b-ensemble	0.8265	0.1410	0.4985	0.7099	0.6386	0.5326

Table 5: Results on AutoScholarQuery test set.

为评估PaSa的性能，研究者在合成数据集AutoScholarQuery和真实数据集RealScholarQuery上进行了实验。实验结果表明，PaSa在多个指标上显著优于现有的基线模型，如Google Scholar和ChatGPT等。具体而言，PaSa在Recall@20和Recall@50的表现上分别提高了37.78%和39.90%。这些结果不仅验证了PaSa在学术搜索中的有效性，也表明其在真实场景中的应用潜力。

实验过程中还对Crawler和Selector的性能进行了详细分析，结果显示，Crawler的回调率在PaSa-7b模型中达到了79.31%，而Selector的F1得分也达到了85%。这表明，PaSa的设计有效地提升了文献检索的准确性和可靠性。

结论与展望

本文介绍了PaSa，一个旨在提高学术文献搜索效率和准确性的先进工具。通过结合强化学习和多种创新技术，PaSa在复杂学术查询的处理上展现了优越的性能。未来，研究者计划进一步优化PaSa的算法，并扩展其应用范围，以满足更广泛的学术需求。此外，随着数据集的不断丰富和算法的迭代，PaSa有望在学术研究中发挥更大的作用，帮助研究者更高效地获取和利用知识。

Multiple Choice Questions: Reasoning Makes Large Language Models (LLMs) More Self-Confident Even When They Are Wrong

2025-01-16 | NUAA, UPM, UC3M, Somos NLP| ▲ 12

http://arxiv.org/abs/2501.09775v1

https://huggingface.co/papers/2501.09775

研究背景与意义

目录

每日任务