Paper Reading 10

沙之洲 2020012408

**John Schulman - Reinforcement Learning from Human Feedback: Progress and Challenges**

1 现在的 LLM 已经具备了人类基本的语言能力，但是并不具备人类的逻辑推理能力。人类的这种因果推理能力应该如何被建模？

2 回顾当我们做选择题的时候，我们是根据一个句子的整体来判断句子的正确性。现有的句子概率是将所有词的概率做简单的乘积，而这并不是一个合理的计算句子概率的方式。有更好的建模整个句子概率的方法吗？

3人类是通过不断的自我反思找到自身知识中的错误并进行修正的，而这期间往往不需要人类和外界进行交互。这种内省的自我反思是否应该加入到 LLM 当中？应该以什么样的方式加入？

**Holistic Evaluation of Language Models**

1 文章最后提出，在一些 Summary 的数据集上，人类写的 ground truth 甚至已经不能在 metrics 上边取得很好的分数。同时，人类目前衡量一个句子或者文章的好坏也不是通过任何量化的指标。这是否意味着 automation 的语言量化评估方式不是一种很好的评价方式？我们应该如何建模，如何理解人类本身对于一篇文章或者一个句子的评价方式？

2 现有关于 model fairness 的评价指标，是期望模型能够只做事实陈述。但是 Reasoning 的评价指标要求模型自己学会推理。而推理会使得模型产生对于事实的看法，也即模型自己的 opinion。这两者是否在对于模型的期待上有所矛盾？

3 现有模型和人类的差距在于，模型产生的回答不会给模型带来感受。如果模型能够对提问者收到回答之后的情感状态进行建模，更好地预期自己的回答对提问者的影响，会不会能够对模型的 fairness 或者 toxicity 有一些正向的作用？