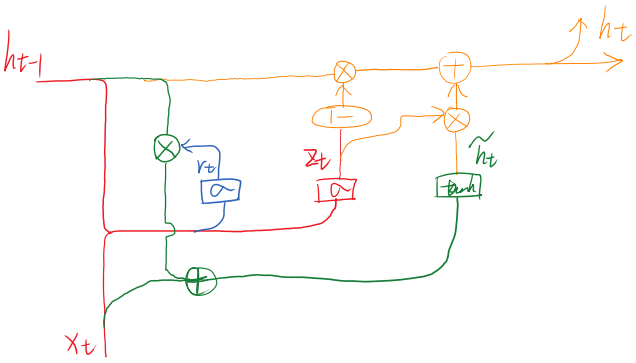
GRU、LSTM

2018年7月31日 13:39

- GRU, Gated Recurrent Neural Networks
  Gated Recurrent Unit



$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r) \quad \text{Reset Gate}$$

$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z) \quad \text{Update Gate}$$

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

$r = h$

$$\tilde{H}_t = \tanh(X_t W_{xh} + R_t \odot H_{t-1} W_{hh} + b_h)$$ 候选 Hidden State

t 时刻输入

$t-1$ 时刻状态

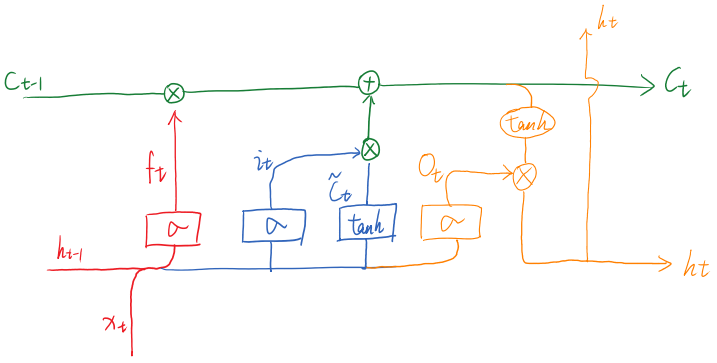
可以控制 $t-1$ 时刻的哪些状态被保留

$z = h$

$$H_t = Z_t \odot \tilde{H}_t + (1 - Z_t) \odot H_{t-1}$$

$\begin{cases} Z_t \to 1, & H_t \approx H_{t-1} \\ Z_t \to 0, & H_t \approx \tilde{H}_t \end{cases}$ ⇒ 解决长期依赖问题

- LSTM, Long Short Term Memory



$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i), \text{ Input gate}$$

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f), \text{ Forget gate}$$

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o), \text{ Output gate}$$

$$\tilde{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c), \text{ 候选状态, 类似于 naive RNN}$$

↑ Cell

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t$$

$F_t \to 0$ "forget" $I_t \to 1$ "input"

$F_t \to 1$ "long term memory"

$$H_t = O_t \odot \tanh(C_t)$$

Control how cell state (C_t) can effect hidden state (H_t)

- $loss = \frac{-1}{N} \sum_{i=1}^{N} \log P_{target_i}$

times of prediction; ground truth of next target word

$perplexity = e^{loss}$

困惑度