5总结与疑问

结果上看，引入”注意力”能使得模型表现得变好。

然后 我想这种模型(*Attention-based RNN model for joint intent detection and slot filling* ) 和 加入注意力的decoder-encoder的模型某种程度上都是为了解决把在识别槽值或者意图时，能让机器在计算单独某个槽值或者意图时看到的”视野”能更广泛一些，但是又不能让训练参数过多，让模型训练变难的一种方法吧，而论文中这种注意力的方法(像解码器把之前输出的h1,h2,h3,h4输入，*Attention-based RNN model for joint intent detection and slot filling也有类似操作*)，这样在每次传递过程中可以减少注意力也就是那些h在传输过程中由于深度过深(相比于像传统神经网络 的那种“头接尾“直接相连的接法)，而导致h的影响力变得过弱。感觉这个有点想那个“千层神经网络”的做法。

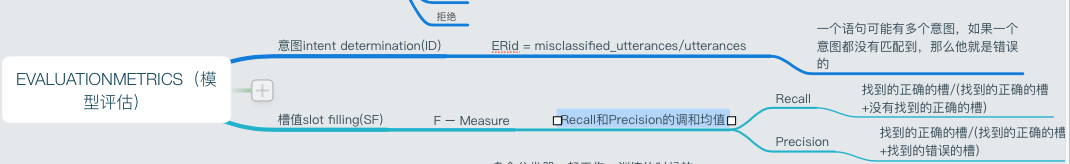
那个不过这种我感觉还是很玄学的，只有自己真正把模型敲出来，然后慢慢测试和感觉才能有更好的体会吧。

然后接下来就是问题了。

1. 实验的时候一次输入128个词汇？这是什么意思？多个句子同时输入?

还有句子是不定长的，怎么能刚刚好组合出128的句子呢？直接放空？

1. 整个模型中各个部分的激活函数的使用设置，以及要训练的参数的初始值设置。
2. adam优化器的其他参数设置？只看到max\_gradient\_norm=5
3. 使用的loss函数是什么？（直接就材料1那样?

）

1. 然后上图的效果是训练了几遍后的？
2. 还有材料1提到的CRF 看不懂。