Paper reading 4 questions

沙之洲 2020012408

CoLT5：

1 感觉 conditional computation 有点类似于卷积神经网络。Light branch 关注的局部信息像是最底层的卷积核， Heavy branch 关注重要的信息像是 CNN 中更高层的卷积核所关注的全局信息。CoLT5 的方法和 CNN 是否有共通之处？是否可以参考像 CNN 那样，堆叠更多的层，逐层筛选最重要的 token 进行 heavy branch 呢？

2 关于 Routing 这部分，本篇文章是对于所有的 token 经过一个 routing 之后选出重要的 token。但是在人类实际阅读的过程中，会对于每一个段落进行总结归纳。是否可以考虑首先将 long token 变成若干的重要的段落，再针对于每一个段落从中 routing 出重要的 token？

Compression for AGI:

1 Neural Network 被认为是一种对于世界中的观测的 compression 机制。这种压缩的方法可以被认为是一种知识，现在这种知识是存在于模型的权重当中的。是否有可能将这种在模型权重中的隐式知识抽取出来，显示地以人类能看懂的形式表述出来？

2 talk 中将模型结构和模型 setting理解为一个模型的先验知识，并且将这种先验知识和人类的 DNA 进行类比。实际上模型结构和模型 setting 很大程度上影响了模型的学习能力，是否可以借助神经网络去搜索更好的模型先验知识，也即模型结构和模型 setting？

3 当下已经出现了针对语言的通用模型，但是没有针对图片或者视频的通用模型。语言和图片的区别在于，语言是人类创造出来的一种抽象结构，而图片和视频是世界的最原始的表示形式。以人类的智慧，到目前为止都没有能够针对图片和视频构造出很好的抽象结构来表示它们。我们是否应该期待 Neural Network 能够从世界的图像和视频表示中找到压缩的方法，毕竟这个事情人类都没有做到？