Paper reading 2 questions

沙之洲

**Instruct-GPT:**

1 文章在 who we are aligning to 一节提到，其实他们 align 的对象是 labeler 和 gpt3 用户群体的价值观的一种平均。是否有可能进一步精简 align 的流程，实现针对每一个用户的快速 align，实现类似推荐系统那样，对于每一个用户进行精准服务？具体来说，是否可以通过训练一个预测模型，将用户在交互中的问题作为模型输入，模型可以从中抽取或者预测用户偏好的特征向量，从而让 Instruct gpt 更好地服务于每一个具体的用户。

2 文章中提到对于 instruct gpt 的衡量方法，目前只包括 human preference 和传统的 NLP dataset 上的 score，但是在这两个指标上的优越并不能保证模型真正理解了人类的语言结构。由于现在模型的语言能力已经非常强大，是否能够通过“教材数据”对模型进行 few shot，同时保证这些基于人类教材数据的 few shot 能够最大程度影响到模型的根本运转模式，增强模型在语言层面的逻辑性。进而是否可以通过这种思路逐渐训练模型的逻辑推理能力？

3 有点没太看懂 Reward Model 排序那部分关于 K=2 会导致训练崩溃的具体逻辑。

**T0:**

1 在读完 Instruct gpt 和 T0 之后，能明显感觉到 instruct gpt 在数据标注过程中，要求 labeler 的人工回答和提问者的意图想匹配，而在T0 中，我们发现模型具有能够揣度基于 task 的 prompt 的意图的能力，这种能力是否能够得到进一步增强？是否有可能通过训练，来省去人工标注提问者意图的这个步骤？

2 在 NLP 任务中，对于一个问题的答案，可能存在多种同义转述，但是给定的判定指标 BLEU 似乎会很大程度上否定同义转述，是否能够将句子 encoder 到 latent space 上，通过在 latent space 上的距离作为模型的 loss 来衡量模型是否生成了正确的回答呢？