

Max Likelihood 的一些相关推导

夏庆荣

October 21, 2016

1 符号定义

d : 表示一个动作 ($d \in \{\text{shift, pop root, left arc, right arc}\}$)。

$\rho(s, d; \theta)$: 表示在某个状态 s , 做了动作 d 的得分, θ 表示模型参数。

$\rho(d_{1:j-1}, d_j; \theta)$: 等价于 $\rho(s, d; \theta)$, s 表示经过动作序列 $d_{1:j-1}$ 得到的状态。

\mathcal{D}_n : 表示所有长度为 n 的有效动作序列。

2 Max Likelihood 概率定义

$$P_G(d_{1:n}) = \frac{e^{\text{score}(d_{1:n}; \theta)}}{Z_G(\theta)} \quad (1)$$

其中

$$\text{score}(d_{1:n}; \theta) = \sum_{j=1}^n \rho(d_{1:j-1}, d_j; \theta) \quad (2)$$

$$Z_G(\theta) = \sum_{d'_{1:n} \in \mathcal{D}_n} e^{\text{score}(d'_{1:n}; \theta)} \quad (3)$$

进而获得概率最高的动作序列:

$$\arg \max_{d_{1:n} \in \mathcal{D}_n} P_G(d_{1:n}) = \arg \max_{d_{1:n} \in \mathcal{D}_n} \sum_{j=1}^n \rho(d_{1:j-1}, d_j; \theta) \quad (4)$$

3 Max Log Likelihood

$d^*_{1:n}$: 表示 gold 动作序列

$$\begin{aligned} L(d^*_{1:n}; \theta) &= -\ln P_G(d^*_{1:n}; \theta) \\ &= -\text{score}(d^*_{1:n}) + \ln Z_G(\theta) \end{aligned} \quad (5)$$

4 Max Likelihood 求导

$$\begin{aligned}
 \frac{\partial L(d_{1:n}^*; \theta)}{\partial \text{score}(d_{1:n}; \theta)} &= \frac{\partial -\text{score}(d_{1:n}^*; \theta) + \ln Z_G(\theta)}{\partial \text{score}(d_{1:n})} \\
 &= \frac{\partial -\text{score}(d_{1:n}^*; \theta)}{\text{score}(d_{1:n}; \theta)} + \frac{\partial \ln Z_G(\theta)}{\partial \text{score}(d_{1:n}; \theta)} \\
 &= \frac{\partial -\text{score}(d_{1:n}^*; \theta)}{\text{score}(d_{1:n}; \theta)} + \frac{\mathbf{e}^{\text{score}(d_{1:n}; \theta)}}{Z_G(\theta)} \\
 &= \frac{\partial -\text{score}(d_{1:n}^*; \theta)}{\text{score}(d_{1:n}; \theta)} + P_G(d_{1:n}) \\
 &= \begin{cases} -1 + P_G(d_{1:n}) & d_{1:n} = d_{1:n}^* \\ P_G(d_{1:n}) & d_{1:n} \neq d_{1:n}^* \end{cases}
 \end{aligned} \tag{6}$$

另外:

$$\frac{\partial \text{score}(d_{1:n})}{\partial \rho(d_{1:j-1}, d_j; \theta)} = 1 \tag{7}$$