



deeplearning.ai

Sequence to sequence models

Refinements to beam search

Length normalization

$$p(y^{(1)} \dots y^{(T_y)} | x) = \frac{P(y^{(1)} | x) P(y^{(2)} | x, y^{(1)}) \dots}{P(y^{(T_y)} | x, y^{(1)}, \dots, y^{(T_y-1)})}$$

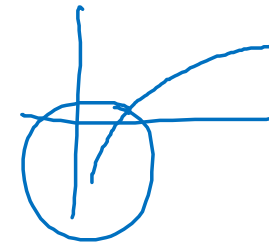
$$\arg \max_y \prod_{t=1}^{T_y} P(y^{(t)} | x, y^{(1)}, \dots, y^{(t-1)})$$

log

$$\arg \max_y \sum_{t=1}^{T_y} \log P(y^{(t)} | x, y^{(1)}, \dots, y^{(t-1)}) \leftarrow$$

$T_y = 1, 2, 3, \dots, 30.$

$$\rightarrow \frac{1}{T_y^\alpha} \sum_{t=1}^{T_y} \log P(y^{(t)} | x, y^{(1)}, \dots, y^{(t-1)})$$



$$\log P(y|x) \leftarrow$$

$$P(y|x) \leftarrow$$

$$\underline{\alpha = 0.7}$$

$$\underline{\alpha = 1}$$

$$\underline{\alpha = 0}$$

Beam search discussion

Beam width B?

$1 \rightarrow 3 \rightarrow 10, 100, 1000 \rightarrow 3000$

large B: better result, slower
small B: worse result, faster

Unlike exact search algorithms like BFS (Breadth First Search) or DFS (Depth First Search), Beam Search runs faster but is not guaranteed to find exact maximum for $\arg \max_y P(y|x)$.