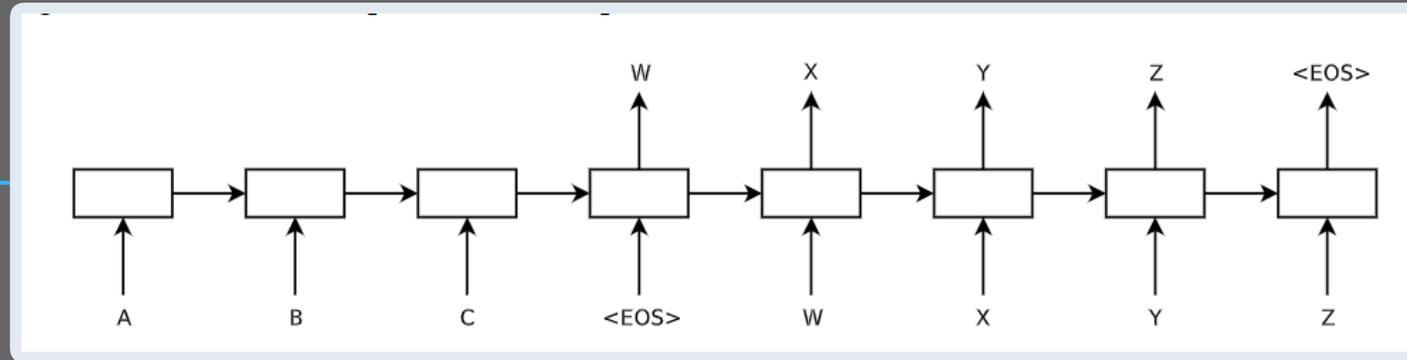


Seq2Seq

提出了一种通用的端到端的序列学习方法

在此之前，神经网络只能应用于输入和目标可以用固定维数的向量编码的问题。

模型架构图



构造序列到序列模型，以循环神经网络LSTM为基础，ABC是原句子，<EOS>是结束符，WXYZ是依次生成的序列，其中上一时刻的输出作为下一时刻的输入，结合中间传导过来的“记忆”一直生成到EOS为止。（原始长度怎么设定？）

关键技术

reversed the order of words in the source sentence

map c, b, a to α, β, γ , where α, β, γ is the translation of a, b, c. This way, a is in close proximity to α , b is fairly close to β

结果确实好了很多，作者认为是头几个单词更近

By doing so, we introduced many short term dependencies that made the optimization problem much simpler

神经网络学习仅仅是建立映射关系，从流程图来看确实有一部分更近了，借由这种思想，那是否可以在合适的位置额外添加结构？

used two different LSTMs: one for the input sequence and another for the output sequence

makes it natural to train the LSTM on multiple language pairs simultaneously

deep LSTMs significantly outperformed shallow LSTMs,

chose an LSTM with four layers

训练

Beam Search

当Beam width $B=2$ 时，每一层保留前两个最有可能的prefix带着往下走，直到到达最大句子长度（词语个数或者说time step）或者中间遇到EOS（也就是预测到这是句尾）

在概率的基础加了一个“对过短的输出句子给出惩罚”

论文声称beam长度取2最佳，实际上开源代码实现偷偷取了4

8个GPU跑了10天，最终实现的速度达到每秒6300个单词

在长句上的表现意外的也好