

基于思维链的大语言模型知识蒸馏

李荣涵¹, 浦荣成¹, 沈佳楠², 李栋栋¹, 苗启广¹

(1. 西安电子科技大学计算机科学与技术学院, 西安 710000; 2. 武警工程大学反恐指挥信息工程教育部重点实验室(立项), 西安 710086)

摘要: 思维链(Chain of thought, CoT)提示使大语言模型能够按照具体推理步骤处理复杂的任务, 让大语言模型在常识推理、数学逻辑推理和可解释性等方面表现出更强的能力。然而, CoT方法的主要缺点在于其对庞大语言模型的依赖, 这些模型通常拥有数百亿的参数, 在大规模部署方面面临挑战。为此, 本文提出一种基于思维链的大模型知识蒸馏方法, 主要目标在于充分利用大型语言模型的思维推理能力, 通过知识蒸馏技术, 引导小模型解决复杂任务。以大型模型为教师模型, 小型模型为学生模型, 通过获取教师模型的推理数据来微调学生模型。通过更改数据生成方式、基于聚类的问答示例采样、示例启发式纠错以及答案的自适应生成等一系列精心设计的方法, 使教师模型的生成过程更高效, 生成的推理数据质量更高、数量更多, 从而更好地微调学生模型, 使其获得强大的推理能力, 实现高效的知识蒸馏。这一研究框架旨在建立一个有效的知识传递机制, 使得大模型的深度思考能够有效指导小模型, 为解决复杂任务提供更为智能且高效的解决方案。通过这种方式, 希望能够克服大模型部署的挑战, 并促进语言模型在现实世界中的应用和进步。

关键词: 思维链; 逻辑推理; 知识蒸馏; 微调

中图分类号: TP391 **文献标志码:** A

Knowledge Distillation of Large Language Models Based on Chain of Thought

LI Ronghan¹, PU Rongcheng¹, SHEN Jianan², LI Dongdong¹, MIAO Qiguang¹

(1. School of Computer Science and Technology, Xidian University, Xi'an 710000, China; 2. Key Laboratory of Counter-Terrorism Command & Information Engineering of Ministry of Education (Approval), Engineering University of PAP, Xi'an 710086, China)

Abstract: The chain of thought (CoT) prompts enable large language models to process complex tasks according to specific reasoning steps, allowing them to demonstrate stronger capabilities in common sense reasoning, mathematical logic reasoning, and interpretability. However, the main drawback of the CoT approach lies in its reliance on massive language models, which typically have billions of parameters and face challenges in large-scale deployment. To address this issue, this paper proposes a large model knowledge distillation method based on the CoT, aiming to fully leverage the thinking and reasoning capabilities of large language models. Through knowledge distillation techniques, the main goal is to guide smaller models in solving complex tasks. This study adopts a large model as the teacher model and a small model as the student model, fine-tuning the student model by acquiring reasoning data from the teacher model. Through a series of carefully designed methods, such as changing data generation methods,

clustering-based sampling of question-answer examples, heuristic correction of examples, and adaptive generation of answers, this study makes the generation process of the teacher model more efficient, resulting in higher-quality and larger quantities of reasoning data. This enables better fine-tuning of the student model, allowing it to acquire strong reasoning capabilities and achieve efficient knowledge distillation. The framework of this study aims to establish an effective knowledge transfer mechanism, allowing the deep thinking of large models to effectively guide smaller models, providing more intelligent and efficient solutions for solving complex tasks. Through this approach, we hope to overcome the challenges of deploying large models and promote the application and advancement of language models in the real world.

Key words: chain of thought; logical reasoning; knowledge distillation; fine-tuning

引言

随着大型语言模型(Large language model, LLM)^[1]的发展,它们在各种复杂任务上展现出了强大的思维推理能力,但同时也带来了巨大的计算开销和部署难度。思维推理能力是指语言模型能够根据给定的问题和背景,利用自身的知识和逻辑,生成一个合理的答案或解决方案的能力。这种能力不仅需要语言模型具备丰富的常识和专业知识,还需要能够进行多步、多层次、多方面的推理,以及处理不确定性、冲突性和复杂性。大型语言模型由于其庞大的参数量和数据量,具有较强的思维推理能力,可以在各种领域和场景中展现出人类水平或超越人类的表现。例如,大型语言模型可以通过思维推理,完成数学证明、编程、写作和对话等任务。

思维推理能力是人类智能的重要组成部分,也是人工智能的核心目标之一。拥有思维推理能力的大型语言模型,可以为人类提供更多的智能服务和辅助,解决一些难以用传统方法解决的问题,创造一些新的价值和可能性。尽管大型语言模型具有较强的思维推理能力,但仍然面临着一些挑战和局限,需要进一步的研究和改进。一方面,大型语言模型的训练和使用需要消耗大量的计算资源和能源,导致高昂的成本和环境影响。另一方面,大型语言模型的思维推理能力并不稳定和可靠,可能会产生一些错误、偏差或不道德的结果,导致不良的后果和风险。例如,大型语言模型可能会因为数据的质量、数量或多样性的问题,而出现一些不符合逻辑、事实或伦理的推理,影响其正确性、可解释性和可信度。

因此,为了充分利用大语言模型的思维推理能力,同时降低部署模型的成本,本文使用一种基于知识蒸馏的方法,通过引导小型模型学习大语言模型的思维链(Chain of thought, CoT)推理能力,从而解决复杂任务。具体来说,首先利用思维链方法从 LLM 中生成多个推理路径,并对其进行筛选和重构,形成高质量的推理样本。然后,使用这些推理样本对小模型进行微调,使其能够模仿 LLM 的推理过程,并在需要时调用 LLM 进行高阶推理。方法实现过程如图 1 所示。

本文在多个复杂任务上进行了广泛实验,结果表明,提出的方法可以显著提升小模型的性能,同时大幅减少了 LLM 的调用次数和计算成本。本文的贡献在于,提出了一种有效的方法,可以通过思维链技术将大型语言模型的思维推理能力迁移到小模型中。这种方法不仅能够充分

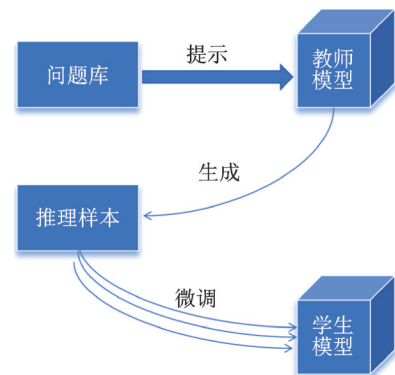


图 1 大模型知识蒸馏方法示意图

Fig.1 Schematic diagram of large model knowledge distillation method

发挥大型语言模型的优势,还能够降低其成本和风险,从而为复杂任务的解决提供一种新的思路和可能性。

1 相关工作

(1)思维链提示。思维链是一种创新的提示工具,旨在辅助语言模型进行深层次的推理和思考。这一技术的核心在于通过分步骤地引导,使模型能够清晰地展示其推理的思路和逻辑关系,从而实现复杂的思维过程。思维链提示的基本理念在于将复杂的推理过程细分为一系列简单的步骤,并在每个步骤中提供明确的指导,以帮助模型逐步建立起完整的推理链条^[2],实现过程如图2所示。

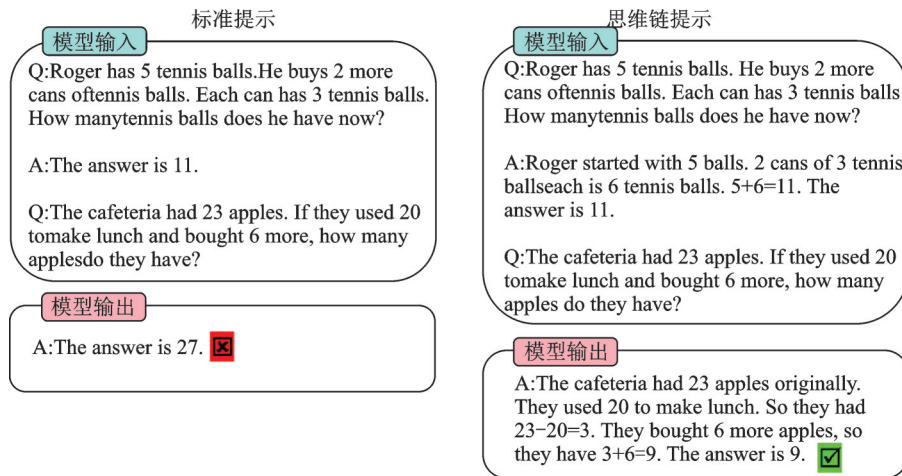


图2 思维链技术示意图

Fig.2 Conceptual diagram of CoT technology

思维链的概念最早由美国谷歌大脑的研究员 Wei等^[3]在2022年1月提出,他发现思维链可以在大语言模型中增强推理能力,特别是在数学和常识推理方面。之后,美国、英国、日本等国家的研究者在思维链的基础上提出了一些改进和扩展的方法,如自洽性提示^[2]、生成知识提示^[4]及最小到最大提示^[5]等,以提高思维链的效率和可解释性。目前,国外的思维链研究主要集中在如何设计有效的提示^[6-7]、如何评估思维链的质量、如何利用思维链进行知识发现和创新等方面。

国内对思维链的研究相对较晚,主要从2022年下半年开始,一些高校和企业的研究者开始关注思维链的概念和应用。国内的思维链研究主要借鉴和参考国外的成果,尝试在国内的数据集和任务上复现和验证思维链的效果,如在数学问题解答、阅读理解和对话生成等方面进行实验。国内的思维链研究还面临一些挑战,如缺乏统一的思维链标准和评价体系、缺乏适合国内语境的思维链数据集和任务、缺乏思维链的理论基础和深入分析等。

为了使小型模型能够进行复杂的推理,来自韩国科学技术院的研究者提出了一种名为 Fine-tune-CoT 的方法^[8],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。Fine-tune-CoT 应用 Zero-shot-CoT 从非常大的教师模型中生成推理,并使用它们来微调较小的学生模型。由于基于语言模型的教师具有显著的零样本推理能力,无需手工制作推理注释及特定任务设置,可以很容易地应用于新的下游任务。从本质上讲,论文的方法保留了基于思维链的多功能性,同时模型规模还不是很大。

(2)知识蒸馏。知识蒸馏是一种利用大型或复杂的模型(称为教师模型)来指导小型或简单的模型

(称为学生模型)学习知识的方法,既可以实现模型的压缩,也可以提升模型的性能^[9-10]。知识蒸馏的研究现状可以从以下几个方面进行概述:

知识的形式。知识蒸馏中的知识可以有不同的形式,包括输出特征知识、中间特征知识、关系特征知识、结构特征知识和参数知识^[11]。输出特征知识是指教师模型的输出层的概率分布或对数概率分布,是最常用的知识形式。中间特征知识是指教师模型的中间层的激活值或特征图,可以捕捉教师模型的内部表征。关系特征知识是指教师模型的不同层或不同样本之间的相互关系,可以反映教师模型的全局信息。结构特征知识是指教师模型的网络结构或参数,可以表示教师模型的先验知识。参数知识是指教师模型的权重或梯度,可以直接传递给学生模型。

学习的方式。知识蒸馏中的学习方式可以有不同的方式,包括知识合并、多教师学习、教师助理、跨模态蒸馏、相互蒸馏、终身蒸馏和自蒸馏^[12]。知识合并是指将多种形式的知识进行融合或加权,以提高知识的多样性和质量。多教师学习是指使用多个教师模型来指导学生模型,以提高知识的丰富性和鲁棒性。教师助理是指使用一个或多个中间规模的模型作为桥梁,以缓解教师模型和学生模型之间的差距。跨模态蒸馏是指使用不同模态的数据或模型来进行知识蒸馏,以提高知识的泛化能力和适应性。相互蒸馏是指使用多个学生模型之间相互学习,以提高知识的共享性和协同性。终身蒸馏是指使用一个或多个教师模型来指导学生模型在多个任务或领域上进行知识蒸馏,以提高知识的连续性和可持续性。自蒸馏是指使用一个模型自身作为教师和学生,以提高知识的自我提升和自我优化。

学习的目的。知识蒸馏中的学习目的可以有不同的目的,包括模型压缩和模型增强^[13]。模型压缩是指使用教师模型来指导学生模型在相同的任务或领域上进行知识蒸馏,以减少模型的大小、参数或计算量。模型增强是指使用教师模型来指导学生模型在不同的任务或领域上进行知识蒸馏,以提高模型的性能、泛化或鲁棒性。

技术的融合。知识蒸馏中的技术融合是指将知识蒸馏与其他技术进行结合,以提高知识蒸馏的效果或拓展知识蒸馏的应用范围。其中,常见的技术有生成对抗网络、神经架构搜索、强化学习、图卷积、其他压缩技术、自动编码器、集成学习和联邦学习等^[14]。

2 基于思维链的大模型知识蒸馏

基于 CoT 方法的主要缺点是它需要依赖于拥有数百亿参数的巨大语言模型,由于计算要求和推理成本过于庞大,这些模型难以大规模部署。因此,为了使小型模型能够进行复杂的推理,已有的 Fine-tune-CoT 方法是一种从非常大的教师模型生成推理样本以微调较小模型的方法,在小模型中实现了很好的推理能力,但其仍存在很多不足。本文主要是在 Fine-tune-CoT 方法的基础上提高教师模型生成推理数据的质量以及提升学生模型的推理能力,首先对问题进行聚类采样,选出最有代表性的问题,然后对选出的问题用思维链提示生成推理过程和答案,最后将生成好的样本输入教师模型中用 few-shot-cot 方式生成对问题的推理过程和答案,以达到提高教师模型生成数据质量的目的,本文将改进后的方法命名为 New-Fine-tune-CoT。总体技术路线如图 3 所示。

2.1 基于聚类的问答示例采样

Fine-tune-CoT 方法教师模型生成的推理及答案,存在数据量小以及错误率高的问题。为了解决这个问题,现决定更改教师模型生成数据的方式,以 few-shot-cot 方式从教师模型生成关于问题的更高质量数据,去微调学生模型,从而实现对模型推理能力的提升,更好地进行大模型的知识蒸馏,本文将该方法命名为 New-Fine-tune-Cot。要更改教师模型的数据生成方式,首先需对问题库进行聚类,然后再从聚类好的问题中进行采样,采样出每一类最具代表性的问题或相同类别的问题,作为后续教师

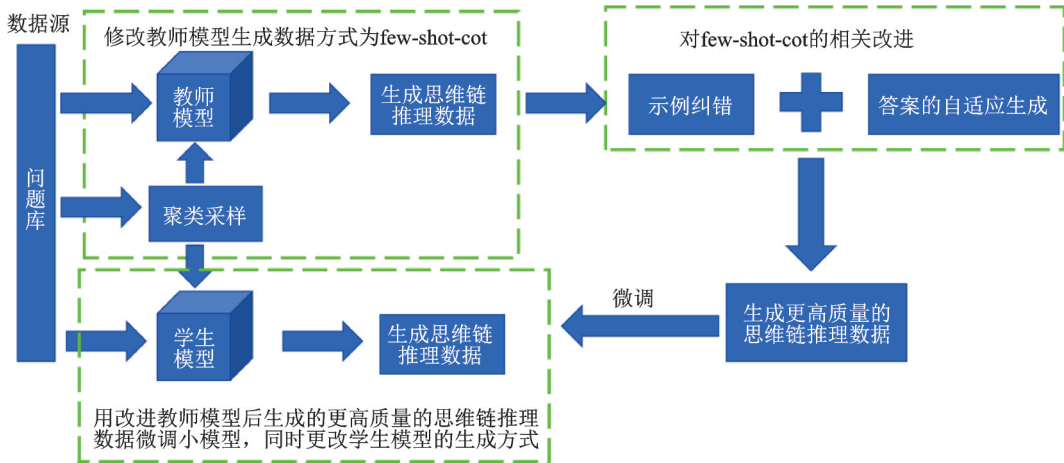


图3 New-Fine-tune-CoT 技术路线图

Fig.3 New-Fine-tune-CoT technology roadmap

模型生成思维链推理过程的样例。

教师模型的生成数据过程分为两步:在第一步中,对问题集中的问题进行聚类采样。这一步骤的目的是识别并选择具有代表性的示例,以便大模型能够基于这些示例生成高质量的推理数据。聚类采样确保了所选示例能够涵盖问题集中的各种特征和复杂性,从而为大模型的 zero-shot-cot 生成提供丰富的上下文信息;在第二步中,使用生成的示例数据对大模型进行 few-shot-cot 提示。这一步骤的关键在于利用少量的示例数据来引导大模型生成问题集中所有问题的推理数据,通过少量的示例数据,模型能够快速适应并生成新的推理数据,从而大大提高了生成效率。

以上措施可以提高教师模型生成的推理数据质量,更好地指导小模型进行复杂问题的推理,教师模型本文使用 gpt-3.5-turbo-instruct,生成推理数据的过程如图4所示。

本文对给定的问题集 Q 进行聚类分析。首先通过 Sentence-BERT 计算 Q 中的每个问题的向量表示^[15],对上下文向量进行平均,形成固定大小的问题表示。然后,通过 k -means 聚类算法对问题表示进行处理,生成 k 个问题聚类。对于每个聚类 i 中的问题,将它们排序到一个列表 $q(i)$,按到聚类 i 中心的距离升序排列。然后采样每个聚类的代表性示例,越靠近聚类中心的问题越早被考虑,用采样得到的示例提示大模型生成问题集问题的推理链,这个问题聚类采样阶段总结如算法1所示。

算法1 问题的聚类采样

- (1) procedure 聚类(Q, k)

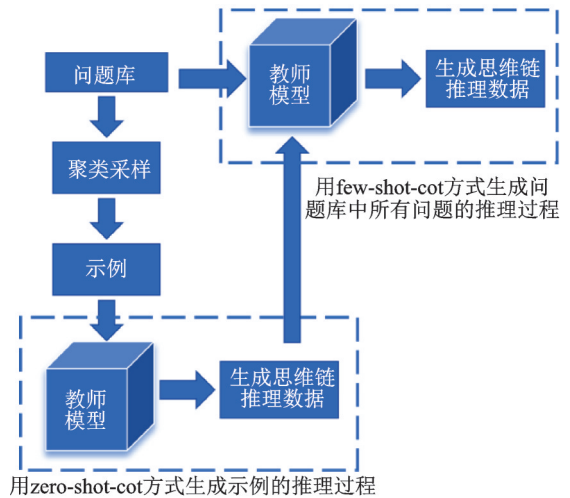


图4 教师模型生成推理数据过程图

Fig.4 Flowchart of the teacher model's reasoning data generation process

Input: 一个问题集 Q 和示例数 k , 一个空示例列表 d

Output: 将每一个聚类 $i(i=1, 2, \dots, k)$ 中的问题排序为 $q(i)=[q1(i), q2(i), \dots]$, 得到一个示例列表 $d=[d(1), d(2), \dots, d(k)]$

- (2) for 问题 q in Q do \Rightarrow 文本编码
- (3) 用 Sentence-BERT 对 q 编码
- (4) end for
- (5) 将所有编码好的问题聚类为 k 个簇
- (6) for 聚类簇 $i=1, 2, \dots, k$ do \Rightarrow 聚类排序
- (7) 将聚类 $i(i=1, 2, \dots, k)$ 中的问题按照距离聚类中心的远近排序为 $q(i)=[q1(i), q2(i), \dots]$
- (8) for $qj(i)$ in $q(i)$ do \Rightarrow 采样示例
- (9) 从聚类中心开始查找符合条件的样本作为示例 $d(i)$
- (10) 将示例 $d(i)$ 加入列表 d 中
- (11) break
- (12) return d
- (13) end procedure

2.2 示例的启发式纠错

当教师模型使用采样的示例进行 few-shot-cot 提示生成推理数据时, 示例的答案正确与否会产生很大影响, 因此需要确保示例答案的正确, 从而提高生成的问题推理链的质量。因之前的实验过程中, 在教师模型第一步生成示例推理过程时, 模型对数据集先采样再生成答案, 因此无法确定第二步生成时示例答案的正确性。本实验采取了一种优化的方法, 即将聚类过程置于初始阶段之后进行。具体而言, 首先指导教师模型生成针对所有提问的推理链数据。待这一步骤完成后, 再对提出的问题进行聚类抽样。得益于预先生成的答案, 此流程允许引入一个额外的判断环节, 即只挑选那些答案准确的问题推理链作为样本。通过这种筛选机制, 在后续的第二步中, 当使用这些抽样示例来指导大型模型对问题集中的各个问题生成推理链时, 所得到的数据的准确性将得到显著提升。修改后的生成过程如图 5 所示。

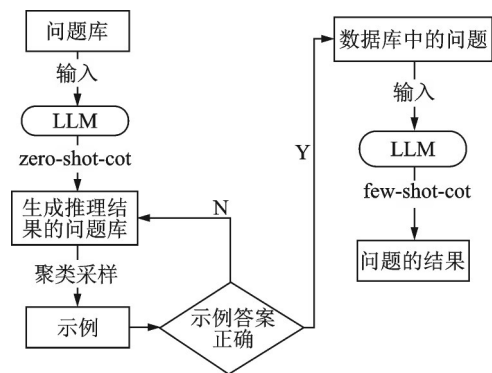


图 5 增加示例纠错之后的生成过程

Fig.5 Generation process after adding example correction

2.3 答案的自适应生成

由于教师模型采用 GPT-3.5 模型, 其推理过程依赖于 OpenAI 的 API 调用, 这无疑增加了实验的成本考量。在传统的 Fine-tune-Cot 方法中, 教师模型的推理过程分为两步: 首先, 通过 API 输入问题生成推理链; 其次, 再次调用 API, 将问题和推理作为输入, 生成答案。

然而, 在本实验中, 发现通过示例提示大模型生成推理链时, 示例的形式为 {问题, 推理, 答案} 的字典。GPT-3.5 的强大学习能力使得在生成推理链的过程中, 推理链中往往已经包含了答案。鉴于此, 不必再进行第二次 API 的调用, 而是直接从推理链中提取答案。

为了优化这一流程, 增加了一步判断过程: 首先, 检查推理链中是否包含答案。如果包含, 则直接提取答案; 如果不包含, 则再次调用 API, 将问题和推理输入教师模型中, 以生成答案。

由于大部分情况下推理链中已经包含了答案,因此本实验只需调用一次 API 即可生成完整的推理链和答案。这一改进不仅将实验的成本降低了一半,还提高了生成数据的正确性。通过这种方法,能够更加高效地利用 GPT-3.5 的强大学习能力,生成高质量且成本效益高的推理数据,为后续学生模型的训练提供了有力支持。答案自适应生成的过程如图 6 所示。

3 实验与分析

为了深入探讨和优化 Fine-tune-CoT 方法,本文对其进行了全面的对比实验和进一步的改进。首先针对教师模型的推理生成方法进行了革新,以期提高其性能。然后通过对改进前后的效果进行细致比较,得以评估这些变更的有效性。

在第一部分实验中,选取了 AddSub^[16]和 Date Understanding^[17]这两个数据集,作为测试平台。通过对这些数据集的深入分析,验证了教师模型推理生成方法改进的有效性,并为进一步的探索奠定了基础。在此基础上,进一步提出了两项重要的改进措施:示例的启发式纠错和答案的自适应生成。这两项改进旨在进一步提高教师模型的生成效率和准确性,从而为后续学生模型的训练提供更优质的素材。将这种新方法命名为 New-Fine-tune-CoT 方法。

在第二部分实验中,扩大了实验范围,选择了 SingleEq^[18]、AddSub^[16]、MultiArith^[19]、SVAMP^[20]、Date Understanding^[17]、Tracking Shuffled Objects^[17]、Last Letter Concatenation^[21]和 Coin Flip^[2]等 8 个数据集进行测试。通过在这些数据集上的实验,得以全面对比 New-Fine-tune-CoT 方法和 Fine-tune-CoT 方法的效果,从而证明本文提出的改进措施的有效性和重要性。

3.1 数据集与训练模型

在涉及 3 类复杂推理的 8 个数据集上评估了本文的方法,包括算数、符号和其他逻辑推理任务,关于这些数据集的简介如下。

对于算法推理,考虑以下 4 个数据集:(1)SingleEq,(2)AddSub,(3)MultiArith,(4)SVAMP。前 3 个来自经典的数学世界问题库^[22],后一个来自最近的基准测试,SingleEq 和 AddSub 包含更简单的问题,它们不需要多步计算来解决任务。MultiArith、AQUA-RAT、GSM8k 和 samp 是更具挑战性的数据集,需要多步推理来解决。

对于符号推理,使用如下两个数据集:Last Letter concatenation 要求模型连接每个单词的最后一个字母,本文为每个样本随机选择了 4 个名字。Coin Flip 让模型回答在人们抛硬币或不抛硬币后,硬币是否仍然是正面朝上的,本文创建了 4 次翻转或不翻转试验的样本。尽管这些任务对人类来说很容易,但 LMs 通常表现出平坦的缩放曲线。

对于其他逻辑推理任务,从 BIG-bench^[23]工作中选择两个评估集:Date Understanding 和 Tracking Shuffled Objects。Date Understanding 要求模型从上下文推断日期。Tracking Shuffled Objects 测试模型在给定初始状态和对象洗牌序列的情况下推断对象最终状态的能力。本文在实验中使用了跟踪 3 个洗牌对象的数据集。

训练集与测试集的划分。本文实验需要将样本划分为训练子集与测试子集,如果数据集原本提供了训练和测试的子集,就使用数据集原本提供的,否则以 7:3 的比例随机划分训练集和测试集。这部分信息如表 1 所示。

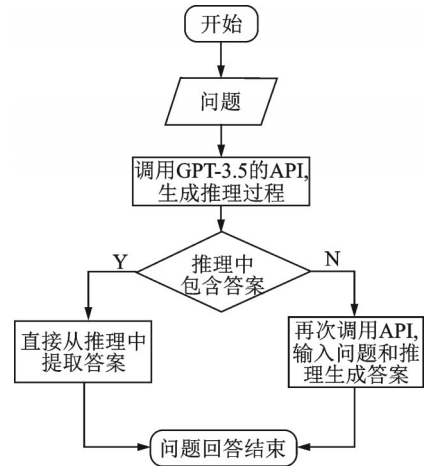


图 6 答案自适应生成流程图

Fig.6 Flowchart of adaptive answer generation

表1 本实验使用的数据集

Table 1 Datasets used in this experiment

数据集	训练样本/个	测试样本/个	数据划分	参考文献
SingleEq	356	152	70:30	[18]
AddSub	276	119	70:30	[16]
MultiArith	420	180	70:30	[19]
SVAMP	700	300	70:30	[20]
Date Understanding	258	111	70:30	[17]
Tracking Shuffled Objects	525	225	70:30	[17]
Last Letter Concatenation	350	150	70:30	[21]
Coin Flip	350	150	70:30	[2]

模型和 API 本文使用 InstructGPT^[24] 作为教师模型, 因为截止本实验开始时, OpenAI 弃用了 text-davinci-001、text-davinci-002、text-davinci-003 等模型, 只留下了一个 gpt-3.5-turbo-instruct 模型以兼容旧的补全端点, 因此本文默认以 gpt-3.5-turbo-instruct 作为教师模型, 本文实验的所有支出在于 API 的调用, 其中 gpt-3.5-turbo-instruct 模型的调用价格为输入 \$1.50/1 M tokens, 输出 \$2.00/1 M tokens。学生模型本文使用 T5、Flan-T5、GPT-2 进行实验, 其中每个模型都有 Small、Base (其中 GPT-2 为 GPT2 和 GPT2-Medium) 两种大小, 具体描述如表 2 所示。

表2 本实验中使用的模型

Table 2 Models used in this experiment

模型	参数量	角色	种类/名称
InstructGPT	175 B	教师	gpt-3.5-turbo-instruct
T5	60 M	学生	Small
T5	220 M	学生	Base
Flan-T5	60 M	学生	Small
Flan-T5	220 M	学生	Base
GPT-2	125 M	学生	Small
GPT-2	255 M	学生	Medium

3.2 第一部分实验

在第一部分实验中, 实验数据集是 AddSub^[16] 和 Date Understanding^[17], 本文首先对教师模型生成推理数据的方式进行了更改, 由原本 Fine-tune-CoT 方法中直接用 zero-shot-cot 的方式生成推理数据, 改为首先用 zero-shot-cot 生成所有问题的推理数据, 然后聚类采样出代表性的示例, 用这些示例对大模型进行 few-shot-cot 提示生成所有问题的推理数据, 再用教师模型生成的数据去微调学生模型, 查看效果, 本文将此次更改后的方法称为第一次更改后的方法。

紧接着在教师模型的生成过程中增加示例的启发式纠错和答案的自适应生成两个改进, 考查教师模型生成数据的质量, 最后将教师模型生成的数据用来微调学生模型, 对比效果, 本文将此次更改后的方法称为第二次更改后的方法, 实验结果分别如表 3、表 4 所示 (表 3 中数值为在 AddSub 数据集上以思维链方式回答问题的准确率)。

在这部分实验中, 本文采取了一种针对每个问题生成一个推理结果的策略。通过这一策略, 不仅验证了各个方法的独立有效性, 还展现了它们之间的协同效应。具体而言, 经过两次对教师

表3 在 Date Understanding 数据集上的第一部分实验结果

Table 3 The first part of experimental results on the Date Understanding dataset

方法	教师模型	GPT2	GPT2-Medium	%
Fine-tune-CoT	76.71	4.2	5.88*	
第一次更改后的方法	83.54	6.7	9.24	
第二次更改后的方法	91.13	8.4	10.1	

注: 标*的数据是本文复现 Fine-tune-CoT 方法而没有达到他论文中准确率的实验, 在这里以本文复现的 Fine-tune-CoT 方法的准确率为准。

模型的更改,在数据集 AddSub 上的准确率实现了显著提升,从 76.71% 提升至 83.54%,再进一步提升至 91.13%。同样,在数据集 Date Understanding 上,准确率也从 67.58% 提升至 76.35%,并最终达到 78.68%。这些结果清楚地表明,本文方法能够有效地提升模型的推理能力,从而为复杂问题的解决提供了坚实的支持。

此外,这些改进在学生模型上(本实验中选择了 GPT2 和 GPT2-Medium 两个学生模型)也取得了显著成效。学生模型的性能得到了大幅提升,进一步验证了本文方法的有效性。

3.3 第二部分实验

本文将两次更改后的方法称为 New-Fine-tune-CoT,在第二部分实验中将实验对象扩大到 Single EQ、AddSub、MultiArith、SVAMP、Last Letter Concatenation、Coin Flip、Date Understanding、Tracking Shuffled Objects 八个数据集上,并在生成推理结果时增加了 Fine-tune-CoT 方法中的多样性推理(Diverse reasoning)策略,用以增加学生模型的训练数据。实验结果如表 5 所示。

多样性推理即为:在 Fine-tune-CoT 方法中,采纳一种直观的教学策略,最大化教学效果,为每个问题生成多个推理解释,从而通过多个推理路径来解决复杂问题^[17]。具体操作上,对于每一个样本,通过随机采样策略,即温度采样,来获取 D 个不同的推理-答案对,而非仅采用贪心策略获取单个推理-答案对。这种方法旨在通过多样化的推理路径,为学生模型提供更全面的理解和更丰富的学习资源。类似的方法也被 Wei 等^[2]、Huang 等^[22]等研究者所采用,他们通过生成多个 CoT 输出并对其进行边缘化处理,以找到最优答案。这种多样化推理的目的是为了增强模型的推理能力和解决问题的多样性。然而,这种方法在并行工作中的影响以及对教学学生模型的实际效果,在现有的研究中尚未得到充分的认识或深入的研究^[23,25-26]。Fine-tune-CoT 方法中,作者对于数据集 MultiArith 和 SVAMP 设置 $D=64$,其他数据集设置 $D=8$,出于实验成本的考虑,本文设置所有的数据集 $D=8$,即执行多样化推理策略时,每一个数据集为每一个问题生成 8 个推理解释。

表 5 在 8 个数据集上的总体实验结果

Table 5 Overall experimental results on eight datasets

方法	参数量	准确率/%							
		Single EQ	Add Sub	Multi Arith	SVAMP	Date Understanding	Tracking Shuffled Objects	Last Letter Concatenation	Coin Flip
Teacher: InstructGPT 175B (GPT-3.5-turbo-instruct)									
Fine-tune-CoT	175 B	81.50	76.71	78.79	64.20	67.58	53.20	57.71	90.04
New-Fine-tune-CoT	175 B	91.33	91.13	98.09	79.28	78.68	63.04	82.00	100
Student: T5-{Small, Base}									
Fine-tune-cot	60 M	2.63	5.04	5.56	9.33	77.48	40.00	29.33	100
	220 M	4.61	7.56	10.56	7.00	80.18	42.67	47.33	98.67
New-Fine-tine-CoT	60 M	3.95	7.56	8.33	7.67	78.40	52.00	44.70	100
	220 M	5.26	10.10	12.20	8.67	79.30	59.60	60.00	100

表 4 在 AddSub 数据集上的第一部分实验结果

Table 4 The first part of experimental results on the AddSub dataset

方法	教师模型	GPT2	GPT2-Medium	%
Fine-tune-CoT	67.58	17.12	20.7	
第一次更改后的方法	76.35	19.8	22.5	
第二次更改后的方法	78.68	23.4	26.1	

续表

方法	参数量	准确率/%							
		Single EQ	Add Sub	Multi Arith	SVAMP	Date Under-standing	Tracking Shuffled Objects	Last Letter Concatenation	Coin Flip
Fine-tune-CoT	60 M	5.26	7.56	—	—	81.08	59.11	46.67	100
(diverse)	220 M	5.26	10.10	—	—	82.88	65.33	60.67	100
New-Fine-tune-	60 M	5.92	10.10	—	—	78.40	69.80	46.70	100
CoT(diverse)	220 M	5.92	13.40	—	—	82.90	75.10	62.70	100
Student: Flan-T5-{Small, Base}									
Fine-tune-CoT	60 M	6.58	5.88	8.33	5.67	80.18	36.00	35.33	100
	220 M	4.61	9.24	12.22	6.00	83.78	48.89	50.00	100
New-Fine-tune-	60 M	7.89	9.24	16.10	8.67	76.60	61.80	46.70	100
CoT	220 M	8.55	11.80	19.40	14.00	79.30	66.70	65.30	100
Fine-tune-CoT	60 M	7.24	10.92	—	—	84.68	62.22	46.00	100
(diverse)	220 M	9.21	10.92	—	—	84.68	67.11	56.67	100
New-Fine-tune-	60 M	7.24	10.90	—	—	78.40	71.10	50.00	100
CoT(diverse)	220 M	10.50	11.80	—	—	81.10	72.90	70.00	100
Student: GPT-2{Small, Medium}									
Fine-tune-CoT	124 M	4.61	4.22	10.00	5.67	17.12	38.67	4.67	88.0
	355 M	3.29	5.88	7.22	7.33	20.70*	35.56	10.67	80.0
New-Fine-tune-	124 M	6.58	8.40	11.10	10.70	23.40	56.00	16.00	92.0
CoT	355 M	7.24	10.10	17.80	9.00	26.10	66.20	14.00	88.0
Fine-tune-CoT	124 M	7.24	9.24	—	—	21.62	57.33	10.67	86.7*
(diverse)	355 M	5.92	9.24	—	—	20.72	56.00	20.00	93.3*
New-Fine-tune-	124 M	7.89	10.10	—	—	21.60	72.00	17.30	92.0
CoT(diverse)	355 M	8.55	14.30	—	—	25.20	72.90	28.00	93.3

注:标*的数据是本文复现Fine-tune-CoT方法而没有达到他论文中准确率的实验,在这里以本文复现的Fine-tune-CoT方法的准确率为准。

表5结果证明,本文方法不仅大大提高了教师模型生成数据的准确率,也提高了学生模型的推理能力。这个结果是令人振奋的,说明大模型的知识蒸馏是一个可行的研究方向,也证明了本文方法的有效性。

4 当前问题与未来研究方向

思维链作为大型模型涌现能力的主要来源,自提出之日起就备受关注。然而,基于思维链的大型模型知识蒸馏的研究却鲜有人涉足。这一现象的根本原因在于人们普遍认为逻辑推理能力只能由百亿级别参数量的大型模型拥有。然而,本文的研究却颠覆了这一观念,证明了即使是规模较小的模型也能通过知识蒸馏的方法获得出色的逻辑推理能力。这一发现不仅拓展了人们对于模型规模与能力关系的认识,更为未来的研究指明了方向。本文的成果为自然语言模型的轻量级部署提供了崭新的解决方案,为构建更加智能和高效的人工智能系统铺平了道路。

但该方向目前仍存在许多问题,因此本文展望未来可研究的方向,以进一步加强现在的工作。

(1) 教师模型的选择。在进行本实验时 OpenAI 将所有的补全节点都弃用了只保留了一个 gpt-3.5-turbo-instruct 模型,为了和 Fine-tune-CoT 方法进行对比,本文只能选择这个模型,但是未来工作可以选择其他或许性能更好的大模型,以进一步提高教师模型生成数据的质量。

(2)生成数据的提升。本实验通过一系列改进大大提高了教师模型生成数据的质量和数量,但依然存在不足,其中最明显的问题就是进行多样化生成时,教师模型对于同一个问题生成多个答案的相似性过高,这影响了对学生模型的训练,虽然生成数据的数量增加很多,但是对于学生模型的训练提升并没有达到相应的效果,这是未来一个重要的研究方向。

(3)学生模型的改进。目前本文的一切工作都是针对教师模型进行的,主要是提升教师模型生成数据的数量和质量,从而为学生模型提供更多更好的训练数据,但是如何提高学生模型本身的推理能力是一个值得思考的方向。这是未来工作的要点。

5 结束语

本文介绍了一种基于思维链的大模型知识蒸馏方法,在前人工作的基础上,通过改进教师模型生成数据的方法,在不改变模型的前提下,实现了性能的大幅度提升。以 Fine-tune-CoT 为基线,采用更改教师模型生成数据的方式,以及增加示例的启发式纠错和答案的自适应生成两步改进,大大提高了教师模型生成数据的质量和数量,从而有效提高了学生模型的思维链推理能力。在多个数据集上进行实验,都取得了很好的效果,这有力地证明了本方法效果的普遍性,同时证明通过知识蒸馏的方法引导小模型进行思维推理是可行的。

参考文献:

- [1] 夏润泽,李丕绩.ChatGPT大模型技术发展与应用[J].数据采集与处理,2023,38(5): 1017-1034.
XIA Runze, LI Piji. Large language model ChatGPT: Evolution and application[J]. Journal of Data Acquisition and Processing, 2023, 38(5): 1017-1034.
- [2] WEI Jason, WANG Xuezhi, SCHUURMANS D, et al. Chain of thought prompting elicits reasoning in large language models [C]//Proceedings of thirty-sixth Conference on Neural Information Processing Systems (NeurIPS 2022). [S.l.]: [s.n.], 2022: 11903.
- [3] WANG X, WEI J, SCHUURMANS D, et al. Self-consistency improves chain of thought reasoning in language models [EB/OL]. (2022). <https://arxiv.org/abs/2203.11171>.
- [4] KOJIMA T, GU S S, REID M, et al. Large language models are zero-shot reasoners[J]. Advances in Neural Information Processing Systems, 2022, 35: 22199-22213.
- [5] ZHOU D, SCHÄRLI N, HOU L, et al. Least-to-most prompting enables complex reasoning in large language models [EB/OL]. (2022). <https://arxiv.org/abs/2205.10625>.
- [6] LI Yifei, LIN Zeqi, ZHANG Shizhuo, et al. On the advance of making language models better reasoners [EB/OL]. (2022). <https://arxiv.org/abs/2206.02336>.
- [7] ZHANG Z, ZHANG A, LI M, et al. Automatic chain of thought prompting in large language models [EB/OL]. (2022). <https://arxiv.org/abs/2210.03493>.
- [8] HO Namgyu, SCHMID Laura, YUN Se Young. Large language models are reasoning teachers[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto: Association for Computational Linguistics, 2023: 14852-14882.
- [9] STANTON S, IZMAILOV P, KIRICHENKO P, et al. Does knowledge distillation really work[J]. Advances in Neural Information Processing Systems, 2021, 34: 6906-6919.
- [10] OJHA U, LI Y, S RAJAN A S, et al. What knowledge gets distilled in knowledge distillation[J]. Advances in Neural Information Processing Systems, 2024, 36: 1-12.
- [11] WANG L, YOON K J. Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 3048-3068.
- [12] XU X, LI M, TAO C, et al. A survey on knowledge distillation of large language models [EB/OL]. (2024). <https://arxiv.org/abs/2402.13116>.
- [13] GOU J, YU B, MAYBANK S J, et al. Knowledge distillation: A survey[J]. International Journal of Computer Vision, 2021,

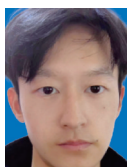
129: 1789-1819.

- [14] PHAM Q T, PHAM T T, NGUYEN T Q. Knowledge distillation: A comprehensive review and new perspectives[J]. *ACM Computing Surveys*, 2020, 53(6): 1-37.
- [15] REIMERS N, GUREVYCH I. Sentence-BERT: Sentence embeddings using siamese BERT-networks[C]//*Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, 2019: 3982-3992.
- [16] 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)*. Doha: Association for Computational Linguistics, 2014: 523-533.
- [17] SRIVASTAVA A, RASTOGI A, RAO A, et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models[EB/OL].(2022).<https://arxiv.org/abs/2206.04615>.
- [18] KONCEL-KEDZIORSKI R, HAJISHIRZI H, SABHARWAL A, et al. Parsing algebraic word problems into equations[J]. *Transactions of the Association for Computational Linguistics*, 2015, 3: 585-597.
- [19] ROY S, ROTH D. Solving general arithmetic word problems[C]// *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon: Association for Computational Linguistics, 2015: 1743-1752.
- [20] PATEL A, BHATTAMISHRA S, GOYAL N. Are NLP models really able to solve simple math word problems?[C]// *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. [S.l.]: Association for Computational Linguistics, 2021: 2080-2094.
- [21] EVANS JS BT. Intuition and reasoning: A dual-process perspective[J]. *Psychological Inquiry*, 2010, 21(4): 313-326.
- [22] 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*. 2016, 10: 1152-1157.
- [23] SRIVASTAVA A, RASTOGI A, RAO A, et al. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models [EB/OL]. (2022). <https://arxiv.org/abs/2206.04615>.
- [24] OUYANG L, WU J, JIANG X, et al. Training language models to follow instructions with human feedback[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 27730-27744.
- [25] HUANG J, GU S S, HOU L, et al. Large language models can self-improve[EB/OL]. (2022). <https://arxiv.org/abs/2210.11610>.
- [26] LI S, CHEN J, SHEN Y, et al. Explanations from large language models make small reasoners better [EB/OL]. (2022). <https://arxiv.org/abs/2210.06726>.

作者简介:



李荣涵(1992-),男,博士,研究方向:自然语言处理、深度学习、大模型, E-mail: lironghan@xidian.edu.cn。



浦荣成(2000-),男,硕士研究生,研究方向:深度学习、自然语言处理, E-mail: purongcheng123@163.com。



沈佳楠(2001-),男,硕士研究生,研究方向:信息化、大语言模型, E-mail: 1120331465@qq.com。



李栋栋(1999-),男,硕士研究生,研究方向:深度学习、大模型, E-mail: 2854712433@qq.com。



苗启广(1972-),通信作者,男,教授,研究方向:计算机视觉、机器学习、大模型, E-mail: qgmiao@xidian.edu.cn。

(编辑:夏道家)