1) Кол-во предложений $- R$

2) Длина i-го предложения : n_i

3) Длина i-го перевода : m_i

4) S_i — i-е предложение

5) S_{ij} — j-е слово в i-м предложении

6) t_{ij} — j-е слово в i-м предложении-переводе

7) $q(A)$ - распределение на скрытых переменных

8) a_{ij} — скрытая переменная (выравнивание) j-го слова в i-м переводе.

9) a_i — выравнивание

$\theta(y|x) = p(y|x)$ - матрица условных вер-тей перевода

вер-ть того, что переводом слова x с исх. языка на целевой явл. слово y (нормировка по целевому языку)

Правдоподобие латентных переменных и предложения на целевом языке в этой модели записыв. так:

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

E-шаг: Получим $q(A) = p(A|T,S)$

номер слова

$$q(a_{ij}=k) = p(a_{ij}=k|t_{ij}, S_i) = \frac{p(a_{ij}=k, t_{ij}|S_i)}{p(t_{ij}|S_i)} =$$

номер предложения

$$= \frac{p(t_{ij}|\overbrace{S_i, a_{ij}=k}^{S_{ik}}) \cdot p(a_{ij}=k|S_i)}{\sum_{d=1}^{n_i} p(t_{ij}, a_{ij}=d|S_i)} = \frac{p(t_{ij}|S_{ik}) \cdot \frac{1}{n_i}}{\sum_{d=1}^{n_i} p(t_{ij}|\underbrace{a_{ij}=d, S_i}_{S_{id}}) \cdot p(a_{ij}=d|S_i)} =$$

$$= \frac{\frac{1}{n_i}\, p(t_{ij}|S_{ik})}{\sum_{d=1}^{n_i} \frac{1}{n_i}\, p(t_{ij}|S_{id})} = \frac{\theta(t_{ij}|S_{ik})}{\sum_{d=1}^{n_i} \theta(t_{ij}|S_{id})}$$

Итого на E-шаге получили что $q^*(a_{ij}=k) = \frac{\theta(t_{ij}|S_{ik})}{\sum_{d=1}^{n_i} \theta(t_{ij}|S_{id})}$

$$\theta(t_{ij}|S_{ik}) = \frac{\theta(t_{ij}, S_{ik})}{\theta(S_{ik})}$$

M-шаг: Нижняя оценка логарифма неполного правдоподобия:

$$L(q,\theta) = \mathbb{E}_{q^*} \log \frac{P(T,A|S)}{q^*(A)}\, dz = \mathbb{E}_{q^*} \log \frac{\prod_{i=1}^{R} P(t_i,a_i|s_i)}{q^*(a_i)} =$$

$$= \sum_{i=1}^{R} \mathbb{E}_{q^*} \log \left(\frac{p(t_i,a_i|s_i)}{q^*(a_i)} \right) = \sum_{i=1}^{R} \mathbb{E}_{q^*} \log \frac{\prod_{j=1}^{m_i} p(t_{ij},a_{ij}|s_i)}{\prod_{j=1}^{m_i} q^*(a_{ij})} =$$

$$= \sum_{i=1}^{R}\sum_{j=1}^{m_i}\sum_{k=1}^{n_i} q^*(a_{ij}=k) \log \left(\frac{\theta(t_{ij}|s_{ik}) \frac{1}{n_i}}{q^*(a_{ij})} \right) = \text{const} + \sum_{i=1}^{R}\sum_{j=1}^{m_i}\sum_{k=1}^{n_i} q^*(a_{ij}=k) \log(\theta(t_{ij}|s_{ik}))$$

$$\Downarrow$$

$$\begin{cases} \sum_{i=1}^{R}\sum_{j=1}^{m_i}\sum_{k=1}^{n_i} q^*(a_{ij}=k) \log(\theta(t_{ij}|s_{ik})) \to \max\limits_{\theta} \\ \sum_{y} \theta(y|x) = 1 \end{cases}$$

(огр на $\theta(y|x) > 0$ не нужно и в конце будет видно почему)

y — по всему мн-ву слов целевого языка

$\exists\ \mathcal{L}$ - ф-ция Лагранжа. Тогда.

$$\mathcal{L} = \sum_{i,j,k} q(a_{ij}=k) \log \theta(t_{ij}|s_{ik}) + \sum_{x} \lambda_x \left(\sum_{y} \theta(y|x) - 1 \right)$$

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \theta(y|x)} = \sum_{i,j,k} q(a_{ij}=k) \dfrac{1}{\theta(y|x)} [t_{ij}=y][s_{ik}=x] + \lambda_x = 0 \\ \dfrac{\partial \mathcal{L}}{\partial \lambda_x} = \sum_{y} \theta(y|x) - 1 = 0. \end{cases}$$

$$\Rightarrow \quad \theta(y|x) = -\frac{1}{\lambda_x} \cdot \left(\sum_{i,j,k} q(a_{ij}=k)[t_{ij}=y][s_{ik}=x] \right) \quad (*)$$

$$-\frac{1}{\lambda_x} \cdot \sum_{i,j,k} \left(q(a_{ij}=k)[s_{ik}=x] \underbrace{\sum_{y}[t_{ij}=y]}_{=1} \right) = 1.$$

т.к. $t_{ij} \in$ словарю target, а y проходится по всему словарю

$$\Downarrow$$

$$-\lambda_x = \sum_{i,j,k} q(a_{ij}=k)[s_{ik}=x]$$

$\Downarrow$ + (*) $\Downarrow$

$$\Theta(y|x) = \frac{\sum\limits_{i,j,k} q(a_{ij}=k)[t_{ij}=y][s_{ik}=x]}{\sum\limits_{i,j,k} q(a_{ij}=k)[s_{ik}=x]}$$

$$\Downarrow$$

$$\Theta(y|x) = \frac{\sum\limits_{i=1}^{R}\sum\limits_{j=1}^{m_i}\sum\limits_{k=1}^{n_i} q(a_{ij}=k)[t_{ij}=y][s_{ik}=x]}{\sum\limits_{i=1}^{R}\sum\limits_{j=1}^{m_i}\sum\limits_{k=1}^{n_i} q(a_{ij}=k)[s_{ik}=x]}$$

!

$\Theta(y|x) \geq 0$

$\Rightarrow$ огр. автоматом выполняется

Явл.-ли максимумом ?

$$d^2\mathcal{L} = -\sum_{i,j,k} q(a_{ij}=k)\frac{1}{\Theta^2(y|x)}[t_{ij}=y][s_{ik}=x]$$

$$\Downarrow$$

$d^2\mathcal{L} < 0 \Rightarrow$ максимум.

Бонус. (обозначения те же что

] 1) Кол-во предложений $-\ell$

2) Длина i-го предложения : n_i

3) Длина i-го перевода : m_i

4) s_i — i-е предложение

5) s_{ij} — j-е слово в i-м предложении

6) t_{ij} — j-е слово в i-м предложении.-переводе

7) $q(A)$ - распределение на скрытых переменных

8) a_{ij} — скрытая переменная (выравнивание) j-го слова в i-м переводе.

9) a_i — выравнивание

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

$$p(a_{ik}=j,t_{ik}\mid s_i)=p(a_{ik}=j\mid m_i,n_i)\cdot p(t_{ik}\mid a_{ik}=j,s_i)=$$

$$=\varphi_{m_in_i}(j\mid k)\cdot\theta(t_{ik}\mid s_{ij})$$

E-шаг : $q^*(a_{ik}=j)=p(a_{ik}=j\mid t_{ik},s_i)=\dfrac{p(a_{ik}=j,t_{ik}\mid s_i)}{p(t_{ik}\mid s_i)}=$

$$=\frac{p(a_{ik}=j,t_{ik}\mid s_i)}{\sum\limits_{j=1}^{n_i}p(a_{ik}=j,t_{ik}\mid s_i)}=\frac{\varphi_{m_i,n_i}(j\mid k)\cdot\theta(t_{ik}\mid s_{ij})}{\sum\limits_{j=1}^{n_i}\varphi_{m_i,n_i}(j\mid k)\,\theta(t_{ik}\mid s_{ij})}$$

M-шаг

Нижняя оценка: $L=\int q^*(A)\cdot\log\dfrac{p(T,A)}{q^*(A)}=\{$ также как и в Word Alignment

$$=\sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_i}q^*(a_{ik}=j)\cdot\log\frac{p(a_{ik}=j,t_{ik}\mid s_i)}{q^*(a_{ik}=j)}\ \textcircled{=}$$

$$\textcircled{=}\ \sum_{i=1}^{\ell}\sum_{j=1}^{n_i}\sum_{k=1}^{m_i}q^*(a_{ik}=j)\log\frac{\varphi_{m_i,n_i}(j\mid k)\,\theta(t_{ik}\mid s_{ij})}{q^*(a_{ik}=j)}=$$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_i} q^*(a_{ik}=j) \log \Phi_{m_i, n_i}(j|k) +$$

$$+ \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_i} q^*(a_{ik}=j) \log \frac{\Theta(t_{ik} | s_{ij})}{q^*(a_{ik}=j)}$$

1

2

Θ и Φ независимы в 1 и 2

⇓

для Θ из Word Aligner получаем,

$$\Theta(y|x) = \frac{\sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_i} q^*(a_{ik}=j) \cdot [s_{ij}=x][t_{ik}=y]}{\sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_i} q^*(a_{ik}=j)[s_{ij}=x]}$$

теперь разберёмся с 1

$$\begin{cases} \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_i} q_{ijk} \log \Phi_{m_i, n_i}(j|k) \to \max \\ \sum_{j=1}^{n_i} \Phi_{m_i n_i}(j|k) = 1 \end{cases}$$

(здесь неотр. не будем требовать, т.к. в оптимизационной з-че макс. будет достигаться в "хорошей" точке)

Лагранжиан:

$$\mathcal{L} = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_i} q(a_{ik}=j) \log \Phi_{m_i, n_i}(j|k) - \sum_{i=1}^{\ell} \sum_{k=1}^{m_i} \lambda_{ik} \sum_{j=1}^{n_i} (\Phi_{m_i, n_i}(j|k) - 1)$$

дифф-м по $\Phi_{md, nd}(b|c)$

$$0 = \frac{\partial \mathcal{L}}{\partial \Phi_{md, nd}(b|c)} = \sum_{i=1}^{\ell} \sum_{j=1}^{n_i} \sum_{k=1}^{m_i} q(a_{ik}=j) \frac{1}{\Phi_{m_i n_i}(j|k)} [n_i = nd][m_i = md] [j=b][k=c] -$$

$$- \sum_{i=1}^{\ell} \sum_{k=1}^{m_i} \lambda_{ik} [n_i = nd][m_i = md][k=c]$$

$$\Phi_{md,nd}(b|c) = \frac{\sum_{i=1}^{e}\sum_{j=1}^{n_i}\sum_{k=1}^{m_i} q(a_{ik}=j)\,[n_i=nd]\,[m_i=md]\,[j=b]\,[k=c]}{\sum_{i=1}^{e}\sum_{k=1}^{m_i} \lambda_{ik}\,[n_i=nd]\,[m_i=md]\,[k=c]}$$

$\Downarrow$

Суммируем обе части по $b \in \overline{1, nd}$

$$\sum_{b=1}^{nd} \Phi_{md,nd}(b|c) = 1 = \frac{\sum_{i=1}^{e}\sum_{j=1}^{n_i}\sum_{k=1}^{m_i} q(a_{ik}=j)\,[n_i=nd]\,[m_i=md]\,[k=c]\overbrace{\sum_{b=1}^{nd}[j=b]}^{1}}{\sum_{i=1}^{e}\sum_{k=1}^{m_i} \lambda_{ik}\,[n_i=nd]\,[m_i=md]\,[k=c]} \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{e}\sum_{k=1}^{m_i} \lambda_{ik}\,[n_i=nd]\,[m_i=md]\,[k=c] = \sum_{i=1}^{e}\sum_{j=1}^{a_i}\sum_{k=1}^{m_i} q(a_{ik}=j)\,[n_i=nd]\cdot[m_i=md]\,[k=c)$$

$$\Phi_{md,nd}(b|c) = \frac{\sum_{i=1}^{e}\sum_{j=1}^{n_i}\sum_{k=1}^{m_i} q(a_{ik}=j)\,[n_i=nd]\,[m_i=md]\,[j=b]\,[k=c]}{\sum_{i=1}^{e}\sum_{j=1}^{n_i}\sum_{k=1}^{m_i} q(a_{ik}=j)\,[n_i=nd]\,[m_i=md]\,[k=c]}$$