Vanishing gradients solved by LSTM.

$$\frac{\partial E_t}{\partial W} = \sum_{i=0}^{t} \frac{\partial E_t}{\partial y_t} \times \frac{\partial y_t}{\partial h_t} \times \frac{\partial h_t}{\partial h_i} \times \frac{\partial h_i}{\partial W}$$

$$\frac{\partial h_t}{\partial h_i} = \frac{\partial h_t}{\partial h_{t-1}} \times \frac{\partial h_{t-1}}{\partial h_{t-2}} \cdots \frac{\partial h_{i+1}}{\partial h_i}$$

$$= \prod_{k=i}^{t-1} \frac{\partial h_{k+1}}{\partial h_k}$$

$$\frac{\partial E_t}{\partial W} = \frac{\partial E_t}{\partial h_t} \times \frac{\partial h_t}{\partial c_t} \times \frac{\partial c_t}{\partial c_{t-1}} \times \cdots \times \frac{\partial c_2}{\partial c_1} \times \frac{\partial c_1}{\partial W}$$

$$= \frac{\partial E_t}{\partial h_t} \times \frac{\partial h_t}{\partial c_t} \left(\prod_{i=2}^{t} \frac{\partial c_i}{\partial c_{i-1}} \right) \frac{\partial c_1}{\partial W}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t$$

this is the key difference. the c_t is a function of 4 elements

$$f_t, c_{t-1}, i_t, \tilde{c}_t$$

Whereas in RNN case it is a function of previous states output function.

You can think that even though C_t is a function of 4 elements, but in those elements it containg the previous states output function.

yes it containg but the diff is not all 4 elements are functions of previous outputs.

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

we want partial derivatives w.r.t C_{t-1}

$$\frac{\partial C_t}{\partial C_{t-1}} = \frac{\partial}{\partial C_{t-1}} \left[f_t \times C_{t-1} + i_t \times \tilde{C}_t \right]$$

$$\frac{\partial C_t}{\partial C_{t-1}} = \frac{\partial}{\partial C_{t-1}}\left[f_t \times C_{t-1}\right] + \frac{\partial}{\partial C_{t-1}}\left[i_t \times \tilde{C}_t\right]$$

$$= \frac{\partial f_t}{\partial C_{t-1}} \times (C_{t-1}) + \frac{\partial C_{t-1}}{\partial C_{t-1}} \times (f_t) + \frac{\partial i_t}{\partial C_{t-1}} \times (\tilde{C}_t) + \frac{\partial \tilde{C}_t}{\partial C_{t-1}} \times (i_t)$$

output f_t

Even though the others can be 0 still there one f_t element which will solve the problem of vanishing gradient.