



赛道三 - 金融领域的长 思维链压缩

——丛雨酱队

Contents

目录 ▶

01

团队介绍

02

方案综述

03

实现方案

04

总结及展望

01 团队介绍

赛题三：金融领域中的长思维链压缩

队名：丛雨酱

成员介绍:



队长-罗铭

新长城科技有限公司

本科就读于电子科技大学，研究生毕业于南洋理工大学，2024年工作至今，大模型应用领域新晋牛马一位。



队员-PC


24G显存笔记本，同2024年工作至今，罗铭同学的犁地工具。
外附4*H100显卡服务器用于LLM模型的训练和微调。



02 方案综述

基于渐进式优化策略的创新方案

——从提示词工程到参数微调到偏好优化



效率问题

- **实时性要求高**：金融咨询需要快速响应，长思维链导致推理时间过长
- **决策时效性**：用户等待数分钟可能错过最佳决策时机
- **用户体验差**：响应延迟严重影响实际应用推广

成本问题

- **资源消耗大**：冗长输出使计算资源消耗增加数倍
- **API成本高**：长文本生成显著提升调用费用
- **运营成本**：高成本限制了模型在金融行业的广泛应用

稳定性问题

- **逻辑偏差**：长文本生成容易出现逻辑不连贯
- **错误累积**：推理链过长导致错误传播和放大
- **决策风险**：错误建议可能影响用户的金融决策

泛化能力

过拟合问题导致在其他问题上效果变差

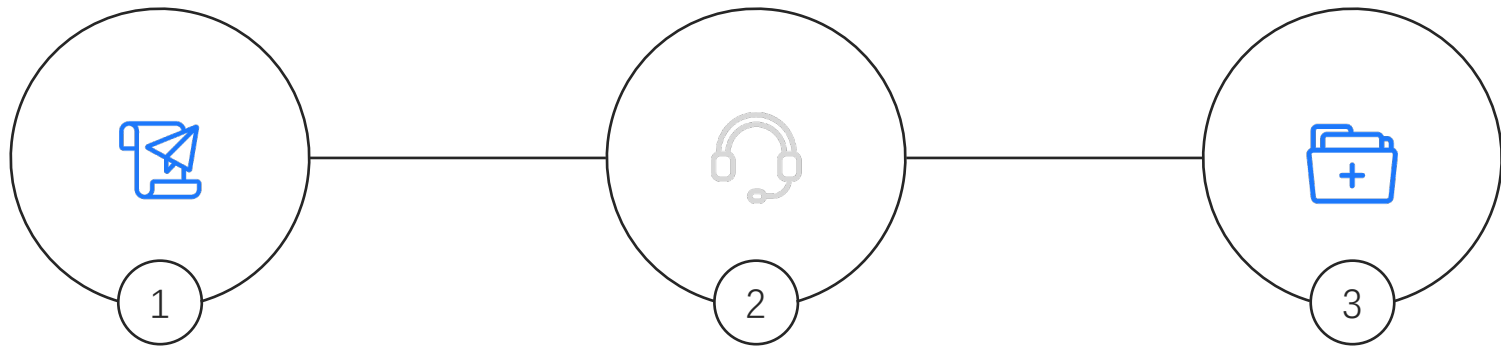
平衡点

问题回复准确率和压缩率难以兼顾

信息损失

压缩思维链条过程中丢失
关键信息





Prompt

- 零参数优化
- 成本最低
- 效果有限

普通lora微调

- 轻量训练
- 适度提升
- 渐进改善

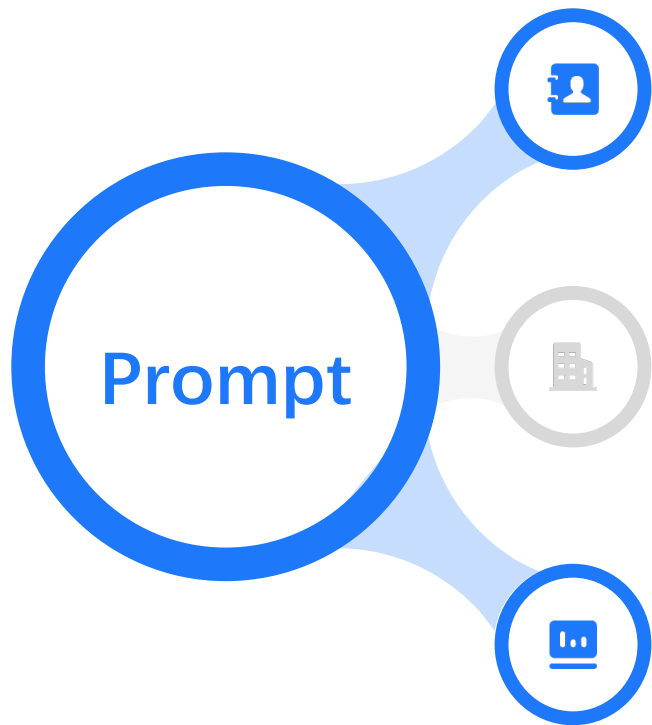
DPO微调

- 核心突破
- 显著效果
- 对比学习

03 实现方案

基于渐进式优化策略的创新方案

——从提示词工程到参数微调到偏好优化



角色定义法：

为模型分配特定的专业角色身份——你是惜字如金的金融领域专家，擅长生成简明扼要的回答。

直接约束法：

明确输出的格式和回答的长度——保证回复在xx字数内，且结构化回复

思维引导法：

引导模型的思考路径和重点——请先快速识别问题的核心，然后直接分析答案选项，不要在选项之间犹豫多次分析，干脆利落地给出分析的结果。

01 实验结果

- 思维链压缩：不显著，仅减少部分推理中的犹疑部分，分析过程仍旧较长。
- 准确率变化：部分回答结果受限于约束会导致错误

02 问题分析

- 模型仍倾向生成冗长完整回答
- 需要更深层的模型行为调整

第二轮优化：LoRA微调



01

框架选择

选择ms-swift框架，有可视化webui，且内置模型微调和驯良方法较完善。

02

数据准备

1000条蒸馏和人工标注得到的基于测试问题的高质量问答对

03

微调参数

- lora_rank: 8
- lora_alpha: 32
- learning_rate: 5e-4
- num_train_epochs: 2

01 实验结果

- 思维链压缩：可观察到压缩效果，效果不够显著，平均约压缩35%
- 准确率变化：准确率提升至80%以上

02 问题分析

- 对基座QWEN-4B模型微调有效，但压缩效果仍需增强，同时准确率需要更加稳定
- 需要尝试其他模型微调的方法

第三轮优化：DPO(Direct Preference Optimization)

DPO

效果显著

DPO在推理和对话任务中表现优异，能够显著直观地展现想要的效果

数据高效

对标注数据要求相对较低，充分利用收集到的有限的金融领域数据

直接优化偏好

无需奖励模型设计，直接学习人类偏好，精准识别冗余推理

稳定性强

避免RL的不稳定性，确保优化局限于压缩cot和提高准确性

偏好数据构建

定义“好”回答

- 推理链简洁但完整
- 逻辑步骤清晰连贯
- 关键信息无遗漏
- 结论准确可靠
- 符合金融专业规范

01

定义“坏”回答

- 推理链冗长繁琐
- 包含无关信息
- 逻辑跳跃或错误
- 重复性描述过多
- 缺乏关键推理步骤

02

数据拓展策略

- 外部金融问题添加
- 保证至少五个reject
- 模型生成对比回答

03

实践方案

- 基于DeepSeek满血蒸馏
- 人工做题得到最精简答案

04

Dpo Beta

- $\beta = 0.05$: 偏好信号弱, 压缩效果不明显
- $\beta = 0.1$: 适中偏好
- $\beta = 0.3$: 过度压缩, 准确率下降

训练轮次

- Epoch 1: 压缩率 45%, 准确率 88%
- Epoch 2: 压缩率 80%, 准确率 95%
- Epoch 3: 压缩率 82%, 准确率 93% (过拟合)

参数配置

学习率

- LR= $5e-5$: 收敛慢, 训练效率低
- LR= $1e-4$: 稳定收敛, 两轮最优
- LR= $5e-4$: 收敛过快, 过拟合

批次与梯度

- BS=4, GA=4: 16批次, 内存占用过高
- BS=1,GA=16: 16批次, 内存占用低, 稳定
- BS=1,GA=32: 32批次, 内存占用低, 收敛过慢

01 实验结果

- 思维链压缩：从平均35%压缩提升至90%以上压缩
- 准确率变化：准确率提升至95%以上
- 泛化能力：在非测试集数据上测试也有较明显效果，非金融问题推理过程也保持精简

02 原因分析

- 偏好数据质量：构建了高质量的好坏回答对比数据
- 参数配置优化：Beta=0.1实现了最佳偏好强度平衡
- 训练策略得当：2轮训练避免了过拟合问题
- 数据多样性：5种不同类型的拒绝回答确保对比效果


优化阶段	方法	思维链压缩	准确率	训练成本	核心突破
第一阶段	提示词工程	不显著	略微降低	无成本	无
第二阶段	Lora微调	35%	约70%	较低成本	验证参数优化可行性
第三阶段	Dpo全量微调	>85%	>95%	较高成本	偏好优化突破



04 总结与展望

基于渐进式优化策略的创新方案

——从提示词工程到参数微调到偏好优化



渐进式优化策略

- 核心理念： 从零参数到全量微调的三阶段递进策略
- 技术路径： 提示词工程 → LoRA微调 → DPO全量微调
- 创新价值： 首次将渐进式思想应用于思维链压缩，有效控制成本与风险

DPO思维链压缩

- 技术突破： 将 Direct Preference Optimization应用于思维链压缩任务
- 方法创新： 构建“简洁vs冗长”的偏好对比数据集
- 核心优势： 无需复杂奖励函数设计，直接优化输出偏好
- 显著成果： 实现85%思维链压缩率，准确率提升至95%+

多样化数据

- 数据工程创新： 自适应温度采样生成多样化拒绝答案
- 迭代优化策略： 数据驱动的持续改进闭环
- 技术实现： 温度递增策略 + 多模式答案提取



数据依赖性挑战

- 高质量标注数据需求量大
- 跨领域适应能力受限
- 持续数据更新成本高



计算效率瓶颈

- DPO训练资源需求高
- 推理速度提升有限
- 大模型扩展成本激增



质量平衡难题

- 压缩率vs准确率权衡
- 复杂问题处理能力受限
- 推理可解释性降低

自适应压缩
根据问题复杂度动态调整压缩策略



混合优化
结合多种微调方法的优势，开发自适应策略



增量学习
实现增量式DPO训练，持续优化模型性能



跨领域迁移
数据集扩展到法律、医疗等其他专业领域



可优化方向

感谢聆听！

— 共创金融AI智能化未来

汇报人：丛雨酱 - 罗铭