



deeplearning.ai

Recurrent Neural Networks

长短期记忆网络

LSTM (long short
term memory) unit

GRU and LSTM

更适用于构建较大网络，效率更高，可训练性更强

GRU

$$\underline{\tilde{c}^{<t>}} = \tanh(W_c[\underline{\Gamma_r} * \underline{c^{<t-1>}}, x^{<t>}] + b_c)$$

$$\underline{\Gamma_u} = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$$

$$\underline{\Gamma_r} = \sigma(W_r[c^{<t-1>}, x^{<t>}] + b_r)$$

$$\underline{c^{<t>}} = \underline{\Gamma_u} * \underline{\tilde{c}^{<t>}} + \underline{(1 - \Gamma_u) * c^{<t-1>}}$$

$\underline{a^{<t>}} = \underline{c^{<t>}}$

更强大的记忆

LSTM (历史更长)

$$\underline{\tilde{c}^{<t>}} = \tanh(\omega_c[a^{<t-1>}, x^{<t>}] + b_c)$$

$$\left. \begin{array}{l} \text{(update)} \quad \Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u) \\ \text{(forget)} \quad \Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f) \\ \text{(output)} \quad \Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o) \end{array} \right\}$$

$$\underline{c^{<t>}} = \underline{\Gamma_u} * \underline{\tilde{c}^{<t>}} + \underline{\Gamma_f} * \underline{c^{<t-1>}}$$

$$\underline{a^{<t>}} = \underline{\Gamma_o} * \underline{c^{<t>}} \rightarrow \tanh c^{<t>}$$

LSTM in pictures

$$\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$$

$$\Gamma_u = \sigma(W_u[a^{<t-1>}, x^{<t>}] + b_u)$$

$$\Gamma_f = \sigma(W_f[a^{<t-1>}, x^{<t>}] + b_f)$$

$$\Gamma_o = \sigma(W_o[a^{<t-1>}, x^{<t>}] + b_o)$$

$$c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + \Gamma_f * c^{<t-1>}$$

$$a^{<t>} = \Gamma_o * c^{<t>}$$

↑
tanh

前个孔
perphole
connection

