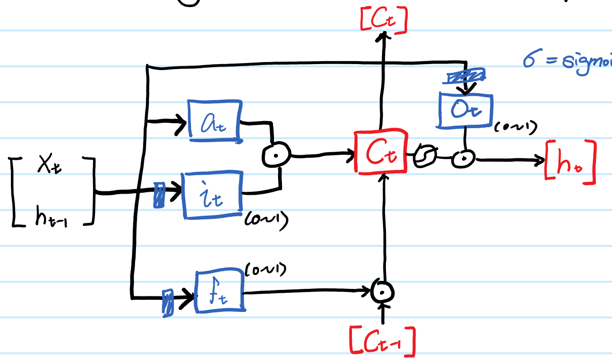
LSTM

2017년 7월 15일 토요일 오전 2:03

LSTM (Long Short-Term Memory) Structure



σ = sigmoid function

Feed Forward

(Input activation) $a_t = \tanh(W_c \cdot x_t + U_c \cdot h_{t-1}) = \tanh(\hat{a}_t)$

(Input gate) $i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1}) = \sigma(\hat{i}_t)$

(Forget gate) $f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1}) = \sigma(\hat{f}_t)$

(Output gate) $O_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1}) = \sigma(\hat{o}_t)$

Ingnores non-linearities

$$z_t = \begin{bmatrix} \hat{a}_t \\ \hat{i}_t \\ \hat{f}_t \\ \hat{o}_t \end{bmatrix} = \underbrace{\begin{bmatrix} W_c & U_c \\ W_i & U_i \\ W_f & U_f \\ W_o & U_o \end{bmatrix}}_{\text{matrix } W} \times \underbrace{\begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix}}_{\text{Vector } I} = W \times I_t$$

present activated value / past value

(Internal state) $C_t = i_t \odot a_t + f_t \odot C_{t-1}$

How much put in... / How much forget...

avoid Gradient vanishing

(output) $h_t = O_t \odot \tanh(C_t)$

Backpropagation Through Time (BPTT)

$$\delta h_t = \frac{\partial E}{\partial h_t}$$

$* \; C_t = i_t \odot a_t + f_t \cdot C_{t-1}$

$$\delta O_t = \frac{\partial E}{\partial O_t} = \boxed{\frac{\partial E}{\partial h_t}} \cdot \frac{\partial h_t}{\partial O_t} = \delta h_t \odot \tanh(C_t)$$

$$\delta C_t = \frac{\partial E}{\partial C_t} = \frac{\partial E}{\partial h_t} \cdot \frac{\partial h_t}{\partial C_t} = \delta h_t \odot O_t \odot (1 - \tanh^2(C_t))$$

$$\delta a_t = \frac{\partial E}{\partial a_t} = \frac{\partial E}{\partial C_t} \cdot \frac{\partial C_t}{\partial a_t} = \delta C_t \odot i_t$$

$$\delta i_t = \frac{\partial E}{\partial i_t} = \frac{\partial E}{\partial C_t} \cdot \frac{\partial C_t}{\partial i_t} = \delta C_t \cdot a_t$$

$$\delta f_t = \frac{\partial E}{\partial f_t} = \frac{\partial E}{\partial C_t} \cdot \frac{\partial C_t}{\partial f_t} = \delta C_t \cdot C_{t-1}$$

$$\delta C_{t-1} = \frac{\partial E}{\partial C_{t-1}} = \frac{\partial E}{\partial C_t} \cdot \frac{\partial C_t}{\partial C_{t-1}} = \delta C_t \odot f_t$$

$$\frac{\partial E}{\partial \hat{a}_t} = \frac{\partial E}{\partial a_t} \cdot \frac{\partial a_t}{\partial \hat{a}_t} = \delta a_t \cdot \frac{\partial \tanh(\hat{a}_t)}{\partial \hat{a}_t}$$

$$\delta \hat{a}_t = \delta a_t \odot (1 - \tanh^2(\hat{a}_t))$$

$$\delta \hat{i}_t = \delta i_t \odot i_t \odot (1 - i_t)$$

$$\delta \hat{f}_t = \delta f_t \odot f_t \odot (1 - f_t)$$

$$\delta \hat{o}_t = \delta O_t \odot O_t \odot (1 - O_t)$$

$$\delta z_t = [\delta \hat{a}_t, \delta \hat{i}_t, \delta \hat{f}_t, \delta \hat{o}_t]^T$$

$$\delta I_t = W^T \times \delta z_t$$

$$I_t = \begin{bmatrix} x_t \\ h_{t-1} \end{bmatrix} \rightarrow \delta h_{t-1} \text{ from } \delta I_t$$

$$\delta W = \delta z_t \times (I_t)^T$$

Can Update all weights

$$\frac{\partial E}{\partial I_t} = \frac{\partial E}{\partial z_t} \cdot \frac{\partial z_t}{\partial I_t}$$

$$= \delta z_t \cdot \frac{\partial [\hat{a}_t, \hat{i}_t, \hat{f}_t, \hat{o}_t]}{\partial [x_t, h_{t-1}]}$$

$$W_o \cdot x_t + U_o \cdot h_{t-1}$$

$$= \begin{bmatrix} W_c & U_c \\ W_i & U_i \\ W_f & U_f \\ W_o & U_o \end{bmatrix} = \text{Matrix } W$$