**标题：**SEQUENCE LEVEL TRAINING WITH RECURRENT NEURAL NETWORKS

**来源：**ICLR2016

**问题：**验证实验包括摘要生成、机器翻译、图像转文字。

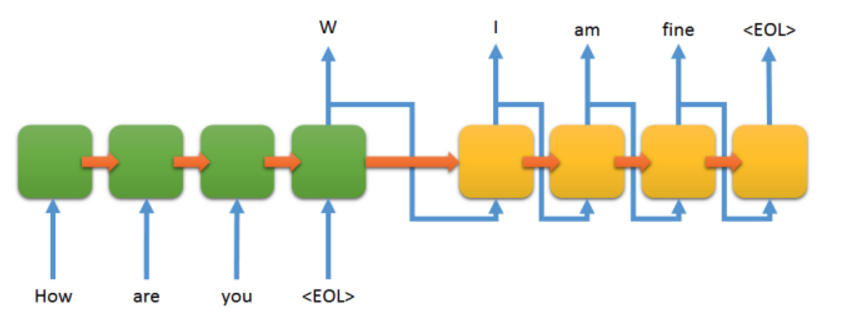
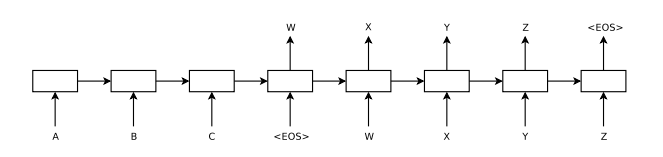
研究目标主要是在模型结构上解决seq2seq任务的exposure bias.

Exposure bias指的是在seq2seq的预测过程中的解码误差传递

造成这个问题的主要原因是由于seq2seq的训练和预测过程在解码部分存在不同的 参数依赖。例如在训练过程，decode过程的输入词是前一个词的期望输出，而在测 试阶段则是前一个词的预测输出。

**相关工作：**

1. 基础网络seq2seq的训练和预测图

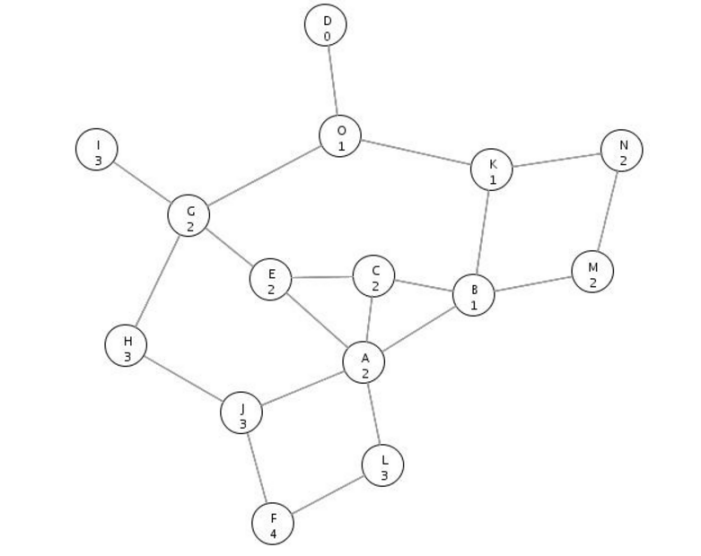


对应Exposure bias，该论文之前有两种解决办法

1. beam search

算法原理：图路径搜索中，每一步深度扩展的时候，剪掉一些质量比较差的结点， 保留下一些质量较高的结点，这样就减少了空间消耗，并提高了时间效率，但缺点就是 有可能存在潜在的最佳方案被丢弃，其中保留的质量较高路径的个数称为beam size.

（走n步吃到嘴多的果实）



应用到seq2seq里就是每一次解码过程看成是一步深度扩展，每一次解码预测概率最大的beam size个词就是候选结点，并累积概率选择概率和最高的beam size条路径，把整个解码过程看成是寻求最大联合概率的图搜索过程。该方法能解决Exposure bias的原因在于它使解码过程不仅仅依赖于前一个词输出，还要满足全局解码概率最大，因而原始模型前一个预测错误而带来的误差传递的可能性就降低了。

1. DAD

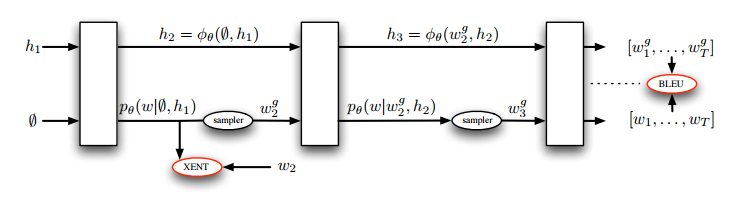
对于Exposure bias造成解码结果不好的原因，bengio解释为由于训练和预测过程在输入数据的分布不同，前者是样本的数据分布，后者则是模型的输出分布。因而解决办法就是保证两个流程在解码的时候输入参数服从相同分布，即都采用前一个词的预测结果作为当前词的输入。为此提出了一种退火算法来解决这个问题，引入一个概率值参数，称其为温度，

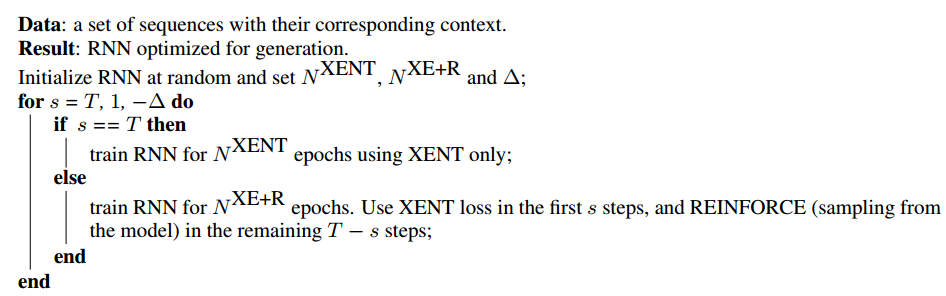
当温度值较大时高概率采用前一个词的期望输出作为输入，当值较小时，则高概率采用预测输出，随着迭代次数的增加，该参数趋近于0.即完全采用前一个词的预测输出。

**主要方法：**

1. 本文采用的方法

作者思考了DAD和beam search的算法后，认为可以通过引入增强学习的思想来进行结 合



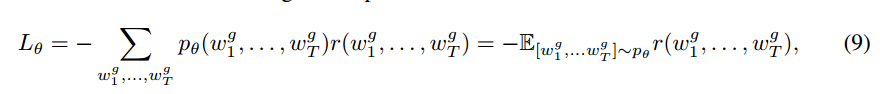


算法流程

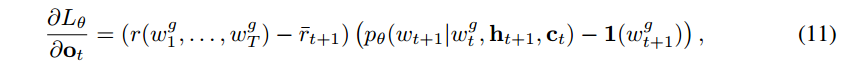
1. 传统的RNN方式进行训练，即解码的前一词输入为期望输入
2. 当迭代到N(XENT)次，逐渐在decode的末尾引入增强学习的decode方式
3. 训练到最后除了第一个timestep采用期望输入其余词都采用增强学习的方式（因为 第一个词的前一个词的期望输入固定的<eos>）

增强学习的应用过程

1. action依旧是每一个timestep的候选词
2. state是每一个timestep的隐藏层状态
3. reward是解码完成以后的bleu

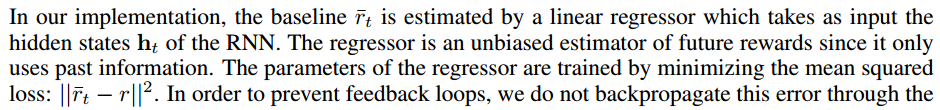


累积的loss如上，单个timestep的loss如下。



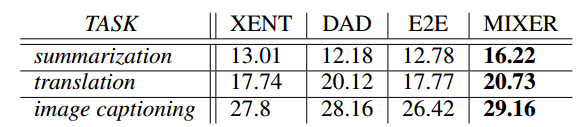


这里的r[t+1]不是指的总体的



单个词的r[t]是怎么算出来的？

F(h1)+F(h2)+F(h3)....F(hT) = r(w(1)w(2)w(..)w(T))????



**简评：**

该篇论文详细描述了所提方案的思考过程，从相关工作一步一步介绍的过程中，不断思考其他模型的不足，最终提出针对性解决问题的本文模型，整个研究过程值得学习。

本文引入增强学习的nlp领域恰到好处，因为rl是面向基于马尔可夫过程的动态规划问题，而RNN的隐层传递确实也符合HMM过程，并且beam search也generative模型的解码过程应用动态规划能有更好的效果。所以在seq2seq的decode部分引入rl是合情合理的

本文的优势在于不仅仅解决了训练和预测分布不同的问题，还将多候选词的探索引入到了训练过程中，因为DAD在解决预测词输入的时候只能选择最高概率词，而beam search这个方法又只能局限在预测阶段。

参考文献：

Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems. 2014: 3104-3112.

Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. Scheduled sampling for sequence prediction with recurrent neural networks. In NIPS, 2015.