Paper Reading 5

沙之洲 2020012408

**Perceiver io:**

1 perceiver io 指出自己的 process 模块能够处理任何模态的信息。但是不同模态之间的知识是否会存在冲突？后续任务的训练是否会对先前任务产生灾难性遗忘？

2 不同任务的 encode 模块的层数应该不尽相同。Encode 一个 image 的模块层数应该少于 encode 一个 video 的模块层数。如何选择一个 general 的 encode 模块层数呢？

**Memory transformer:**

1 memory transformer 中的 external memory 是从上下文中提取的。是否可以通过预训练构造一个更加 general 的 external memory？对于具体的 case，通过上下文微调这个 external memory 呢？

2 memory transformer 提供的这个视角认为，真正对解决实际问题有用的知识更多的包含在 external memory 当中。而之前的 scaling up 的工作使通过增加 transformer 的层数，将更多的知识存在模型权重里。所以，模型从语料中抽取的知识究竟应该保存在权重当中，还是应该保存在 external memory 当中？

3 是否可以扩展 memory transformer 的工作，通过增加限定，让模型把学到的知识存储到 external memory 当中。这样是否可以实现一个类似“通用权重”的概念，也就是一个模型拿到之前没有见过上任务的 external memory，通过external memory 的利用，获取解决这个未知任务的能力？