用于解析生成的大规模通用语言模型

Teacher LM: A Generalized Reasoning Model

赵晨阳 清华大学计算机系

动机

- 选题介绍
  - GPT-3 模型发布后，研究者已经证明了参数固定情况下的 In-Context Learning 能够解锁大规模语言模型的推理能力，然而小规模语言模型难以在参数固定情况下利用 prompt 完成推理任务。
  - 放宽条件，允许参数更新的情况下，带有推理的 CoT Instruction Finetuning 也被证明能够提升大规模语言模型的推理能力，然而需要大量带有推理过程的数据才能提升小规模模型的推理效果。
  - 即便如此，使用 CoT Instruction Finetuning 提升了模型的推理能力，模型的表现仍旧不佳；甚至 CoT Instruction Finetuning 的效果往往不如单纯使用 Instruction Finetune 而不加入推理过程。
  - 因此，我们希望提出全新的专用于生成推理过程的语言模型 Teacher LM，为现有推理任务提供模型生成的高质量的解析，取代人工标注的解析。最终，利用我们的解析进行 CoT Instruction Finetune，以此提升小规模语言模型的推理能力。

模型与方法

- 模型定义
  - Chain-of-Thought Instruction Prompting：思维链（CoT），即一系列的从输入到输出的推理步骤。将传统 Finetune 时输入给模型的 Input-Output Pair 更换为 Input-Output-Prompt-CoT Chain 可显著地提升预训练模型在数学问题、常识推理和符号处理等复杂推理任务上的表现。
  - Teacher LM ：在中英题库数据集上训练的预训练语言模型组，含有 7B1 与 176B 两个大小，可对 Input-Output Pair 输出解析、相关知识和易错点。
- 研究方法
  - Instruction finetuning：使用 Teacher LM 为其他问答数据集生成解析、相关知识和易错点。利用增强后的数据集分别微调 Bloomz, Bloom, OPT 等下游小规模语言模型，提升其推理能力。

模型与方法

- 相关工作
  - Finetuned Language Models are Zero-Shot Learners
  - Scaling Instruction-Finetuned Language Models
  - Self-Consistency Improves Chain of Thought Reasoning in Language Models
  - STaR: Self-Taught Reasoner Bootstrapping Reasoning With Reasoning
- 工作亮点
  - 提出并开源了第一组专用于解析生成的通用推理语言模型 Teacher LM
  - 证明了小规模语言模型在少量带有模型生成解析的数据上 Finetune 也能解锁复杂的推理能力
  - 最后，我们基于 Teacher-Student Learning 与知识蒸馏，提出了训练友好的解锁语言模型推理能力的新范式

实验结果

- 数据集
  - 数据集简介
    - Reasoning: 包括 9 个涉及到常识、推理、数学计算等领域的 Qusetion-Answer 数据集（GSM8K, StrategyQA, AQuA, Creak, ECQA, ESNLI, QASC, QED, SenseMaking）。每条数据由问题、答案以及人工标注的解析构成。
    - PromptSource：包括 12 组 55 个 Qusetion-Answer Task。每条数据由问题、答案以及人工设计的 Instruction Template 构成。

实验结果

人工评测

左侧为人工标注解析与 ChatGPT 生成的解析，而右侧为 Teacher LM 成的解析

Problem

Question: What are is most likely to have a kosher deli?

(A) food court **(B)** jewish neighborhoods **(C)** jewish community **(D)** italian restaurant **(E)** jerusalem

Answer: The answer is (C).

Human

Analysis: Jewish Community means a community of Jewish people originating from the Israelites and Hebrews of historical Israel and Judah. Jewish community is most likely to have a kosher deli ... Jerusalem is the city and businesses are run by people.

流畅性：5　事实性：5　信息充分性：4

ChatGPT

Analysis: A kosher deli is a type of restaurant that serves kosher food, which is food that follows Jewish dietary laws ... However, the most likely place to find a kosher deli would be in a Jewish community or Jewish neighborhood.

流畅性：5　事实性：5　信息充分性：4

SenseModel

Analysis: Kosher food, which is food that has been prepared according to Jewish Law, is prepared and inspected for quality and safety by a certified kosher inspector.

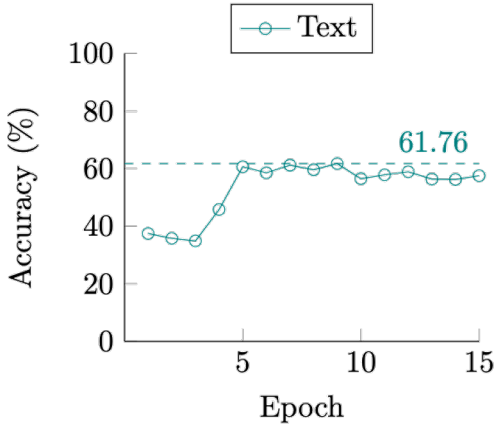
Foundamental: Various regions and states have their way of cooking according to their foods and the population of their areas ... Moreover, the differences in culture also define the differences between regions and areas.

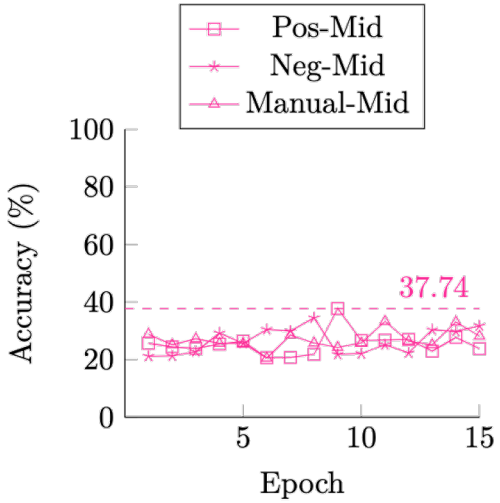
Error: Assuming that food can simply be prepared is incorrect. Because, food must be produced following certain religious rules and guidelines to be fit for consumption.

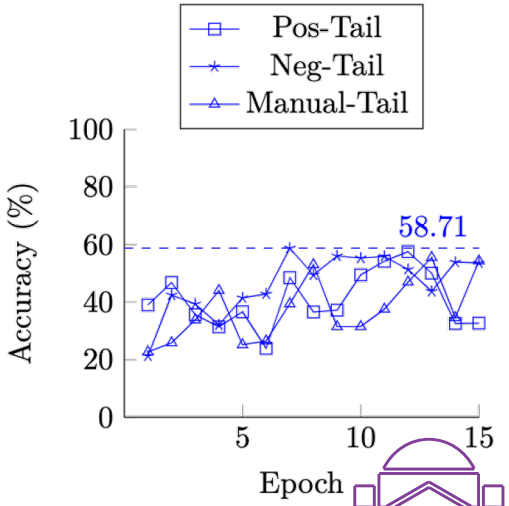
流畅性：5　事实性：5　信息充分性：5

实验结果

- 结果呈现
  - 以难度中等的选择题数据集 ECQA 举例。
  - 每个样例带有问题、答案、五个选项、正确答案解析（pos）、错误答案解析（neg）和整体解析（manual）这六个字段。SenseModel 为其生成了模型解析（cot）、易错点（error）和相关知识（fundamental）三个字段。
  - Text 代表 Finetune 时仅输入 question + answer；Manual-Mid 代表输入 question + manual + answer；Manual-Tail 表示输入 question + answer + manual，其余以此类推。

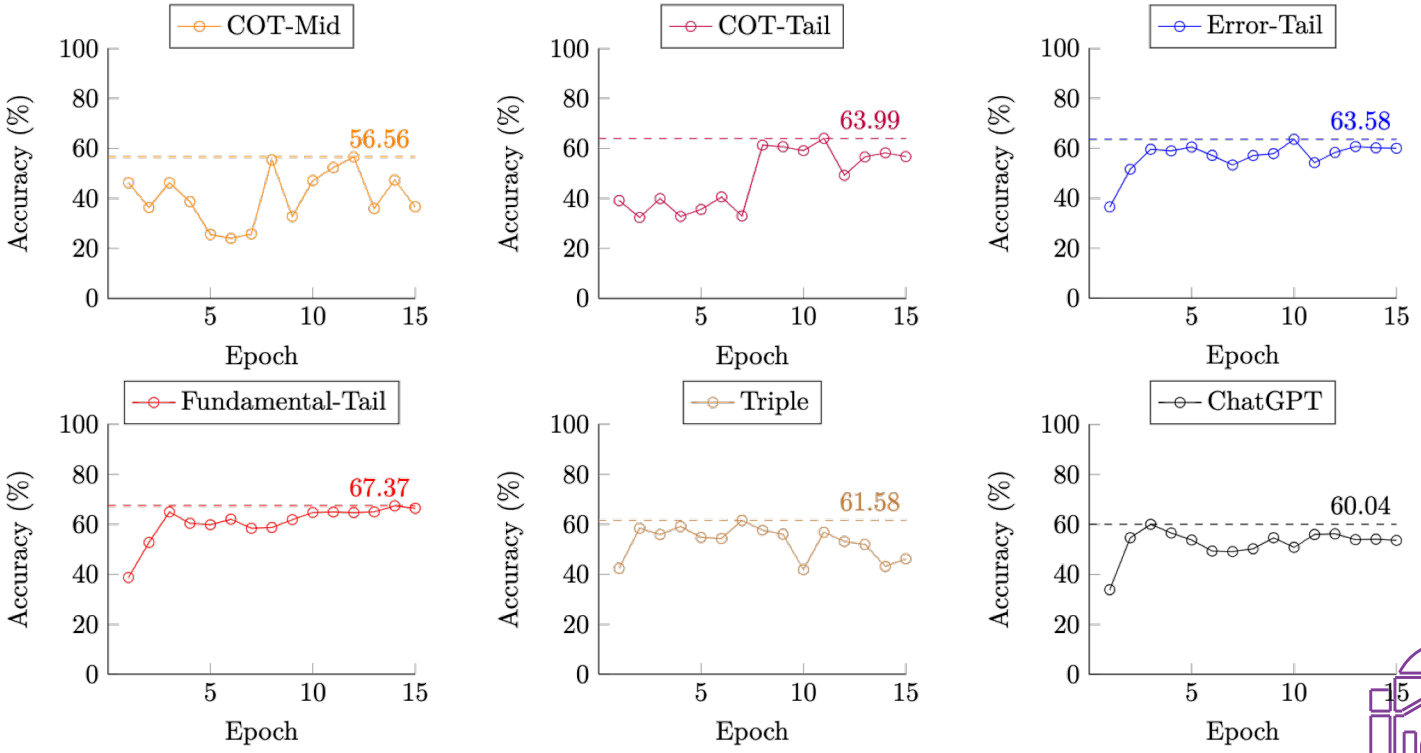






实验结果

- 结果呈现
  - CoT-Tail 表示输入 question + answer + cot；Triple 表示将 Error-Tail、CoT-Tail 和 Fundamental-Tail 直接叠加，共同训练；ChatGPT 表示将 CoT-Tail 的解析更换为 ChatGPT 的解析。



实验结果

- 结果探讨
  - 人工解析与 7B1 的 SenseModel 的解析在人工评测中效果不分上下，然而小规模模型 Finetune 后，SenseModel 的解析显著优于人工解析；
  - Tail-Training 的效果实际上不如 Middle-Training 结合 Self-Consistency，倘若采用后者，模型的表现还能更上一层楼；
  - Common Sense Task 的效果对于 Reasoning Ability 的反映不如 Math Word Problem，因为前者不单单依赖推理能力，更需要模型客观记忆了相关的常识；
  - 最后，CoT hurts 的情况还屡有发生，这可能是受到了模型规模较小且数据集人工标注的解析质量不佳的双重影响。

实验总结

- 未来工作
  - 进一步利用 176B 的完整 SenseModel 的 Reasoning Ability；
  - 探索如何混合利用 SenseModel 的三种解析字段以减缓 CoT hurts；
  - 验证更多下游模型，譬如 Encoder-Decoder 类型的 T5 模型；
  - 通过 CoT Instruction Finetune 的方式，试图系统性减轻其他模型的 Toxiclity and Bias。

谢谢！

赵晨阳 清华大学计算机系