

N -gram: 基于概率论的自然语言生成

$$\text{next word} = \max_{m \in V} P(w | \text{previous words})$$

由前面的单词进行预测, 取条件概率最大的一个词

例如:

青椒 炒 肉片
青椒 炒 牛肉
青椒 炒 肉片
茄子 炒 肉沫
茄子 炒 豆角

$$P(\text{肉片} | \text{青椒炒}) = \frac{2}{3}$$

$$P(\text{牛肉} | \text{青椒炒}) = \frac{1}{3}$$

$$P(\text{豆角} | \text{青椒炒}) = 0$$

当 $n=2$ 时 previous words = 青椒炒

则得到预测肉片

$$n=1 \text{ 时 } P(\text{炒} | \text{青椒}) = 1$$

$$P(\text{豆角} | \text{炒}) = \frac{1}{5}$$

步骤:

① 看前 $n-1$ 个词

② 得到下个词的分布(离散)

③ 根据分布采样

采样算法: ① 贪婪(取最大概率)

② 取 top-k 然后重要性采样

截断分布

③ Beam search 树搜索

✓

有一个束, 例如 $k=2$, 翻译了 3 个词

$k=2$ 中词典: 我爱 你

英词典: I L U

当到“我”时 概率为 $\begin{cases} I = 0.8 \\ L = 0.15 \\ U = 0.05 \end{cases}$

选择最大 2 个 I, L

... $I I = 0.2$ {40%}

当到“受”时. 概率为 $\left\{ \begin{array}{l} IL: 0.7 \\ IU: 0.1 \end{array} \right\}$ 和为1

选择最大2个 IL, LI $\left\{ \begin{array}{l} LI: 0.7 \\ LL: 0.2 \\ LU: 0.1 \end{array} \right\}$ 和为1

当到“你”时. 概率为 $\left\{ \begin{array}{l} ILL: 0.1 \\ ILL: 0.3 \\ ILL: 0.6 \end{array} \right\}$ 和为1

选择最大一个 ILL $\left\{ \begin{array}{l} LLL: 0.6 \\ LLL: 0.3 \\ LLL: 0.1 \end{array} \right\}$ 和为1

N-gram 特点:

- ① 训练速度快
- ② 具备一定的多样性
- ③ 可以学得基本语法、标点
- ④ 效果受到语料窗口大小的明显限制
- ⑤ 具有马尔可夫属性

$$P(S^{<t+1>} | S^{<t>}) = P(S^{<t+1>} |$$

$$S^{<0>}, S^{<1>}, S^{<2>}, \dots, S^{<t>})$$