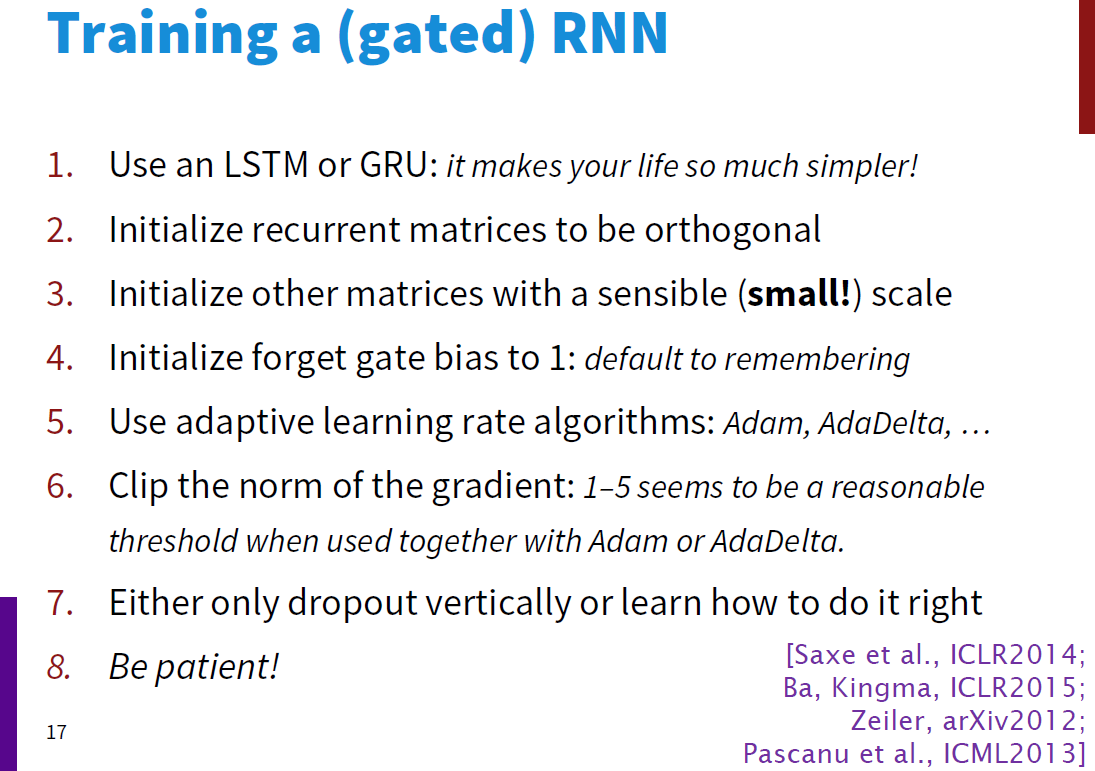
1、训练RNN初始化权重

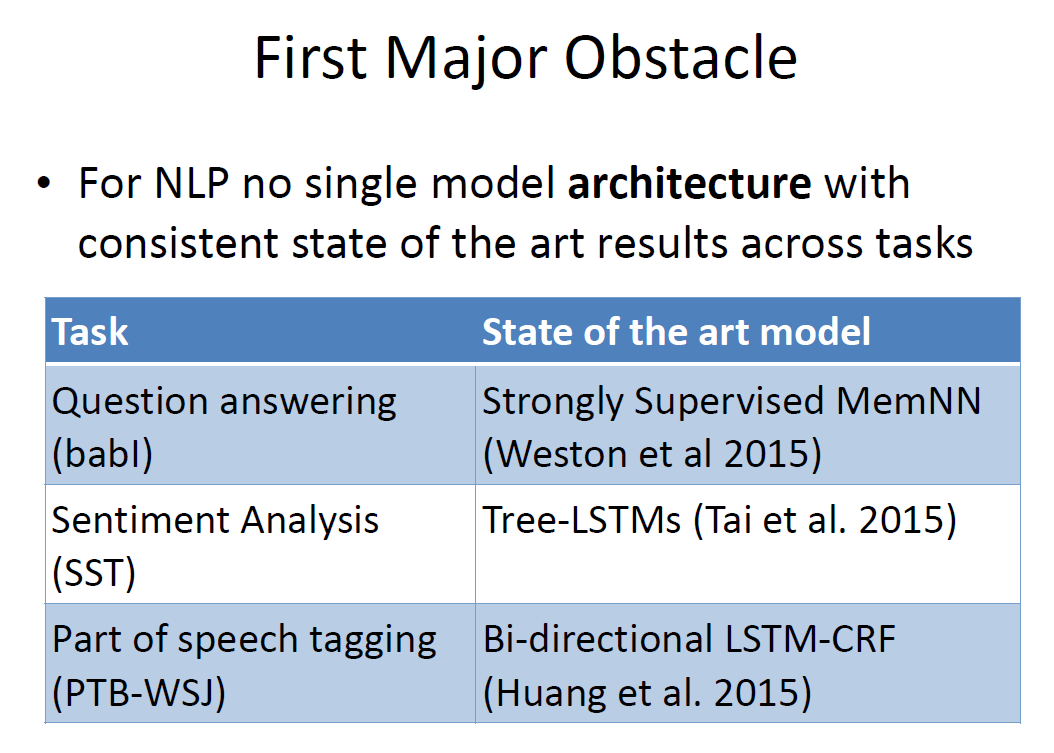


2、lecture8 SyntaxNet中对于无标签数据，使用两个非常不同的句法分析器进行依存句法分析，如果得到的结果相同，就把这个句子放入有标签的数据集中，从而提高了监督学习的数据集。可以获得相当大的提升。

因此，对于测试集，也可以用两个差异较大的方法进行训练，如果得到的结果是相同的，就可以把它放入后选的训练集，从而提高训练集的数量。

3、训练的时候nimi\_batch的size不能太大，因为这样会减弱noise的影响。很多函数是非凸优化的，所以引入适当的噪声可以防止停滞在局部最优解。包括dropout也可以看做是在引入噪声，以更可能获得全局最优解。

4、一些state-of-art的做法



5、对于图像问答系统，只给图像预测答案和只给问题预测答案也能得到一个很高的分数，说明数据集里面有一些特殊的pattern，或者leak