Paper Reading 12

沙之洲 2020012408

**Tree of Thought：**

1 ToT 的主要思想就是模拟人类的思维中的慢速系统。这是一种通过模拟人类而通向通用人工智能的方式，但是有时候人类自己的思维方式也是存在局限性的，我们是否应该考虑不把人类固有的思维模式强加给模型。具体来说，如果我们给模型一块内存，然后让模型分若干步来解决一个复杂推理的问题，来观察模型本身是否会在那一块内存上有什么更惊艳的表现，这样是否会更好？

2 文中提到让模型自己 evaluate 自己写的东西，但是如果模型本身存在思维惯性或者思维局限的话，并不能矫正其中的错误。而在现实世界中的更好的方法是，让一个人去做，让另外一个人去检查，而现在市面上已经出现了来源各不相同的 LM，类比于不用的人，是否可以让 GPT4 生成结果，让其他训练方式的 LM 去评判。这样可以最小程度减少陷入思维误区的可能性。

将任务拆分，让模型对每一步进行打分

相当于用模型对 A\* 算法的 evaluation function 进行评分

ToT 的模型搜索成本是不能忽略的

GSM8K 数学数据集

ToT 过程很像 RL 中 MCTS 的过程，两者的区别在于 MCTS 的 value function 是人为给定的，而 ToT 是模型实现 value function

对 AlphaZero 是用 value net 实现这一步 ToT 现在是在一个更高的维度上做 evaluation，有可能 rollout 的效果会更好

LLM 现在可以将复杂任务树形化，必然也可以实现任务拆分的扁平化，最终必然会实现成为图灵机的状态

Less Is More for Alignment

1 这篇工作通过使用 1000 个优质问答增加了模型多伦对话的能力。那么模型所缺少的关于一些专业领域的知识，是否也可通过这种高质量例子 finetune 的方式获取呢？

2 文中提到了模型的知识都是在预训练过程中获取的，而 1000 个例子的 finetune 只是教会了模型如何进行表达，那么如果将高质量问答的数量 scale up 或者增加其训练的权重，也即对模型关于高质量问答的 loss 进行放大，是否能实现通过高质量回答，将更多的知识注入模型呢？

为什么 supervise learning 能实现小规模样本微调 但是 RLHF 却不行

对于数据质量的定义 更少的人力获得更多高质量的数据