6. 门控循环单元网络 (GRU)

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot g(x_t, h_{t-1}; \theta)$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

$$\tilde{h}_t = g(x_t, h_{t-1}; \theta)$$

$$= \tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$

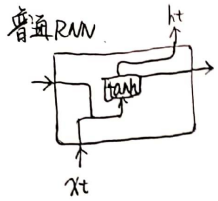
r_t 为重置门，是否清除 h_{t-1}

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r)$$

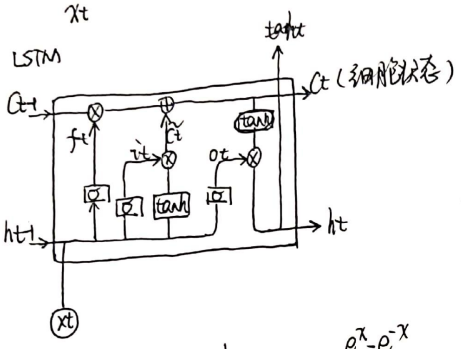
LSTM(长短时间记忆网络)

避免长期依赖问题

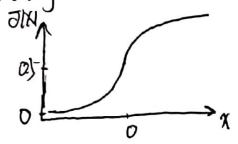
普通RNN



LSTM



1) σ: sigmod $\sigma(x)=\frac{1}{1+e^{-x}}$, $\tanh x=\frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$

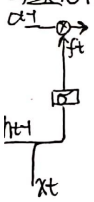


$1-2\sigma(x)=-\tanh(\frac{x}{2})$

⊗: 按位乘法

⊕: 按位加

2)遗忘门(在记输入中Ct中不必要的信息)

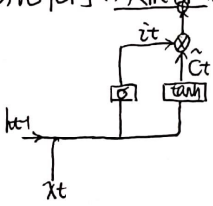


$f_t=\sigma(W_f\cdot[h_{t-1},x_t]+b_f)$

$f_t\in[0,1]$

$C_t=f_t\cdot C_{t-1}$ 按位乘

3)记忆门(决定ht-1和Xt哪些需要保留)



$i_t=\sigma(W_i\cdot[h_{t-1},x_t]+b_i)$

$\tilde{C}_t=\tanh(W_C\cdot[h_{t-1},x_t]+b_C)$

$C_t=C_{t-1}+i_t\cdot\tilde{C}_t$

更新细胞状态

$C_t=f_t*C_{t-1}+i_t*\tilde{C}_t$

3)输出门

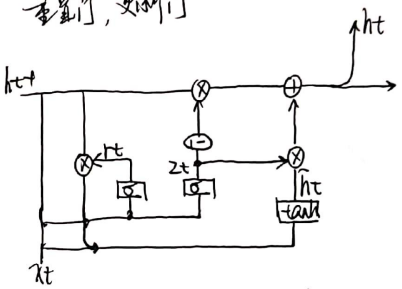
$o_t=\sigma(W_o[h_{t-1},x_t]+b_o)$

$h_t=o_t*\tanh(C_t)$

变体:

GRU:将合了LSTM的Ct和ht

重置门,更新门



$z_t=\sigma(W_z\cdot[h_{t-1},x_t])$

$r_t=\sigma(W_r\cdot[h_{t-1},x_t])$

$\tilde{h}_t=\tanh(W\cdot[r_t*h_{t-1},x_t])$

$h_t=(1-z_t)*h_{t-1}+z_t*\tilde{h}_t$