基于大语言模型蒙特卡洛树搜索的智算网络故障根因分析系统

Fault Root Cause Analysis System of Intelligent Computing Networks Based on Large Language Models and Monte Carlo Tree Search

罗子秋/LUO Ziqiu，苗宇铠/MIAO Yukai，李丹/LI Dan
（清华大学，中国 北京 100080）
(Tsinghua University, Beijing 100080, China)

DOI：10.12142/ZTETJ.202502004
网络出版地址：http://kns.cnki.net/kcms/detail/34.1228.TN.20250424.0902.002.html
网络出版日期：2025-04-24
收稿日期：2025-02-20

摘要：提出了一种基于大语言模型（LLM）进行蒙特卡洛树搜索的智算网络故障根因分析系统（RCA-MCTS）。利用LLM推理研究领域在蒙特卡洛树搜索上的前沿研究，面向智算网络复杂故障场景，设计了适用于故障根因分析任务的多策略提示语扩展机制，并基于与故障模拟环境交互反馈的方式设计了模拟机制，使得LLM推理时的蒙特卡洛树搜索过程适配于故障根因分析任务场景。实验表明，RCA-MCTS在故障根因分析任务准确率上提升33%～43%，在故障推理动作序列平均匹配度上提升18%～34%。

关键词：智算网络；故障根因分析；大语言模型；蒙特卡洛树搜索

Abstract: A fault root cause analysis (RCA) system of intelligent computing networks based on Monte Carlo tree search (MCTS) and large language models (LLM), named RCA-MCTS, is proposed in this paper. By leveraging cutting-edge research on MCTS in the domain of LLM reasoning, a multi-strategy prompt expansion mechanism is designed for fault root cause analysis tasks in intelligent network fault scenarios. Additionally, a simulation mechanism is developed based on feedback interactions with the fault environment, enabling the MCTS process during LLM reasoning to be adapted to the fault root cause analysis task. Experimental results show that RCA-MCTS improves the accuracy of fault root cause analysis by 33%-43%, and enhances the average matching degree of fault inference action sequences by 18%-34%.

Keywords: intelligent computing network; fault root cause analysis; large language model; Monte Carlo tree search

引用格式：罗子秋，苗宇铠，李丹. 基于大语言模型蒙特卡洛树搜索的智算网络故障根因分析系统 [J]. 中兴通讯技术, 2025, 31(2):21-30. DOI: 10.12142/ZTETJ.202502004

Citation: LUO Z Q, MIAO Y K, LI D. Fault root cause analysis system of intelligent computing networks based on large language models and Monte Carlo tree search [J]. ZTE technology journal, 2025, 31(2): 21-30. DOI: 10.12142/ZTETJ.202502004

为了支撑日益增长的通信和计算需求，现代网络系统不断向更庞大的规模、更复杂的架构演进。自2022年ChatGPT[1]发布后，大语言模型（LLM）的应用成为热门话题，由此带来的算力需求爆炸式地增长，也对支撑LLM训练和推理的网络系统提出了更高的要求。人工智能算力网络（简称智算网络）是一种专门服务于大规模LLM训练和推理的新型网络系统，提供了灵活调度、资源共享、统一服务的能力。智算网络往往具有复杂的拓扑设计，集成了大量异构的新型算力设施，这给网络运维带来了新的挑战。目前智算网络的运维主要依赖于传统网络运维的专家经验，可能存在对新型算力设施的运维经验不足、不同技术背景的运维人员之间协调效率低等问题。面对可能存在的故障，发现不及时、响应慢、解决不彻底，将会对LLM的训练和推理任务产生严重的影响。

针对传统人工运维方法的不足，学术界已提出一系列基于人工智能的运维（AIOps）方法，可应用于故障检测、故障定位、故障根因分析（RCA）、故障恢复等多种任务[2-3]。本工作聚焦于故障根因分析任务，研究基于人工智能模型分析故障信息和环境数据，输出可能的故障根因，指导运维决策。近年来，由于LLM在理解和生成自然语言以及执行复杂的推理任务上表现出了超越以往智能模型的卓越能力，学术界已有将LLM应用于AIOps的工作。例如，RCACopilot使

用LLM生成结构化告警数据的文本摘要，并根据语义相似度从历史告警的故障根因中检索出可能的根因[4]；RCAgent利用LLM生成与故障告警信息高度相关的根因分析等内容，在生成过程中通过多路径并行推理的一致性检查，提升输出内容的可靠性[5]。然而，这些方法完全依赖LLM的领域知识，将各项结构化告警信息以提示语的形式输入LLM后，只通过单步推理就直接获取模型预测的故障根因，这与网络系统故障运维的实际情况存在较大差异。在实际故障运维场景中，运维人员需要根据故障处理经验与环境进行多轮交互来确定故障根因。在每一轮交互中，运维人员要根据故障信息和当前状态，决定下一步要执行的运维动作，随后执行此动作，并获取环境反馈，以帮助下一轮的决策。因此，我们认为故障根因应由故障告警信息和若干<运维动作，环境反馈>二元组来确定，即：$RC = F(Alert, (Action, Observation) \times N)$，其中$N$需要足够大，以唯一地确定故障根因。

根据上述讨论，如图1所示，我们提出一个基于LLM的自动与环境进行多轮交互并分析故障根因的智能体。该智能体以LLM为核心，配备智算网络运维知识库，在获得输入的故障告警信息后，通过多次调用运维工具与故障环境进行交互，最终定位故障根因。

RCA任务对LLM的能力提出了很高的要求：LLM不仅要具备基本的知识问答能力，还要具备复杂领域知识的理解能力和多步迭代的推理能力。现有的很多LLM尽管在各类通用任务上具备强大的能力，但在完成复杂推理任务时仍表现不佳。例如，Mistral-7B等主流LLM在使用链式思维等技术[6-8]增强推理能力的情况下，在GSM8K等需要多步推理的数据集上的最高准确率仅为36.5%。计划推理（RAP）[9]等改进的链式推理框架采用了一种自我探索的方案，通过自我奖励反馈，迭代地提高LLM的推理表现。但是，这种推理方式难以有效地探索解空间，即使经过多次尝试，也常常困在一个低质量的区域中。LLM多步推理能力的不足显著阻碍了RCA智能体的开发和应用。

近段时间以来，以OpenAI o1模型为代表的LLM多策略推理技术获得了较大的进展，为RCA智能体推理能力的实现提供了思路。面对GPT-4、GPT-4o等模型都难以回答的奥林匹克数学竞赛题目，o1借鉴人类思考方式，自主优化思考过程，尝试多种策略，识别思考过程中的错误，最终完成复杂逻辑计算的推理。相比于链式推理以及基于链式推理数据的微调，这种推理方式能够探索更大范围的解空间，提高推理出最终正确结果的概率。不过OpenAI并没有公开o1推理框架的具体实现方法，一系列复现o1的开源研究[10-17]大多使用蒙特卡洛树搜索（MCTS）这一经典算法来实现类似o1的推理框架。基于MCTS算法，即使在模型基础能力不足、数据集质量不佳的情况下，也能够增强LLM的推理能

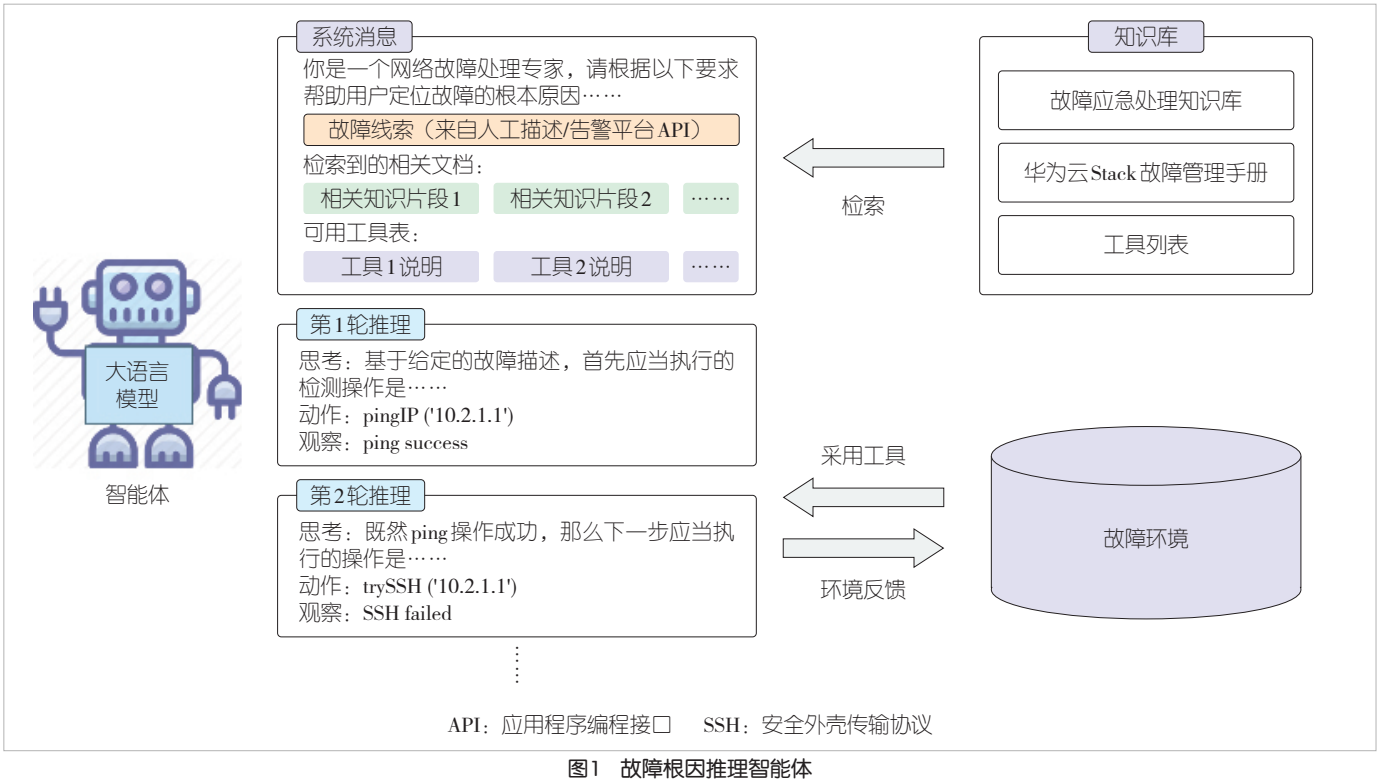


图1 故障根因推理智能体

力。但是这些工作的应用场景多为求解数学问题，与需要多轮环境交互的RCA场景有很大区别，直接套用现有的MCTS推理框架并不可行。针对上述挑战，本文提出了适用于RCA任务的MCTS推理框架RCA-MCTS，通过设计符合RCA场景特征的MCTS扩展策略和环境交互机制，显著提升了7B参数规模的LLM在RCA任务上的端到端表现。

1 RCA-MCTS研究动机

MCTS核心思想是将统计模拟（Monte Carlo方法）和树搜索（Tree Search）相结合，平衡“探索（Exploration）”与“利用（Exploitation）”，逐步建立和扩展决策树，并覆盖多路高价值的决策路径。MCTS以模型初始输入信息作为根节点，迭代式地进行多轮计算，每轮过程包含选择、扩展、模拟、回传4个步骤。在选择阶段，算法从根节点出发，根据平衡探索与利用的策略选择出一条路径，路径的终点作为当前节点；在扩展阶段，算法在当前节点基于动作扩展策略添加一个新的子节点；在模拟阶段，算法以扩展的新节点为起点进行多次模拟过程，每次模拟过程通过一步步的随机采样探索出一条新的子路径，直到达到预设的终止状态，并根据终点的正确性获得奖励分数；在回传阶段，算法将多次模拟的奖励回传给路径上的所有节点，更新节点的价值，并以此影响下一轮的选择过程。在现有的基于MCTS的LLM多策略推理研究中，MCTS搜索树的每一条连接父子节点的边对应LLM的一步推理，多种提示语策略构成了MCTS扩展阶段的动作空间，LLM在MCTS的指引下提高了生成正确答案的概率。

然而，在上述基于MCTS的LLM推理研究中，绝大多数选择求解数学问题作为任务场景，这与RCA任务存在如下差异：第一，数学题目的解决过程只需要根据推理历史选择合适的推理策略，现有方案的MCTS扩展仅需遵循静态的、预定义的规则。而RCA任务则需要考虑环境的反馈信息，即使在推理历史相同的情况下，故障环境反馈仍存在多种可能，对于不同的环境反馈信息，应采取不同的故障分析策略。因此，LLM需要对推理历史和环境交互历史信息进行综合处理，不断动态调整策略。第二，诸如GSM8K的数学题目数据集中，每道题均存在唯一的、正确的数值解，在MCTS模拟评估阶段可以直接基于规则提取数值答案，通过数值计算、置信度评估、一致性检查或基于外部LLM判断等方式，验证答案的正确性，并以此计算奖励分数。而RCA任务的最终输出为文本形式的故障根因描述，不能直接利用数值计算或置信度评估来验证正确性，一致性检查或者基于外部LLM判断则需要评判模型是否具有RCA领域专业知识和推理能力。第三，主流的LLM的预训练数据集普遍包含了海量的数学相关数据，这使得LLM本身已具备较强的数学基础能力，可通过MCTS推理框架更进一步增强数学推理能力。而RCA任务的专业性相对强得多，且LLM预训练过程中缺少海量高质量的RCA数据集，导致模型本身RCA基础能力较弱。MCTS推理框架虽然能在一定程度上提升推理表现，但绝对的端到端准确率仍然较低。

基于上述讨论，我们列出实现RCA-MCTS的主要挑战：

1）故障环境反馈的不确定性。我们需要针对可变的环境反馈，设计MCTS扩展阶段的推理策略。

2）故障推理中奖励分数计算的复杂性。我们需要设计有效的奖励函数，为模拟阶段产生的推理节点计算奖励分数。

3）LLM故障运维基础能力的不足。我们需要提升LLM在RCA任务领域的基础能力，从而进一步提升使用MCTS方法后的端到端效果。

针对上述挑战，本文设计了RCA-MCTS故障根因推理框架，使得LLM在RCA任务场景下进行多策略并行式推理，输出准确的故障根因。RCA-MCTS在MCTS扩展阶段运用适配RCA场景的提示语策略，综合分析推理历史数据和故障环境反馈信息，合理规划下一步推理行动。在模拟阶段，RCA-MCTS结合故障环境交互和我们基于过程标注数据训练的过程奖励模型（PRM）计算推理节点的奖励分数，从而引导MCTS的路径搜索。我们构建了一个故障向量知识库，为LLM推理提供故障运维示例，并基于故障告警信息设计故障环境，为LLM采取的故障运维动作提供唯一确定的故障反馈。模拟过程中，RCA-MCTS支持LLM与故障环境持续交互，直至输出根因或者达到指定的最大搜索深度，基于故障知识库对输出根因进行验证。同时，受到AlphaGo工作的启发[18]，我们训练了PRM模型，用于评估MCTS中间节点对终局推理结果的贡献价值，以便提供更可靠的奖励反馈。最后为了进一步提升模型在推理框架下的故障推理能力，我们构建了ReAct格式的标准推理数据集，以此对模型进行了微调训练。微调后的模型增强了RCA领域的基础能力，结合RCA-MCTS推理框架展现出更强的故障根因分析能力。

结合以上所述，本研究的主要贡献包括：

1）构建RCA-MCTS推理框架，针对故障环境反馈设计MCTS动态扩展策略。

2）构建故障知识库、故障模拟环境，作为RCA-MCTS推理框架的外部依赖，支持LLM与环境的交互、输出根因的验证评估，模拟现实运维场景，客观评估RCA-MCTS推理框架性能。

3）训练面向RCA任务奖励计算的PRM模型，用于辅助

MCTS模拟评估。

4）基于自动化流程构造ReAct格式故障根因推理数据集，训练核心LLM，提升模型故障根因推理基础能力。

5）通过实验证明了RCA-MCTS故障根因推理框架的效果。以Qwen2-7B-Instruct模型为基座模型时，RCA-MCTS在故障根因推理数据评测集的端到端根因分析准确率、动作决策序列合理匹配度两个指标上分别取得了43.6%和34.0%的提升。对于其他7B模型，我们的方法也显著提升了推理表现。

本文将在第2节介绍RCA-MCTS故障根因推理框架的整体架构，在接下来的第3节重点论述推理框架的扩展策略和模拟过程设计以及相关模型的训练，并在第4节中给出实验设置、实验结果和结果分析。

2 RCA-MCTS推理框架架构设计

如图2所示，RCA-MCTS推理框架以原始故障告警信息作为输入，基于语义相似度在故障运维知识库中匹配可采取的故障运维动作集合、待确定的故障根因范围；初始提示语模版对告警输入、运维动作集合、可能的根因范围进行文本组合，并在系统提示语中定义模型推理范式遵守ReAct规则以确保模型每步推理的规范性和有效性[19-20]，组合后的文本加上系统提示语构成初始提示语；推理框架的核心LLM接收初始提示语作为输入，在基于面向RCA任务场景设计的MCTS扩展策略指导下，综合分析推理历史和故障环境反馈信息，合理选择提示语模版进行下一步推理；RCA-MCTS基于推理内容在故障知识库提供的运维动作集合中匹配运维动作，并采取运维动作与故障模拟环境交互获得反馈信息并记录，产生新的MCTS推理节点；随后，RCA-MCTS基于新节点按照规则与故障模拟环境模拟交互至终局状态，并结合PRM完成节点评估，回传节点分数，完成MCTS一轮过程；在进行多轮MCTS构建过程后，核心LLM输出故障告警对应的RCA推理树，对终节点的各项属性进行综合分析，基于多数投票的方式决定唯一的故障根因作为推理框架的最终输出。

在进行故障根因推理之前，RCA-MCTS预先构建一个故障运维知识库和一个故障模拟环境作为依赖，前者用于提供推理框架下故障处理可使用的动作描述集合和待分析的故障根因范围，后者用于接收模型推理过程中采取的运维动作，提供环境反馈信号。对于每一条故障告警输入，故障模拟环境对于故障运维动作集合提供的反馈信号是唯一确定的。若干<故障运维动作，故障环境反馈>信息对结合原始故障告警输入，对应唯一的故障根因。核心LLM接受原始故障告警后，基于RCA-MCTS推理框架不断通过推理获取动作和环境反馈信息对，最终推理出故障根因。

3 RCA-MCTS方法

3.1 RCA-MCTS外部依赖构建

RCA-MCTS外部依赖包含故障运维知识库和故障模拟

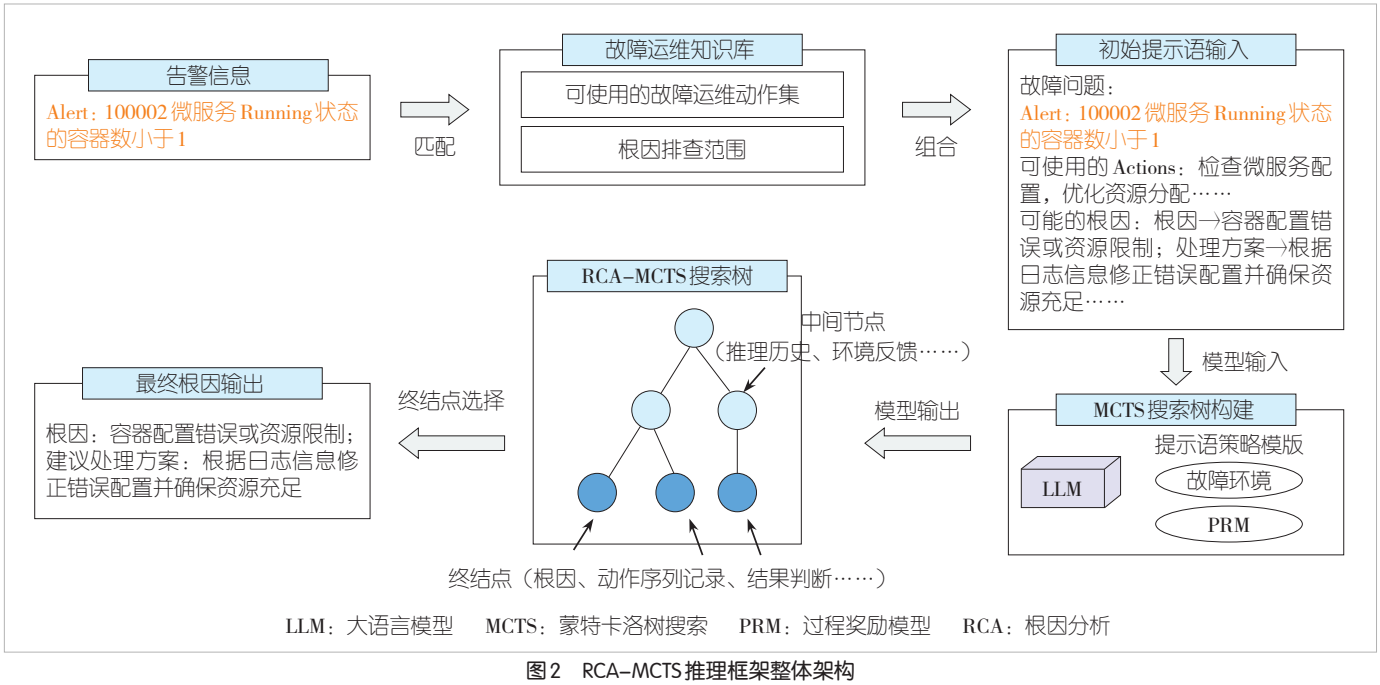


图2　RCA-MCTS推理框架整体架构

环境。故障运维知识库提供可使用的运维动作描述和候选的根因集合范围，并将它们作为初始输入的系统提示语部分；故障模拟环境在MCTS扩展阶段对核心LLM采取的运维动作进行反馈，在MCTS模拟评估阶段对最终输出的根因进行验证评估。二者均是完成完整MCTS推理过程的关键组件，需要在执行MCTS算法之前构建。

为了构建故障运维知识库，我们通过工具爬取了17个PDF格式的华为云故障文档和1个Word格式的云网络环境故障处理知识库，作为构建故障运维知识库的数据源。我们采取一系列文档处理规则对文档内容进行格式清洗和关键内容提取，挖掘出3 203个故障处理文本描述块。接下来，我们使用GPT-4在自设计提示语方法下，将每个故障文本描述块转化成ReAct格式的故障根因推理思维链数据，每条数据包含告警信息、故障根因信息，以及多步故障运维动作和故障模拟环境反馈。一般情况下，同一个故障告警在不同故障模拟环境反馈下对应于不同故障根因。我们进一步对这些数据进行清洗，滤除格式不合规的数据，最终构造了包含1 507个故障告警信息和3 818条ReAct格式的故障根因推理思维链数据集。我们通过对思维链数据的处理步骤、故障根因的整理和向量化，构造了故障运维动作集合和故障根因集合的向量数据库，供RCA-MCTS在开始阶段基于故障告警输入进行检索增强（RAG）以生成系统提示语。

故障运维思维链数据集除了用于构建故障运维知识库，还用于PRM和核心LLM的训练评估以及推理框架端到端评测实验。我们基于该数据集进行采样、划分、预处理工作，这些工作服务于对应的模型训练和评测环节，数据处理的细节将在后续小节中论述。

在构建故障模拟环境时，基于$RC = F(Alert, (Action, Observation) \times N)$的模式定义，对于每一条故障告警输入，我们在故障模拟环境中为每个运维动作设置了唯一确定的动作反馈，并针对原始故障告警输入设置唯一确定的故障根因作为答案。当核心LLM经过多轮推理达到终局状态时，我们基于文本规则提取出根因答案作为最终根因输出，在可能故障根因范围中匹配根因，随后基于故障知识库提供的验证方法，在故障模拟环境中验证核心LLM推理出的根因的正确性。

3.2 面向RCA任务的MCTS过程设计

如图3所示，RCA-MCTS推理过程遵循MCTS算法步骤多次循环，构建故障根因推理搜索树。在选择阶段RCA-MCTS沿用了MCTS经典算法的默认设置，基于“探索-利用”准则递归地选择待扩展的叶子节点，但在MCTS扩展阶段和模拟评估阶段采用了适用于RCA任务的设计。这些设计有助于RCA-MCTS推理框架在故障推理任务场景下不断动态调整策略和并行式推理，在故障推理数据集上的端到端准确率提升方面起到了关键作用。

1）RCA-MCTS扩展过程

RCA-MCTS扩展阶段的策略综合考虑了推理历史和故障模拟环境反馈。RCA-MCTS扩展动作定义为一系列适用于核心LLM进行故障根因推理的提示语，包括以下几类：（1）提示核心LLM对过去推理历史进行归纳总结，提炼当前故障根因推理思路，专注于异常的故障模拟环境反馈信息；（2）提示核心LLM判断故障模拟环境反馈信息是否异

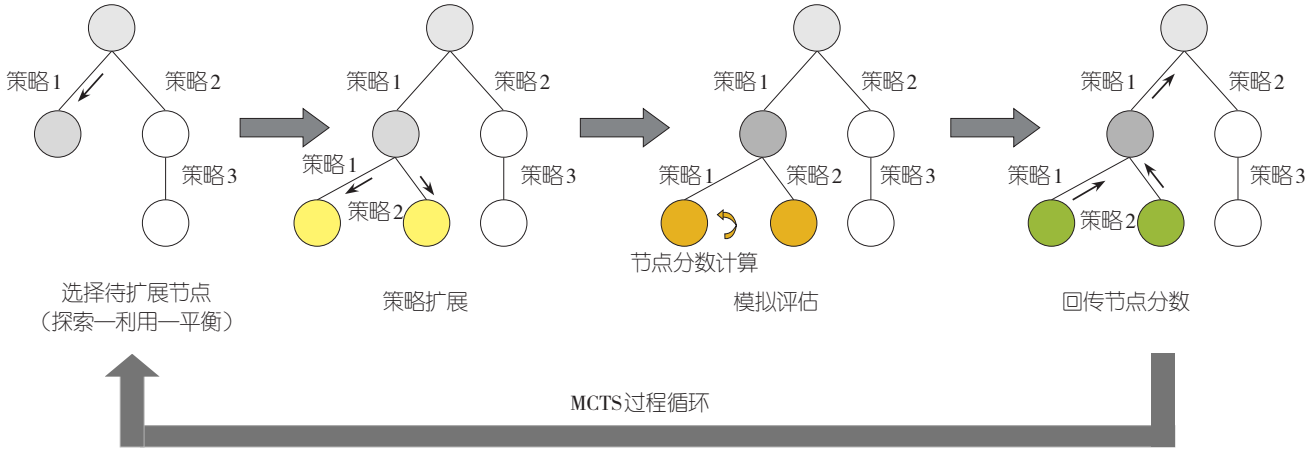


图3 RCA-MCTS搜索树构建过程

常，结合推理历史思考下一步应当采取的动作，以获取更多的环境反馈作为推理依据；（3）提示核心LLM根据过去推理历史，判断故障告警最可能的根本原因；（4）提示核心模型回顾ReAct格式的推理历史，反思采取故障运维动作是否合理、故障模拟环境反馈是否与根因匹配。如图4所示，MCTS节点基于当前节点状态（如推理链条长度、推理历史、过去推理所采取的推理策略等）选择多个提示语策略模版，每个被选择的提示语模版与节点内记录的系统提示语、推理历史组成扩展提示语输入；核心LLM接受每个扩展提示语进行单步扩展推理，输出扩展推理内容，扩展产生新的MCTS推理节点。特别地，当扩展策略中包含故障运维动作时，RCA-MCTS提取LLM输出的故障运维动作内容片段，随后基于语义相似度，在故障知识库提供的故障运维动作集合中匹配最近似的动作，采用动作与故障模拟环境进行交互并获取环境反馈，构成新的推理历史。基于这种扩展方式，RCA-MCTS产生多个新的MCTS子节点，在多种思维策略方向上并行推理，扩大故障根因搜索空间。随后RCA-MCTS基于故障模拟环境多轮交互过程和过程奖励模型，对扩展节点进行模拟评估，为下一轮MCTS过程中选择更有价值的节点提供判断依据，提高推理出正确根因的可能性。

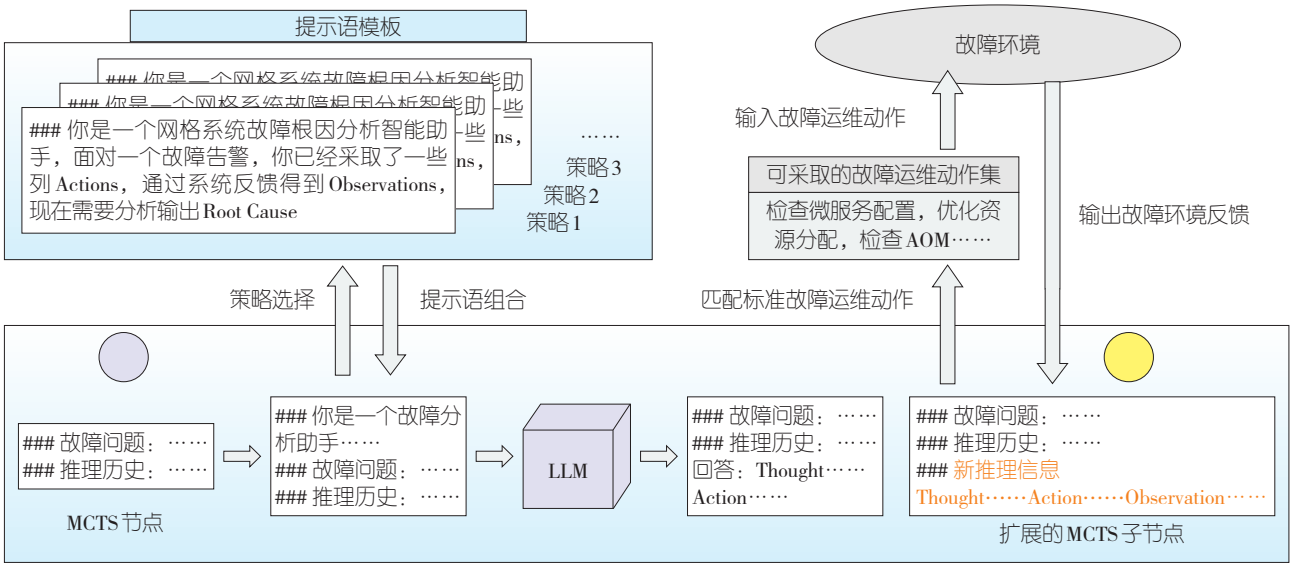
2）RCA-MCTS模拟评估过程

RCA-MCTS模拟评估环节对扩展产生的MCTS节点进行评分，如图5所示。节点分值由基于故障模拟环境交互过程的模拟结果评分和PRM评分共同决定。其中，故障模拟环境交互模拟是指：从当前节点出发，进一步选择策略进行核心LLM推理，与故障模拟环境交互后产生新的MCTS扩展节点，在扩展节点中随机选择节点进一步推理产生新节点，如此推理多次直到能从核心LLM输出中提取故障根因，或者达到规定的树最大深度。当MCTS的扩展策略中包含故障根因推断策略时，RCA-MCTS从核心LLM输出中提取故障根因答案，在故障模拟环境中验证答案的正确性并记录，完成一次故障模拟环境交互模拟。这样的模拟会进行多次，推断出正确根因的模拟次数所占比例即为故障模拟环境交互模拟结果评分。

基于故障模拟环境交互的模拟过程中，由于搜索空间无法穷尽，因此MCTS将随机选择节点进行扩展，选用多次模拟的统计结果作为环境交互模拟评分。这种统计采样方法产生的评分具有一定的随机性，特别是在推理成本不足导致模拟次数有限的情况下，分数的不可靠性会被进一步放大。为此，我们引入参数化PRM对扩展节点进行价值判断，来平衡模拟统计方法的随机性。在不用耗费大量推理成本增大模拟次数的情况下，PRM基于当前节点推理历史输入，判断当前核心模型能最终推理出正确根因的概率。将模拟交互得分和PRM得分的结果加权相加，最终得到了RCA-MCTS模拟评估阶段的节点分数。

3） PRM训练

RCA-MCTS依赖的PRM根据当前推理历史判断已采取



AOM：应用运维管理服务　LLM：大语言模型　MCTS：蒙特卡洛树搜索

图4　RCA-MCTS扩展过程

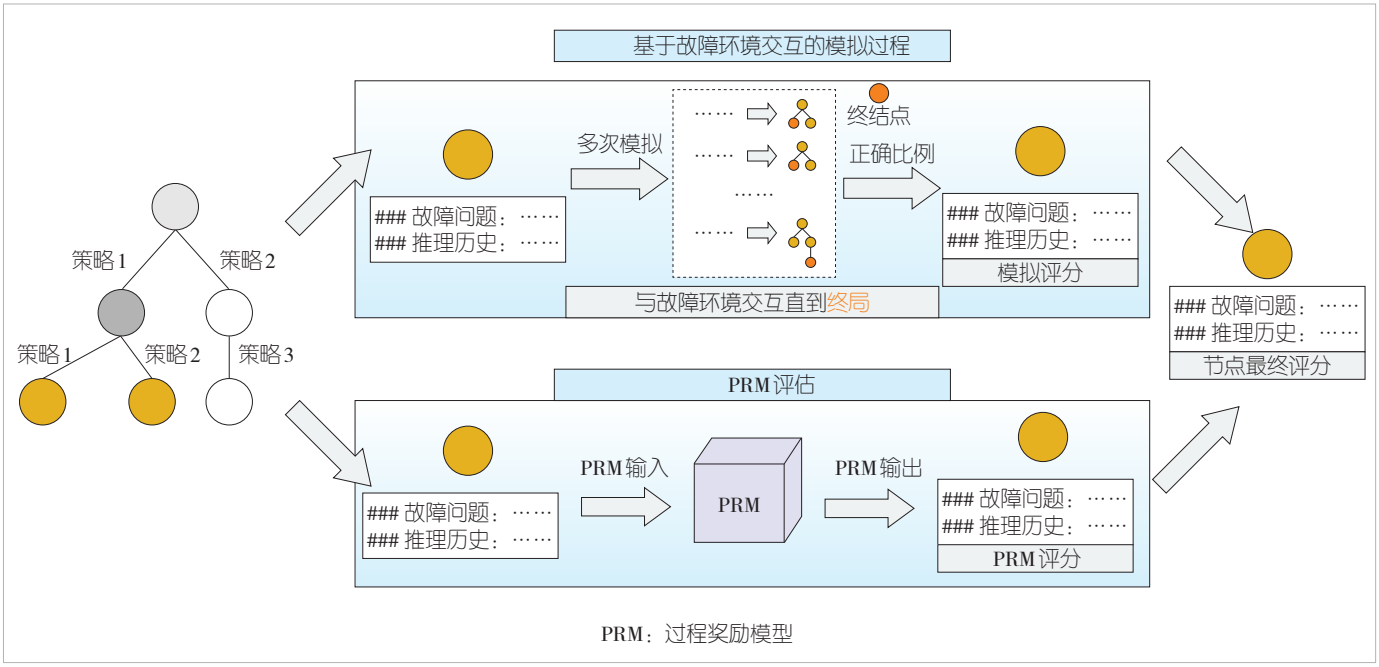


图5 RCA–MCTS模拟评估过程

的故障运维动作的合理性，在不需要LLM与故障模拟环境交互至终止状态的情况下，评估当前推理动作最终推理出正确的故障根因的可能性，其输出值的范围为[0,1]连续区间。训练PRM依赖于带有推理步骤标签的故障根因推理数据集。为了构建该数据集，我们基于构建故障运维知识库过程中产生的ReAct多步推理格式的故障根因推理思维链数据集，从中采样了500个故障告警对应的1 382条推理数据，通过打乱推理顺序、重复步骤、引入其他告警运维动作等方式，替换掉推理思维链数据中的某些步骤环节，并在错误的推理步骤处标注负标签，在剩余步骤处标注正标签。基于该构造方式，我们获得了包含6 742条故障根因推理思维链数据（对应32 633个推理步骤标签）的ReAct故障根因推理数据集，用于训练PRM对每个RCA–MCTS中间推理节点的有效评估能力，服务于RCA–MCTS的模拟评估过程。

3.3 核心模型RCA基础能力微调训练

前述开发LLM推理框架的工作专注于通过设计多策略推理方法激发模型智能、提高推理效率，并不涉及模型参数微调。在模型故障运维领域基础能力本身缺失的情况下，推理框架能够激发的表现有限，导致MCTS搜索树构建过程中获取的正反馈稀疏，降低搜索方法的效率。为了增加与故障模拟环境得到的正反馈，提升RCA–MCTS通过高价值推理节点迭代策略的效率，我们进一步使用故障运维专业领域数据集微调核心LLM。在构造专业数据集工作上，我们基于构建故障运维知识库中产生的ReAct数据集，从中提取了1 000条ReAct格式的多步故障思维链数据。在训练时，如图6所示，

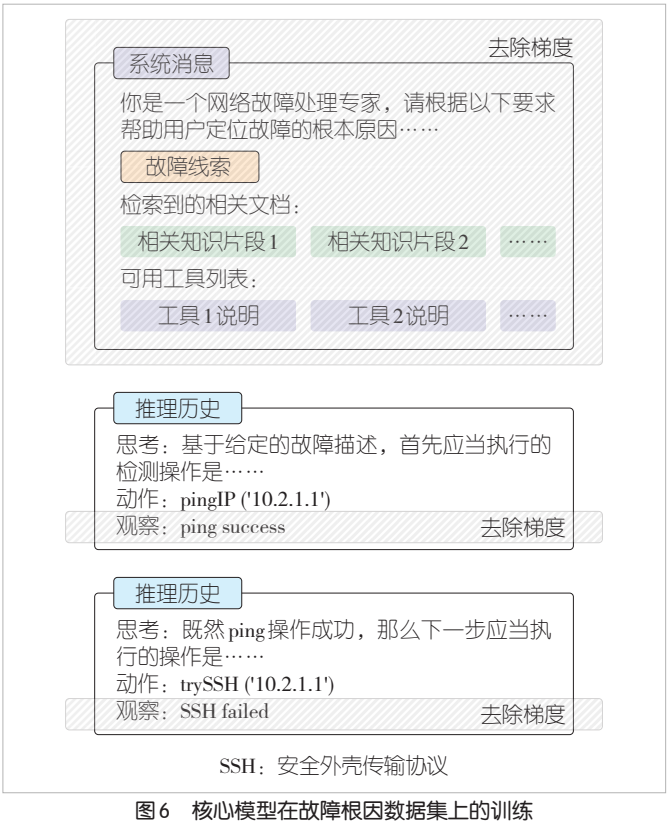


图6 核心模型在故障根因数据集上的训练

我们去除了系统消息前缀和故障反馈环境的梯度，使得核心LLM专注于正确的故障根因分析思考和故障运维动作决策，提升了核心LLM在故障根因推理专业领域的基础能力。

4 RCA-MCTS评估

4.1 评测数据集

我们基于前述从故障运维文档提取到的ReAct推理数据集构造评测数据集，用于RCA-MCTS推理框架性能端到端评估。该评测数据集包含300条故障告警，每条故障告警对应一系列参考运维动作序列和唯一的故障根因。推理框架的目标是输出唯一正确的根因，在推理过程中产生的运维动作序列不必完全遵照参考序列。需要注意的是，该评测数据集与上一节中提到的用于训练PRM、核心LLM的数据集互不包含，以确保评测数据集评估核心LLM在推理框架下展现出的故障推理能力，而非单纯对推理数据集进行拟合。

4.2 实验设置

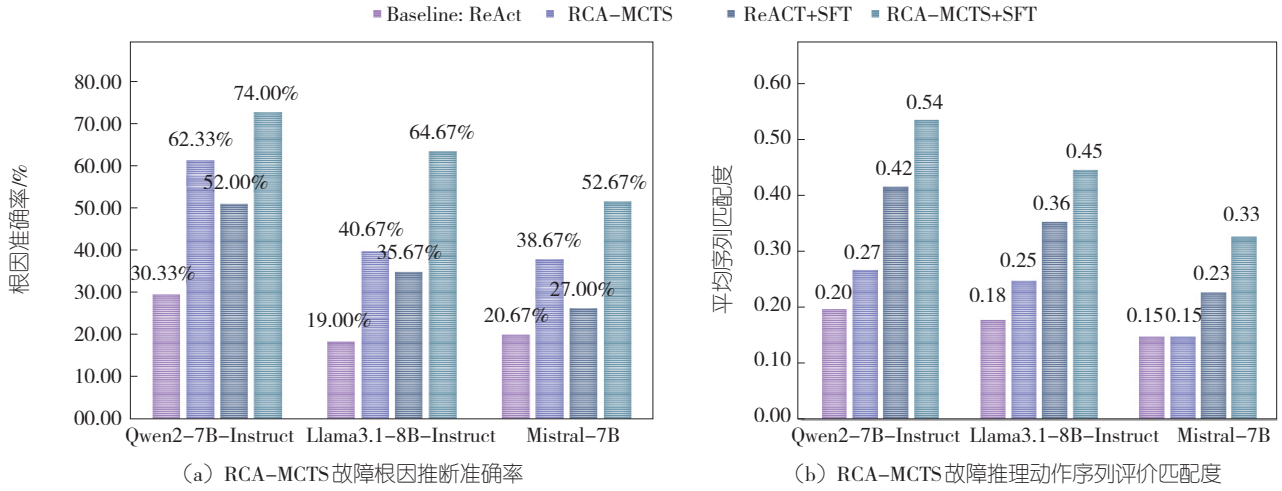
我们采用RCA-MCTS推理框架在评测集上的故障根因推断准确率，评估其故障根因推断能力。另外，为了评估RCA-MCTS在推理过程中采取运维动作的合理性，我们设计了故障推理动作序列匹配度指标。该指标通过计算运维动作中属于评测集参考运维动作序列集的比例获得。我们将ReAct方法作为baseline，对比体现RCA-MCTS推理框架的有效性。

在评测模型方面，我们使用Qwen2-7B-Instruct、Llama3.1-8B-Instruct、Mistral-7B等一系列参数量规模较小的大语言模型作为RCA-MCTS的核心LLM，评估RCA-MCTS推理框架在低模型参数成本场景下的故障根因推理能力。同时我们选择Qwen2-7B模型作为PRM训练的基座模型，在3.3所述含推理步骤标签数据集上训练2个epoch。在匹配故障知识库过程中，我们使用BGE-m3模型评估语义相似度。在推理框架参数设置方面，RCA-MCTS搜索树最大深度设值为6，评测数据集中所有故障的参考运维动作序列长度均小于该值，确保框架推理表现不会受限于树深度上限；设置MCTS过程循环次数为8，单次推理最大token输出为2 048。上述PRM训练、核心LLM在故障根因推理数据集上的有监督微调（SFT）、推理框架的实现均在8张NVIDIA H100 GPU上进行。

4.3 端到端表现

我们评估了RCA-MCTS推理框架的端到端故障根因推理效果。如图7（a）所示，同基于ReAct格式的链式推理框架相比，3个模型均通过RCA-MCTS获得了根因推断准确率的大幅提升。其中，Qwen2-7B-Instruct获得的提升幅度最大，这与其在Baseline上表现最好、展现的故障根因推断基础能力最高相关。同时，基于ReAct故障推理数据集格式的SFT也提升了根因推断准确率，其中Llama3.1-8B-Instruct通过SFT将故障推理能力提升至接近RCA-MCTS的水平。我们进一步将微调和推理框架结合，3个模型均基于SFT+RCA-MCTS的范式达到最佳故障根因推理效果。

我们通过故障推理动作序列匹配度进一步评估RCA-MCTS推理过程的合理性。如图7（b）所示，相比于RCA-MCTS推理框架，SFT更能提升其序列匹配度。这是因为SFT



（a）RCA-MCTS故障根因推断准确率　　（b）RCA-MCTS故障推理动作序列评价匹配度

MCTS：蒙特卡洛树搜索　RCA：根因分析　SFT：有监督微调

图7　RCA-MCTS推理框架端到端评测

数据集包含大量动作序列信息供LLM学习。RCA-MCTS推理框架本身不学习这些序列信息，而是有效扩展搜索空间，提高对正确故障根因覆盖可能性。在实际的故障运维场景中，对于唯一确定的故障根因可能存在多条故障运维动作序列。相较于采用何种序列，更关键的是准确找到故障根因，这更凸显了RCA-MCTS的实际应用优势。

4.4 更难评测集上的表现

同一故障告警输入可能对应于多个故障根因，由运维动作序列和故障模拟环境反馈决定。一般情况下，同一故障告警针对不同故障根因进行处理的运维动作集是相似的。尽管前述SFT训练集与评测集互不包含以确保能够客观评估框架的推理能力，但分属两个数据集的两个不同的故障根因可能对应于相同的故障告警信息。LLM通过在SFT训练集中学到运维动作知识，将有助于其在评测集中，针对同一故障告警输入但涉及另一故障根因的场景进行推理。然而，实际复杂网络系统运维场景中故障告警类别繁多，容易出现模型经验范围以外的全新故障告警，这是智能运维面临的一大挑战。为了评估RCA-MCTS在面对知识范围以外的故障告警下的性能，我们重构SFT训练集，使训练集中不包含评测集中的告警信息，更贴合实际运维场景。我们对Qwen2-7B-Instruct在不同SFT训练集上的训练后使用推理框架的表现做了对比。如表1所示，在SFT训练集的故障告警与评测集故障告警不重合时，使用RCA-MCTS推理框架的故障根因准确率低于重构SFT训练集前的最佳表现，这是评测难度提升所致。但重构SFT训练集后RCA-MCTS推理框架相对于baseline的ReAct方法仍然提升了15%的故障根因推断准确率，进一步体现了RCA-MCTS推理框架在更贴合实际运维场景的情况下对LLM故障根因推断能力的提升作用。

4.5 局限性

实际运维场景中，人工运维的推理策略根据复杂网络系统的迭代而动态调整，一些推理的策略和经验可能随着系统组件变化而与之前的知识矛盾。要想将RCA-MCTS应用于动态变化的实际运维场景，还需要设计更复杂的扩展策略和故障模拟环境模拟机制。另外，RCA-MCTS最终只利用了构建的故障搜索树的高反馈推理路径，对于其他推理数据缺乏应用。实际上，其他低反馈推理数据可能用于拒绝采样、策略梯度强化学习等进一步训练LLM以提高其故障根因分析能力，这有待未来研究探索。

表1 Qwen2-7B-Instruct在不同难度SFT训练集下的故障根因推理准确率

	SFT + ReAct	SFT + RCA-MCTS
训练集中包含部分评测集告警	52%	74%
训练集中不包含评测集告警	33%	48%

MCTS：蒙特卡洛树搜索 RCA：根因分析 SFT：有监督微调

5 结束语

通过设计适用于故障根因推理的扩展策略和模拟机制，本文得以将基于MCTS的LLM推理框架应用于RCA任务，提升了端到端故障根因推理准确率。此项成果为包括智算网络系统在内的复杂网络系统高效自动化运维做出了一定贡献，为未来自动化运维的研究探索了一种具有潜力的方向。

致谢

感谢北京中关村实验室副研究员陈力对本研究工作的大力支持！

参考文献

[1] OpenAI. OpenAI-ChatGPT [EB/OL]. [2025-02-25]. https://chatgpt.com
[2] JIN P X, ZHANG S L, MA M H, et al. Assess and summarize: improve outage understanding with large language models [C]// Proceedings of the 31st ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering. ACM, 2023: 1657-1668. DOI: 10.1145/3611643.3613891
[3] AHMED T, GHOSH S, BANSAL C, et al. Recommending root-cause and mitigation steps for cloud incidents using large language models [C]//Proceedings of IEEE/ACM 45th International Conference on Software Engineering (ICSE). IEEE, 2023: 1737-1749. DOI: 10.1109/ICSE48619.2023.00149
[4] CHEN Y F, XIE H B, MA M H, et al. Automatic root cause analysis via large language models for cloud incidents [C]//Proceedings of the Nineteenth European Conference on Computer Systems. ACM, 2024: 674-688. DOI: 10.1145/3627703.3629553
[5] WANG Z F, LIU Z C, ZHANG Y Y, et al. RCAgent: cloud root cause analysis by autonomous agents with tool-augmented large language models [C]//Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. ACM, 2024: 4966-4974. DOI: 10.1145/3627673.3680016
[6] WEI J, WANG X Z, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2201.11903v6
[7] YAO S Y, ZHAO J, YU D, et al. ReAct: synergizing reasoning and acting in language models [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2210.03629v3
[8] SHINN N, CASSANO F, GOPINATH A, et al. Reflexion: language agents with verbal reinforcement learning [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2303.11366
[9] HAO S B, GU Y, MA H D, et al. Reasoning with language model is planning with world model [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2305.14992v2
[10] WU J Y, FENG M K, ZHANG S, et al. Beyond examples: high-

level automated reasoning paradigm in in-context learning via MCTS [EB/OL]. [2025-02-25]. https://arxiv. org/abs/2411.18478v1
[11] MIN Y Q, CHEN Z P, JIANG J H, et al. Imitate, explore, and self-improve: a reproduction report on slow-thinking reasoning systems [EB/OL]. [2025-02-25]. https://arxiv. org/abs/2412.09413v2
[12] ZHANG D, WU J B, LEI J D, et al. LLaMA-berry: pairwise optimization for O1-like Olympiad-level mathematical reasoning [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2410.02884v2
[13] ZHAO Y, YIN H F, ZENG B, et al. Marco-o1: towards open reasoning models for open-ended solutions [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2411.14405v2
[14] KANG J K, LI X Z, CHEN X, et al. MindStar: enhancing math reasoning in pre-trained LLMs at inference time [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2405.16265v4
[15] QI Z T, MA M Y, XU J H, et al. Mutual reasoning makes smaller LLMs stronger problem-solvers [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2408.06195v1
[16] ZHANG Y X, WU S X, YANG Y Q, et al. o1-coder: an o1 replication for coding [EB/OL]. [2025-02-25]. https://arxiv. org/abs/2412.00154v2
[17] TIAN Y, PENG B L, SONG L F, et al. Toward self-improvement of LLMs via imagination, searching, and criticizing [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2404.12253v2
[18] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search [J]. Nature, 2016, 529: 484-489. DOI: 10.1038/nature16961
[19] YAO S Y, ZHAO J, YU D, et al. ReAct: synergizing reasoning and acting in language models [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2210.03629v3
[20] HAO S B, GU Y, MA H D, et al. Reasoning with language model is planning with world model [EB/OL]. [2025-02-25]. https://arxiv.org/abs/2305.14992v2

作 者 简 介

罗子秋，清华大学网络空间与网络科学研究院在读博士研究生；主要研究方向为网络系统智能运维、网络系统管理智能体开发等。

苗宇锴，清华大学教授，中关村实验室助理研究员；主要研究领域为自然语言处理；2022 年获 SIGCOMM 最佳论文奖；发表论文 10 余篇。

李丹，清华大学教授、博士生导师，教育部长江学者特聘教授，IEEE Fellow，北京高校卓越青年科学家计划项目负责人，国家“973”计划项目首席科学家，国家重点研发计划项目负责人，国家“十四五”重点研发计划“网络空间安全治理”专家组副组长；主要从事计算机网络领域的研究工作；曾获教育部“青年科学奖”，以第一完成人获中国通信学会技术发明一等奖、中国电子学会技术发明一等奖；发表论文 100 余篇，获授权专利 50 余项。