

基于大语言模型的多路径一致性蒸馏

蔡丽琼¹ 霍雨佳^{1*} 袁赛仙¹
CAI Liqiong HUO Yujia YUAN Saixian

摘要

提升小型语言模型在复杂推理任务中的表现,是当前自然语言处理领域的一个关键挑战。基于此,文章提出了一种基于知识蒸馏的推理增强方法(CGRD),旨在提升学生模型在复杂推理任务中的表现。具体而言,采用 GPT-3.5-turbo-instruct 作为教师模型,通过 API 调用生成多样化的推理路径,从中筛选出一致性最高的路径用于指导学生模型的训练。实验结果表明,CGRD 方法在 AddSub 和 Date Understanding 两个数据集上均显著优于传统的方法。还使教师模型在 AddSub 和 Date Understanding 数据集上的准确率分别提升了 15.11 和 10.05 个百分点,更重要的是,在 GPT-2 和 T5 等不同架构的一系列小规模学生模型上均取得了一致性的性能提升,充分证明了 CGRD 在增强小模型推理能力方面的有效性,通过提升教师模型生成推理路径的质量,有效增强了学生模型的推理能力和任务表现。

关键词

蒸馏;大型语言模型;思维链;一致性推理

doi: 10.3969/j.issn.1672-9528.2026.01.029

0 引言

近年来,随着大规模语料库上预训练的 Transformer 模型的广泛应用,预训练语言模型(PLM)在自然语言处理(NLP)领域展现了显著的优势^[1-2]。研究表明,当模型的参数规模超过一定阈值时,这些大规模语言模型在性能上实现了显著提升,并展现出小规模模型(如 BERT^[3])无法实现的能力,例如上下文学习和推理能力。为区分不同规模的语言模型,学术界和工业界引入了“大语言模型”(LLMs)这一术语,这类模型通常拥有数百亿至数千亿参数。在人工智能的发展中,专有的大语言模型(如 GPT-3.5、GPT-4、Gemini 和 Claude 作为开创性技术,重塑了对自然语言处理(NLP)的理解。

其核心优势在于突显能力^[4-5],即模型展示出超出明确培训目标的能力,使其能够高效处理各种复杂任务。然而,尽管像 GPT-4、GPT-4o 和 Gemini 等专有 LLM 展示了卓越的性能,也存在明显缺点,尤其是在考虑开源模型的优势时。使用这些模型往往伴随着高昂的费用和访问限制,且数据隐私和安全问题^[6]也成为一个重要考量。

与此相比,开源模型,如 LLaMA 在可访问性和适应性方面具有显著优势。未许可使用或使用限制,开源模型使更多的用户能够受益,从个人研究人员到小型组织都可以轻松使用。然而,开源 LLM 也存在一些不足,主要体现在模型规模和资源上,相较于专有模型,通常具备较少的参数,这导致其在处理复杂任务时性能较弱^[7]。

因此,知识蒸馏作为一种有效的模型压缩技术,逐渐成为提升大语言模型应用效率的关键方法之一。通过蒸馏,小型学生模型能够借助教师模型的知识,接近甚至超越大模型的性能,从而在减少计算成本的同时提高推理效率。Bucilă

1. 贵州民族大学数据科学与信息工程学院 贵州贵阳 550025
[基金项目] 贵州省教育厅自然科学研究项目(黔教技[2022]047号)(贵州省高等学校智慧教育工程研究中心支持完成);
贵州民族大学自然科学研究基金(GZMUZK[2021]YB22)

- [8] LI J Q, WANG D H. 2D convolutional stochastic configuration networks[J/OL]. Knowledge-based systems, 2024[2025-12-31]. <https://dl.acm.org/doi/10.1016/j.knosys.2024.112249>.
- [9] DENG X G, ZHAO Y, ZHANG J, et al. A holistic global-local stochastic configuration network modeling framework with antinoise awareness for efficient semi-supervised regression[J]. Information science, 2024, 661:120132.

【作者简介】

袁赛仙(1999—),女,贵州毕节人,硕士,研究方向:多源数据融合、小样本学习。

姚雪梅(1985—),通信作者(email:yaomei0119@126.com),女,云南大理人,博士,副教授、硕士生导师,研究方向:多源数据融合及大数据挖掘与分析。

(收稿日期:2025-09-18 修回日期:2026-01-12)

等^[8]首次提出了模型压缩的概念,随后 Hinton 等^[9]在 2015 年进一步定义了知识蒸馏,通过教师模型输出的软标签与真实标签共同监督学生模型的训练。

在推理能力方面,思维链(chain-of-thought, CoT)提示是一种通过引导模型一步步生成推理过程的技术,其实现过程如图 1 所示。思维推理能力不仅是人类智能的核心特征,也是人工智能领域长期追求的关键目标。随着大型语言模型逐步展现出此类能力,思维推理能力在人机协作中发挥越来越重要的作用,为实现更深层次的智能服务与决策辅助提供了可能。思维链的概念由 Wei 等^[4]于 2022 年提出,研究表明思维链能够显著增强大语言模型的推理能力,尤其在数学推理和常识推理任务中表现突出。近年来,随着大语言模型(LLMs)在推理任务中的表现逐步得到认可,思维链技术被广泛应用于各种推理任务中。本文的目标是提升教师模型生成推理数据的质量,并增强学生模型的推理能力,从而缩小教师模型与学生模型之间的差距。Ho 等^[10]提出了一种名为 Fine-tune-CoT 的方法,该方法旨在利用非常大的语言模型的思维链推理能力来指导小模型解决复杂任务,为本研究提供重要借鉴。

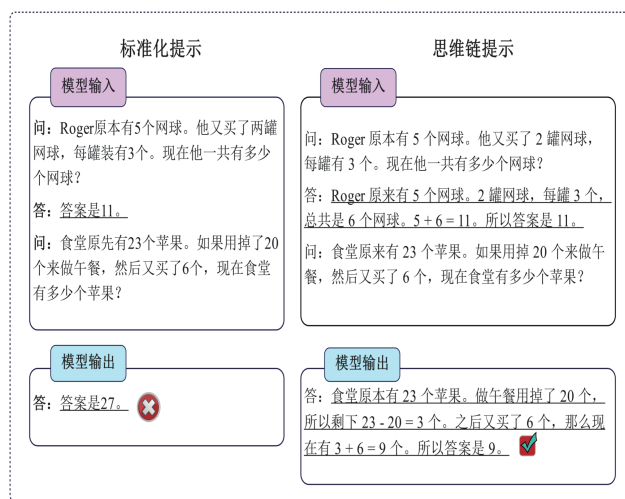


图 1 思维链提示技术图

1 模型描述

CGRD 方法的目标是在保证轻量级模型效率的同时,尽可能保留大规模语言模型(如 GPT-3.5-turbo-instruct)所具备的复杂推理能力,从而提升学生模型在各类推理任务中的表现。CGRD 的核心思路是利用结构化提示引导教师模型针对同一问题生成多条不同的推理路径,再通过一致性投票机制筛选出最可信的结果,并将其作为高质量监督信号用于学生模型的训练。这样不仅能够为蒸馏过程提供更丰富的语义层次信息,还在一定程度上提高了模型训练的稳定性与可解释性。

1.1 思维链提示

本文从问题库中先进行随机采样选取少量样本,对这些样本生成思维链去提示教师模型。为确保提示信息的准确性,每个生成的推理示例都会经过正确性验证;如果发现推理错误,则重新调用接口生成新推理。

1.2 自治性方法与答案聚合

本文对采样出的样本思维链推理来提示教师模型。教师模型通过温度采样生成一组不同的推理路径,为了从多个推理路径中选择最优推理路径和推理答案,本文引入了自治性方法^[11],该方法通过评估多个推理路径的一致性来选择最终的推理路径和推理答案。具体过程如下:

对于每个问题,教师模型生成的推理路径与答案属于一个固定的答案集 $A = \{a_1, a_2, \dots, a_m\}$, 其中 $i \in \{1, 2, \dots, m\}$ 表示从解码器中采样的 m 个候选输出。

每个推理路径都对应着一个潜在变量,该潜在变量表示推理路径的标记序列。对于每一个推理路径,得到一个对应的答案,即:

$$P(r_i, a_i | p, q) = P(r_i | p, q) \cdot P(a_i | r_i, p, q) \quad (1)$$

式中: $P(r_i | p, q)$ 表示推理路径的条件概率; $P(a_i | r_i, p, q)$ 表示给定推理路径后生成推理答案的概率。

根据每个推理路径与答案对的联合概率 $P(a_i | r_i, p, q)$, 对每个推理路径及其推理答案进行加权,得到最终的聚合答案。最终的答案可以通过归一化每个推理路径的条件概率来计算:

$$P(r_i, a_i | p, q) = \frac{P(r_i | p, q) \cdot P(a_i | r_i, p, q)}{\sum_{i=1}^m P(r_i | p, q) \cdot P(a_i | r_i, p, q)} \quad (2)$$

根据加权后的概率值,选择一致性最高的推理路径和答案对参与后续实验。

1.3 多任务学习

本文将这些生成的推理路径和答案作为学生模型的训练数据。为了有效地微调学生模型,本文采用了多任务学习:推理路径损失、答案损失。这两个任务的损失共同作用,指导学生模型在训练过程中掌握与教师模型一致的推理路径。

推理路径损失度量了学生模型生成的推理路径与教师模型生成的推理路径之间的差异。

$$L_r = -\sum_{t=1}^T \log P_s(w_t | w_{1:t-1}) \cdot P_t(w_t | w_{1:t-1}) \quad (3)$$

式中: $P_s(w_t | w_{1:t-1})$ 和 $P_t(w_t | w_{1:t-1})$ 分别是学生模型和教师模型在时间步生成的 token 分布。

答案损失度量了学生模型生成的答案与教师模型生成的答案之间的差异。本文使用交叉熵损失函数来度量学生生成

的答案与目标答案之间的差异:

$$L_a = -\log P_s(a^*|q) \quad (4)$$

式中: a^* 是教师模型生成的答案; $P_s(a^*|q)$ 是学生模型生成答案的概率。

最终目标函数通过两个部分损失来加权和:

$$L_t = \lambda_1 L_r + \lambda_2 L_a \quad (5)$$

式中: λ_1 、 λ_2 是超参数。

2 实验与结果分析

在本节中,在两种复杂的推理任务中评估了本文方法。通过实验验证了本文方法的有效性。首先,证明了本文方法对蒸馏性能的影响。其次,与普通蒸馏方法相比,本文方法相较于 Ft-CoT 方法获得了更好的性能,并减少了教师和学生模型之间的分布差距。

2.1 数据集

本文在两个数据集上进行了实验,来衡量模型的推理性能。这两个数据集分别是:“AddSub”“Date Understanding”。AddSub^[12]数据集用于算术推理;Date Understanding^[13]用于逻辑推理。如表 1 所示。

表 1 本文实验数据集的完整统计信息

数据集	总样本数	License	参考文献
AddSub	395	Unspecified	[13]
Date Understanding	369	Apache-2.0	[14]

具体来说,AddSub 数据集来自经典的数学世界问题库^[13],包含更简单的问题,不需要多步计算来解决任务;Date Understanding 数据集是从 BIG-bench 工作^[14]中选择一个评估集,Date Understanding 此任务旨在测试语言模型通过询问一天的日期来理解日期的能力,要求模型从给定的上下文中推断日期。尽管这些任务较人工而言很容易,但对许多语言模型来说却很难。

对于每个数据集,本文以 7:3 的比例来划分数据集为训练集和测试集。划分 70% 的数据用于模型训练,剩下的 30% 作为测试集。

2.2 实验设置

在本实验中,使用 GPT-3.5-turbo-instruct 作为教师模型来生成正确的推理数据。从温度 $T \in [0.6, 0.8]$ 进行取样。考虑到实验成本问题,在本实验中,每个训练样本,本文采样 $N=5$ 个推理路径。学生模型使用了不同尺寸大小的 GPT-2、T5,分别是 GPT2_Small、GPT2_Medium、T5_small、T5_Base,模型介绍如表 2 所示。

表 2 模型配置

模型类型	名称	参数量
教师模型	GPT-3.5-turbo-instruct	175 B
	T5_small	60 MB
学生模型	T5_Base	220 MB
	GPT2_Small	124 MB
	GPT2_Medium	355 MB

2.3 实验细节

为了评估模型的性能,本文使用准确率(ACC)作为性能指标。批量大小设置为 32,训练 50 个 epoch,使用 AdamW 优化器^[15]更新参数,模型的学习率设置为 $\{1e-4, 2e-4, 3e-4, 5e-4\}$ 。本文使用单个 NVIDIA RTX L20 GPU 实现实验,使用 Python 3.8 和 PyTorch 2.0.0 深度学习框架,并在 CUDA11.8 环境中针对 GPU 性能进行了优化。

2.4 实验结果

为全面评估 CGRD 方法的有效性,本研究在 AddSub 和 Date Understanding 两个推理数据集上进行了实验。实验结果如表 3~4 所示,CGRD 方法在所有测试的学生模型上均显著优于传统的微调思维链基线,证明了 CGRD 方法的有效性和通用性。由表 3~4 中可以看出,本文提出的 CGRD 方法都表现出了最佳的性能。在 AddSub 和 Date Understanding 两个数据集上都优于 Ft-CoT 方法。

表 3 在 GPT2 上的实验结果(ACC: %)

模型	方法	AddSub	Date Understanding
GPT-3.5-turbo-instruct	Ft-CoT	76.71	66.31
	ours	91.82	76.36
GPT2-Small	Ft-CoT	4.22	17.12
	ours	5.85	33.3
GPT2-Medium	Ft-CoT	5.88	28.83
	ours	8.66	34.35

表 4 在 T5 上的实验结果(ACC: %)

模型	方法	AddSub	Date Understanding
T5_Small	Ft-CoT	5.04	77.48
	ours	8.36	80.18
T5-Base	Ft-CoT	7.56	80.18
	ours	9.24	82.88

在学生模型上,GPT2-Small 与 GPT2-Medium 在 AddSub 数据集上分别达到 5.85% 与 8.66%,在 Date Understanding 数据集上分别达到 33.3% 与 34.35%;T5_Small 与 T5-Base 在 AddSub 数据集上分别达到了 8.36%、9.24%,Date

Understanding 数据集上分别达到了 80.18%、82.88%。从实验结果可以看出, 均优于基准方法, 证明了本文的 CGRD 方法的有效性。

3 总结

本文提出的 CGRD 方法, 通过引入一种基于高质量推理路径的知识蒸馏框架, 显著提升了学生模型在数学推理和日期理解任务上的性能。首先从问题数据库中随机采样样本, 利用教师模型生成多样化的推理路径, 并通过一致性筛选机制选择最优路径用于蒸馏过程。实验结果显示, CGRD 方法在 AddSub 和 Date Understanding 数据集上均取得了显著提升, 这些结果充分证明了本文提出的 CGRD 方法在提升小模型推理能力方面的有效性和优越性。

参考文献:

- [1] QIU X P, SUN T X, XU Y G, et al. Pre-trained models for natural language processing: a survey[J]. Science China technological sciences, 2020, 63(10): 1872-1897.
- [2] HAN X, ZHANG Z Y, DING N, et al. Pre-trained models: past, present and future[J]. AI open, 2021, 2: 225-250.
- [3] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C] // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Brussels: ACL, 2019: 4171-4186.
- [4] WEI J, WANG X Z, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[C] // NIPS'22: Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM, 2022: 24824-24837.
- [5] XU X H, TAO C Y, SHEN T, et al. Re-reading improves reasoning in large language models[C] // Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Brussels: ACL, 2024: 15549-15575.
- [6] WU X D, DUAN R, NI J B. Unveiling security, privacy, and ethical concerns of ChatGPT[J]. Journal of information and intelligence, 2024, 2(2): 102-115.
- [7] ZHENG L M, CHIANG W L, SHENG Y, et al. Judging llm-as-a-judge with MT-bench and chatbot arena[J]. Advances in neural information processing systems, 2023, 36: 46595-46623.
- [8] BUCILĂ C, CARUANA R, NICULESCU-MIZIL A. Model compression[C] // KDD'06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2006: 535-541.
- [9] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-09)[2025-12-02]. <https://doi.org/10.48550/arXiv.1503.02531>.
- [10] HO N, SCHMID L, YUN S Y. Large language models are reasoning teachers[C] // Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Brussels: ACL, 2023: 14852-14882.
- [11] WANG X Z, WEI J, SCHUURMANS D, et al. Self-consistency improves chain of thought reasoning in language models[EB/OL]. (2023-03-07)[2025-12-02]. <https://doi.org/10.48550/arXiv.2203.11171>.
- [12] HOSSEINI M J, HAJISHIRZI H, ETZIONI O, et al. Learning to solve arithmetic word problems with verb categorization[C] // Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). Brussels: ACL, 2014: 523-533.
- [13] KONCEL-KEDZIORSKI R, ROY S, AMINI A, et al. MAWPS: a math word problem repository[C] // Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: human language technologies. Brussels: ACL, 2016: 1152-1157.
- [14] SRIVASTAVA A, RASTOGI A, RAO A, et al. Beyond the imitation game: quantifying and extrapolating the capabilities of language models[EB/OL]. (2023-06-12)[2025-12-02]. <https://doi.org/10.48550/arXiv.2206.04615>.
- [15] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[EB/OL]. (2019-01-04)[2025-12-02]. <https://doi.org/10.48550/arXiv.1711.05101>.

【作者简介】

蔡丽琼 (2000—), 女, 云南昆明人, 硕士, 研究方向: 大语言模型。

霍雨佳 (1982—), 通信作者 (email: huo.yujia113@foxmail.com), 男, 贵州遵义人, 博士, 副教授、硕士生导师, 研究方向: 深度学习、智慧教育、自然语言处理。

袁赛仙 (1999—), 女, 贵州毕节人, 硕士, 研究方向: 多源数据融合、小样本学习。

(收稿日期: 2025-09-23 修回日期: 2026-01-06)