

受 Dilated CNN 的启发

Dilated RNN

主要两个方面

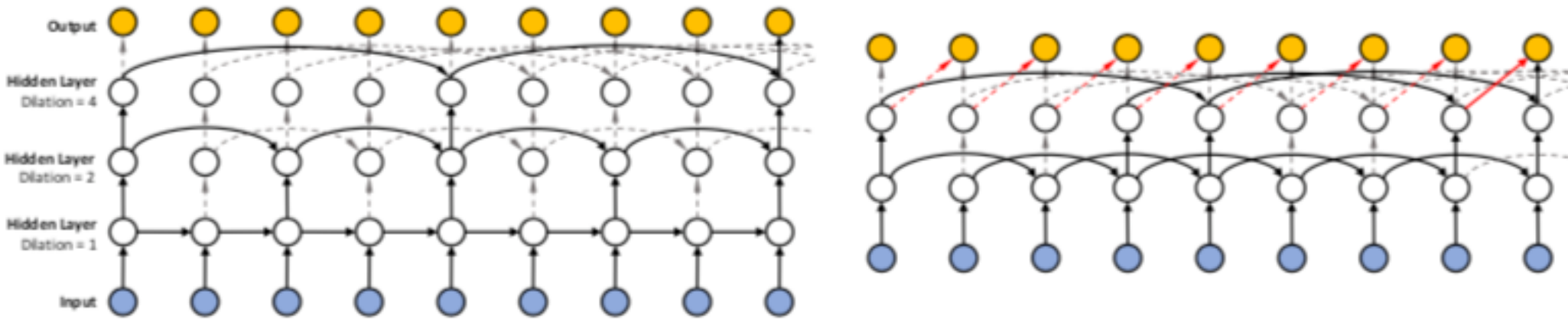
- Dilated recurrent skip connection
- exponentially increasing dilation 指数级膨胀

Dilated Recurrent Skip Connection

dilated skip connection 的公式如下

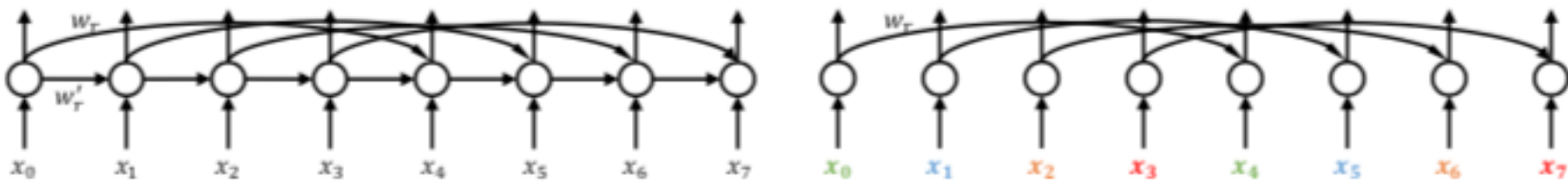
$$c_t^{(l)} = f(x_t^{(l)}, c_{t-s^{(l)}}^{(l)})$$

其中 $s^{(l)}$ 是第 l 层的skip length 跳跃长度， x_t^l 是 l 层时间 t 的输入， $f()$ 是任何 RNN 细胞结构，如原版 RNN、LSTM、GRU 等



如图，左图为dilation 分别为 1，2，4 的网络结构，右图为单独 dilation 为 2 的网络结构，但是缺少了相邻时间的依赖性，所以在红色箭头处添加

Dilated Recurrent Skip Connection vs regular Skip Connection



该版本去除了同层之前时间点的输入($c_{t-1}^{(l)}$), 仅保留前一个跳跃点($c_{t-s^{(l)}}^{(l)}$)的输入

Dilated Recurrent Skip Connection 的并行计算

该网络结构有助于并行计算，例如在 dilation = 4 时， $c_{4t}, c_{4t+1}, c_{4t+2}, c_{4t+3}$ 的计算不依赖于不共用数值，因此可以并行处理，理论上l 层的计算速度提高了 s^l 倍

Exponentially Increasing Dilation

通过叠加 recurrent layers 来构建 Dilated RNN

$$s^{(l)} = M^{l-1}, l = 1, 2, \dots, L$$

优点

- 一方面，叠加多层 RNN 层提高了模型复杂度
- 另一方面，使用指数级增长，既可以让不同的层聚焦于不同的tenporal resolutions
- 又可以减少了节点之间的平均路径，让 RNN 可以提取长期相关依赖性，同时避免了梯度消失和梯度爆炸

Generalized Dilated RNN

$$s^l = M^{l-1+l_0}, l = 1, \dots, L, l_0 \geq 0$$

即 dilation 不从 1 但是从 M^{l_0} 开始，但是这样可能会损失一些短于 M^{l_0} 的依赖性，所以要在最后输出层加上，如上图 1 的右侧