

每日小结

	周一	周二	周三	周四	周五
早	论文阅读	上课	上课	阅读文献	上课
中	上课	论文阅读，上课	论文阅读	上课	上课，跑代码
晚	跑代码	上课	新生讲课	上课	上课

注：简单表述当前时间段工作，如看文献1，整理数据等

科研详情

文献阅读

文献1

题目：Block-Recurrent Transformers

作者：DeLesley Hutchins, Imanol Schlag, Yuhuai Wu, Ethan Dyer, Behnam Neyshabur

出处：arXiv 2022

方法：本文提出 Block-Recurrent Transformers，结合了 LSTM 和 Transformer 优点，该模型的主要突破是循环单元：他是一个修改的 Transformer 层，但是它以循环的方式工作。

以递归方式沿序列应用 Transformer 层，相对序列长度具有线性复杂性。所提出的递归单元对 token 块而不是单个 token 进行操作，利用块内的并行计算，有效利用加速器硬件。该单元本身非常简单，仅仅是一个 Transformer 层：用自注意力和交叉注意力有效计算一大组状态向量和 token 的循环函数。设计部分受到了 LSTM 单元的启发，采用 LSTM 风格的门，但将典型的 LSTM 单元放大了几个数量级。对递归的实现在计算时间和参数数量上与传统的 Transformer 层有相同的成本，但在非常长序列上的语言建模任务中提供了极大的改进。

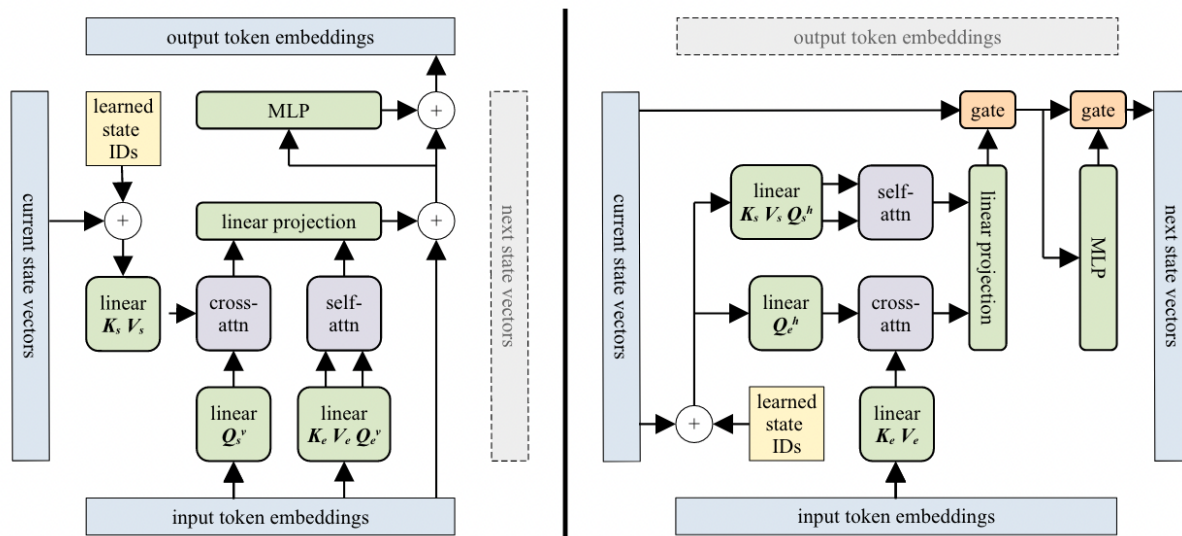
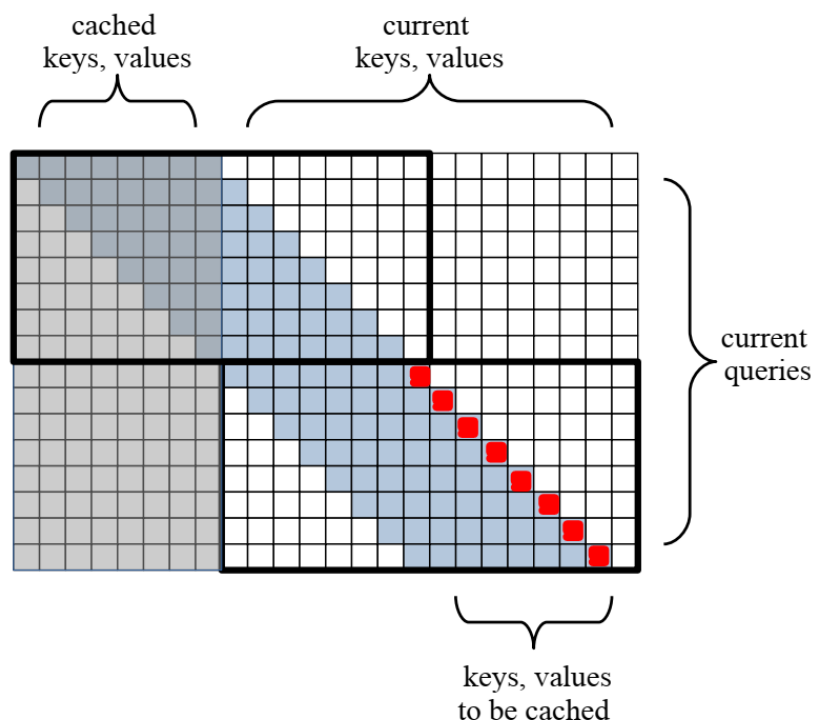


Figure 1: Illustration of our recurrent cell. The left side depicts the vertical direction (layers stacked in the usual way) and the right side depicts the horizontal direction (recurrence). Notice that the horizontal direction merely rotates a conventional transformer layer by 90° , and replaces the residual connections with gates.

滑动注意力机制：本文采用的是一种滑动窗口注意力机制，一种专门针对长文档场景的技术。由于文本过长，让每个 token 都 attend 到整个文本中的所有 token 难以实现。在滑动窗口注意力机制中：每个 token 只需要 attend 到 token。在本文中，滑动窗口长度与每个循环单元所需处理的文本长度相等，即：



启发:

1. 循环单元支持两种类型的操作:自注意力和交叉注意力。更具体地说: Self-Attention 在相同嵌入(K、V 和 Q 矩阵)中生成的键、值和查询上执行。Cross-Attention 是从其他嵌入生成的 K、V 中执行的本身嵌入生成 Q 进行查询。
回想一下原始的 Transformer 编码器-解码器模型, 编码器执行自注意力, 而解码器中的“编码器-解码器注意力”层执行交叉注意力。这是因为查询 Q 来自前一个 Decoder 层, 而 K 和 V 来自 Encoder 输出。循环单元在同一层执行这两种操作。
换句话说: 循环单元同时进行自我注意(编码)和交叉注意(解码)
2. 工程实现上的一些问题。比如说, 模型训练的时候是否会像传统 RNN 一样遇到梯度消失的问题? 如果有, 该如何解决?
3. github 上有一些其他人复现的网络, 还在逐段看代码详细理解, 有一些细节还是需要看代码帮助理解

文献2

题目: MsIFT: Multi-Source Image Fusion Transformer

作者: Xin Zhang , Hangzhi Jiang, Nuo Xu , Lei Ni , Chunlei Huo, Chunhong Pan

出处: Remote Sensing

方法:

在本文中, 提出了一种新颖的多源图像融合变换器 (MsIFT)。由于 Transformer 固有的全局注意力机制, MsIFT 具有非局部融合感受野, 并且对空间错位具有更强的鲁棒性。MsIFT 由 CNN 特征提取器、特征融合变换器和任务预测器组成。在 MsIFT 中, 图像对在特征融合之前被编码两次, 即基于局部和基于全局的特征融合分别由 CNN 特征提取器和融合变换器编码器实现。融合变换器中的解码器进行多源非局部特征融合。任务预测器为每个任务生成最终标签。此外, 三个任务共享 MsIFT 中的多源特征融合模块。提出了一种具有编码器-解码器风格的特征融合变换器 (FFT), 用于多源特征级融合; 全局注意力机制有利于缓解由于配准不准确引起的语义偏差。FFT 允许特征执行全局查询, 激发每个查询特征聚合与其语义信息相似的全局特征。

