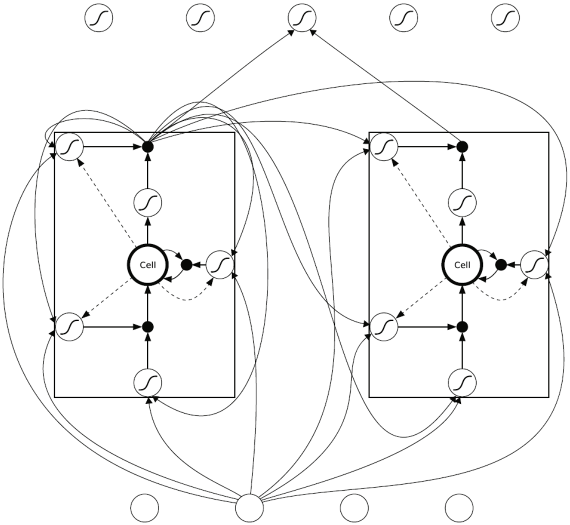
Deep Learning Technology and Application

Ge Li

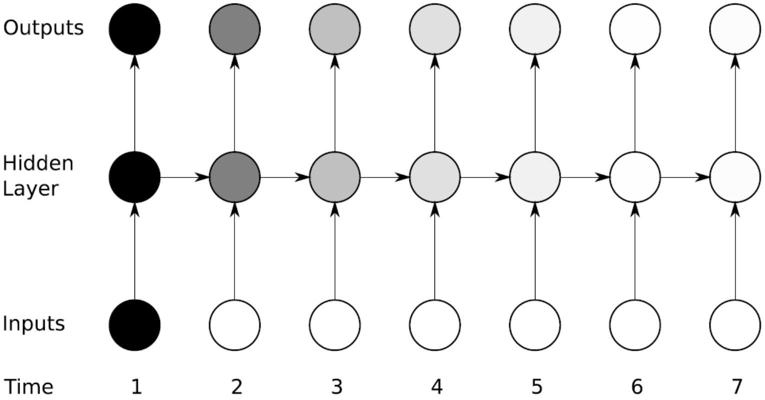
Peking University

Table of contents

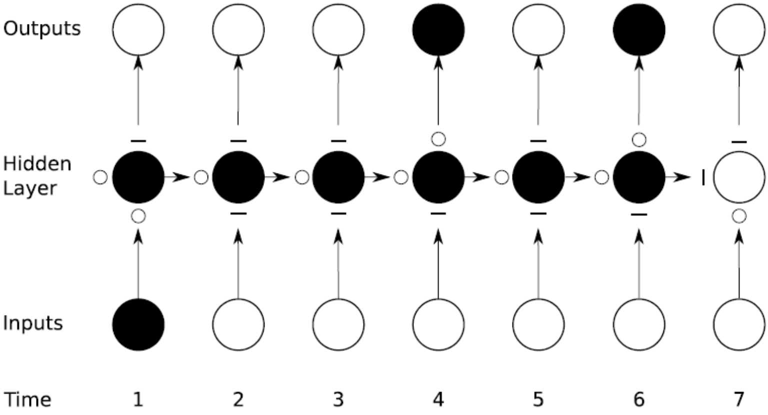
Recurrent Neural Network ——LSTM



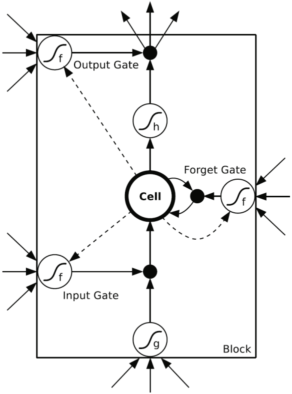
LSTM 的提出



LSTM 的提出



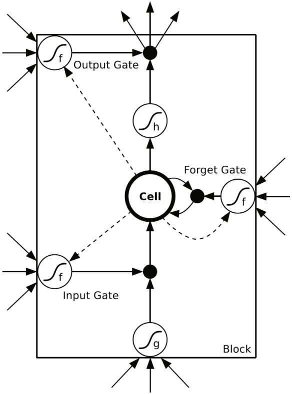
LSTM 单元



- 设 LSTM 隐藏层共包含 H 个神经元，下标 h 表示其中之一；
- 设 LSTM 隐藏层共包含 C 个 Cell，下标 c 表示某个 Cell；
- 当前的 LSTM 单元中 Input Gate,Forget Gate, Output Gate 分别用下标 α, β, γ 标识；
- LSTM 单元在 t 时刻的输入：z_h^t， t 时刻的输出：a_h^t；

  对于仅包含一个 Cell 的 LSTM 单元，$z_h^t = z_c^t, a_c^t = a_h^t$；

LSTM 单元



LSTM 单元的输入:

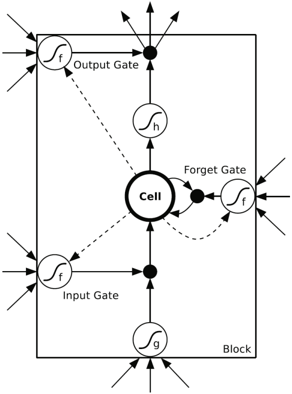
$$z_h^t = \sum_{i=1}^{I} w_{ci} x_i^t + \sum_{h=1}^{H} w_{ch} a_h^{t-1}$$

注意：a_h 表示来自于其他 LSTM 单元的输出 a_c；

LSTM 单元的输出:

$$a_h^t = a_c^t$$

LSTM 单元



Input Gate:

$$z_{\alpha}^{t} = \sum_{i=1}^{I} w_{\alpha i} x_{i}^{t} + \sum_{h=1}^{H} w_{\alpha h} a_{h}^{t-1} + \sum_{c=1}^{C} w_{\alpha c} s_{c}^{t-1}$$

$$a_{\alpha}^{t} = f(z_{\alpha}^{t})$$

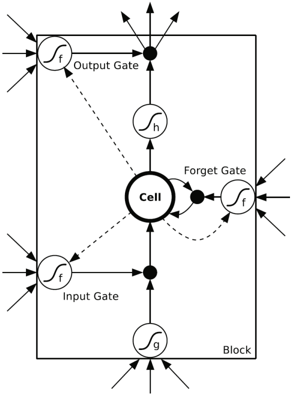
注意：a_h **表示来自于其他 LSTM 单元的输出** a_c；

Forget Gate:

$$z_{\beta}^{t} = \sum_{i=1}^{I} w_{\beta i} x_{i}^{t} + \sum_{h=1}^{H} w_{\beta h} a_{h}^{t-1} + \sum_{c=1}^{C} w_{\beta c} s_{c}^{t-1}$$

$$a_{\beta}^{t} = f(z_{\beta}^{t})$$

LSTM 单元



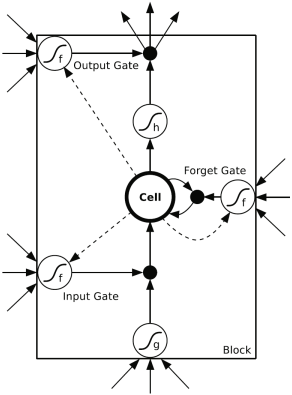
Cells:

$$z_c^t = \sum_{i=1}^{I} w_{ci} x_i^t + \sum_{h=1}^{H} w_{ch} a_h^{t-1}$$

注意：a_h 表示来自于其他 LSTM 单元的输出 a_c；

$$s_c^t = a_\alpha^t g(z_c^t) + a_\beta^t s_c^{t-1}$$

LSTM 单元



Output Gate:

$$z_\gamma^t = \sum_{i=1}^{I} w_{\gamma i} x_i^t + \sum_{h=1}^{H} w_{\gamma h} a_h^{t-1} + \sum_{c=1}^{C} w_{\gamma c} s_c^{t-1}$$

$$a_\gamma^t = f(z_\gamma^t)$$

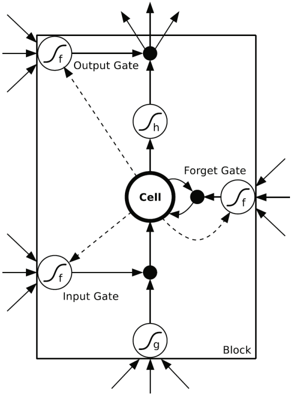
Cell Outputs:

$$a_c^t = a_\gamma^t h(s_c^t)$$

RNN Outputs:

$$z_k^t = \sum_{c=1}^{C} w_{kc} a_h^t$$

LSTM 单元



统计要计算的参数：

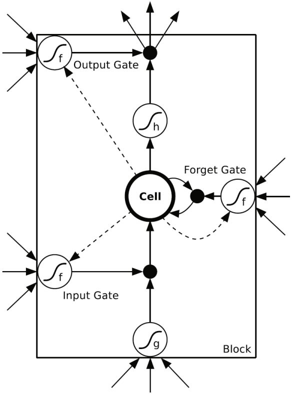
$$w_{\alpha i}, w_{\alpha h}, w_{\alpha c}$$

$$w_{\beta i}, w_{\beta h}, w_{\beta c}$$

$$w_{\gamma i}, w_{\gamma h}, w_{\gamma c}$$

$$w_{ci}, w_{ch}, w_{kc}$$

LSTM 单元



$$w_{\alpha i}: \frac{\partial J(W,b)}{\partial w_{\alpha i}} = \frac{\partial J(W,b)}{\partial z_\alpha^t}\frac{\partial z_\alpha^t}{\partial w_{\alpha i}} = \frac{\partial J(W,b)}{\partial z_\alpha^t} x_i^t$$

$$w_{\alpha h}: \frac{\partial J(W,b)}{\partial w_{\alpha h}} = \frac{\partial J(W,b)}{\partial z_\alpha^t}\frac{\partial z_\alpha^t}{\partial w_{\alpha h}} = \frac{\partial J(W,b)}{\partial z_\alpha^t} a_h^{t-1}$$

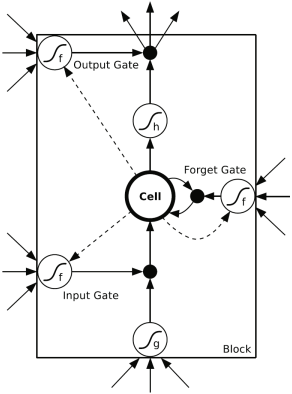
$$w_{\alpha c}: \frac{\partial J(W,b)}{\partial w_{\alpha c}} = \frac{\partial J(W,b)}{\partial z_\alpha^t}\frac{\partial z_\alpha^t}{\partial w_{\alpha c}} = \frac{\partial J(W,b)}{\partial z_\alpha^t} s_c^{t-1}$$

$$w_{\beta i}: \frac{\partial J(W,b)}{\partial w_{\beta i}} = \frac{\partial J(W,b)}{\partial z_\beta^t}\frac{\partial z_\beta^t}{\partial w_{\beta i}} = \frac{\partial J(W,b)}{\partial z_\beta^t} x_i^t$$

$$w_{\beta h}: \frac{\partial J(W,b)}{\partial w_{\beta h}} = \frac{\partial J(W,b)}{\partial z_\beta^t}\frac{\partial z_\beta^t}{\partial w_{\beta h}} = \frac{\partial J(W,b)}{\partial z_\beta^t} a_h^{t-1}$$

$$w_{\beta c}: \frac{\partial J(W,b)}{\partial w_{\beta c}} = \frac{\partial J(W,b)}{\partial z_\beta^t}\frac{\partial z_\beta^t}{\partial w_{\beta c}} = \frac{\partial J(W,b)}{\partial z_\beta^t} s_c^{t-1}$$

LSTM 单元



$$w_{\gamma i}\colon \frac{\partial J(W,b)}{\partial w_{\gamma i}} = \frac{\partial J(W,b)}{\partial z_\gamma^t}\frac{\partial z_\gamma^t}{\partial w_{\gamma i}} = \frac{\partial J(W,b)}{\partial z_\gamma^t} x_i^t$$

$$w_{\gamma h}\colon \frac{\partial J(W,b)}{\partial w_{\gamma h}} = \frac{\partial J(W,b)}{\partial z_\gamma^t}\frac{\partial z_\gamma^t}{\partial w_{\gamma h}} = \frac{\partial J(W,b)}{\partial z_\gamma^t} a_h^{t-1}$$

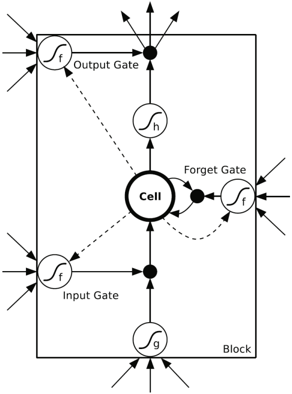
$$w_{\gamma c}\colon \frac{\partial J(W,b)}{\partial w_{\gamma c}} = \frac{\partial J(W,b)}{\partial z_\gamma^t}\frac{\partial z_\gamma^t}{\partial w_{\gamma c}} = \frac{\partial J(W,b)}{\partial z_\gamma^t} s_c^{t-1}$$

$$w_{ci}\colon \frac{\partial J(W,b)}{\partial w_{ci}} = \frac{\partial J(W,b)}{\partial z_h^t}\frac{\partial z_h^t}{\partial w_{ci}} = \frac{\partial J(W,b)}{\partial z_c^t} x_i^t$$

$$w_{ch}\colon \frac{\partial J(W,b)}{\partial w_{ch}} = \frac{\partial J(W,b)}{\partial z_h^t}\frac{\partial z_h^t}{\partial w_{ch}} = \frac{\partial J(W,b)}{\partial z_h^t} a_h^{t-1}$$

$$w_{kc}\colon \frac{\partial J(W,b)}{\partial w_{kc}} = \frac{\partial J(W,b)}{\partial z_k^t}\frac{\partial z_k^t}{\partial w_{kc}} = \frac{\partial J(W,b)}{\partial z_k^t} a_h^t$$

LSTM 单元



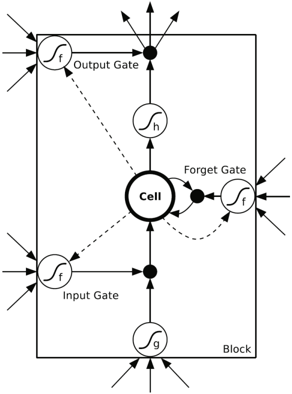
统计要计算的梯度：

$$\frac{\partial J(W,b)}{\partial z_\alpha^t} \quad \frac{\partial J(W,b)}{\partial z_\beta^t} \quad \frac{\partial J(W,b)}{\partial z_\gamma^t}$$

$$\frac{\partial J(W,b)}{\partial z_c^t} \quad \frac{\partial J(W,b)}{\partial z_h^t} \quad \frac{\partial J(W,b)}{\partial z_k^t}$$

LSTM 单元

梯度计算之一：



若 t=T，则：

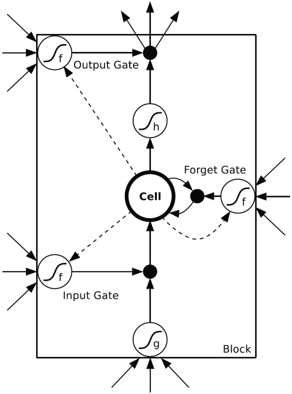
$$\frac{\partial J(W,b)}{\partial z_k^T} = \frac{\partial J(W,b)}{\partial a_k^T}\frac{\partial a_k^T}{\partial z_k^T} = \frac{\partial J(W,b)}{\partial a_k^T} output'(\cdot)$$

否则：

$$\frac{\partial J(W,b)}{\partial z_k^t} = \frac{\partial J(W,b)}{\partial a_k^t}\frac{\partial a_k^t}{\partial z_k^t} = \sum_h^H \frac{\partial J(W,b)}{\partial z_h^{t+1}}\frac{\partial a_k^t}{\partial z_k^t} = \sum_h^H \frac{\partial J(W,b)}{\partial z_h^{t+1}} output'(\cdot)$$

LSTM 单元

梯度计算之二：



$$\frac{\partial J(W,b)}{\partial z_\gamma^t} = \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial a_c^t} \frac{\partial a_c^t}{\partial a_\gamma^t} \frac{\partial a_\gamma^t}{\partial z_\gamma^t}$$

因为：$a_c^t = a_\gamma^t h(s_c^t)$

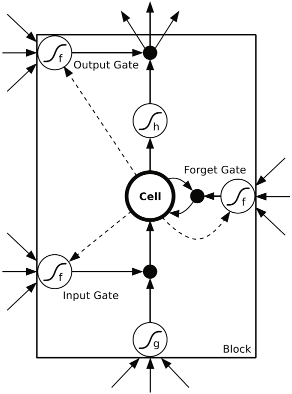
所以：

$$= \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial a_c^t} h(s_c^t) f'(z_\gamma^t)$$

$$= f'(z_\gamma^t) \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial a_c^t} h(s_c^t)$$

LSTM 单元

梯度计算之三：



$$\frac{\partial J(W,b)}{\partial a_c^t} = \sum_k^K \frac{\partial J(W,b)}{\partial z_k^t}\frac{\partial z_k^t}{\partial a_c^t} + \sum_h^H \frac{\partial J(W,b)}{\partial z_h^{t+1}}\frac{\partial z_h^{t+1}}{\partial a_c^t}$$

因为：

$$z_h^{t+1} = \sum_{i=1}^I w_{ci} x_i^{t+1} + \sum_{h=1}^H w_{ch} a_h^t$$

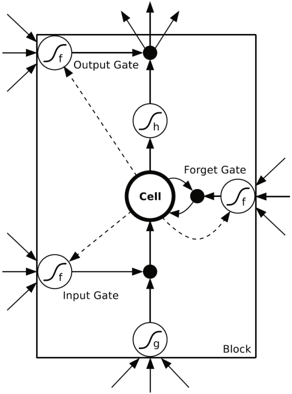
$$z_k^t = \sum_{k=1}^K w_{kc} a_h^t$$

所以：

$$\frac{\partial J(W,b)}{\partial a_c^t} = \sum_k^K \frac{\partial J(W,b)}{\partial z_k^t} w_{kc} + \sum_h^H \frac{\partial J(W,b)}{\partial z_h^{t+1}} w_{ch}$$

LSTM 单元

梯度计算之四：



$$\frac{\partial J(W,b)}{\partial z_h^t} = \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} \frac{\partial s_c^t}{\partial z_h^t}$$

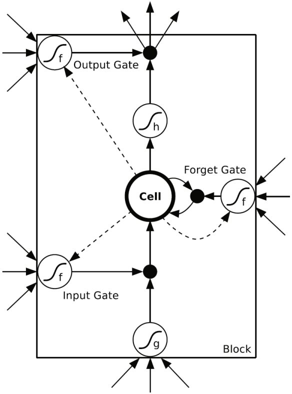
因为：$s_c^t = a_\alpha^t g(z_c^t) + a_\beta^t s_c^{t-1}$

所以：

$$\frac{\partial J(W,b)}{\partial z_h^t} = \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} a_\alpha^t g'(z_c^t)$$

LSTM 单元

梯度计算之五：



$$\frac{\partial J(W,b)}{\partial z_\beta^t} = \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} \frac{\partial s_c^t}{\partial z_\beta^t}$$

因为：$s_c^t = a_\alpha^t g(z_c^t) + a_\beta^t s_c^{t-1}$

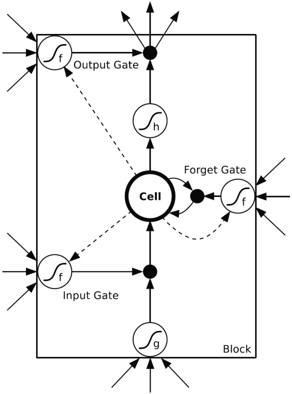
所以：

$$\frac{\partial J(W,b)}{\partial z_h^t} = \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} \frac{\partial a_\beta^t}{\partial z_\beta^t} s_c^{t-1}$$

$$= f'(z_\beta^t) \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} s_c^{t-1}$$

LSTM 单元

梯度计算之六：



$$\frac{\partial J(W,b)}{\partial z_\alpha^t} = \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} \frac{\partial s_c^t}{\partial z_\alpha^t}$$

因为：$s_c^t = a_\alpha^t g(z_c^t) + a_\beta^t s_c^{t-1}$

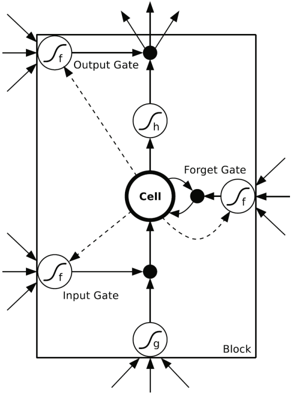
所以：

$$\frac{\partial J(W,b)}{\partial z_h^t} = \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} \frac{\partial a_\alpha^t}{\partial z_\alpha^t} g(z_c^t)$$

$$= f'(z_\alpha^t) \sum_{c=1}^{C} \frac{\partial J(W,b)}{\partial s_c^t} g(z_c^t)$$

LSTM 单元

梯度计算之七：



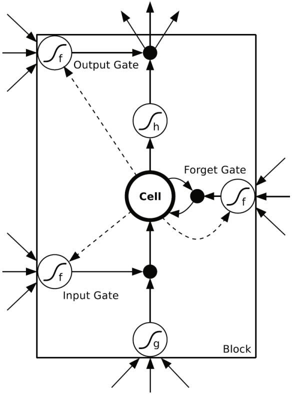
焦点集中在：$\dfrac{\partial J(W,b)}{\partial s_c^t}$

因为：
$$\frac{\partial J(W,b)}{\partial s_c^t} = \frac{\partial J(W,b)}{\partial a_c^t}\frac{\partial a_c^t}{\partial s_c^t} + \frac{\partial J(W,b)}{\partial s_c^{t+1}}\frac{\partial s_c^{t+1}}{\partial s_c^t} + \frac{\partial J(W,b)}{\partial z_\alpha^{t+1}}\frac{\partial z_\alpha^{t+1}}{\partial s_c^t} + \frac{\partial J(W,b)}{\partial z_\beta^{t+1}}\frac{\partial z_\beta^{t+1}}{\partial s_c^t} + \frac{\partial J(W,b)}{\partial z_\gamma^{t+1}}\frac{\partial z_\gamma^{t+1}}{\partial s_c^t}$$

所以：
$$\frac{\partial J(W,b)}{\partial s_c^t} = \frac{\partial J(W,b)}{\partial a_c^t} a_\gamma^t h'(s_c^t) + \frac{\partial J(W,b)}{\partial s_c^{t+1}} a_\beta^{t+1} + \frac{\partial J(W,b)}{\partial z_\alpha^{t+1}} w_{\alpha c} + \frac{\partial J(W,b)}{\partial z_\alpha^{t+1}} w_{\alpha c} + \frac{\partial J(W,b)}{\partial z_\alpha^{t+1}} w_{\alpha c}$$

LSTM 单元

最常见的表述：



$$
\begin{aligned}
i_t &= \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right)\\
f_t &= \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right)\\
c_t &= f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right)\\
o_t &= \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o\right)\\
h_t &= o_t \tanh(c_t)
\end{aligned}
$$

Thanks.