

实现细节

- 结构采用Softmax输出作为目标词汇的概率表示

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}}$$

- Encoder和Decoder是两个不同的LSTM
- 深层LSTM性能要好于浅层LSTM，文中采用了四层LSTM结构
- 目标函数

$$p(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod_{t=1}^{T'} p(y_t | v, y_1, \dots, y_{t-1})$$