

$S = (s_1, \dots, s_n)$ - исходное предложение

$T = (t_1, \dots, t_m)$ - его перевод

Параметры перевода: выраж. $A = (a_1, \dots, a_m)$,
где $a_i \in \{1, \dots, n\}$

Параметры модели: $\theta(y|x) = p(y|x)$ -
вероятность того, что перев. слова x является
словом y (нормировка по словам целевого
языка)

$$p(A, T | S) = \prod_{i=1}^m p(a_i) p(t_i | a_i, S) = \\ = \prod_{i=1}^m \frac{1}{n} \theta(t_i | s_{a_i})$$

Трудоемкость вычисли шах ЕМ-алгоритма

и получение выражения для подсчета

инши. оценки правдоподобия (L)

В кортеже R предложений. Для i -ой

пары n_i - длина предложения, m_i - длина
перевода.

$$p(a_i = j | S) = \frac{1}{n} \quad \forall j = 1, \dots, n \}$$

$$p(a_i, t_i | S) = p(a_i | S) \cdot p(t_i | a_i, S) = \\ = \frac{1}{n} \Theta(t_i | S_{a_i}) - \text{было замечано}$$

что было времени A .

Вывод E -максимизации:

$$q^*(A) = p(A | T, S, \theta) = \frac{p(A, T | S, \theta)}{p(T | S, \theta)}$$

$$p(T | S, \theta) = \sum_{\substack{\text{но было} \\ \text{быстро}}} p(A, T | S, \theta) =$$

$$= \sum_{a_1=1}^n \sum_{a_2=1}^n \dots \sum_{a_m=1}^m \prod_{i=1}^m p(a_i) \cdot p(t_i | a_i, S) = \\ = \prod_{i=1}^m \frac{1}{n} \cdot \sum_{a_i=1}^n p(t_i | a_i, S) = \prod_{i=1}^m \frac{1}{n} \sum_{j=1}^n$$

$$p(t_i | j, S) = \prod_{i=1}^m \frac{1}{n} \sum_{j=1}^n \Theta(t_i | S_j)$$

$$\Rightarrow q^*(A) = \frac{\prod_{i=1}^m \frac{1}{n} \cdot \Theta(t_i | S_{a_i})}{\prod_{i=1}^m \frac{1}{n} \cdot \sum_{j=1}^n \Theta(t_i | S_j)} =$$

$$= \prod_{i=1}^m \frac{\Theta(t_i | S_{a_i})}{\sum_{j=1}^n \Theta(t_i | S_j)}$$

Мы рассматриваем случай, где наборы.

корпус состоят из одной пары. Обобщим результаты на R пар:

$$\{S^{(r)}, T^{(r)}\}_{r=1}^R, \text{ где } S^{(r)} = (S_1^{(r)}, \dots, S_{n_r}^{(r)})$$
$$T^{(r)} = (t_1^{(r)}, \dots, t_{m_r}^{(r)}).$$

$A^{(r)}$ -выраб. для r-ой пары.

$ct = \{A^{(r)}\}_{r=1}^R$ - выраб. для всех корпусов.

$$q^*(ct) = \prod_{r=1}^R \prod_{i=1}^{m_r} \frac{\theta(t_i^{(r)} | S_{a_i^{(r)}}^{(r)})}{\sum_{j=1}^{n_r} \theta(t_i^{(r)} | S_j^{(r)})}$$

Выбор N-шага:

Пусть $T = \{\bar{T}^{(r)}\}_{r=1}^R, S' = \{S^{(r)}\}_{r=1}^R$

$$E_{ct \sim q^*(ct)} \log p(ct, T | S', \theta) \rightarrow \max_{\theta}$$

$$p(ct, T | S', \theta) = \prod_{r=1}^R p(A^{(r)}, \bar{T}^{(r)} | S', \theta)$$

$$= \prod_{r=1}^R \prod_{i=1}^{m_r} \frac{1}{n_r} \theta(t_i^{(r)} | S_{a_i^{(r)}}^{(r)}).$$

$$IE_{A \sim q^*(A)} \log p(A, T | S, \theta) = \sum_{r=1}^R \sum_{i=1}^{m_r} \cdot$$

$$\frac{\Theta(t_i^{(r)} | S_{a_i^{(r)}}^{(r)})}{\sum_{j=1}^{n_r} \Theta(t_i^{(r)} | S_j^{(r)})} \cdot \log \frac{1}{n_r} \Theta(t_i^{(r)} | S_{a_i^{(r)}}^{(r)})$$

$\rightarrow \text{MAX}$

$$Q(\theta) = \sum_{r=1}^R \sum_{i=1}^{m_r} \cdot \frac{\Theta(t_i^{(r)} | S_{a_i^{(r)}}^{(r)})}{\sum_{j=1}^{n_r} \Theta(t_i^{(r)} | S_j^{(r)})} \cdot (\log \Theta(t_i^{(r)} | S_{a_i^{(r)}}^{(r)}) - \log n_r)$$

$$-\log n_r) \rightarrow \max_{\theta} \sum_t \Theta(t|s) = 1 \forall s$$

Зависимость от правильного выделения
награды:

$$L(\theta, \lambda) = \sum_{r=1}^R \sum_{i=1}^{m_r} q_i^*(A^{(r)}) \cdot$$

$$(\log \Theta(t_i^{(r)} | S_{a_i^{(r)}}^{(r)}) - \log n_r) + \sum_s \lambda(s) \cdot \left(\sum_t \Theta(t|s) - 1 \right)$$

Перенесение суммы через $\sum_s \sum_t$:

$$L(\theta, \lambda) = \sum_s \sum_t \sum_{r=1}^R \sum_{i=1}^{m_r} q_i^*(t^{(r)}) \cdot$$

- $\delta(t, t_i^{(r)}) \cdot \delta(s, S_{a_i^{(r)}}^{(r)}) \cdot (\log \theta(t|s) - 1)$
- $\log n_r + \sum_s \lambda(s) \left(\sum_t \theta(t|s) - 1 \right)$

Заменим $g(s, t) = \sum_{r=1}^R \sum_{i=1}^{m_r} q_i^*(t^{(r)}) \cdot$

$\delta(t, t_i^{(r)}) \cdot \delta(s, S_{a_i^{(r)}}^{(r)})$

$$L(\theta, \lambda) = \sum_s \sum_t g(s, t) \cdot \log \theta(t|s)$$

$$+ \sum_s \lambda(s) \left(\sum_t \theta(t|s) - 1 \right)$$

Удалим $\log n_r$ - не влияет
на оптимизацию

$$\frac{\partial L(\theta, \lambda)}{\partial \theta(t|s)} = \frac{g(s, t)}{\theta(t|s)} + \lambda(s)$$

$$\Rightarrow \theta(t|s) = -\frac{g(s, t)}{\lambda(s)}$$

Воспользуемся тем, что:

$$\sum_t \theta(t|s) = 1 \Rightarrow \sum_t \left(-\frac{g(s, t)}{\lambda(s)} \right) = 1$$

$$\Rightarrow \lambda(s) = -\sum_t g(s, t)$$

$$\Rightarrow \theta(t|s) = \frac{g(s,t)}{\sum_{\tau} g(s,\tau)}$$

\Rightarrow кинетика Оценка правдоподобия - это
и есть $Q(\theta)$

$$h = \sum_s \sum_t g(s,t) \cdot \log \frac{g(s,t)}{\sum_{\tau} g(s,\tau)} + \text{const}$$

$$\text{const} = - \sum_{r=1}^R \sum_{i=1}^{m_r} q_i^r (A^{(r)}) \log n_r$$