Long Short Term Memory Neural Network 的相关推导

夏庆荣

April 12, 2016

1 前向传播

$$
\begin{aligned}
i\_z_j^t &= \sum_k W_{jk}^{ix} x_k^t + \sum_k W_{jk}^{ih} h_k^{t-1} + \sum_k W_{jk}^{ic} c_k^{t-1} \\
i_j^t &= \sigma(i\_z_j^t) \\
f\_z_j^t &= \sum_k W_{jk}^{fx} x_k^t + \sum_k W_{jk}^{fh} h_k^{t-1} + \sum_k W_{jk}^{fc} c_k^{t-1} \\
f_j^t &= \sigma(f\_z_j^t) \\
cell\_z_j^t &= \sum_k W_{jk}^{cx} x_k^t + \sum_k W_{jk}^{ch} h_k^{t-1} \\
c_j^t &= f_j^t c_j^{t-1} + i_j^t \phi(cell\_z_j^t) \\
o\_z_j^t &= \sum_k W_{jk}^{ox} x_k^t + \sum_k W_{jk}^{oh} h_k^{t-1} + \sum_k W_{jk}^{oc} c_k^t \\
o_j^t &= \sigma(o\_z_j^t) \\
h_j^t &= o_j^t \phi(c_j^t)
\end{aligned} \tag{1}
$$

定义：

$$\delta_{z\_j^t} = \frac{\partial C}{\partial z_j}$$

输出层：

$$z_j^L = \sum_k W_{jk}^L h_k^t + b_j^L$$

2 反向传播

$$
\begin{aligned}
\frac{\partial C}{\partial h_l^t} &= \sum_j \frac{\partial C}{\partial z_j^L} \times \frac{\partial z_j^L}{\partial h_l^t} + \sum_j \frac{\partial C}{\partial o\_z_j^{t+1}} \times \frac{\partial o\_z_j^{t+1}}{\partial h_l^t} \\
&+ \sum_j \frac{\partial C}{\partial cell\_z_j^{t+1}} \times \frac{\partial cell\_z_j^{t+1}}{\partial h_l^t} + \sum_j \frac{\partial C}{\partial f\_z_j^{t+1}} \times \frac{f\_z_j^{t+1}}{\partial h_l^t} \\
&+ \sum_j \frac{\partial C}{\partial i\_z_j^{t+1}} \times \frac{\partial i\_z_j^{t+1}}{\partial h_l^t} \\
&= \sum_j \delta_{z_j^{t+1}} \times W_{jl}^L + \sum_j \delta_{o\_z_j^{t+1}} \times W_{jl}^{oh} + \sum_j \delta_{cell\_z_j^{t+1}} \times W_{jl}^{ch} \\
&+ \sum_j \delta_{f\_z_j^{t+1}} \times W_{jl}^{fh} + \sum_j \delta_{i\_z_j^{t+1}} \times W_{jl}^{ih} \\
\frac{\partial C}{\partial h^t} &= \sum_l \frac{\partial C}{\partial h_l^t} \\
&= (W^L)^T \delta_{z^L} + (W^{oh})^T \delta_{o\_z^{t+1}} + (W^{ch})^T \delta_{cell\_z^{t+1}} \\
&+ (W^{fh})^T \delta_{f\_z^{t+1}} + (W^{ih})^T \delta_{i\_z^{t+1}}
\end{aligned} \tag{2}
$$

下面，依次求出 lstm 的各个偏导数：

$$
\begin{aligned}
\frac{\partial C}{\partial o\_z_l^t} &= \sum_j \frac{\partial C}{h_j^t} \frac{\partial h_j^t}{\partial o\_z_l^t} \\
&= \sum_j \delta_{h_j^t} \times \frac{\partial h_l^t}{\partial o_l^t} \times \frac{\partial o_l^t}{\partial o\_z_l^t} \\
&= \sum_j \delta_{h_j^t} \times \phi(c_l^t)\sigma'(o\_z_l^t) \\
\frac{\partial C}{\partial o\_z^t} &= \sum_l \frac{\partial C}{\partial o\_z_l^t} \\
&= \sum_l \sum_j \delta_{h_j^t} \times \phi(c_l^t)\sigma'(o\_z_l^t) \\
&= \delta_{h^t} \odot \phi(c^t)\sigma'(o\_z^t)
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
\frac{\partial C}{\partial c_l^t} &= \sum_j \frac{\partial C}{\partial h_j^t} \times \frac{\partial h_j^t}{\partial c_l^t} + \sum_j \frac{\partial C}{\partial o\_z_j^t} \times \frac{\partial o\_z_j^t}{\partial c_l^t} + \sum_j \frac{\partial C}{\partial c_j^{t+1}} \times \frac{\partial c_j^{t+1}}{\partial c_l^t} \\
&+ \sum_j \frac{\partial C}{\partial f\_z_j^{t+1}} \times \frac{\partial f\_z_j^{t+1}}{\partial c_l^t} + \sum_j \frac{\partial C}{\partial i\_z_j^{t+1}} \times \frac{\partial i\_z_j^{t+1}}{\partial c_l^t} \\
&= \sum_j \delta_{h_j^t} \times \phi'(c_l^t) \times o_l^t + \sum_j \delta_{o\_z_j^t} \times \times W_{jl}^{oc} + \delta_{c_l^{t+1}} \times f_l^{t+1} + \sum_j \delta_{f\_z_j^{t+1}} \\
&+ \sum_j \delta_{f\_z_j^{t+1}} \times W_{jl}^{fc} + \sum_j \delta_{i\_z_j^{t+1}} \times W_{jl}^{ic} \\
\frac{\partial C}{\partial c^t} &= \sum_l \frac{\partial C}{\partial c_l^t} \\
&= \sum_l \sum_j \delta_{h_j^t} \times \phi'(c_l^t) o_l^t + \sum_l \sum_j \delta_{o\_z_j^t} \times W_{jl}^{oc} + \sum_l \delta_{c_l^{t+1}} \times f_l^{t+1} \\
&+ \sum_l \sum_j \delta_{f\_z_j^{t+1}} \times W_{jl}^{fc} + \sum_l \sum_j \delta_{i\_z_j^{t+1}} \times W_{jl}^{ic} \\
&= \delta_{h^t} \odot \phi'(c^t) \odot o^t + (W^{oc})^T \delta_{o\_z^t} + \delta_{c^{t+1}} \odot f^{t+1} + (W^{fc})^T \delta_{f\_z^{t+1}} + (W^{ic})^T \delta_{i\_z^{t+1}}
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
\frac{\partial C}{\partial cell\_z_l^t} &= \sum_j \frac{\partial C}{\partial c_j^t} \times \frac{\partial c_j^t}{\partial cell\_z_l^t} \\
&= \sum_j \delta_{c_j^t} \frac{\partial c_j^t}{\partial cell\_z_j^t} \\
&= \delta_{c_l^t} \times \phi'(cell\_z_l^t) \times i_l^t \\
\frac{\partial C}{\partial cell\_z^t} &= \sum_l \frac{\partial C}{\partial cell\_z_l^t} \\
&= \delta_{c^t} \odot \phi'(cell\_z^t) \odot i^t
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
\frac{\partial C}{\partial f\_z_l^t} &= \sum_j \frac{\partial C}{\partial c_j^t} \times \frac{\partial c_j^t}{\partial f\_z_l^t} \\
&= \frac{\partial C}{\partial c_l^t} \times \frac{\partial c_l^t}{\partial f_l^t} \times \frac{\partial f_l^t}{\partial f\_z_l^t} \\
\frac{\partial C}{\partial f\_z^t} &= \sum_l \frac{\partial C}{\partial f\_z_l^t} \\
&= \delta_{c^t} \odot c^{t-1} \odot \sigma'(f\_z^t)
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
\frac{\partial C}{\partial i\_z_l^t} &= \sum_j \frac{\partial C}{\partial c_j^t} \times \frac{\partial c_j^t}{\partial i\_z_l^t} \\
&= \frac{\partial C}{\partial c_l^t} \times \frac{\partial c_l^t}{\partial i_l^t} \times \frac{\partial i_l^t}{\partial i\_z_l^t} \\
&= \delta_{c_l^t} \phi(cell\_z_l^t) \sigma'(i\_z_l^t) \\
\frac{\partial C}{\partial i\_z^t} &= \sum_l \frac{\partial C}{\partial i\_z_l^t} \\
&= \delta_{c^t} \odot \phi(cell\_z^t) \odot \sigma'(i\_z^t)
\end{aligned} \tag{7}
$$

最后，是对输入 **x** 的求导：

$$
\begin{aligned}
\frac{\partial C}{\partial x_l^t} &= \sum_j \frac{\partial C}{\partial o_j^t} \times \frac{\partial o\_z_j^t}{\partial x_l^t} + \sum_j \frac{\partial C}{\partial cell\_z_j^t} \times \frac{\partial cell\_z_j^t}{\partial x_l^t} \\
&+ \sum_j \frac{\partial C}{\partial f\_z_j^t} \times \frac{\partial f\_z_j^t}{\partial x_l^t} + \sum_j \frac{\partial C}{\partial i\_z_j^t} \times \frac{\partial i\_z_j^t}{\partial x_l^t} \\
\frac{\partial C}{\partial x^t} &= \sum_l \frac{\partial C}{\partial x_l^t} \\
&= (W^{ix})^T \delta_{o\_z}^t + (W^{cx})^T \delta_{cell\_z}^t + (W^{fx})^T \delta_{f\_z}^t + (W^{ix})^T \delta_{i\_z}^t
\end{aligned} \tag{8}
$$

至此，所有的公式已经推导完毕！