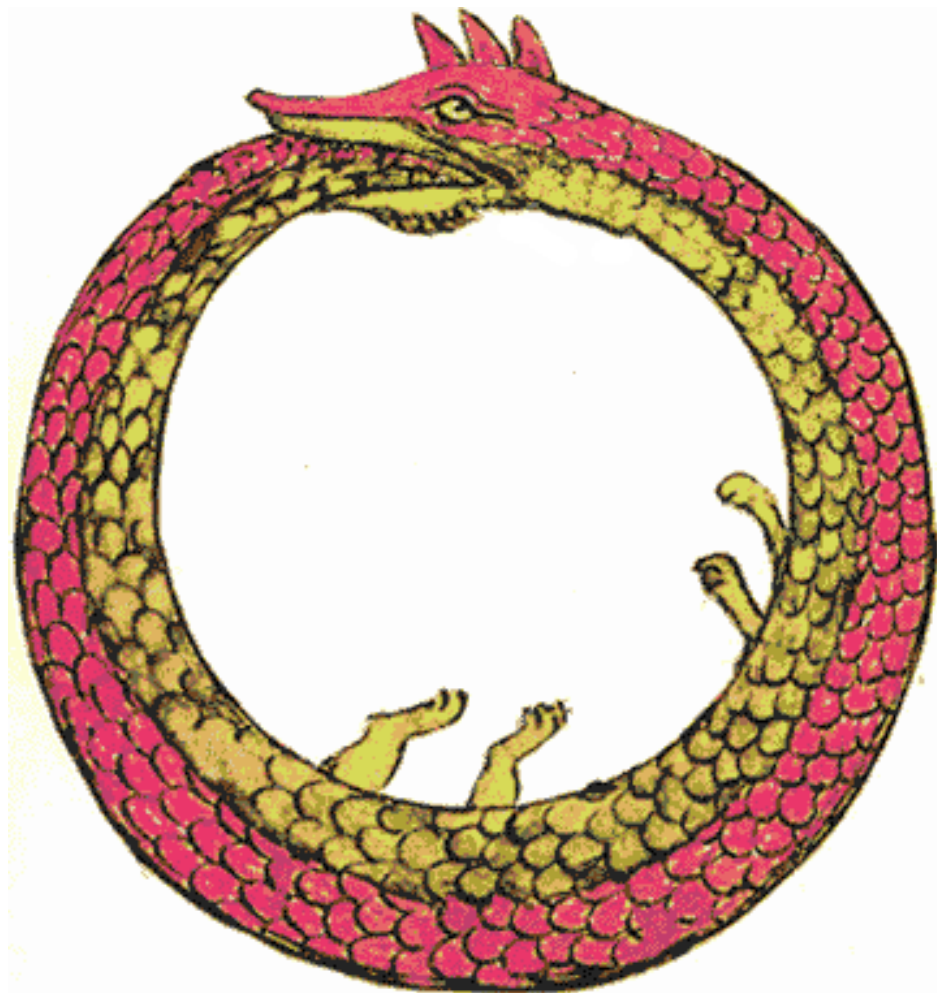


20 高级循环神经网络

概要

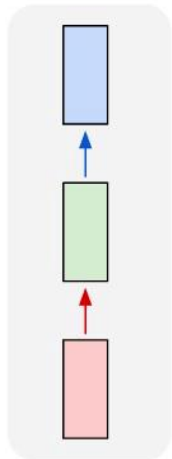
- 深度循环神经网络
- 双向循环神经网络
- 循环神经网络结合
 - 残差网络 (ResNet)
 - 稠密连接网络 (DenseNet)
- 循环神经网络的正则化

深度循环神经网络

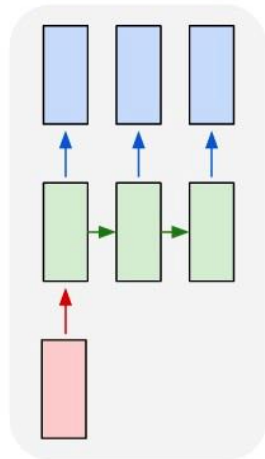


使用循环神经网络

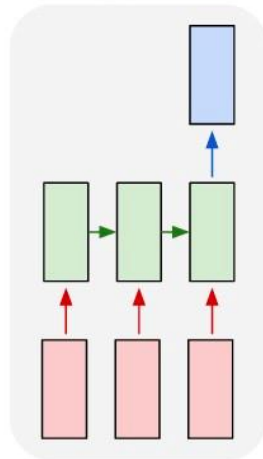
➤ 序列到序列
➤ one to one



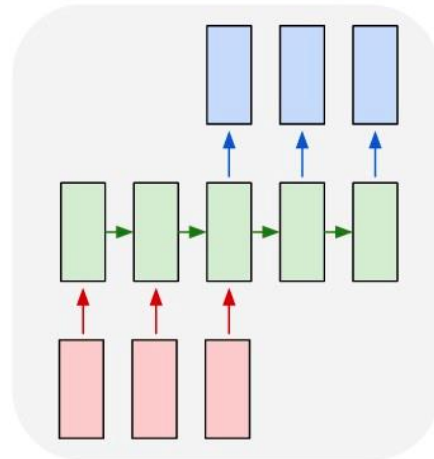
➤ one to many



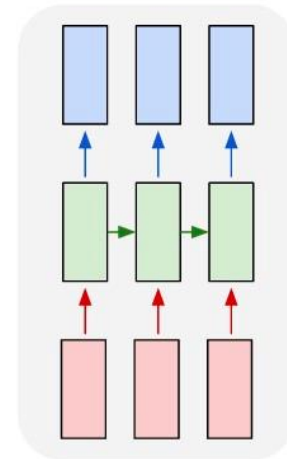
➤ many to one



➤ many to many



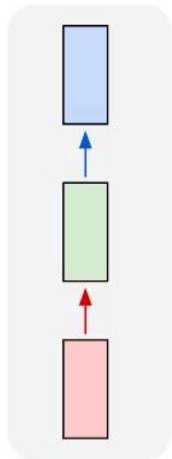
➤ many to many



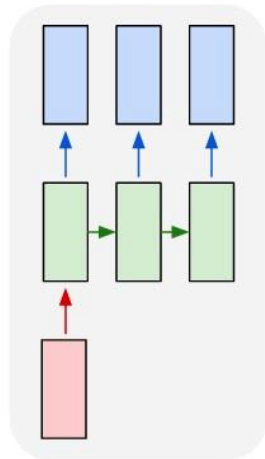
(image courtesy of [karpathy.github.io](https://github.com/karpathy))

使用循环神经网络

one to one

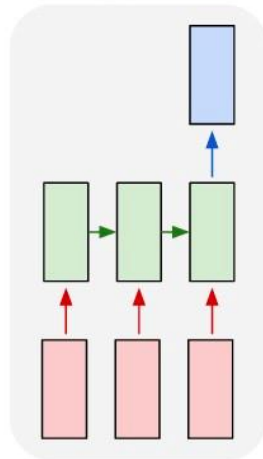


one to many



编写诗歌

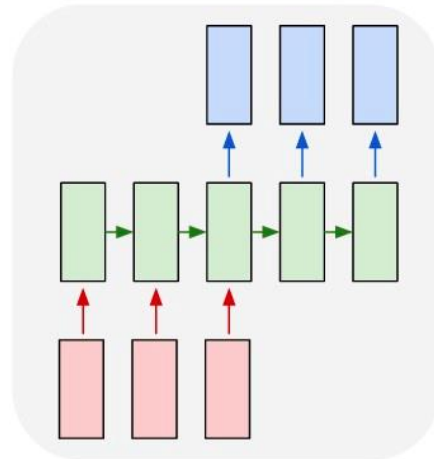
many to one



情感分析

文本分类

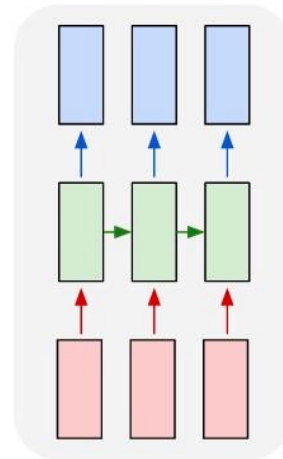
many to many



自动问答

机器翻译

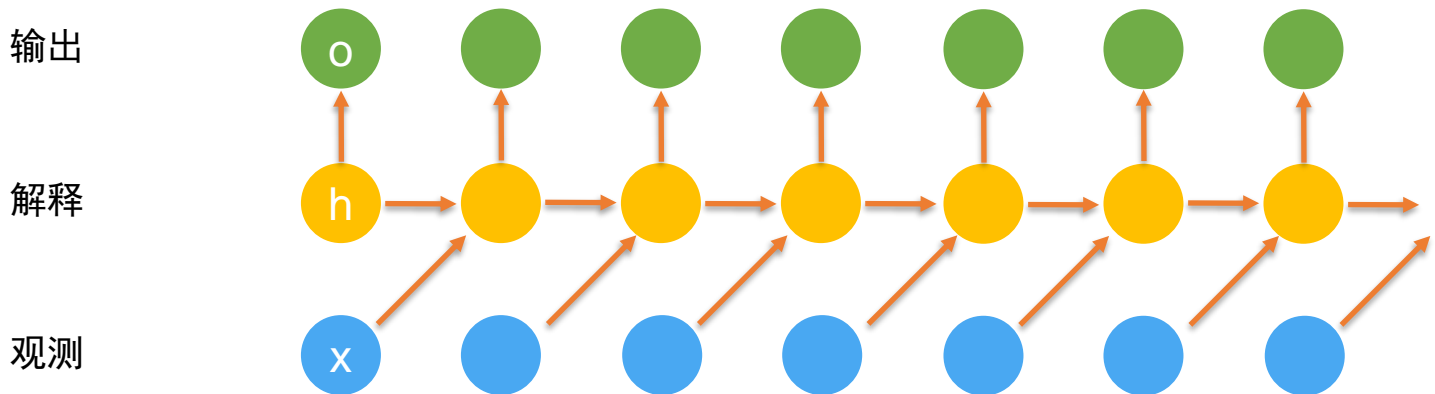
many to many



命名实体标记

(图片来源: karpathy.github.io)

回顾 - 循环神经网络



➤ 隐含状态更新

$$\mathbf{h}_t = \phi(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{hx}\mathbf{x}_{t-1} + \mathbf{b}_h)$$

➤ 观测更新

$$\mathbf{o}_t = \phi(\mathbf{W}_{ho}\mathbf{h}_t + \mathbf{b}_o)$$

➤ 怎么添加更多非线性的层？

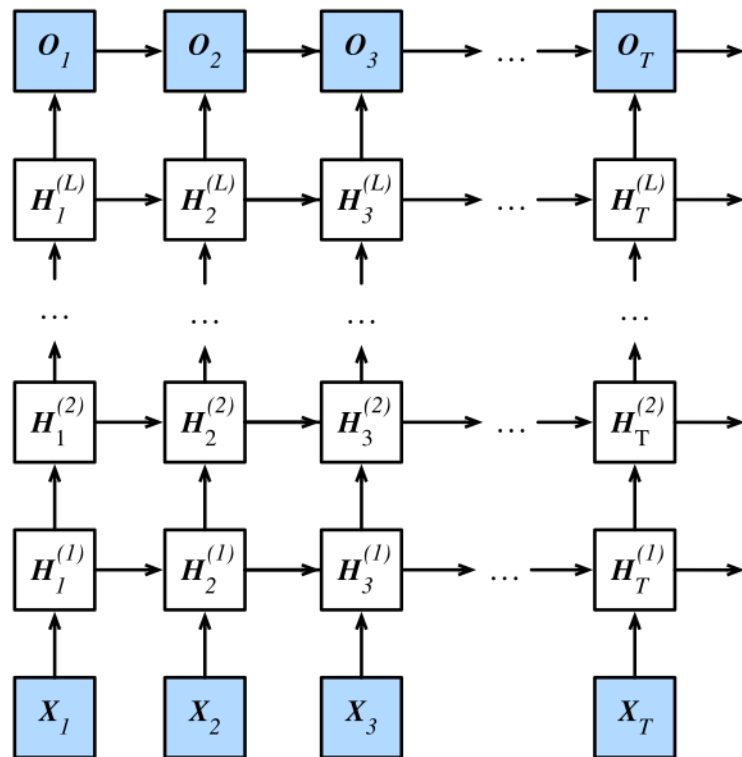
深度循环神经网络

浅度循环神经网络

- 输入
- 隐含层
- 输出

深度循环神经网络

- 输入
- 隐含层
- 隐含层
- ...
- 输出

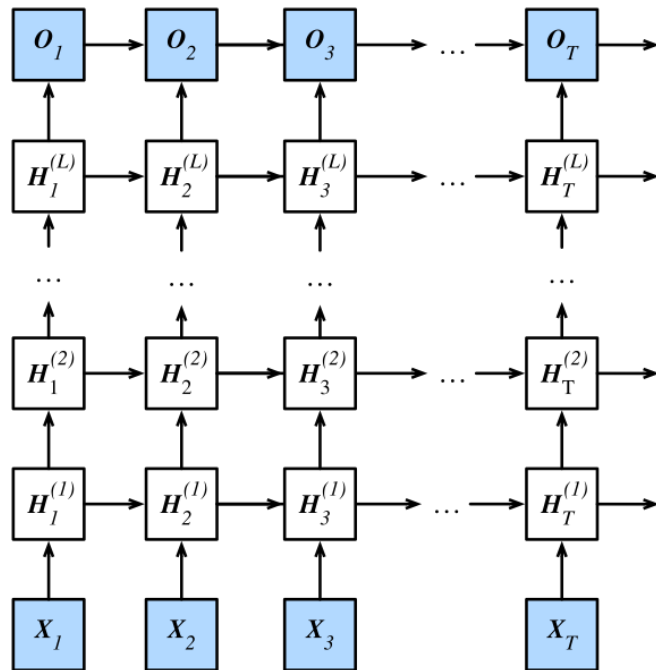


深度循环神经网络

➤ $\mathbf{H}_t^1 = f_1(\mathbf{H}_{t-1}^1, \mathbf{X}_t)$

➤ $\mathbf{H}_t^j = f_j(\mathbf{H}_{t-1}^j, \mathbf{H}_t^{j-1})$

➤ $\mathbf{O}_t = g(\mathbf{H}_t^L)$



双向循环神经网络



“未来”的重要性

- 我 _____ 。
- 我 _____ 非常饿。
- 我 _____ 非常饿，我可以吃下一只猪。

“未来”的重要性

- 我 饿 。
- 我 不 是 非 常 饿 。
- 我 非 常 非 常 饿 ， 我 可 以 吃 下 一 只 猪 。

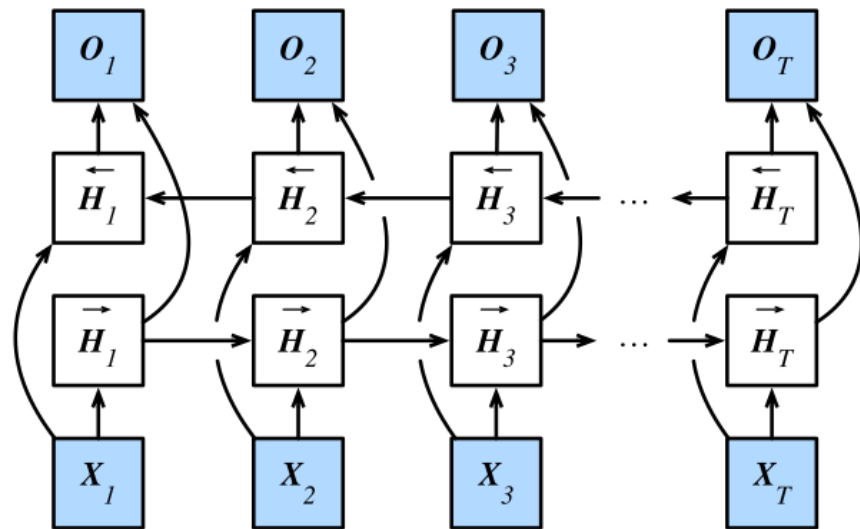
“未来”的重要性

- 我 饿 。
- 我 不是 非常饿。
- 我 非常 非常饿，我可以吃下一只猪。

- 填写完全不同的单词，取决于单词的上下文
- 到目前为止，RNN 只关注上文
- 也可以使用下文来填写

双向循环神经网络

- 一个 RNN 正向传递
- 另一个 RNN 反向传递
- 将两种隐含状态组合在一起以生成输出



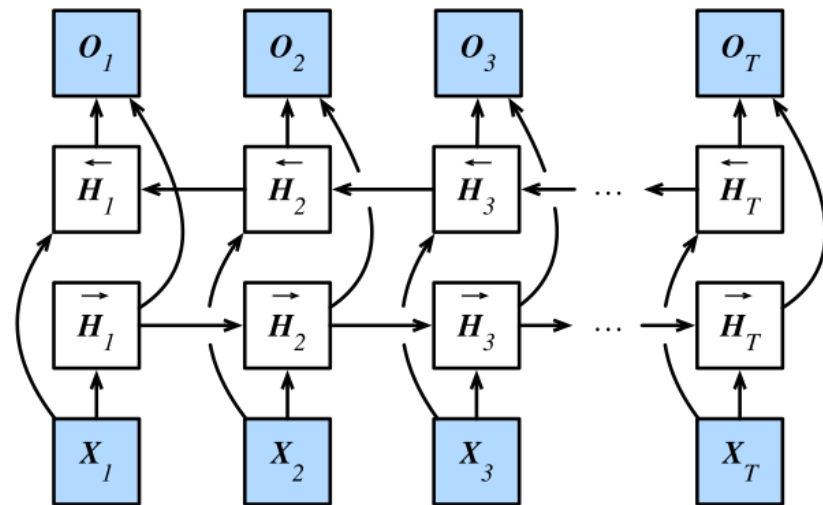
双向循环神经网络

$$\vec{\mathbf{H}}_t = \phi \left(\mathbf{X}_t \mathbf{W}_{xh}^{(f)} + \vec{\mathbf{H}}_{t-1} \mathbf{W}_{hh}^{(f)} + \mathbf{b}_h^{(f)} \right)$$

$$\leftarrow \mathbf{H}_t = \phi \left(\mathbf{X}_t \mathbf{W}_{xh}^{(b)} + \leftarrow \mathbf{H}_{t+1} \mathbf{W}_{hh}^{(b)} + \mathbf{b}_h^{(b)} \right)$$

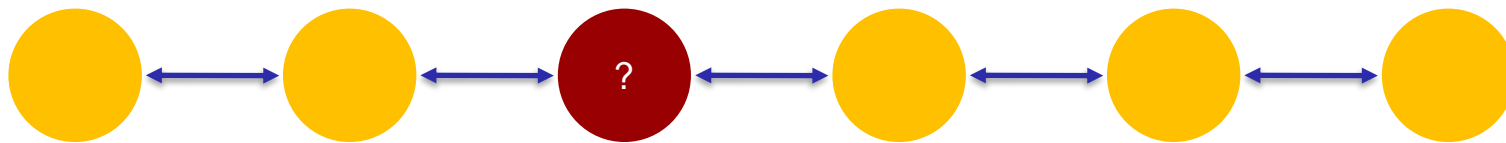
$$\mathbf{H}_t = [\vec{\mathbf{H}}_t, \leftarrow \mathbf{H}_t]$$

$$\mathbf{O}_t = \mathbf{H}_t \mathbf{W}_{hq} + \mathbf{b}_q$$

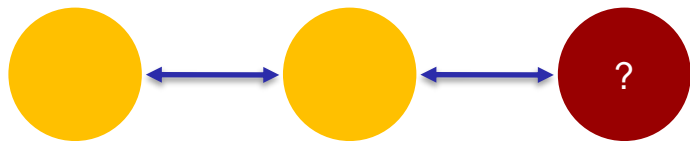


推理

➤ 训练阶段



➤ 推理（不合适）



➤ 仍然可以使用它来编码序列

总结

- 双向循环神经网络通过反向更新的隐藏层来利用方向时间信息
- 通常用来对序列抽取特征、填空,而不是预测未来