

序列预测问题与循环神经网络

目录

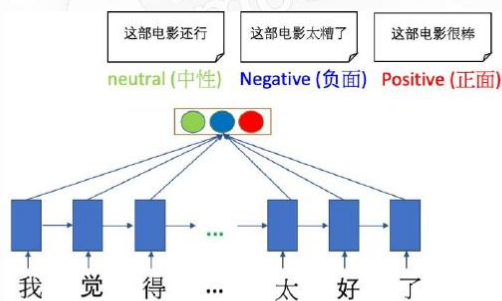
◆ 序列预测问题

◆ 循环神经网络

序列预测问题

典型序列预测问题-分类

- ◆ 输入不定长序列，给出类别预测，常采用Sequence-to-vector结构



情感分类



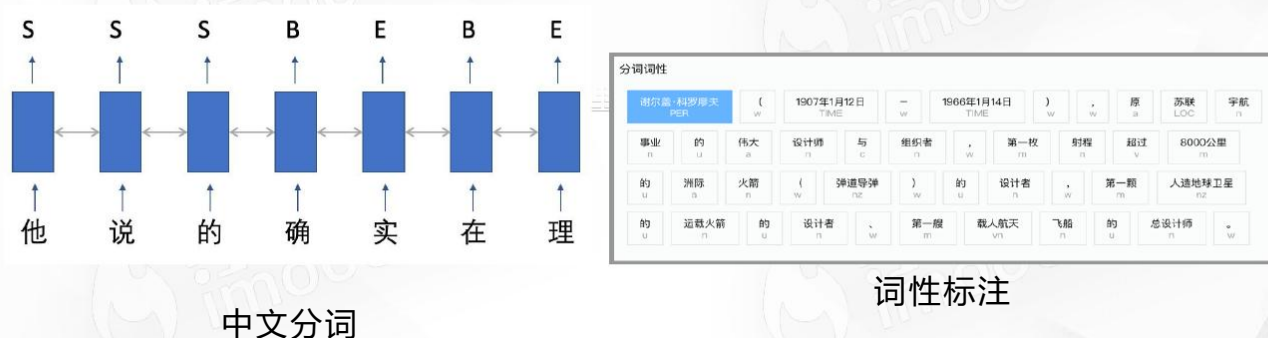
视频分类



语音分类

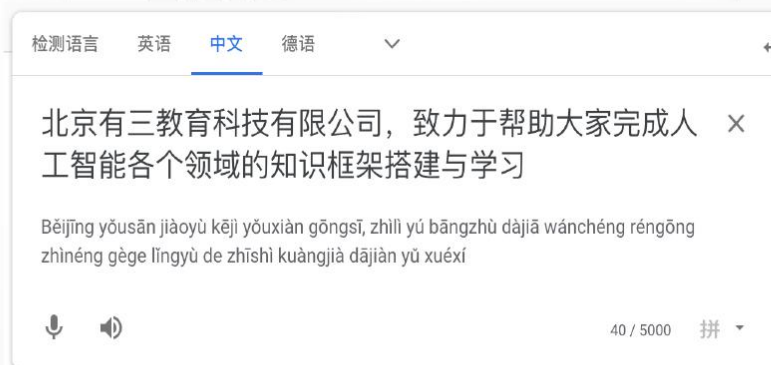
典型序列预测问题-同步序列预测

- ◆ 输入不定长序列，输出同样长度的标注结果，常采用Sequence-to-sequence结构



典型序列预测问题-异步序列预测

- ◆ 输入不定长序列，输出不同长度的预测结果，常采用Encoder-Decoder结构



机器翻译

典型序列预测问题-序列生成

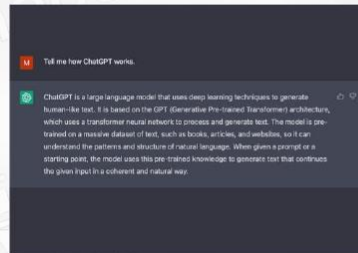
◆ 预测不定长的输出序列，常采用vector-to-Sequence结构



语音生成



图像标注

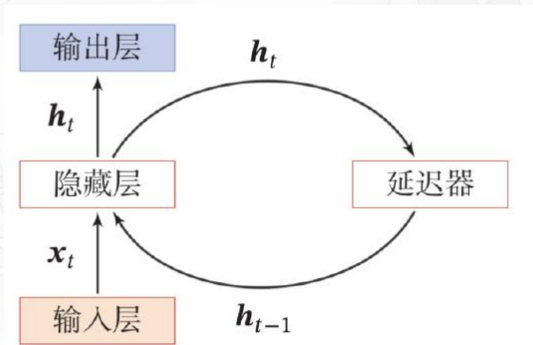


文本生成

循环神经网络

循环神经网络

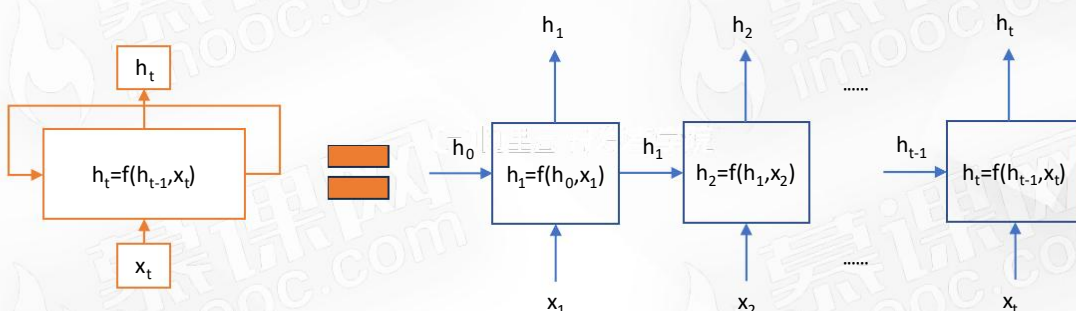
- ◆ Recurrent Neural Networks, 简称RNN, 一种带自反馈的神经网络, 能够处理任意长度的时序数据



循环神经网络比前馈神经网络更加符合生物神经网络的结构, 被广泛应用于语音识别、语言模型等任务上。

RNN模型展开

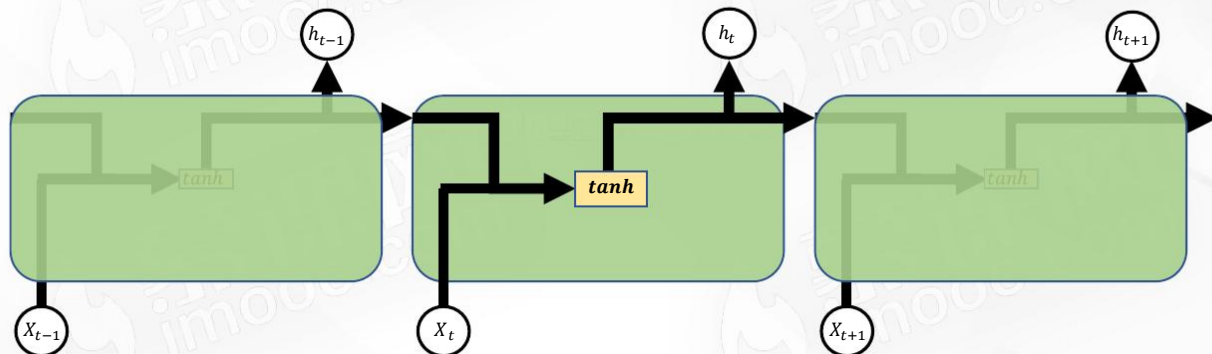
- ◆ 将RNN按照时间维度展开



输入是 $\{x_1, x_2, \dots, x_t\}$, 网络内部隐藏单元为 $h_t = f(h_{t-1}, x_t)$, 输出是一个序列 $\{h_1, h_2, \dots, h_t\}$, 对于序列的每一个输入, f 都是相同的, h_t 与序列 $\{x_1, x_2, \dots, x_t\}$ 中每一个元素有关系。

RNN内部结构单元的计算

- ◆ 令 x 是长度为 I 的向量， h_t 是长度为 H 的向量，将其拼接成 $(I + H)$ 的向量作为输入，对应权重系数 W 的维度是 $H(I + H)$

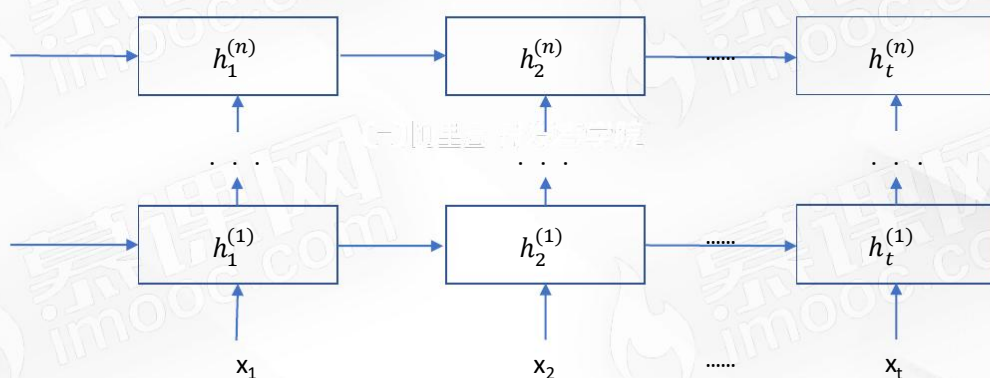


$$h_t = \tanh\left(W \begin{pmatrix} x_t \\ h_{t-1} \end{pmatrix}\right) = \tanh(W_{xh}x_t + W_{hh}h_{t-1})$$

输出 t 时刻的预测 $y_t = \tanh(W_{ho}h_t)$, W_{ho} 的维度是 $K \times H$, K 是输出维度

深层RNN模型

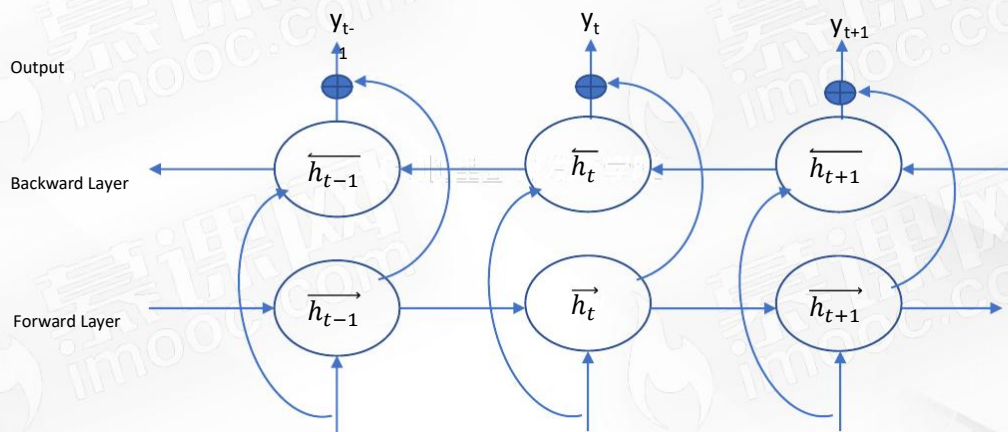
- ◆ 包含了多个隐藏层，获得更复杂的表达能力



每个隐藏状态不断传递到当前层的下一个时间步以及当前时间步的下一层

双向RNN模型

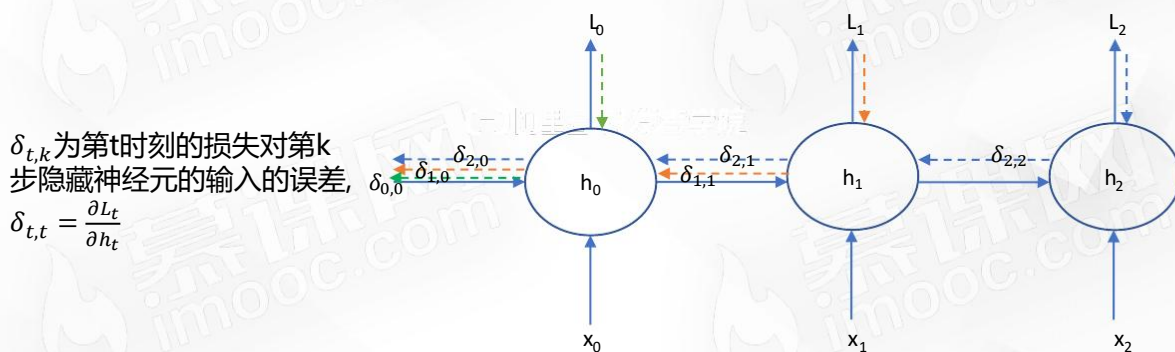
- ◆ 不仅包括正向过程，也包含反向过程



双向的RNN获得更加完整的“上下文”信息，具有更强大的表达能力，在语音识别，机器翻译任务中都得到了比单向RNN更好的性能

RNN的参数学习

- ◆ 随时间反向传播算法BPPT(backpropagation through time), 误差不仅依赖于当前时刻 t ，也依赖于之前时刻 k , $0 < k < t$ 。



第0个时刻得到第 t 个时刻传来的导数 $\delta_{t,t} \cdot \left(\frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial h_{t-2}} \dots \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial h_0} \right)$

RNN梯度问题

- ◆ 存在长距离依赖问题（梯度爆炸或消失），参数更新主要依靠当前时刻的相邻状态

第0个时刻得到第t个时刻传来的导数 $\delta_{t,t} \cdot \left(\frac{\partial h_t}{\partial h_{t-1}} \cdot \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdots \frac{\partial h_2}{\partial h_1} \cdot \frac{\partial h_1}{\partial h_0} \right)$

$$h_t = \tanh(W_{xh}x_t + W_{hh}h_{t-1}) \quad \frac{\partial h_t}{\partial h_{t-1}} = W_{hh}(1 - (\tanh(W_{xh}x_t + W_{hh}h_{t-1}))^2)$$

- (1) 连乘存在 W_{hh} 的t次幂，如果矩阵特征值远大于1，多次连乘会造成梯度爆炸；如果远小于1，会造成梯度消失；
- (2) Tanh激活函数如果进入饱和区间，连乘后数值变得很小；

下次预告：长短时记忆网络与门控循环单元