基于指令微调的数学推理任务探索

Huazheng Zeng
School of Computer Science and Technology
Fudan University
220 Handan Rd, Shanghai, 200433
22302010022@m.fudan.edu

Yang Wang
School of Computer Science and Technology
Fudan University
220 Handan Rd, Shanghai, 200433
223020100@m.fudan.edu

Yuxuan Xie
School of Computer Science and Technology
Fudan University
220 Handan Rd, Shanghai, 200433
223020100@m.fudan.edu

Abstract

数学推理是评估人类智力基本认知能力的基石。近年来，针对自动解决数学问题的大型语言模型 LLMs 得到了显著发展。我们基于 Qwen2.5-0.5B-instruction，探索了全参数 SFT，全参数 LoRA 指令微调的效果，同时，通过数据集增强，提高了模型的性能；我们还在解码阶段加入多轮投票，得到了相关的测试结果。

1 Introduction

数学推理是评估人类智力的核心任务之一。随着大规模语言模型（LLMs）的不断发展，基于自然语言处理的数学推理任务已成为一个重要的研究领域。尽管传统的数学推理方法依赖于符号计算和规则推导，近年来，深度学习，尤其是大型语言模型，已经在自动推理方面取得了显著的进展。

本研究基于 Qwen2.5-0.5B-instruction 模型，探讨了指令微调（instruction fine-tuning）方法在数学推理任务中的应用，重点研究了全参数微调（SFT）和 LoRA 微调两种技术的效果。此外，我们还结合了数据增强策略，通过 DeepSeek API 扩展训练数据集，以提升模型在复杂数学推理任务中的表现。为进一步增强模型的推理能力，我们在解码阶段引入了多轮投票机制，旨在通过集成多次推理结果来提高推理的稳定性和准确性。

本文的主要贡献包括：提出了 Qwen2.5 模型在数学推理任务中的应用，探索了不同微调技术的效果，并评估了数据增强和投票机制在该任务中的有效性。

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

2 RelatedWork

3 Method

在本研究中，我们采用了多种方法来提升 Qwen 系列模型在数学推理任务上的表现。我们的方法结合了以下主要技术：

- 全参数微调，增强模型在特定任务上的适应能力；
- LoRA 微调，优化计算资源效率的同时能取得较好的性能；
- 使用 DeepSeek 在原有数据集上生成回答以对其进行扩充；
- 多数投票推理机制，以期望提升推理过程的稳健性；
- 数据增强策略，进一步提升模型的泛化能力。

3.1 **全参数微调**

全参数微调对模型的全部参数进行优化，以提升模型在特定任务 (如数学推理任务) 上的适应能力：

$$\mathcal{L}_{full} = \frac{1}{N}\sum_{i=1}^{N} \ell(f(X_i;\Theta), y_i)$$

其中，X_i 表示输入问题，y_i 表示期望输出，Θ 为模型的全部参数。

3.2 LoRA **微调**

为减少全参数微调带来的计算成本，同时降低其灾难性遗忘风险，以提高模型对新任务的迁移能力，我们采用了 LoRA 技术进行微调。该技术在微调过程中固定大部分预训练参数，仅更新小规模的适配器矩阵：

$$\Theta' = \Theta + A \cdot B$$

假设 $\Theta \in \mathbb{R}^{d\times k}$，则 $A \in \mathbb{R}^{d\times r}$、$B \in \mathbb{R}^{r\times k}$，且 $r \ll min(d,k)$。

3.3 DeepSeek **生成训练数据**

DeepSeek 是一个较为强大的大语言模型，其在数学推理方面的能力也较为出色。故我们尝试调用 DeepSeek api 对 GSM8K 和 MATH 训练集生成回答以扩充训练数据。由于 DeepSeek 生成的回答能够包含较为详细的解答步骤，故我们期望通过这种类似蒸馏的过程来指导 Qwen2.5B 提升数学推理方面的能力。

3.4 **多数投票推理机制**

在推理阶段，我们通过多数投票机制以期增强模型推理的鲁班性。假设模型的正确率为 p，在多数投票推理阶段生成 r 次回答，则期望的正确回答数量为 $p \times r$。若考虑最佳情况，其余错误回答均不一致，那么希望从该机制中获得正确回答，则需要 $p \times r \geq 2$，即 $p \geq \frac{2}{r}$。但在实际测试中，该方法反而降低了模型的准确性，这可能是由于模型生成的回答与最佳情况的假设不符，生成了相同的错误答案。

3.5 **数据增强**

由于数据在模型微调中占有较为重要的作用，所以我们采用了《MuggleMath: Assessing the Impact of Query and Response Augmentation on Math Reasoning》一文中提供的 AugGSM8K 和 AugMATH 数据集对模型进行微调，并对其进行过滤以保留高质量训练数据，以此让模型更好地学习数学推理能力，并取得了较好的效果。

4 Evaluation

4.1 基于 SFT 与 LoRA 的微调比对

在本部分中，我们比较了全参数微调（SFT）和 LoRA 微调的效果。我们基于 Qwen2-0.5B 模型对 SFT 和 LoRA 两种微调方法进行了测试，并对比了显存占用、训练时间和训练效果。实验结果如下表所示：

Table 1: 基于 GSM8K 数据集的 SFT 与 LoRA 微调对比

指标	SFT	LoRA
显存占用	较高	较低
训练时间	较长	较短
训练效果	较好（在大规模任务中）	较好（在小规模任务中，适应性强）

分析：全参数微调（SFT）在训练大规模任务时效果较好，但由于需要更新模型的所有参数，其显存占用较高且训练时间较长。相比之下，LoRA 微调仅更新少部分适配器矩阵，显著降低了显存占用和训练时间，同时在小规模任务中表现优异，且能够较好地适应不同任务。

4.2 基于增强数据集训练与正常训练的微调比对

在这部分实验中，我们对比了基于增强数据集训练和正常训练的微调效果。我们通过 DeepSeek 生成的扩充数据集以及 AugGSM8K 和 AugMATH 数据集进行了微调，并评估了训练效果。

Table 2: 基于增强数据集训练与正常训练的微调比对

指标	**正常训练**	**增强数据集训练**
显存占用	较低	较低
训练时间	较短	较长
训练效果	较好（在常规数据上）	较好（在数据多样性和鲁棒性上）

分析：增强数据集的使用能够显著提升模型在数据多样性和鲁棒性上的表现。尽管增强数据集训练的时间较长，但其在训练效果上超越了常规训练，尤其是在数学推理等复杂任务中，增强数据集对提升模型的泛化能力具有显著作用。

4.2.1 DeepSeek 生成的数据集

我们使用 DeepSeek 生成的回答来扩充 GSM8K 和 MATH 训练集。DeepSeek 能够生成带有详细解答步骤的回答，这种类似蒸馏的过程帮助 Qwen2.5B 模型提升了数学推理方面的能力。经过扩充数据集的训练后，Qwen2.5B 模型在数学推理任务上获得了较好的效果。

4.3 Scaling Law 的验证

为了验证模型性能随模型规模增长的趋势，我们对比了 0.5B 与 3B 模型的表现。实验结果显示，随着模型规模的增加，训练效果得到了提升，但相应的训练时间和显存占用也显著增加。

Table 3: Scaling Law 的验证（0.5B 与 3B 模型对比）

指标	0.5B **模型**	3B **模型**
显存占用	较低	较高
训练时间	较短	较长
训练效果	较好（在小规模任务中）	较好（在大规模任务中）

分析：通过 Scaling Law 的验证，我们观察到更大的模型（如 3B 模型）在大规模任务中表现出了更好的训练效果。然而，随着模型规模的增加，显存占用和训练时间也相应增加，限制了模型的实际应用。因此，在选择模型规模时，需要权衡效果和资源消耗。

4.4 **多轮投票机制的比对**

在多轮投票机制中，我们期望通过生成多个回答并进行投票来提升推理的鲁棒性。实验结果表明，多轮投票机制并未显著提升模型的准确性。反而，由于生成的回答存在重复的错误答案，投票机制的效果不如预期。

Table 4: 多轮投票机制的比对（有无）

指标	**有多轮投票机制**	**无多轮投票机制**
显存占用	较高（需要存储多个投票结果）	较低
训练时间	较长（多轮投票需要更多计算）	较短
训练效果	略微提升（但效果不明显）	较差（可能会受到单次预测的影响）

分析：尽管多轮投票机制理论上可以提高鲁棒性，但在实际实验中，该机制未能显著提升模型的准确性。我们认为，这可能是因为多个投票生成的答案存在重复的错误，从而影响了整体效果。此外，多轮投票增加了显存占用和计算开销，未能带来足够的性能提升。

5 Conclusion

本研究探索了多种方法提升 Qwen 系列模型在数学推理任务上的表现。实验结果表明，基于全参数微调（SFT）和 LoRA 微调的模型在数学推理任务中表现出色，尤其是在特定任务下，通过优化计算资源能够显著提升模型的推理能力。同时，数据增强策略有效提升了模型在复杂任务中的泛化能力，尤其是在基于 DeepSeek 生成的扩充数据集进行训练后，模型的数学推理能力得到了明显提升。

然而，实验结果也表明，多轮投票机制在提高推理精度方面的效果并不显著，可能是由于生成的多个答案存在重复的错误。因此，在模型推理过程中的策略选择仍需进一步优化。此外，随着模型规模的增加，显存占用和训练时间也呈现上升趋势，如何在性能和资源消耗之间找到最佳平衡仍然是一个值得关注的研究方向。

未来的研究可以从以下几个方面进行拓展：进一步优化微调策略，探索不同类型的数据增强方法，提升多轮投票机制的鲁棒性，探索如何结合符号推理与深度学习模型来增强数学推理能力，以及多步骤的微调。

6 References

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), Advances in Neural Information Processing Systems 7, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System. New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. Journal of Neuroscience 15(7):5249-5262.