EM - алгоритм для задачи перевода

Обозначения: за R число предложений в корпусе;

n_i - длина i-го предложения

m_i - длина i-го перевода

$S_i = (S_{1i}, \dots, S_{n_i i})$ - исходное i-е предложение

$T_i = (T_{1i}, \dots, T_{m_i i})$ - его перевод

$A_i = (a_{1i}, \dots, a_{m_i i})$ - выравнивание каждого слова в i-м предложении (скрытые переменные)

E шаг:

$$p(a_i = z \mid t_i, S_i) = \frac{P(a_i = z, t_i \mid S_i)}{P(t_i \mid S_i)} =$$

$$= \frac{P(a_i = z, t_i \mid S_i)}{\sum_{m=1}^{n_i} P(a_i = m, t_i \mid S_i)} = \frac{\theta(t_i \mid S_{iz})}{\sum_{m=1}^{n_i} \theta(t_i \mid S_{im})} = q_i(z)$$

M шаг:

$$\mathcal{L}(q; z) = \sum_{z=1}^{n} \sum_{i=1}^{m} q(z) \cdot \log \frac{P(a_i = z, t_i \mid S)}{q_i(z)} =$$

$$= \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) \cdot \log \frac{\theta(t_i \mid S_k)}{n\, q_i(z)} =$$

$$= \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) \cdot \log \theta(t_i \mid S_k) - \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) \cdot \log n\, q_i(z) \to \max_{\theta}$$

$\Rightarrow$ задача сводится к $\sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) \log \theta(t_i \mid S_k) \to \max_{\theta}$

Лагранжиан:

$$L = \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) \cdot \log \theta(t_i \mid S_k) - \sum_{x \in X} \lambda_x \left(\sum_{y \in Y} \theta(y \mid x) - 1 \right)$$

$$\frac{\partial L}{\partial z} = \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) \frac{\partial \log \theta(t_i \mid S_k)}{\partial \theta(y \mid x)} - \frac{\partial \sum_{x \in X} \left(\sum_{y \in Y} \theta(y \mid x) - 1 \right)}{\partial \theta(y \mid x)} =$$

$$= \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) \cdot \frac{[t_i = y]\,[S_k = x]}{\theta(y \mid x)} - \lambda_x = 0$$

Нужно найти λ_x. Воспользуемся св-ми вер-ти:

$$\sum_{y \in Y} \Theta(y|x) = 1 = \sum_{y \in Y} \frac{1}{\lambda_x} \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) [t_i = y] [S_k = x] =$$

$$= \frac{1}{\lambda_x} \sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) [S_k = x] \underbrace{\sum_{y \in Y} [t_i = y]}_{=1}$$

$$\Rightarrow \Theta^*(y|x) = \frac{\sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) [t_i = y] [S_k = x]}{\sum_{z=1}^{n} \sum_{i=1}^{m} q_i(z) [S_k = x]}$$