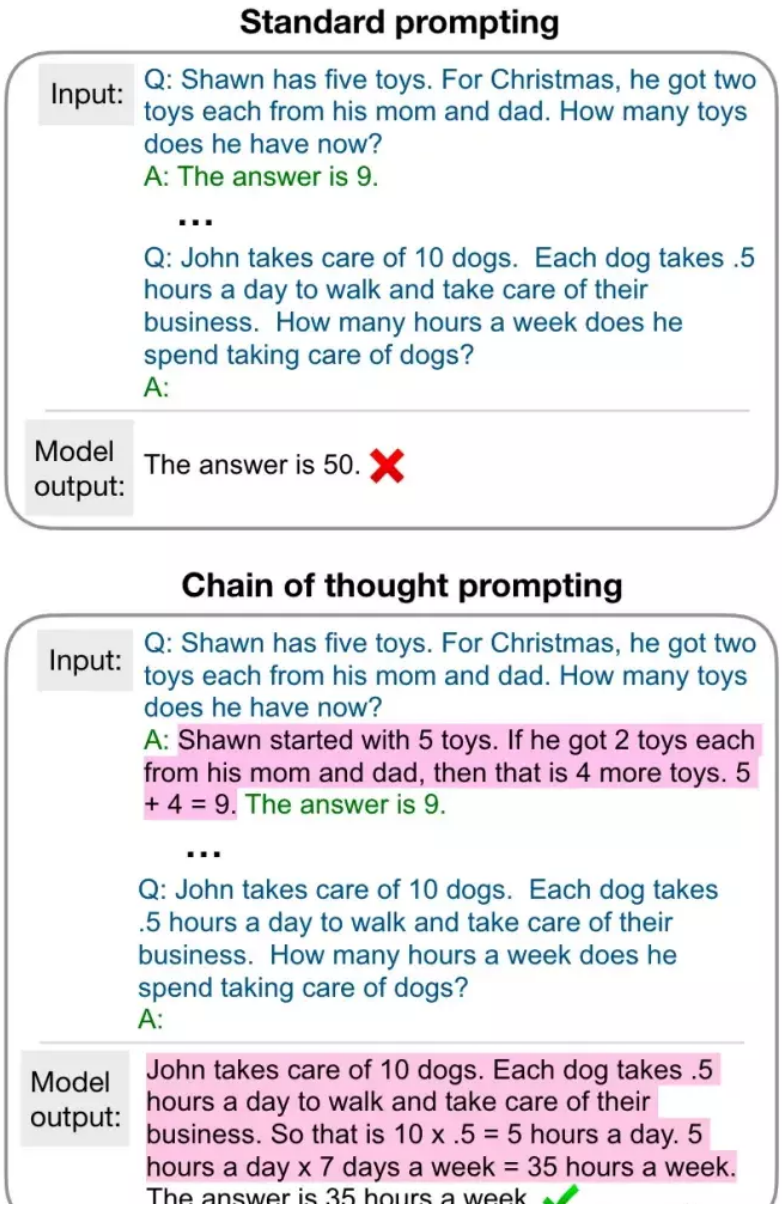
思维链（Chain-of-thoughts）

研究背景

故事发生在 2022 年 1 月份，逐渐被大家意识到是在 2022 年的 2 月到 3 月之间。
2021 年一年中，提示学习（prompt learning）浪潮兴起，**以离散式提示学习（提示词的组合）为起点，连续化提示学习（冻结大模型权重 + 微调较小参数达到等价性能）**为复兴，几乎是在年末达到了研究的一个巅峰。
但在 2022 年开始，逐渐有很多人意识到连续化提示学习其中的一些好处伴随的一些局限性，比如伪资源节约，不稳定等等。很多研究者拒绝陪玩，虽认同提示学习将会带来下一代 NLP 界的革命，但是认为拒绝做他人大模型的附庸，开始探索大模型的训练技术，并且训练自己的大模型；而手头上暂时没有掌握资源的研究者研究单位则开始再次**将研究重心从连续化学习转移到离散式提示学习**上去，将研究聚焦于特定的大模型 GPT3 上。
此时，距离 175B 的 GPT3 模型被发布和上下文学习被发现过去了不到 2 年，热度经历了高潮与低谷，经历了深度学习流派关于连接学派和符号学派的辩论和是否具有意识和推理能力的讨论，一些基础的玩法在被开发之后就被搁置了一段时间直到提示学习的兴起。2022 年 1 月，OpenAI 通过强化学习调试模型，使用强化学习调试更新了他们的模型到了第二代，LLM 肉眼可见地变得更好提示，很多任务的性能也显著提升，尤其是一些之前没有办法很好进行的任务被显著地提高了起来。
思维链系列工作就是在这样一个大环境下产生的。

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models

思维链概念的开山之作
这篇文章是现任谷歌大脑研究员的 Jason Wei 在 22 年 1 月放到 arxiv 上面的文章，在上文所说的大背景下提出了思维链这个概念。简单来说，思维链是一种离散式提示学习，更具体地，大模型下的上下文学习（即不进行训练，将例子添加到当前样本输入的前面，让模型一次输入这些文本进行输出完成任务），相比于之前传统的上下文学习，即通过 $x_1, y_1, x_2, y_2, \ldots x_{test}$ 作为输入来让大模型补全输出 y_test，思维链多了中间的一些闲言碎语絮絮叨叨，以下面这张图为例子：



这个例子选择自一个数据集叫 GSM8K，每一个样例大概就是一个小学一二年级的看几句话（基本都是三句）写算式然后算答案的难度，但是 GPT3 通过我们刚刚说的最简单的提示方法曾经只能在这个数据集上做到 6% 左右的准确度。由此可见，直接预测 y 是一个非常不太行输出空间。

思维链的絮絮叨叨即不直接预测 y，而是将 y 的 “思维过程” r（学术上有很多学者将这种过程统称为 relationale）也要预测出来。当然最后我们不需要这些 “思维过程”，这些只是用来提示获得更好的答案，只选择最后的答案即可。作者对不同的数据集的原本用于上下文学习的提示标注了这些思维链然后跑了实验，发现这么做能够显著的提升性能（左图），且这种性能的提升是具有类似于井喷性质（右图）的（后来他们发文号称这种性质叫涌现性，我们这里先按下不表）

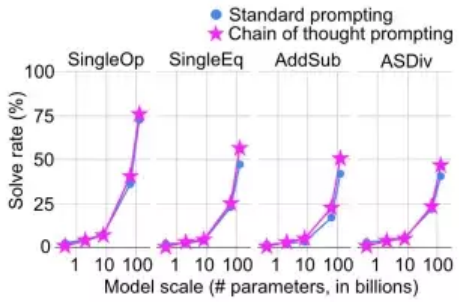


Figure 2. When scaling up the model alone already allows models to solve math word problems, chain of thought prompting does as well or better.

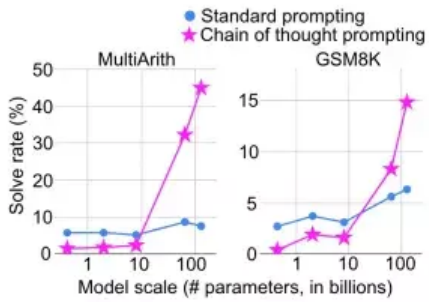


Figure 3. Employing chain of thought enables language models to solve problems for which standard prompting has a flat scaling curve.

Self-Consistency Improves Chain of Thought Reasoning in Language Models

思维链初代文章很快的一个跟进工作，是思维链系列文章版图的重要一步，在 2022 年 3 月在 arxiv 上被放出来。**这篇文章几乎用的和初代思维链文章完全一样的数据集和设置，主要改进是使用了对答案进行了多数投票（majority vote），并且发现其可以显著地提高思维链方法的性能。**

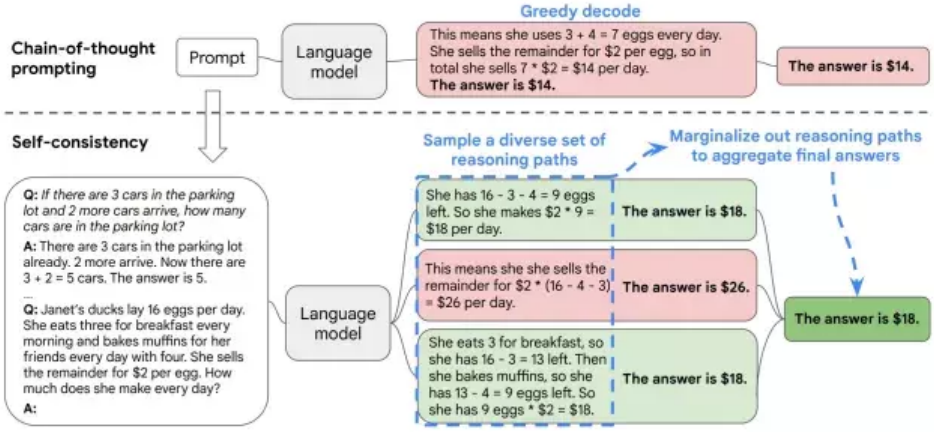


Figure 1: The self-consistency method contains three steps: (1) prompt a language model using chain-of-thought (CoT) prompting; (2) replace the "greedy decode" in CoT prompting by sampling from the language model's decoder to generate a diverse set of reasoning paths; and (3) marginalize out the reasoning paths and aggregate by choosing the most consistent answer in the final answer set.

将 greedy search 变成了 sample+vote

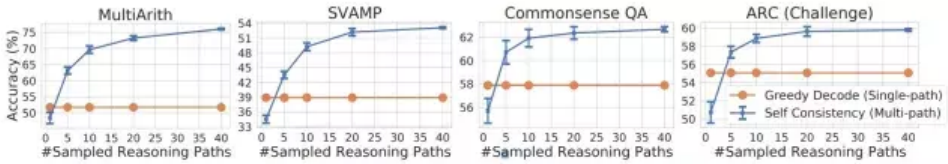


Figure 2: Self-consistency (blue) significantly improves accuracy over CoT-prompting with greedy decoding (orange) across arithmetic and commonsense reasoning tasks, over LaMDA-137B. Sampling a higher number of diverse reasoning paths consistently improves reasoning accuracy.

一开始不如 cot（因为 temperature），后面大幅度超过

这里面的一个 takeaway 是：可以将贪婪搜索（greedy search），即将 GPT 模型的 temperature 从 0 设置为某个数值，比如说 0.4，然后 sample 多个按照 y 进行投票，会显著地提升性能。

STaR: Self-Taught Reasoner Bootstrapping Reasoning With Reasoning

提出了一种 boost 方法，**让中小模型也可以通过训练具有思维链能力**