本科毕业论文（设计）开题报告

|  |  |  |  |
| --- | --- | --- | --- |
| 毕业论文（设计）题目 | 基于微调大语言模型与多Agent协作的数学推理工作流系统研究 | | |
| 学 院 | 人工智能学院 | 专 业 | 人工智能 |
| 学生姓名 | 陈柏政 | 指导教师 | 王曼 |
| 一、选题目的与意义  1、选题目的  近年来，自然语言生成式大语言模型[1]在各个应用领域展现出强大的推理和理解能力，尤其是在知识问答、数学解题以及代码编写等复杂任务上得到了广泛应用。然而，当前大语言模型在高难度推理任务中的表现仍有提升空间，特别是在数学领域例如抽象代数、离散数学等复杂抽象的问题，从基础的算术到奥林匹克竞赛级别的难题之间存在显著差距，尤其是结合了具体上下文情景后的表现。由此大模型是否具备真正的推理能力目前是大模型研究领域的热点问题[2]，为此，本设计旨在探讨如何通过微调与优化，探索如何提升大语言模型在逻辑推理和数学解题方面的能力，并开发一套简单可行的多Agent协作[3]的系统应用。  在现实应用中，数学解题能力的提升不仅仅是从小学到高等数学的跨越，还需要模型能够解决结合具体上下文语境的复杂问题以及部分抽象概念的数学证明。这要求模型具备极强的推理深度和跨领域的知识储备。此外，还需要微调一个具备一定的世界知识和一阶推理能力[4]的模型，确保系统不仅能够处理数学问题，还能够应对涉及多学科知识的问题分解任务。因此，本研究计划构建一个基于微调强化的大模型多Agent数学问题分析工作流系统，该系统不仅能提高解题效率，还能够处理从简单到复杂的多种数学任务，在实际应用中表现出强大的灵活性和智能性。  结合当前大模型的实际应用，这一系统的构建将进一步推动数学教育、科研等领域的智能化发展，为不同层次的学习者提供更精准严谨的数学解题指导思路，也能为未来大模型推理能力的提升方案路线提供一定的借鉴意义。  2、选题意义  本课题的研究意义在于深入探讨和提升大模型在数学问题推理方面的能力，进而对于提高大模型在复杂抽象情境下的逻辑推理能力有一定意义上的借鉴作用。近年来，研究者通过提示词工程(Prompt Engineering) [5]、检索增强生成(RAG, Retrieval-Augmented Generation) [6]等大模型在许多复杂任务中展现了强大的推理和理解能力，但在高难度推理任务，尤其是在处理数学问题，特别是抽象代数、离散数学等复杂数学问题时尤其是证明类的问题，依然存在显著不足。通过本研究，我旨在强化大模型在数学问题推理方面的能力，特别是在结合具体上下文情境的情况下。首先，提升大模型的推理能力是本研究的核心目标。我将借鉴当前专门设计于进行复杂逻辑推理的大模型，并采用思维链(COT, Chain of Thought) [7]和工具集成推理(TIR, Tool-integrated reasoning))[8]两阶段数据训练方法，以显著增强模型在复杂推理场景中的表现。同时，微调一个具备广泛世界知识和一阶推理能力的模型，将确保系统能够处理多学科知识的问题分解任务，从而提升其在实际应用中的表现。  其次，数学解题能力的提升不仅涉及从小学到高等数学的知识体系跨越，还需要模型能够解决具体上下文中的复杂问题以及抽象概念的数学证明。这就要求模型具备较强的推理深度和跨领域的知识储备。因此，本研究计划构建一个基于微调增强的多Agent数学解题工作流系统，利用大模型参数高效微调技术(PEFT, Parameter-Efficient Fine-Tuning)[9]和开源训练工具框架高效进行模型训练，并结合开源推理优化工具以及开源应用服务框架进行服务部署简单使用，必要时会加入RAG技术与外部知识库[10]提高‌大语言模型（‌LLM）的生成质量和可靠性‌。通过多模型协同工作，该系统能够有效帮助用户分解和求解复杂的数学问题，提升解题效率，并且也具备正常的基础对话能力，展现出强大的灵活性和智能性。  最后，结合当前大模型的现实应用，这一系统的构建将推动数学教育和科研领域的智能化发展，为不同层次的学习者提供更精准严谨的解题指导。同时，本研究的成果将为未来大模型推理能力的提升方案提供借鉴意义和实践经验，进一步推动大模型技术的进步和应用。 | | | |
| 二、研究现状与文献综述  1、大语言模型推理能力强化的典型范式与技术背景  在大语言模型推理能力的研究中，针对数学和逻辑推理的专门强化逐渐成为研究热点，尤其是在自然语言处理与数理分析结合的场景下，不同模型在数学推理上的表现呈现出多样化特征。目前，学界和工业界已研发出多种针对数学推理的典型模型，为提升模型的推理能力提供了参考。首先，模型是否具备推理能力，已成为生成式大语言模型领域的研究热点，这涉及模型在特定问题场景中对上下文语境的理解与不同学科知识运算的精确性。例如，国内阿里大模型算法团队通义千文发布了Qwen Math[11]的一系列开源模型，经过专门的近端策略优化算法(PPO, Proximal Policy Optimization Algorithms)的训练流程[12]，以应对数学问题的推理需求，而其Coder系列进一步融入了清洗好的各种代码数据进行训练，使得模型不仅可以应对数学推理，还能针对问题生成算法代码解决代码问题[11]。另一方面，Numina团队研发的Numina-Math的数学解题模型则专注于数学竞赛和数学奥赛等高难度问题的解决，通过对逻辑链条和外部工具结合Python代码推理的强化，使用COT和TIR两阶段数据[13]训练的方式，展示了在复杂推理中的显著进展，以29/50的回答正确率成功获得了第一届人工智能数学奥林匹克竞赛(AIMO, AI Math Olympiad)第一名。此外，Deepseek-Math系列模型[14]在深度数学推理上具有突出表现，特别是在高阶数学和物理问题的解答中，展现出较强的演绎推理能力，同时其Coder系列[15]也针对代码生成做了独特的优化。而OpenAI团队近期发布O1系列模型则通过链式推理与数理分析的结合，在复杂高效推理任务中表现优异[16]，成为该领域中的一个重要代表，通过将推理思维链嵌入到训练过程中，让用户可以简化复杂的Prompt提示词解决深度推理问题。这些模型的多样化应用和在数学推理中的表现，不仅为大语言模型推理能力的研究奠定了基础，也为未来的发展提供了重要方向。  2、增强大语言模型推理能力的研究进展  在增强大语言模型推理能力的研究中，提示词工程、链式思维推理、少样本推理(Few-shot prompting)[17]等技术日益成为提升模型表现的重要方法。这些技术通过细化问题的理解过程，再通过部分示例样本促使模型能够更系统性地分步推理，从而显著提升其在复杂任务中的表现。COT推理通过内嵌推理链条，引导模型逐步分解任务，以实现更连贯的推理路径，尤其适用于多步骤逻辑推理与数学问题。在此基础上，少样本推理策略(如TOT , Tree of Thought) [18]通过有限的样本展示推理路径，有效减少了对大规模标注数据的依赖，同时使模型能够更灵活地应对多样化问题。  进一步而言，内置多步连贯推理结构(MCTs, Multi-step Coherent Thought Structures) [19]的引入也为推理能力提升提供了新的技术支撑。MCTs通过构建多步骤的内嵌思维结构，使模型能够自适应地生成逻辑清晰的解题过程，大幅提升了其在抽象与高复杂度问题上的处理能力例如代码生成、数学问题求解等。这些方法在提示词工程中得到了广泛应用，不仅提升了模型对输入提示的响应效果，还显著增强了模型的系统性推理能力，为大语言模型在高阶推理任务中的应用拓展了更为坚实的技术基础。  3、推理能力大语言模型的训练方法与核心技术理念  在大语言模型的训练方法与核心技术理念中，通义千问团队的Qwen2-Math系列，他们在提升模型数学推理能力方面做出了创新性探索。团队首先通过700亿Token的数学预训练数据训练出Qwen2-Math的基座预训练模型，为模型奠定了扎实的数学知识基础。接着，团队引入了COT思维链数据，进一步构建强化的COT推理能力，并在此基础上开展有监督微调（SFT, Supervised Fine-Tuning）[20]，团队还同步训练了强化奖励模型，通过奖励打分机制鼓励模型输出更加合理的解题过程[21]，再进行群体相对策略优化(GRPO, Group Relative Policy Optimization)[14]，进一步提升了模型的决策策略和准确性，最后产出后续的一系列Qwen2-Math-Instruct模型。而在其近期发布的Qwen2.5-Math系列模型[22]中，其预训练语料扩展到了1000亿Token的规模，在保持相同的训练策略的情况下，采用COT和TIR的混合数据加以外部Python编译器进行训练，在其设定的各个评估指标上均达到了SOTA。  Numina团队研发的Numina-Math的数学解题模型采用两阶段训练方法[23]，第1阶段通过在自然语言“数学题 + 解答”的大规模、多样化数据集上微调基础模型，其中每个解答都需套用思维链模板以促使 LLM 进行推理。第 2 阶段则是在工具整合推理的合成数据集上微调第 1 阶段得到的模型，其中每个数学题都分解为一系列推理、Python程序及其输出。其遵循微软的ToRA[8]的做法，提示GPT-4以ToRA格式生成带有代码执行反馈的解答。对这些数据进行微调会产生一个推理代理，它可以通过将自然语言推理和使用Python REPL来计算中间结果结合起来以解决数学问题。  在Quiet-Star[24]的训练过程中，尝试调用一部分推理头生成上下文作为真正的训练输入，同时内置的损失函数修改为DPO[25]相关损失策略函数，在某种层面上在训练过程中已经将推理思维链嵌入到模型的理解语义空间中。M Renze, E Guven等又提出了Self-Reflection的理念[26]，在训练过程中引入特殊Token结合具有反思阶段的数据思维链，让模型学会自我判断即自我反思生成的内容，基于背后的隐藏生成Token进行隐式思考后输出反思评估后的内容。  4、大语言模型的实际应用场景与范例  大语言模型在实际应用中的发展，结合了检索增强生成（RAG）、外部知识库的接入及多Agent协同工作流等技术，为多个领域带来了显著成效。检索增强生成（RAG）通过实时联接大语言模型与外部知识库，使得模型能够在生成过程中获取最新、准确的外部信息，克服了内置知识的时效性局限。例如在医疗问诊和法律咨询等场景中，RAG使模型能实时调用专业知识库，提供更具权威性、针对性的回答，从而提高了回答的精度和可信度。  外部知识库的引入还可以优化模型在特定领域的信息生成过程。通过结构化知识库的预先构建，模型能够更精准地定位所需信息，例如在金融数据分析和技术支持等领域快速响应复杂问题，为用户提供更深层次的分析和咨询服务。  与此同时，多Agent协同工作流的应用使得大语言模型在多步骤任务的处理上更为高效。通过分布式的智能Agent，每个Agent被指定不同的任务模块，如信息检索、数据分析和决策推荐等，形成模块化智能系统。例如在科研助理和自动化流程管理中，协同工作流使系统能够动态调用多个Agent，共同完成复杂任务，提高了处理的流畅性和准确性。RAG、外部知识库与多Agent智能系统的，极大地拓宽了大语言模型的应用场景，为各行业提供了更为全面的技术支撑和应用潜力。 | | | |
| 三、研究的主要内容、预期目标与创新点  1、研究的主要内容  本研究旨在基于微调开源大模型的基础上构建和优化一个多Agent工作流系统，提升大语言模型在推理和问答任务中的综合能力，研究内容包括基座模型的选取、训练数据的准备、训练计划的制定、评估流程的设计以及多Agent系统的构建。  1.1 基座模型的选取  本研究首先调研了各类现有模型的性能和资源需求，最终选取了Qwen2-1.5B-Instruct[27]作为基座模型。Qwen2-1.5B-Instruct在模型规模和计算资源上的平衡性使其在处理复杂推理任务和知识问答任务中表现出色，同时具有较强的扩展性和适应性，而且具备中文能力，为研究搭建了稳固的基础。  1.2 训练数据的准备  为提升模型在不同任务上的表现，研究整合了多源数据，构建了涵盖数学解题COT思维链数据和数学解题TIR数据和一般知识问答的多样化数据集。推理模型的数据主要来源于数理逻辑推理数据集，以确保模型能够精准应对多步骤推理任务；知识问答模型的数据集则聚焦于世界知识、学科基础知识等广泛领域，以确保模型能高效应答多种知识性问题。  1.3 训练计划的制定  在训练计划中，研究将Qwen2-1.5B-Instruct分别进行推理模型和知识问答模型的定向微调。推理模型通过思维链（Chain of Thought, COT）和自回归推理链的数据进一步增强其分步推理能力，帮助其在复杂解题过程中实现逐步推理。而知识问答模型则依托知识对齐强化数据训练或检索增强生成（RAG）策略，通过引入外部知识库提高其问答准确性和知识广度。  1.4 评估模型流程：确定相关指标和计划  研究制定了全面的评估指标，针对推理能力确定数学相关测试指标、针对世界知识的认知设置多维度指标。推理模型的评估重点在于推理的逻辑性和准确率，知识问答模型则关注知识覆盖率和问答的准确性性。多指标的评估方案为模型的迭代优化提供了科学依据。同时评估模型的吞吐速率  1.5 多Agent工作流系统的构建  在多Agent工作流系统的搭建中，研究将推理模型和知识问答模型与外部闭源大语言模型的API接口相结合，使各模型根据任务需求互相协作。系统根据任务属性动态调用相应模型，实现了复杂任务的模块化处理与集成，为系统的可扩展性和高效性提供了坚实的支持。  2、研究的预期目标  本研究旨在实现高效的多Agent工作流系统，在自行构建的评估指标体系中，相较于baseline取得性能提升。该系统应具备高灵活性，能够动态调用多个大模型以适应不同任务需求。  3、研究的创新点  研究创新点主要体现在以下三个方面：   * 自行构建综合性能评估指标，为模型表现提供科学评估标准。 * 自行构建部分数学推理数据集，并通过外部LLM增强MMLU数据集，提升模型推理和知识覆盖能力。 * 基于自训大模型与外部LLM，搭建简易且实用的多Agent工作流系统，为复杂任务处理提供灵活的解决方案。 | | | |
| 四、研究思路、研究方法与研究计划  1、研究思路  本研究的整体思路旨在构建一个高效、灵活的多Agent工作流系统，以提升大  语言模型在知识问答与复杂推理任务中的表现，具体步骤如下：  1.1确定基座模型与评估指标  首先选择适用的基座模型（如Qwen2-1.5B），并制定评估模型表现的综合指  标。评估体系将涵盖逻辑推理、知识覆盖率、任务响应速度等多个方面，为模型在知识问答与一阶推理任务中的表现提供清晰的基准。选择Qwen2-1.5B作为基座模型基于其在推理、问答任务的性能平衡性，能够为后续训练奠定稳定的技术基础。  1.2强化训练世界知识与一阶推理能力模型训练  在训练初期，通过大量包含世界知识和基本逻辑推理的数据集，进行多样  化的知识问答训练基座模型，以提升模型在不同任务场景中的知识广度与推理精准度。这一步为模型奠定了良好的知识基础，提升其一阶推理能力，为后续复杂推理训练打下坚实的基础。  1.3推理能力模型的CoT数据训练  针对推理任务，通过思维链（Chain of Thought, CoT）数据集训练基座模  型，使其在应对多步骤推理任务中表现出逐步推理的能力。CoT数据训练增强了模型在复杂问题上的分步推理能力，使其能够更好地应对多步骤解题和逻辑推演。  1.4推理能力模型的TIR数据深化训练  在经过CoT数据训练后，进一步使用任务导向推理（TIR）数据集对1.3训  练后的模型进行强化训练，以提高模型在任务拆解、问题求解方面的能力。TIR训练旨在让模型在推理过程中更具适应性，能够快速应对复杂的多任务场景。  1.5模型的能力评估  对上述训练好的两个模型设置评估指标框架，大概先以MMLU、GSM8K、HumanEval，从世界知识、数学、代码能力上评估模型能力，对于推理模型可能会涉及到Math等复杂度略高的数学测评集测试推理能力。  1.6部署VLLM模型推理服务与工作流应用服务  当模型具备稳定的推理和问答能力后，用推理加速开源工具，如VLLM[28]，Sglang[29]等，构建稳定、低延迟的推理服务，确保其能够实时响应用户的多类型任务需求。同时调研相关的应用服务开源框架，用于后续系统部署应用的可视化界面和用户操作。  1.7完善工作流整体流程  在整个多Agent工作流系统中，确定整个工作流需要的agent数量以及各个agent需要完成的任务，同时最后引入外部LLM（如GPT、智谱AI等）作为最终审核（critic）Agent，确保在遇到复杂任务时能够进一步优化模型的输出，提高多Agent协作系统的性能和结果准确性。  2、研究方法  本研究采用多层次的优化与协作方法，围绕数据处理、模型训练、推理加速、多  Agent分工协作等方面，提升大语言模型的整体性能和推理效率。具体方法包括：  2.1数据改写、增强、划分与抽样  通过数据改写和增强技术构建高质量的多样化训练数据集，并进行有效的数据划分和抽样，以确保模型在知识问答和推理任务上的泛化能力与适应性。这种数据预处理为模型的准确推理与知识覆盖提供了基础保障。  2.2高效训练框架的使用  采用Llama Factory[30]训练框架工具，加速模型的预训练与微调过程同时简化工程的代码设计工作，通过此框架的资源优化能力和多种类模型适配能力，使模型在多任务训练中保持高效运转，同时有效管理计算资源。  2.3 LoRA 参数高效微调  使用LoRA（Low-Rank Adaptation）[31]进行参数微调，有效节省显存资源并减少硬件开销，从而确保大型语言模型能够在有限的显存条件下达到较高的推理性能。  2.4 高效推理机制  调研采用高效推理机制相关工具，通过减少冗余内存调用加快推理过程，同时在显存利用率方面进行优化，提升推理速度并降低硬件资源需求。  2.5 大模型应用服务部署开源工具  利用大模型应用服务部署开源工具进行模型服务部署，为模型的在线推理提供稳定、高效的部署环境，使其能够在实际应用中提供低延迟的服务响应。  2.6 Prompt Engineering  使用Prompt Engineering技术，并结合Few-shot提示方式，为模型生成更精准的推理路径和响应结果，优化其对特定任务的适应性和解题思路。  2.7自定义评测框架评估模型  通过OpenCompass[32]大模型高效测评框架对模型性能进行全面评估，确保模型的任务完成度和效果。此外，针对复杂数学问题，使用Omni-Math[33]评测框架并进行改写适配，以更好地评价模型在数学推理任务中的表现。  2.8 多Agent协作工作流的计划制定  构建多Agent工作流，不同模型在系统中各司其职：如问题拆分、分析、知识问答和答案审核等。最终输出交由Judge LLM进行进一步审核，以确保高质量的多  骤推理。  2.9量化技术的应用  在推理阶段通过量化技术在略微牺牲精度的前提下加速推理过程，进一步提高推理速度，适应对响应速度要求较高的实际应用场景。  3、研究计划   * 第一阶段，模型训练，选取基座模型后，进行两个模型的训练 * 第二阶段，探索采样参数和Prompt的选择，在构建的测评框架下进行测试 * 第三阶段，调研相关推理工具，衡量部署难度和推理速度 * 第四阶段，调研开源模型框架服务，学习部署相关内容以及私有模型的接入调用 * 第五阶段，完善整个系统的具体流程走向 | | | |
| 五、参考文献  [1] 文森, 钱力, 胡懋地, and 常. J. 数据分析与知识发现, "基于大语言模型的问答技术研究进展综述," p. 1, 2024.  [2] Z. Chen, Y. Li, and K. Wang, "Optimizing reasoning abilities in large language models: A step-by-step approach," 2024.  [3] C.-M. Chan *et al.*, "Chateval: Towards better llm-based evaluators through multi-agent debate," 2023.  [4] L. Huang *et al.*, "A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions," 2023.  [5] J. White *et al.*, "A prompt pattern catalog to enhance prompt engineering with chatgpt," 2023.  [6] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive nlp tasks," vol. 33, pp. 9459-9474, 2020.  [7] J. Wei *et al.*, "Chain-of-thought prompting elicits reasoning in large language models," vol. 35, pp. 24824-24837, 2022.  [8] Z. Gou *et al.*, "Tora: A tool-integrated reasoning agent for mathematical problem solving," 2023.  [9] N. Ding *et al.*, "Parameter-efficient fine-tuning of large-scale pre-trained language models," vol. 5, no. 3, pp. 220-235, 2023.  [10] J. Cui, Z. Li, Y. Yan, B. Chen, and L. J. a. p. a. Yuan, "Chatlaw: Open-source legal large language model with integrated external knowledge bases," 2023.  [11] J. Bai *et al.*, "Qwen technical report," 2023.  [12] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. J. a. p. a. Klimov, "Proximal policy optimization algorithms," 2017.  [13] J. Li *et al.*, "Numinamath: The largest public dataset in ai4maths with 860k pairs of competition math problems and solutions," 2024.  [14] Z. Shao *et al.*, "Deepseekmath: Pushing the limits of mathematical reasoning in open language models," 2024.  [15] Q. Zhu *et al.*, "DeepSeek-Coder-V2: Breaking the Barrier of Closed-Source Models in Code Intelligence," 2024.  [16] J. de Winter, D. Dodou, and Y. B. J. a. p. a. Eisma, "System 2 thinking in OpenAI's o1-preview model: Near-perfect performance on a mathematics exam," 2024.  [17] Y. Wang, Q. Yao, J. T. Kwok, and L. M. J. A. c. s. Ni, "Generalizing from a few examples: A survey on few-shot learning," vol. 53, no. 3, pp. 1-34, 2020.  [18] S. Yao *et al.*, "Tree of thoughts: Deliberate problem solving with large language models," vol. 36, 2024.  [19] M. DeLorenzo *et al.*, "Make every move count: Llm-based high-quality rtl code generation using mcts," 2024.  [20] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. E. J. A. i. n. i. p. s. Hinton, "Big self-supervised models are strong semi-supervised learners," vol. 33, pp. 22243-22255, 2020.  [21] H. Touvron *et al.*, "Llama 2: Open foundation and fine-tuned chat models," 2023.  [22] A. Yang *et al.*, "Qwen2. 5-math technical report: Toward mathematical expert model via self-improvement," 2024.  [23] S. Yin, W. You, Z. Ji, G. Zhong, and J. J. a. p. a. Bai, "MuMath-Code: Combining Tool-Use Large Language Models with Multi-perspective Data Augmentation for Mathematical Reasoning," 2024.  [24] E. Zelikman, G. Harik, Y. Shao, V. Jayasiri, N. Haber, and N. D. J. a. p. a. Goodman, "Quiet-star: Language models can teach themselves to think before speaking," 2024.  [25] R. Rafailov, A. Sharma, E. Mitchell, C. D. Manning, S. Ermon, and C. J. A. i. N. I. P. S. Finn, "Direct preference optimization: Your language model is secretly a reward model," vol. 36, 2024.  [26] M. Renze and E. J. a. p. a. Guven, "Self-Reflection in LLM Agents: Effects on Problem-Solving Performance," 2024.  [27] A. Yang *et al.*, "Qwen2 technical report," 2024.  [28] W. Kwon *et al.*, "Efficient memory management for large language model serving with pagedattention," in *Proceedings of the 29th Symposium on Operating Systems Principles*, 2023, pp. 611-626.  [29] L. Zheng *et al.*, "Efficiently programming large language models using sglang," p. arXiv: 2312.07104, 2023.  [30] Y. Zheng *et al.*, "Llamafactory: Unified efficient fine-tuning of 100+ language models," 2024.  [31] E. J. Hu *et al.*, "Lora: Low-rank adaptation of large language models," 2021.  [32] P. A. Buitrago and N. A. Nystrom, "Open compass: accelerating the adoption of AI in open research," in *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning)*, 2019, pp. 1-9.  [33] B. Gao *et al.*, "Omni-math: A universal olympiad level mathematic benchmark for large language models," 2024. | | | |
| 指导教师的具体意见：  签名： 年 月 日 | | | |
| 院系意见：  负责人签名（或盖章）： 年 月 日 | | | |