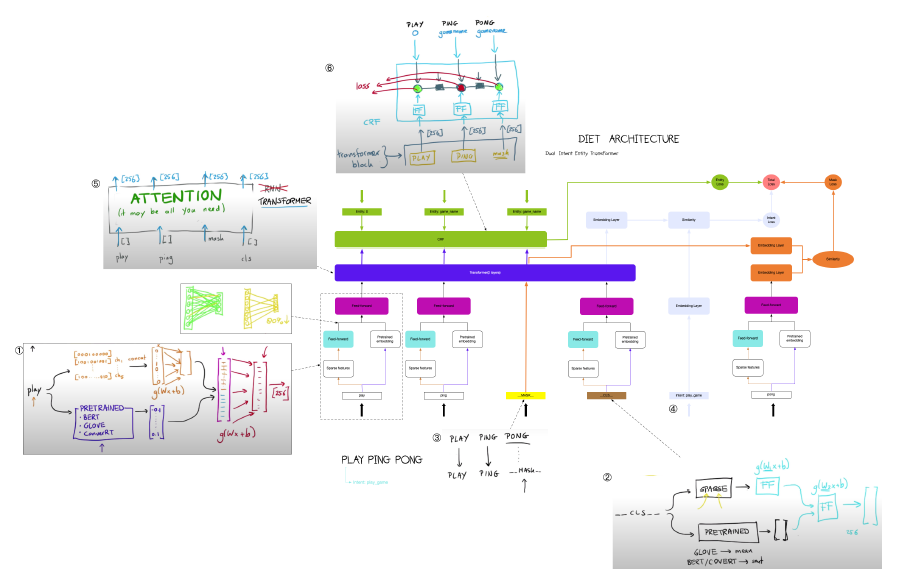
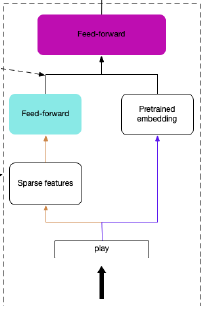
关于Rasa中DIET和DST理解

RASA中NLU的设计



整体结构



结构1

从结构1的token block可以看到它有两条链路，橙色是将token获取它的稀疏特征，然后通过一全连接层，紫色是它通过一个它的预训练的embedding向量。然后将他们的concat起来，然后通过一个全连接层(这其中有个drop out 随机去掉80%)获得一个256的向量

它对输入的每个词进行token时随机进行了mask,所以它的输入token包括了mask和cls。cls是对整句语句的理解信息。cls也会经过一个token block模块获得一个256向量，然后放入transformer中进行特征交叉得到一个整句话信息。然后经过一个embedding层。由于cls代表是一整句话的信息，所以将intent经过一个embedding层。然后让这两个embedding做一个相似度计算，让它们接近。这其中思想是一种星空算法思想。

Mask也是先通过transformer然后经过一个embedding层与原本的token经过token block得到一个256向量然后embedding统一一个维度，然后采用星空算法思想相似。采用MASK的原因一方面对话机器人输入预料不一定规范。去学习这种场景表达。另一方面是增加了模型难度，学习更泛化的特征。

Sloty识别是一个序列标注任务，transformer得到的带有上下文的向量，然后通过CRF来识别Sloty。

整个模型的loss包括intent loss，mask loss和entity loss 然后将它们相加作为整损失，然后反向传播去学习参数。

如何选择

对于简单的例如FQ,no entities,no mask,larger embedding

Hard的选择mask,bert,bigger FF,增加transformer的layer

RPI的NO PRETRAIN，更大的dropout，更小的embedding.

DST

获取多轮对话的intent和slot,我们需要自动识别hidden的Slot

Multi-turn中如何获得intent,需要多轮引导获得所有的intent

DST处理的Memory问题，多长多少次。这个需要设置。

以及Slot with Slot、change Slot、重复的intent问题、系统状态。