基于课程思维链的大语言模型蒸馏方法

余 翔[1] 霍雨佳[1]
YU Xiang HUO Yujia

摘 要 当前大语言模型（LLMs）虽擅长复杂推理，但高计算需求限制其实际落地，知识蒸馏成为将 LLMs 能力迁移至 T5 等小模型的关键技术。然而传统蒸馏多聚焦教师模型最终输出，忽略思维链这一核心内部推理过程，导致小模型仅“记答案”不“懂逻辑”，推理性能受限。同时，课程学习“先易后难”的特性虽能优化知识吸收，却鲜与思维链结合用于知识蒸馏，尤其在高质量真实值标注稀缺场景，现有方案效果不足。基于此，文章提出一种思维链课程蒸馏方法，结合思维链引导与课程学习旨在优化蒸馏过程。以 T5-base、T5-large 为实验模型，在 4 个推理基准数据集对比标准答案蒸馏与微调，实验结果表明，课程 - 思维链蒸馏方法可显著提升 770 百万规模 T5 模型的性能，该模型在 SVAMP 数据集上的表现达到 540 亿教师模型的 94%。该蒸馏方法的效果持续优于仅提取最终答案的蒸馏方法，且模型在微调阶段的性能甚至超越了教师模型。该研究证明知识蒸馏与课程学习结合的有效性，为标注稀缺场景提供高效训练方案，且所有方法均随模型规模增大而性能提升。

关键词 深度学习；自然语言处理；大语言模型；课程学习；知识蒸馏

doi：10.3969/j.issn.1672-9528.2026.02.045

0 引言

以 GPT、LLaMA、通义千问、深度求索系列为代表的大语言模型（LLMs）[1]，近年来在自然语言理解与生成领域取得突破性进展，展现出惊人能力，尤其在需要复杂推理的任务中表现突出。这类模型擅长生成详细的中间推理步骤，从而有效解决数学解题、常识问答、逻辑推理等难题。然而，LLMs 的庞大规模使其在资源受限环境中直接部署面临巨大挑战。

目前，最先进的模型参数量通常超过 5 000 亿，所需的内存与计算资源让大多数用户难以实现部署。为解决大模型的部署难题，常采用更小的专用模型，这类模型通常通过微调或知识蒸馏技术训练。基于人工标注数据进行微调以匹配大模型性能，往往需要承担高昂的人工标注成本 [2]。知识蒸馏技术则通过 LLMs 生成的标签训练更小的学生模型，是模型压缩与知识迁移的常用手段。传统蒸馏方法引导学生模型模仿教师模型的最终输出（概率或标签），虽在部分任务中有效，却常忽略 LLMs 解决复杂问题时关键的内部推理过程。尤其在推理任务中，LLMs 生成的思维链推理过程蕴含丰富的逻辑结构与解题策略 [3]。仅蒸馏最终答案可能无法充分迁移这类推理知识。

由此推测，引导学生模型学习教师模型的推理过程，将更有效地提升其推理能力。本研究探索了一种能更充分利用 LLMs 推理能力的知识蒸馏策略。同时将教师 LLMs 生成的思维链推理过程及其最终答案作为监督信号，用于学生模型训练。此外，为应对推理过程学习本身的复杂性，融入了课程学习机制 [4]。该机制按“由易到难”的顺序组织训练样本，将这种融合思维链与课程学习的集成式知识蒸馏方法命名为“课程 - 思维链蒸馏”。

1 方法概述

本文提出了一种新颖的渐进式课程蒸馏框架，该框架利用大型教师模型从未标记数据中生成伪标签，以实现学生模型的高效训练。其具体流程为：首先由教师模型生成响应及附带的推理过程，阐明每个未标记样本的推理路径；随后，每个样本的难度由充当评估器的 GPT 模型依据推理过程的特征进行评估 [5]；基于这些难度分数，训练样本被进一步整理为课程体系；在学生模型的训练阶段，更具挑战性样本的占比会在预设阶段逐步提升，从而为模型实现稳定且渐进地学习提供支持，该框架的整体架构如图 1 所示。

1. 贵州民族大学 贵州贵阳 550000

[基金项目] 贵州省教育厅（黔教技〔2022〕47 号）（贵州省高等学校智慧教育工程研究中心、贵州思索电子有限公司支持完成）

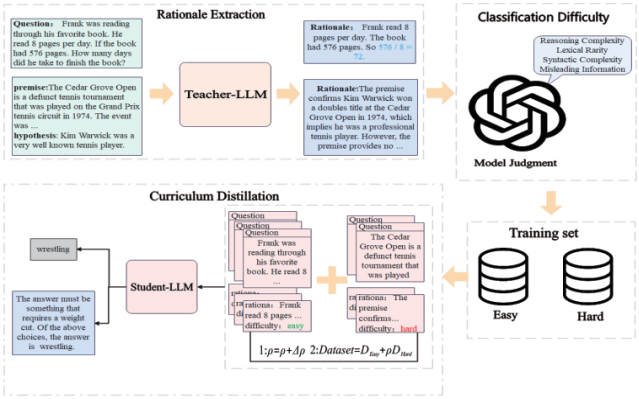

图 1 蒸馏框架

1.1 教师模型知识提取

近期研究表明，大型语言模型（LLMs）不仅能生成最终预测结果，还能生成自生成的推理过程以解释其输出[6]。现有研究通常将这些推理过程用于可解释性分析或伦理分析，而本文方法则将其作为训练学生模型的额外监督信号。具体而言，对于一个无标记数据集 D，采用一个提示模板 P。该模板指示教师 LLMs 求解出数据集中的每个问题并生成相应的推理过程。每个问题 q 被嵌入到提示模板 p 中并输入教师模型，后者会生成一个元组 $(q_i, a_i, r_i, \text{label}_i)$，其中 r 为序列推理过程；a 为基于 r 得到的最终答案；label 为可选的特定任务真实标签（尽管 LLMs 生成的数据是主要关注对象）。教师模型的生成过程可形式化为：

$$(r_i, a_i) = \text{Model}_\text{T}(P(q_i)) \tag{1}$$

式中：$P(q_i)$ 为经提示的输入。

推理过程 r 以序列形式生成，而答案 a 通常基于 q 和 r 推导得出。通过同时提取推理过程 r 和最终答案 a，学生模型既能从教师模型的直接输出中学习，也能从其内部推理过程中学习，这有助于实现更丰富的知识迁移 —— 这对于推理密集型任务至关重要，因为仅蒸馏最终答案可能会遗漏关键的问题解决结构。

1.2 问题难度定义

近年来，大型语言模型（LLMs）作为评估器的应用获得了广泛关注，这主要得益于其成本效益，以及其评估结果与人类标注者判断之间观察到的强相关性。尽管有研究指出基于 LLMs 的评估存在潜在偏差，但采用 LLMs 作为自动评判者的范式已被广泛接受[7]。在本研究中，为定量评估思维链（CoT）推理的难度，采用 GPT-3.5 Turbo 作为专门的评分代理。借鉴文本复杂度分析领域的先前研究，明确了难度的四个核心维度：

（1）推理复杂度（R）：评估思维链轨迹的逻辑深度，包括推理步骤数量、逻辑分支的存在与否以及抽象程度。

（2）词汇稀有度（V）：衡量低频术语和领域特定术语的密度。LLMs 基于其内部语料库估计稀有度，并评估其对理解的影响。

（3）句法复杂度（S）：评估句子层面的结构复杂度，包括句子长度、句法嵌套以及语法多样性（例如简单句与并列复合句结构的差异）。

（4）误导性信息（M）：检测干扰性或模糊信息的存在，包括无关内容、逻辑陷阱以及可能阻碍正确推理的模糊表述。

为确保评估的一致性，本文设计了结构化提示模板，该模板为每个维度定义了评分标准和等级量表。每个维度独立按照预定义量表进行评分。整体思维链难度得分记为 D_{score}，通过加权平均计算得出：

$$D_{\text{score}} = \boldsymbol{W}_R \cdot R + \boldsymbol{W}_V \cdot V + \boldsymbol{W}_S \cdot S + \boldsymbol{W}_M \cdot M \tag{2}$$

式中：R、V、S、M 分别为四个维度的得分；$\boldsymbol{W}$ 为其相应的权重。考虑到成本因素将权重设置相同。

1.3 知识蒸馏框架

采用知识蒸馏（KD）策略[8]，以预训练大型语言模型（LLMs）作为教师模型，为规模更小的学生模型提供监督信号用于训练。基于可用的监督信息，设计了 3 种不同的配置方案：

真实标签模式：学生模型 MS 使用数据集 D 中的真实标签 Y，并可选地使用该数据集中的真实推理过程 r 进行训练。其 优化目标为最小化损失为：

$$\mathcal{L}_{\text{GT}} = \text{CE}(M_S(x), y) + \lambda \cdot \text{CE}(M_S(x), r) \tag{3}$$

式中：x 为输入样本；CE 为交叉熵损失；λ 为平衡答案监督与推理过程监督的权重参数。若推理过程不可用或不用于训练，则 λ 取值为 0。

LLMs 生成标签模式：学生模型 MS 使用教师模型生成的伪标签$\hat{y}$，并可选地使用其生成的伪推理过程 $\hat{r}$ 进行训练。该模式充分利用了 LLMs 的丰富输出信息，在真实推理过程稀缺的场景下尤为适用。其优化目标为：

$$\mathcal{L}_{\text{LLM}} = \text{CE}(\text{MS}(x), \hat{y}) + \lambda \cdot \text{CE}(\text{MS}(x), \hat{r}) \tag{4}$$

即便没有完整的真实标注，该模式也能让学生模型从教师模型的结构化输出中获益。

任务前缀适配：尝试在输入样本 x 前添加指令性前缀 t，通过该前缀引导学生模型生成推理过程或最终答案。其损失函数为：

$$\mathcal{L}_{\text{prefix}} = \text{CE}\big(\text{MS}(p \parallel x), y_{\text{target}}\big) \tag{5}$$

式中：符号“ ||”表示拼接操作；y_{target} 表示目标输出，具体取值取决于具体任务需求与可用数据类型。

1.4 课程蒸馏策略

为进一步提升学生模型的训练效率与稳定性，融入了课程学习策略[9]。该策略的核心原则为按难度递增的结构化顺序向模型输送训练样本，从简单样本开始，逐步引入复杂度更高的样本。基于前文节定义的难度得分 D_{score}，将训练数据集划分为两个互斥的子集：简单样本集 D_{easy} 与困难样本集 D_{hard}。训练过程以离散阶段推进，其节奏由受超参数 curriculum_steps () 调控的课程进度函数控制。初始化阶段具体流程如下：

阶段 1：训练仅使用简单样本集 D_{easy} 启动。

阶段 p（p>1）：在预设的训练步数 steps 处，进入下一个课程阶段。每个阶段会引入更高比例的困难样本，阶段 p 中困难样本的占比由公式确定为：

$$\rho_p = \min\left(1, \frac{p \cdot y}{P}\right) \tag{6}$$

式中：p 为当前阶段索引；P 为课程总阶段数。

在不同阶段中，困难样本集的样本数量为：

$$|D_{\text{hard}}^{(p)}| = \rho_p \cdot |D_{\text{hard}}| \tag{7}$$

该阶段的训练数据集为：

$$D^{(p)} = D_{\text{easy}} \cup D_{\text{hard}}^{(p)} \tag{8}$$

此课程学习过程持续推进，直至所有困难样本均被纳入训练，以确保模型逐步接触复杂推理任务。

值得一提的是，按难度划分数据集时，初始的简单样本数量相对有限，在所有情况下，这类样本占总数据集的比例均不足 40%。经过多轮测试，确定难度系数参数$\dot{y}$ = 0.2为最优值。

2 实验与结果分析

本节将全面概述为评估所提方法及其组成模块有效性而设计的实验。首先，将详细说明实验设置，涵盖所用数据集、所采用的模型、所选用的评估指标及关键实现细节；随后，将设计并阐述 5 种不同的实验设置，用于开展对比分析；最后，将呈现并分析在 4 个基准数据集上的评估结果。

2.1 数据集

在 4 个广泛使用且涵盖多种推理任务的数据集上开展实验，具体包括：SVAMP[10] 数学应用题数据集，专门用于评估数值推理能力；CQA[11] 常识问答数据集，旨在评估通用世界知识与常识推理技能；ANLI[12] 自然语言推理 NLI 数据集，含更具模糊性和挑战性的任务，需模型具备溯因推理能力以及 e-SNLI[13] 自然语言推理 NLI 数据集，整合带人工标注的自然语言解释，可评估模型执行可解释推理并表述推理过程的能力。实验中，教师模型采用 PaLM-540B，其核心任务是为训练样本生成最终答案及对应的思维链（CoT）推理过程；学生模型则选取 T5 模型进行评估。

表 1 T5 模型在 4 个数据集上蒸馏和微调及课程蒸馏的结果

Method	e-SNLI	ANLI	CQA	SVAMP	Avg
Teacher PALM	0.743	0.658	0.767	0.740	0.727
Standard Fine-tuning	0.914	0.612	0.724	0.741	0.748
CoT Fine-tuning	0.924	0.688	0.740	0.803	0.789
CoT Fine-tuning(Ours)	0.921	**0.692**	**0.743**	**0.822**	**0.794**
Teacher PaLM	0.743	0.658	0.767	0.740	0.727
Standard Distillation	0.829	0.522	0.679	0.563	0.648
COT Distillation	0.833	0.535	0.703	0.575	0.661
CoTCD(Ours)	**0.849**	**0.617**	**0.703**	**0.585**	**0.689**

2.2 评价指标

本研究采用“结果准确率 + 推理质量评分”的双维度评价指标体系，以全面评估模型在推理任务上的性能，核心指标包括准确率 ACC 与 GPT 辅助打分。其中，ACC 作为基础量化指标，用于衡量模型预测结果的正确性，计算方式为模型正确预测的样本数量占总测试样本数量的比例；该指标直接对应各数据集（如 e-SNLI、ANLI、CQA、SVAMP）的任务目标，例如在 e-SNLI 中匹配模型预测的推理标签与人工标注标签，在 SVAMP 中验证数值计算结果与真实答案的一致性，能直观反映模型对任务核心需求的满足程度。

对于生成的思维链推理过程，注重推理过程而非单纯的结果。引入 GPT-3.5 Turbo 作为自动评估器进行辅助打分。该打分机制聚焦推理过程的 3 个核心维度：逻辑连贯性、步骤完整性、表述准确性；评估器依据预设的结构化评分标准 1~5 分制，5 分为最对每个样本的 CoT 推理过程独立打分，最终取所有样本的平均分作为推理质量评分。

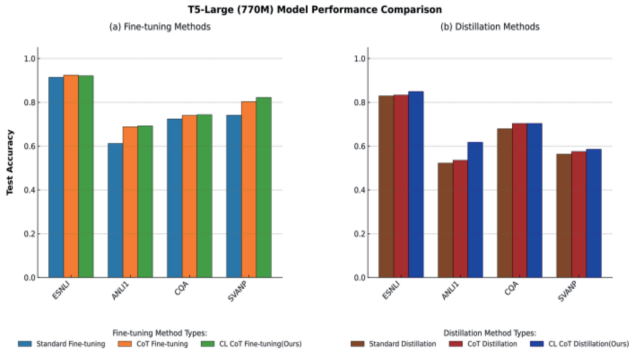

图 2 T5 模型在 4 个数据集上的结果

2.3 预处理

选择了初始简单样本的 40% 作为第一阶段的训练，后续每一轮训练加入困难样本的训练过程中，一共进行了 4 轮训练选择了恒定学习率计划，并将学习率设置为 2×10^{-5}。所有实验均在 A100 GPU 上进行，以确保足够的计算能力和效率。

实验结果后，如表 1 和图 2 所示，在微调类方法的性能对比中，本文提出的课程学习方法表现最优，其在 4 个基准数据集及平均值上的具体得分如下：

e-SNLI 数据集为 0.850，ANLI 数据集为 0.692，CQA 数据集为 0.743，SVAMP 数据集为 0.822，平均得分为 0.794。相较于传统 Fine-tuning 方法及不含思维链的微调变体，在所有推理任务中均实现性能提升，其中在对逻辑深度要求较高的 ANLI 数据集较传统微调提升 0.042，在数值推理密集的 SVAMP 数据集较传统微调提升 0.037，优势尤为明显，体现了 CoT 对复杂推理能力的增强作用。

在蒸馏类方法的对比中，本文提出的课程蒸馏方法展现出蒸馏类最优性能，其各数据集得分具体为：e-SNLI 数据集为 0.849、ANLI 数据集为 0.635、CQA 数据集为 0.698、

SVAMP 数据集为 0.774，平均得分为 0.689。与传统知识蒸馏方法相比，课程蒸馏在所有数据集上的得分均高于传统 KD，其中在 e-SNLI 数据集（提升 0.039）和 SVAMP 数据集（提升 0.044）上的性能增益最为显著，证明了将 CoT 推理过程融入蒸馏框架，能有效缓解轻量化模型的性能损耗。

从整体实验结果来看，微调类方法的平均性能（0.747~0.794）普遍高于蒸馏类方法（0.657~0.689），符合轻量化学生模型在知识蒸馏过程中因参数规模限制存在一定性能损耗的常见规律。但核心结论在于，无论是微调类还是蒸馏类，结合 CoT 推理过程的方法均在各自类别中实现最优性能，且较不含 CoT 的同类方法平均提升 0.04~0.05，充分验证了本文所提框架通过引入 CoT 推理信号，能有效提升模型在多类推理任务上的综合能力。

3 总结

本研究提出了思维链课程蒸馏方法，是一种整合了思维链提示与课程学习的新型知识蒸馏技术。该方法通过课程学习方式，从大型语言模型（LLMs）中蒸馏出思维链推理过程与最终答案，用于训练规模更小的学生模型。此方法的性能显著优于标准蒸馏方法，在真实标签稀缺的复杂推理任务中表现尤为突出，且能达到与标准微调相甚至更优的性能。实验结果验证了思维链与课程学习在该框架中的有效性。

本研究存在若干局限性，具体包括：依赖教师 LLMs 推理过程的质量；当真实推理过程本身已较为丰富时，课程学习的效果会相对温和；样本难度评估指标可能不具备通用性；所测试的模型类型范围有限；以及生成推理步骤需承担较高的计算成本。未来工作将围绕以下方向展开：解决上述局限性；将测试场景扩展到更多样化的推理任务与模型类型；并探索准确率之外的其他评估指标。

参考文献：

[1]CHOWDHERY A, NARANG S, DEVLIN J, et al. PaLM: scaling language modeling with pathways[J]. Journal of machine learning research, 2023, 24(240): 111-113.

[2]FU X Y, LASKAR M T R, KHASANOVA E, et al. Tiny titans: can smaller large language models punch above their weight in the real world for meeting summarization?[EB/OL]. (2024-04-15)[2025-09-15].https://doi.org/10.48550/arXiv.2402.00841.

[3]DENG Y T, PRASAD K, FERNANDEZ R, et al. Implicit chain of thought reasoning via knowledge distillation[EB/OL]. (2023-11-02)[2025-12-22].https://doi.org/10.48550/arXiv.2311.01460.

[4]YUE Y H, WANG C Y, HUANG J, et al. Distilling instruction-following abilities of large language models with task-aware curriculum planning[EB/OL].(2024-10-03)[2025-12-26]. https://doi.org/10.48550/arXiv.2405.13448.

[5]SAHA S, LI X, GHAZVININEJAD M, et al. Learning to plan & reason for evaluation with thinking-LLM-as-a-judge[EB/OL].(2025-07-08)[2025-12-26].https://doi.org/10.48550/arXiv.2501.18099.

[6]CHIA Y K, CHEN G Z, TUAN L U U, et al. Contrastive chain-of-thought prompting[EB/OL].(2023-11-15)[2025-12-26]. https://doi.org/10.48550/arXiv.2311.09277.

[7]YU J C, SUN S N, HU X H, et al. Improve LLM-as-a-judge ability as a general ability[EB/OL].(2025-09-07)[2025-12-26]. https://doi.org/10.48550/arXiv.2502.11689.

[8]HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL].(2015-03-09)[2025-12-26]. https://doi.org/10.48550/arXiv.1503.02531.

[9]BENGIO Y, LOURADOUR J, COLLOBERT R, et al. Curriculum learning[C]//Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09). NewYork:ACM,2009: 41-48.

[10]PATEL A, BHATTAMISHRA S, GOYAL N. Are NLP models really able to solve simple math word problems?[EB/OL].(2021-04-15)[2025-12-26].https://doi.org/10.48550/arXiv.2103.07191.

[11]TALMOR A, HERZIG J, LOURIE N, et al. CommonsenseQA: a question answering challenge targeting commonsense knowledge[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.Brussels: ACL, 2019: 4149-4158.

[12]NIE Y X, WILLIAMS A, DINAN E, et al. Adversarial NLI: a new benchmark for natural language understanding[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.Brussels:ACL,2020:4885–4901.

[13]CAMBURU O M, ROCKTÄSCHEL T, LUKASIEWICZ T, et al. e-SNLI: natural language inference with natural language explanations[EB/OL].(2018-12-06)[2025-12-25].https://doi.org/10.48550/arXiv.1812.01193.

【作者简介】

余翔（2001—），男，湖南怀化人，硕士，研究方向：大语言模型。

霍雨佳（1982—），通信作者（email:huo.yujia@gzmu.edu.cn），男，贵州遵义人，硕士研究生，研究方向：深度知识追踪、大语言模型。

（收稿日期：2025-10-19 修回日期：2026-02-04）