

$$\textcircled{2} = \frac{\partial E_k}{\partial h_k} \cdot \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k [A_t + B_t + G_t + D_t] \right) \cdot \frac{\partial c}{\partial w}$$

BPTT of LSTM 不易发生梯度消失/爆炸的原因:

- (1) 连乘项中, $B_t = f_t \in (0, 1)$, 只要控制好 f_t 的取值使其不至于太小/太大, 梯度消失/爆炸不会出现
- (2) 连乘项是四项加和 (additive property) 不再是单值, 四项加和不至于趋近于 0, 可以通过控制使其接近 1.