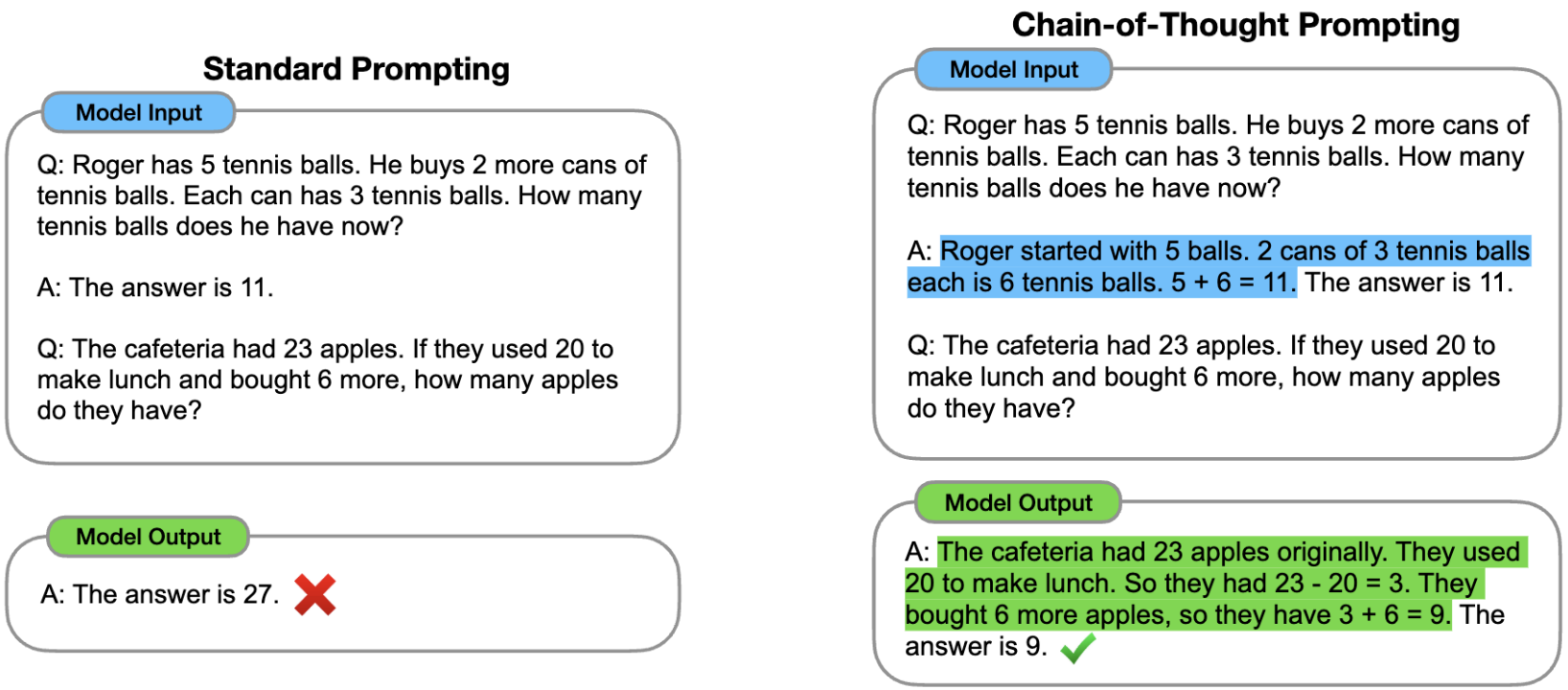
# 思维链提示

《大语言模型》编写团队：李军毅

# 思维链提示

AI Box

- 格式：<输入，思维链，输出>
- 思维链：中间推理步骤，建立输入与输出的联系




Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. **NIPS 2022.**

# 思维链提示

AI Box

- 少样本思维链 vs 零样本思维链
  - 现有主流模型往往可以自发触发思维链模式

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. 5 + 6 = 11. The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A:

---

*(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are 16 / 2 = 8 golf balls. Half of the golf balls are blue. So there are 8 / 2 = 4 blue golf balls. The answer is 4.* ✓

少样本思维链

(通过示例让模型学会生成思维链)

注：通过微调能够让大模型无需特殊提示即可生成思维链

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: ***Let's think step by step.***

---

*(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls.* ✓

零样本思维链

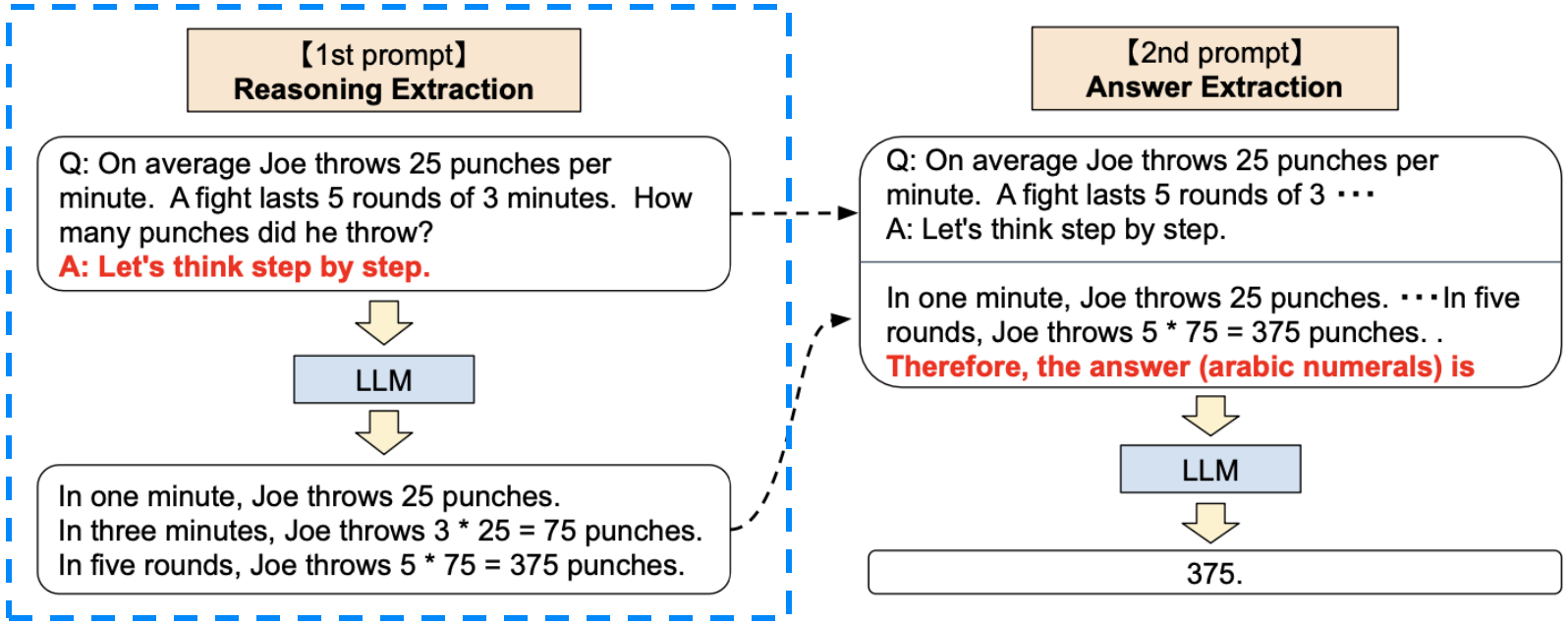
(通过提示让模型学会生成思维链)

Large Language Models are Zero-Shot Reasoners. **NIPS 2022.**

# 零样本思维链

AI Box

- 工作流
  - 模型在 Let’s think step by step 后生成推理步骤

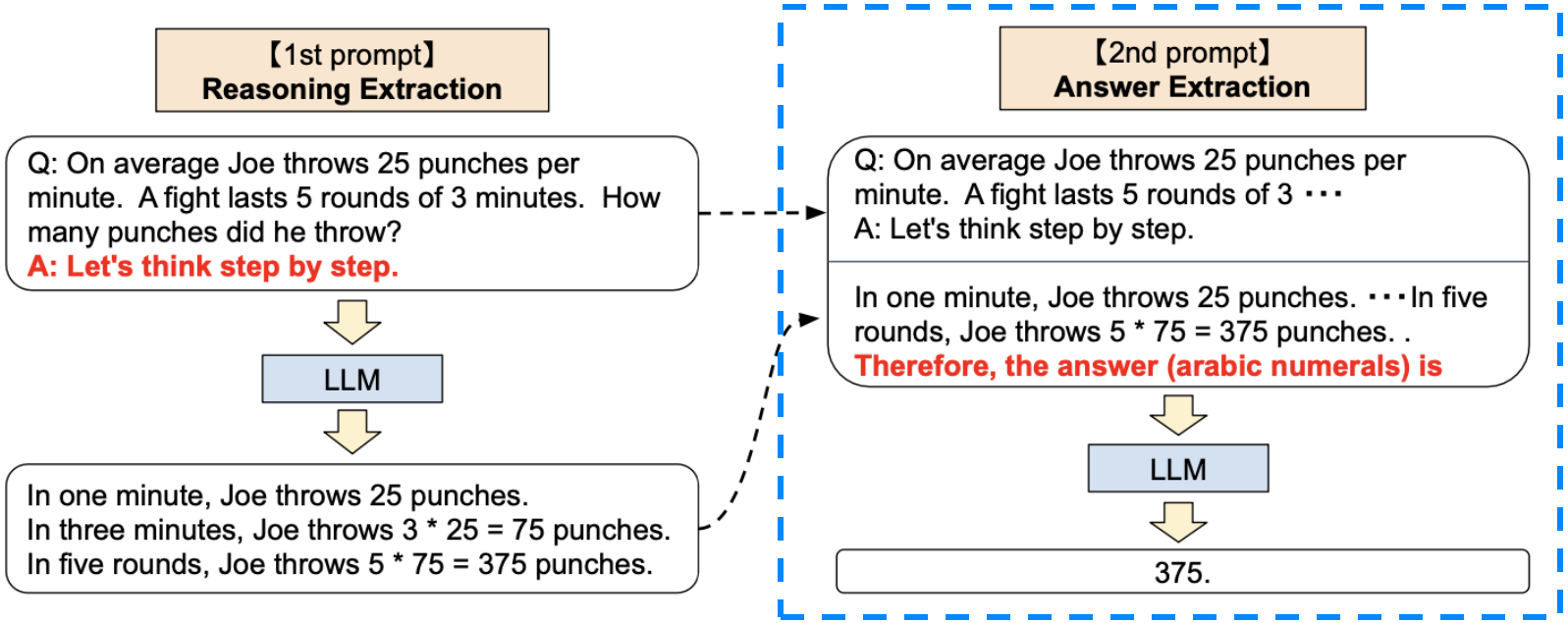



Large Language Models are Zero-Shot Reasoners. **NIPS 2022.**

# 零样本思维链

AI Box

- 工作流
  - 模型在 Let’s think step by step 后生成推理步骤
  - 把推理步骤再作为输入，模型生成答案

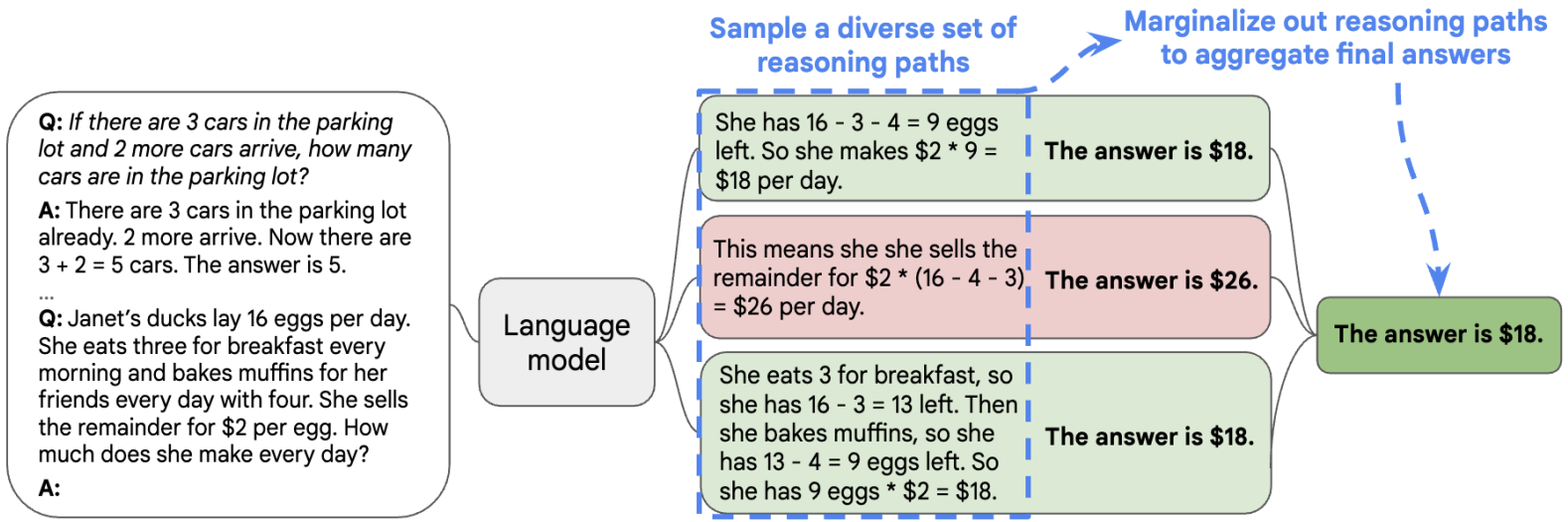



Large Language Models are Zero-Shot Reasoners. **NIPS 2022.**

# 改进思维链生成

AI Box

- 基于采样的方法
  - 问题：在使用单一思维链时，一旦中间步骤出错，容易导致最终答案也出错
  - Self-consistency: 生成多条推理路径和对应的答案，然后基于这些答案进行集成（例如选择出现频率最高的答案）并获得最终的答案，也称为majority vote

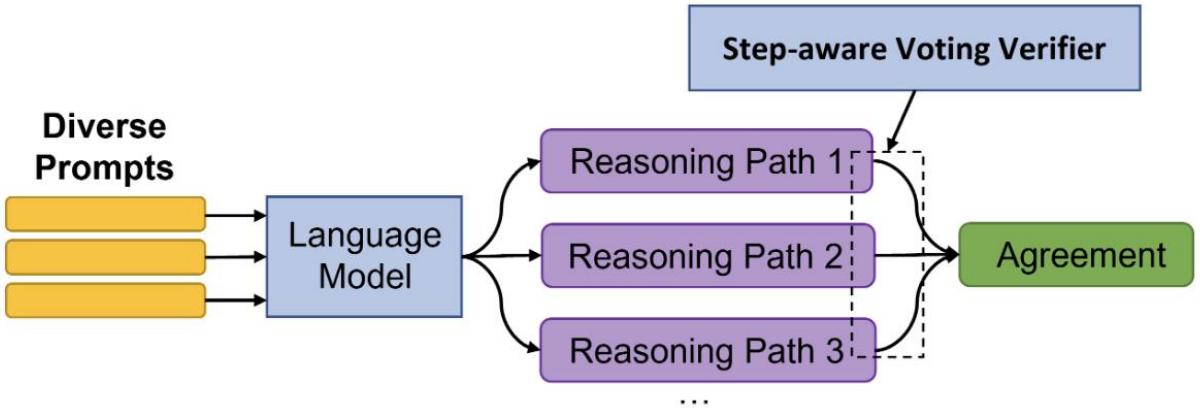


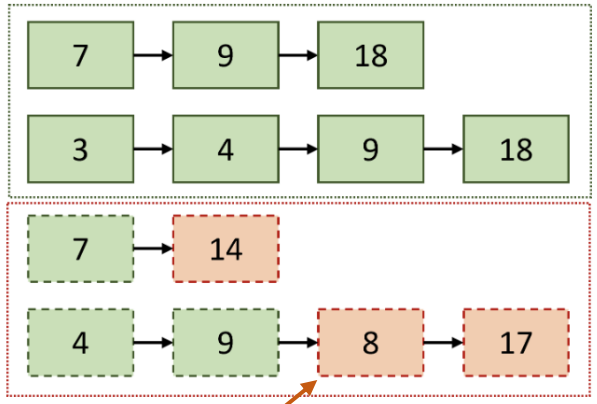

Self-Consistency Improves Chain of Thought Reasoning in Language Models. **ICLR 2023**

# 改进思维链生成

AI Box

- 基于验证的方法
  - 问题：思维链的顺序推理本质可能导致推理过程出现错误传递或累计现象
  - DIVERSE: 构造针对中间推理步骤的正负数据训练打分模型



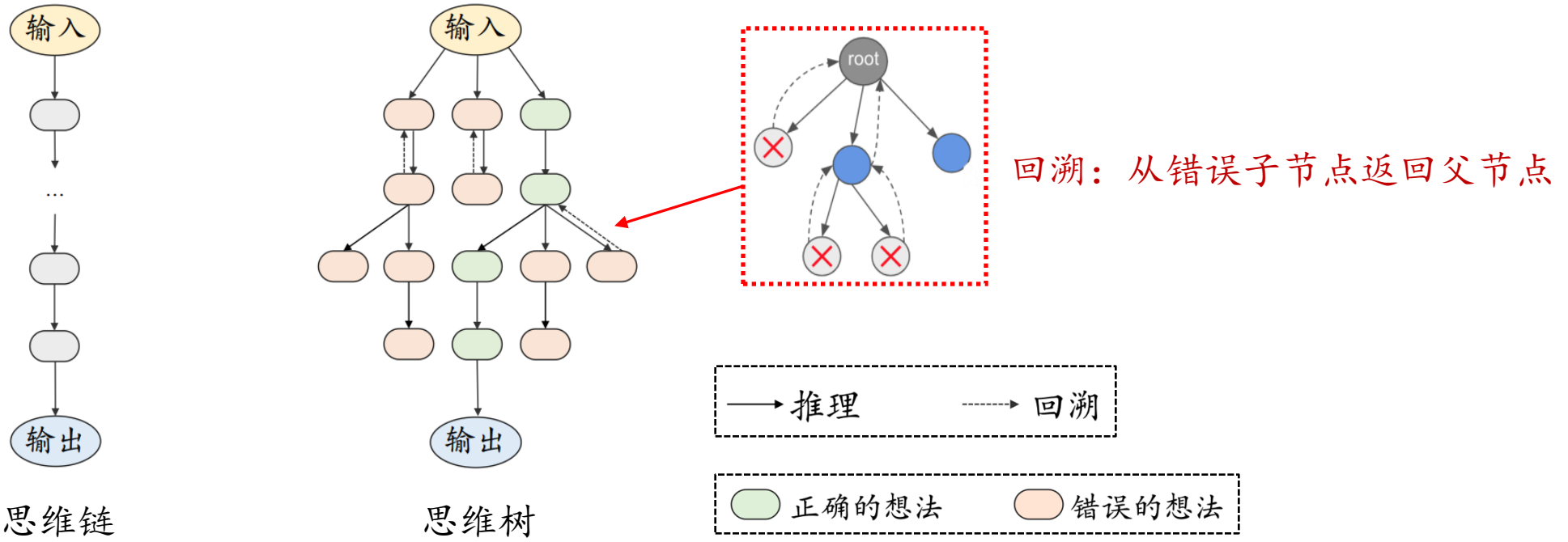

推理得到正确答案的路径

推理得到错误答案的路径

如果某一步没有在正确路径中出现，则标记为错误步骤

# 改进思维链结构

- 思维树 (Tree of Thought, ToT)
  - 每个节点对应一个思考步骤，父节点可以生成多个子节点
  - 优点：当子节点出错时具备回溯到父节点的能力



# 改进思维链结构

AI Box

- 思维树 (Tree of Thought, ToT)
  - 算法流程

**Algorithm 1** ToT-BFS$(x, p_\theta, G, k, V, T, b)$

**Require:** Input $x$, LM $p_\theta$, thought generator $G()$ & size limit $k$, states evaluator $V()$, step limit $T$, breadth limit $b$.

$S_0 \leftarrow \{x\}$

**for** $t = 1, \cdots, T$ **do**

$S'_t \leftarrow \{[s, z] \mid s \in S_{t-1}, z_t \in \mathrm{G}(p_\theta, s, k)\}$

$V_t \leftarrow V(p_\theta, S'_t)$

$S_t \leftarrow \arg\max_{S \subset S'_t, |S|=b} \sum_{s \in S} V_t(s)$

每次保存b个最佳状态

**end for**

**return** $G(p_\theta, \arg\max_{s \in S_T} V_T(s), 1)$

基于广度优先搜索 (BFS) 的思维树

**Algorithm 2** ToT-DFS$(s, t, p_\theta, G, k, V, T, v_{th})$

**Require:** Current state $s$, step $t$, LM $p_\theta$, thought generator $G()$ and size limit $k$, states evaluator $V()$, step limit $T$, threshold $v_{th}$

**if** $t > T$ **then** record output $G(p_\theta, s, 1)$

**end if**

**for** $s' \in G(p_\theta, s, k)$ **do** ▷ sorted candidates

**if** $V(p_\theta, \{s'\})(s) > v_{thres}$ **then** ▷ pruning

DFS$(s', t+1)$

**end if**

**end for**

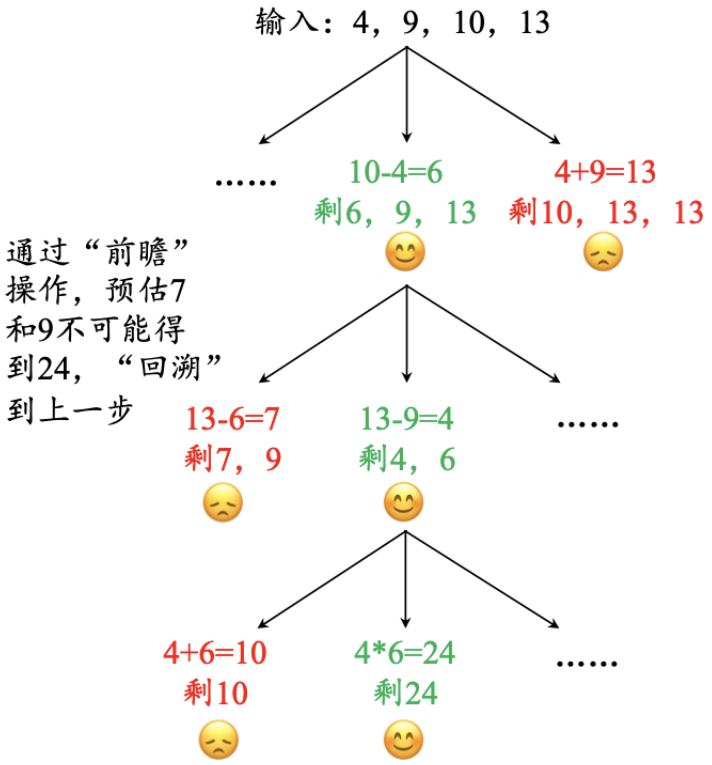
每次对最好的状态进行持续探索

基于深度优先搜索 (DFS) 的思维树

# 改进思维链结构

AI Box

- 思维树 (Tree of Thought, ToT)
  - 以 24 点游戏为例 (给定4个数，通过四则运算得到24)




任务：给定四个数4、9、10、13，如何通过加减乘除四则运算得到24？

① 生成多个可能的初始步骤，例如10-4=6，4+9=13……

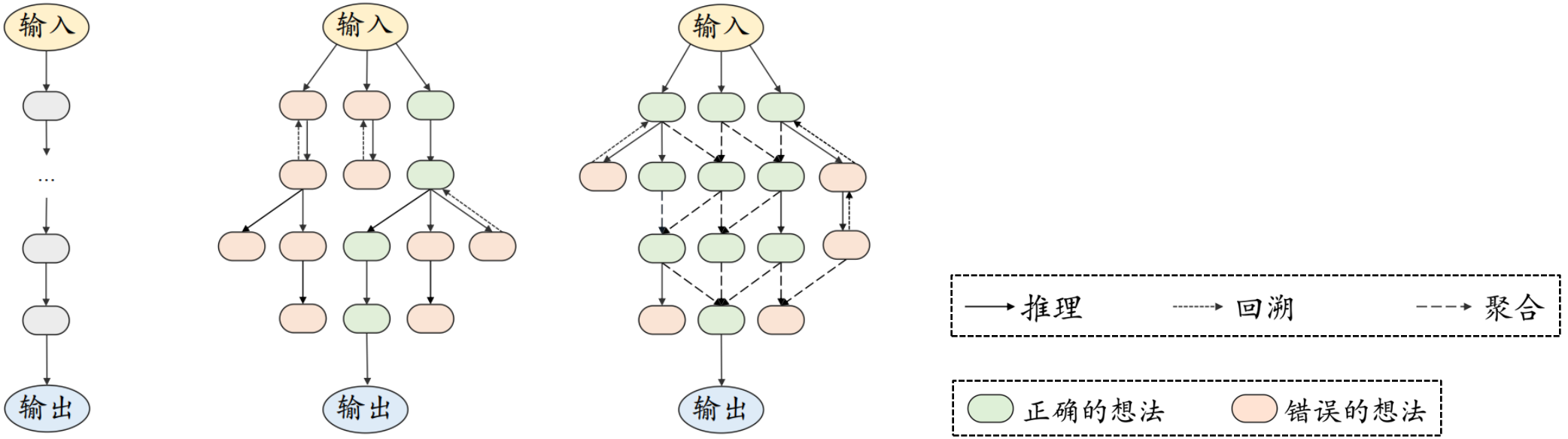
② 每一个步骤都会生成多个下一步骤，例如10-4=6之后可以生成13-6=7或13-9=4。

③ 对中间步骤进行打分。这里可以对当前思考步骤进行“前瞻”，例如在当前思考步骤剩下7和9时，能前瞻性地得知无法得到24，应该得到一个低分。

④ 如果当前节点不太可能得到最终结果，那么“回溯”到上一节点，选择其他路径。例如从13-6=7的节点回溯到父节点，然后前进走到13-9=4节点。

# 改进思维链结构

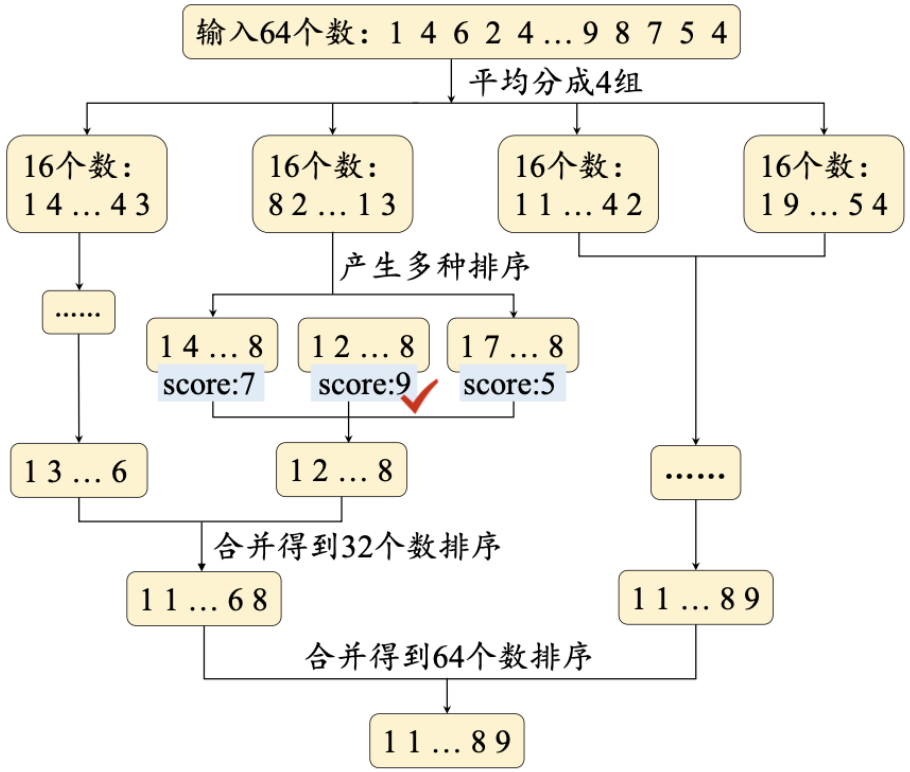
- 思维图 (Graph of Thought, GoT)
  - 任意两个节点可以相连，表示思考步骤之间的依赖关系
  - 图结构可以刻画更为复杂的拓扑结构，支持更复杂的推理关系




思维链　　思维树　　思维图

# 改进思维链结构

- 思维图 (Graph of Thought, GoT)
  - 以含有重复数字的 0~9 数组排序为例




任务：对64个1～9之间的数字进行排序

① 平均分成4组，每组16个数。

② 对每一组而言，让大语言模型排序多次得到多个可能的结果。

③ 对中间步骤进行打分，保留得分最高的方案。

④ 对中间步骤两两合并，得到最终结果。

注：在这个场景下，大语言模型难以对长数组进行准确排序，但是短数组排序对于模型来说更为简单

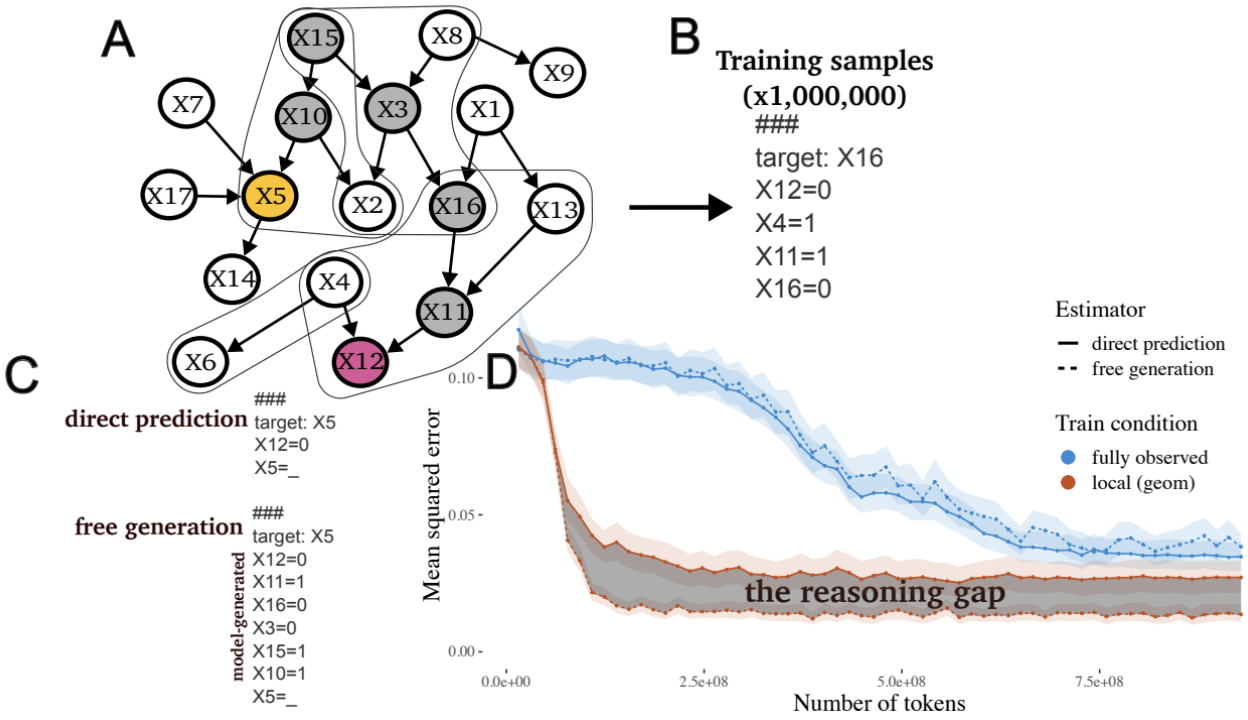
# 思维链推理能力的来源

AI Box

- 为什么思维链可以提升复杂推理任务?
  - 训练数据中存在相互重叠且互相影响的局部变量空间

A: 贝叶斯网络，红圈表示已观测变量，黄圈表示目标变量，灰圈表示中间变量

C: 比较直接预测目标变量和基于中间变量预测目标变量




B: 构建具有链式结构的贝叶斯网络，合成包含互相关联的变量的训练样本

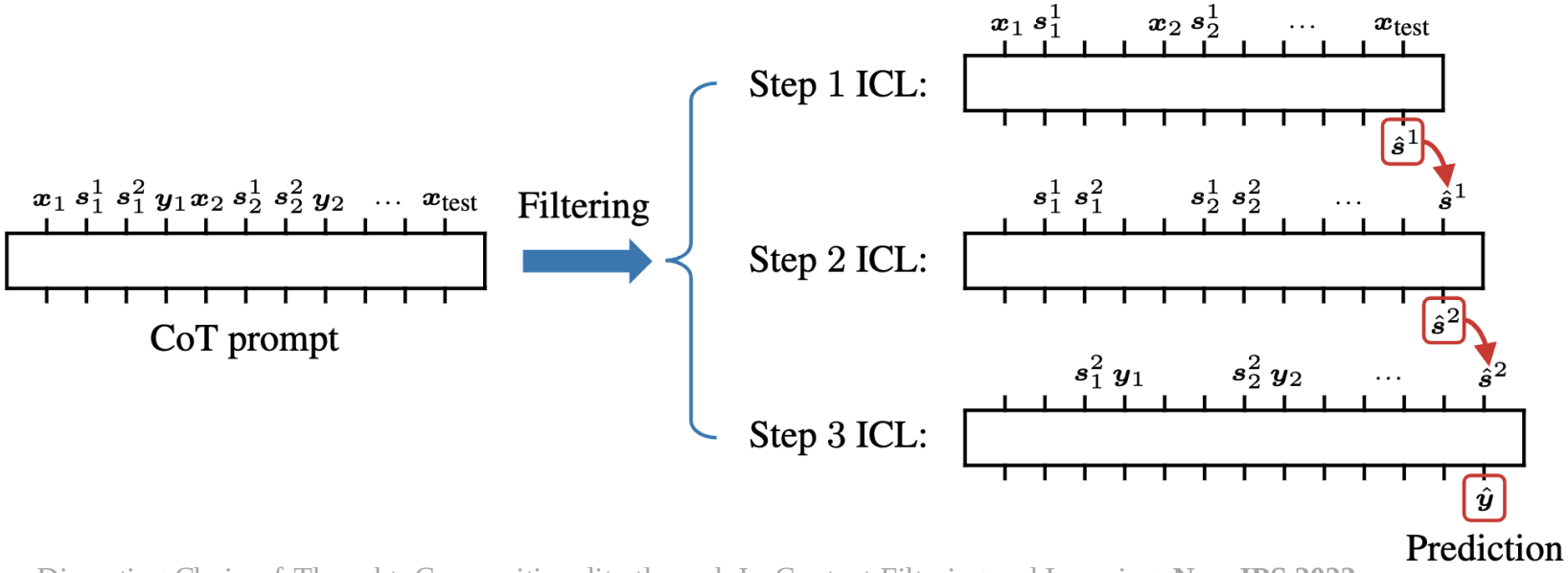
D: 当两个变量不经常在数据中共现时，直接预测条件概率与真实概率存在一定偏差；使用中间变量进行推理预测可以减小偏差

Why think step by step? Reasoning emerges from the locality of experience, **NeurIPS 2023**

# 思维链推理能力的来源

AI Box

- 思维链可以分解为两个阶段
  - 信息聚焦/过滤：模型聚焦于思维链提示中与推理步骤相关的信息
  - 上下文学习：基于过滤的信息，通过上下文学习生成一个推理步骤 (即组合函数的单步解)，迭代这一过程获得最终答案 (即组合函数的解)




Dissecting Chain-of-Thought: Compositionality through In-Context Filtering and Learning. **NeurIPS 2023**

# 思维链提示对模型推理的影响

AI Box

- (少样本) 思维链示例的两个重要组件
  - 符号 (Symbols)，例如数学题的数字、问题中的实体
  - 模式 (Patterns)，例如数学题中的算式、问题的模板

| |
|---|
| ◄ **MATHEMATICAL** ► Solve a grade-school level math reasoning problems |
| **Question:** Shawn has five toys. For Christmas, he got two toys each from his mom and dad. How many toys does he have now?<br>**Thought:** Shawn started with 5 toys. If he got 2 toys each from his mom and dad, then that is 4 more toys. 5 + 4 = 9.<br>**Symbols:** Numbers: **5**, **4**, **9**<br>**Patterns:** Equations: **5 + 4 = 9**. The equations typically appear at the end of the thought, and are almost always involved in generating the final answer. |
| ◄ **COMMONSENSE** ► (**SPORTS**) Verify the accuracy of a statement linking an athlete with a sport. |
| **Question:** Is the following sentence plausible? "Jamal Murray was perfect from the line."'<br>**Thought:** Jamal Murray is a basketball player. Being perfect from the line is part of basketball.<br>**Symbols:** Person and activity: **Jamal Murray**, **Being perfect from the line**<br>**Patterns:** Consistent sentence structure **PERSON *belongs to* SPORT. ACTIVITY *belongs to* SPORT**, where *belongs to* is a phrase that connects a sports personality with an activity. The answer is yes if both the person and the activity are associated with the same sport. |

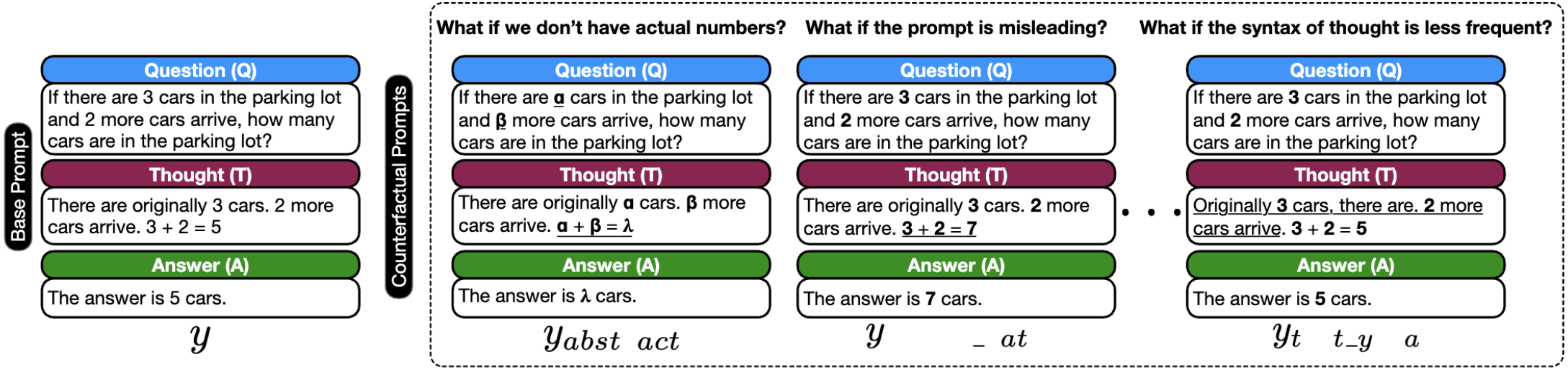
不同任务中思维链示例的符号和模式

What Makes Chain-of- Thought Prompting Effective? A Counterfactual Study. **EMNLP 2023 Findings**

# 思维链提示对模型推理的影响

AI Box

- 探究思维链对模型推理性能的影响
  - 反事实提示技术 (类似控制变量法)
    - 将原始提示的符号或者模式进行修改，以观察模型性能变化




原始提示　　数字替换为希腊字母　　使用错误的数字　　使用不常见回答句式

# 思维链提示对模型推理的影响

AI Box

➢ 探究符号和模式对思维链性能的影响

| Question / Thought | Prompt Type | Solve Rate |
|---|---|---|
| ◄ MATHEMATICAL ► (DIRECT = 10.11%, COT = 27.37%) | | |
| **Thought:** Shawn started with $\alpha$ toys. If he got $\beta$ toys each from his mom and dad, then that is $\lambda$ more toys. $\alpha + \lambda = \pi$. | $C_{\text{symb\_abs}}(p)$ (Table 25) | 25.70% |
| **Thought:** Shawn started with **5.5** toys. If he got **2.5** toys each from his mom and dad, then that is **5** more toys. **5.5** + **5** = **10.5**. | $C_{\text{symb\_ood}}(p)$ (Table 30) | 28.20% |
| ◄ COMMONSENSE ► (SPORTS) (DIRECT = 71.08%, COT = 93.67%) | | |
| **Thought:** **Jamal Murray** is a basketball player. **Being ACTIVITY** is part of basketball. | $C_{\text{symb\_abs}}(p)$ (Table 28) | 92.11% |
| **Thought:** **Adair Foster** is a basketball player. **Juggling the paper cups** is part of basketball. | $C_{\text{symb\_ood}}(p)$ (Table 32) | 79.72% |
| ◄ COMMONSENSE ► (DATE) (DIRECT = 31.61%, COT = 45.18%) | | |
| **Thought:** Today is **DATE**. 24 hours later is one day after today, which would be **DATE**. | $C_{\text{symb\_abs}}(p)$ (Table 24) | 37.41% |
| **Thought:** Today is **04/30/3069**. 24 hours later is one day after today, which would be **04/31/3069**. | $C_{\text{symb\_ood}}(p)$ (Table 31) | 44.50% |
| ◄ SYMBOLIC ► (SORTING) (DIRECT = 46.0%, COT = 60.6%) | | |
| **Thought:** $\varsigma < \phi < \gamma < \delta < \zeta < \chi < \epsilon < \pi < \upsilon$ | $C_{\text{symb\_abs}}(p)$ (Table 26) | 61.8% |
| **Thought:** **11** < **23** < **34** < **48** < **56** < **63** < **72** < **85** < **95** | $C_{\text{symb\_ood}}(p)$ (Table 33) | 80.0% |

将符号替换为抽象的占位符影响很小

What Makes Chain-of- Thought Prompting Effective? A Counterfactual Study. **EMNLP 2023 Findings**

# 思维链提示对模型推理的影响

AI Box

- 探究符号和模式对思维链性能的影响

| Question / Thought | Prompt Type | Solve Rate |
|---|---|---|
| ◄ MATHEMATICAL ► (DIRECT = 10.11%, COT = 27.37%) | | |
| **Thought:** Shawn started with 5 toys. If he got 2 toys each from his mom and dad, then that is 4 more toys. | $C_{\texttt{pat\_inconsistent}}(p)$ (Table 39) | 21.46% |
| **Thought:** 5 + 4 = 9. | $C_{\texttt{pat\_only}}(p)$ (Table 40) | 10.01% |
| **Thought:** Shawn started with 5 toys. If he got 2 toys each from his mom and dad, then that is 4 more toys. **5 + 4 = 7.** | $C_{\texttt{pat\_wrong}}(p)$ (Table 37) | 24.39% |
| ◄ COMMONSENSE ► (SPORTS) (DIRECT = 71.08%, COT = 93.67%) | | |
| **Thought:** Jamal Murray and being perfect from the line are both part of basketball. | $C_{\texttt{pat\_inconsistent}}(p)$ (Table 45) | 79.01% |
| **Thought:** Both are part of the same sport. | $C_{\texttt{pat\_only}}(p)$ (Table 41) | 74.13% |
| **Thought:** Jamal Murray is a soccer player. Being perfect from the line is part of soccer. | $C_{\texttt{pat\_wrong}}(p)$ (Table 46) | 46.02% |
| ◄ COMMONSENSE ► (DATE) (DIRECT = 31.61%, COT = 45.18%) | | |
| **Thought:** Today is 04/19/1969. | $C_{\texttt{pat\_inconsistent}}(p)$ (Table 44) | 34.19% |
| **Thought:** ⟨calculation⟩ Today = 04/19/1969. 24 hours = 1 day. ⟨output⟩ 04/19/1969 + 1 = 04/20/1969. | $C_{\texttt{pat\_only}}(p)$ (Table 42) | 33.52% |
| **Thought:** ⟨calculation⟩ Today is 04/19/1969. 24 hours later is one day after today, which ⟨output⟩ would be 03/20/1969. | $C_{\texttt{pat\_wrong}}(p)$ (Table 36) | 44.84% |
| ◄ SYMBOLIC ► (SORTING) (DIRECT = 46.0%, COT = 60.6%) | | |
| **Thought** 9 > 8 > 7 > 6 > 5 > 4 > 3 > 2 > 1 | $C_{\texttt{pat\_inconsistent}}(p)$ (Table 43) | 45.0% |
| **Thought:** — (similar to DIRECT) | $C_{\texttt{pat\_only}}(p)$ | 46.0% |
| **Thought:** 1 < 2 < 3 < 4 < 7 < 6 < 5 < 8 < 9 | $C_{\texttt{pat\_wrong}}(p)$ (Table 47) | 64.8% |

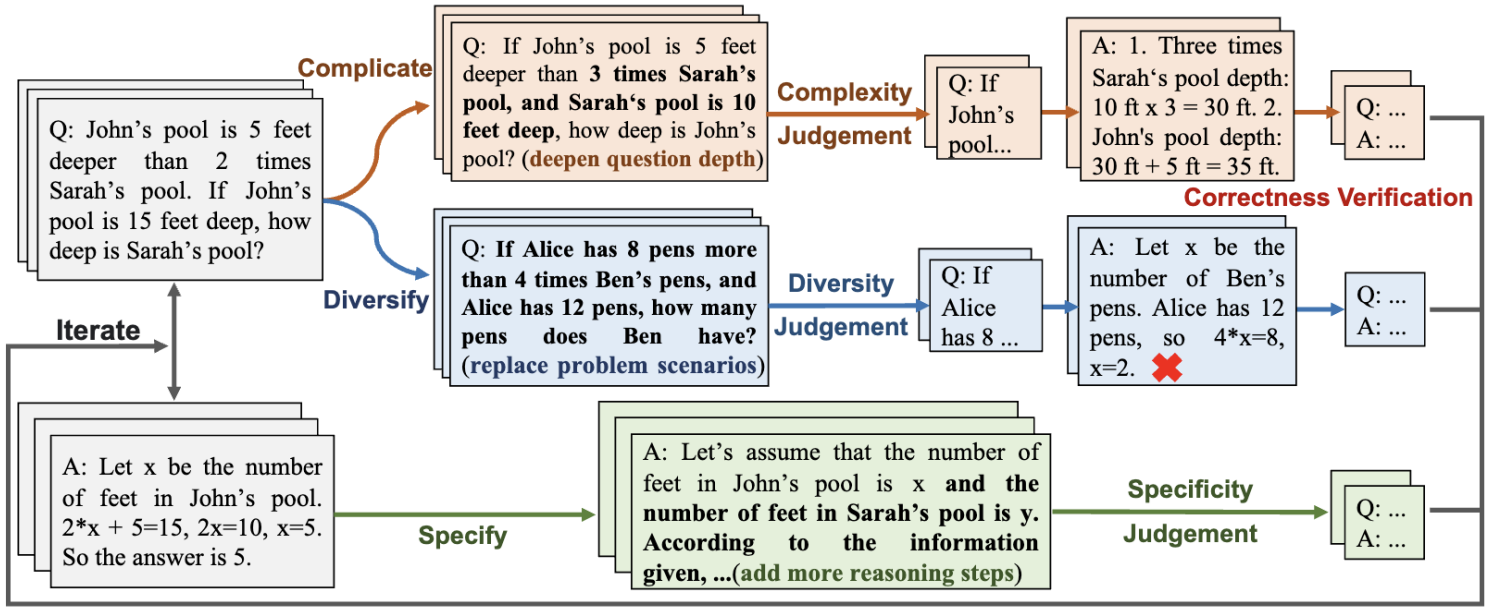
模式与推理过程的一致性和逻辑对思维链性能产生一定影响

What Makes Chain-of- Thought Prompting Effective? A Counterfactual Study. **EMNLP 2023 Findings**

# 如何增强模型的思维链能力？

AI Box

- 思维链数据增强
  - 复杂化 (例如增加条件限制、提高问题深度)、多样化 (例如转换问题背景、提问主题)、具体化 (例如补充或重写推理步骤)

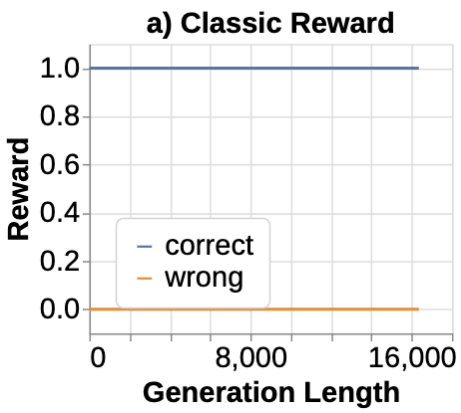



ChainLM: Empowering Large Language Models with Improved Chain-of-Thought Prompting. **COLING 2023**

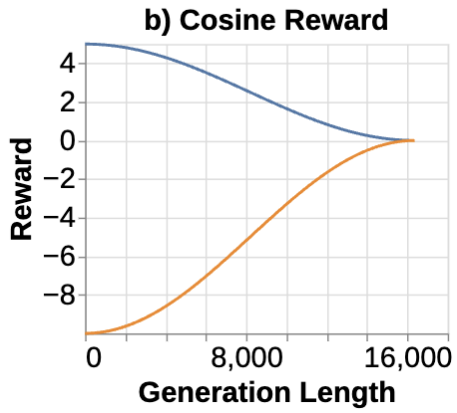
# 如何增强模型的思维链能力？

AI Box

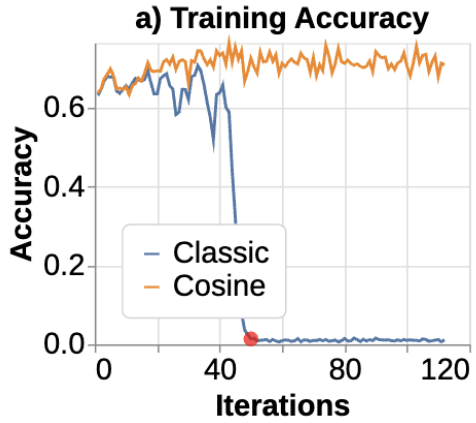
- 强化学习
  - 奖励设计
    - 传统固定奖励 v.s. Cosine奖励
    - 优点：Cosine奖励使得 RL 更加稳定，帮助控制思维链长度，提升推理准确性



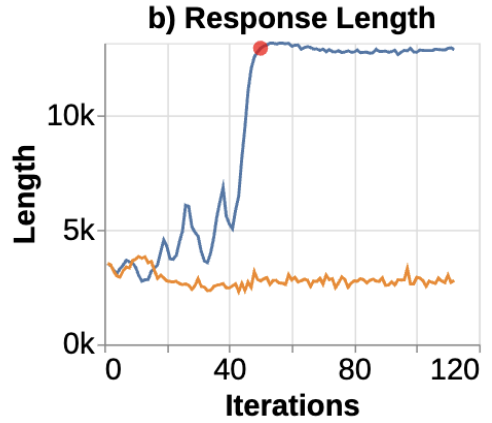

传统固定奖励




Cosine奖励




准确性变化




思维链长度变化

Demystifying Long Chain-of-Thought Reasoning in LLMs. **arXiv 2025**

# 如何增强模型的思维链能力？

AI Box

- 强化学习
  - 验证器选择
    - 基于规则的验证器 v.s. 基于模型的验证器
    - 优点：基于规则的验证器实现简单，搭配相应的过滤机制效果更好

| Prompt Set | Verifier Type | MATH 500 | AIME 2024 | Theo. QA | MMLU Pro-1k |
|---|---|---|---|---|---|
| MATH Baseline | | 59.4 | 4.0 | 25.2 | 34.6 |
| SFT Initialization | | 46.6 | 1.0 | 23.0 | 28.3 |
| Unfiltered | Rule-Based | 45.4 | 3.3 | 25.9 | 35.1 |
| | Model-Based | 47.9 | 3.5 | 26.2 | 40.4 |
| Filtered | Rule-Based | **48.6** | 3.3 | **28.1** | **41.4** |
| | Model-Based | 47.9 | **3.8** | 26.9 | **41.4** |

Demystifying Long Chain-of-Thought Reasoning in LLMs. **arXiv 2025**

# 谢谢