26.10.18 dl

$$p(x_{t+1} \mid y_t) = N(x_{t+1} \mid \mu(y_t), \Sigma(y_t))$$

$$\Sigma = AA^T \qquad \sigma = \exp(\hat{\sigma})$$

$$\sigma = \log(1 + \exp(\hat{\sigma}))$$

$(\triangle^{\circ})$

смесь $p(x_{t+1} \mid y_t) = \sum_k w_k N(x_{t+1} \mid \mu_k(y_t), \Sigma_k(y_t))$

Механизм внимания

$C = [c_1 \dots c_u]$

Воспроизведение рукописного шрифта

$$w_t = \sum_{i=1}^{u} \varphi(t,i)\, c_i$$

$$\varphi(t,i) = \sum_j \alpha_j^t \exp\left(-\beta_j^t (\kappa_j^t - i)^2\right)$$

$$\hat{\alpha}^t, \hat{\beta}^t, \hat{\kappa}^t = W h_t + b$$

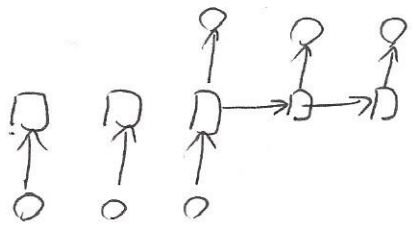
$$\alpha_j^t = \exp(\hat{\alpha}_j^t), \quad \beta_j^t = \exp(\hat{\beta}_j^t), \quad \kappa_j^t = \kappa_j^{t-1} + \exp(\hat{\kappa}_j^t)$$

$$w_j = \frac{\exp(\hat{w}_j (1+a))}{\sum_k \exp(\hat{w}_k (1+a))}$$

$Z \sim Discrete(\alpha_1 \dots \alpha_L)$

$$\alpha_i = \frac{\exp(score(h, f_i))}{\sum_{j=1}^{L} \exp(score(h, f_j))}$$

Машинный перевод



$PE_{pos,2i} = \sin(pos/10000^{2i/d_{model}})$ $\quad 1 \to bs$ $\quad (bs, ch, hw) \times (bs, hw, ch)$

$PE_{pos,2i+1} = \cos(pos/10000^{2i/d_{model}})$ $\quad (bs, ch, ch)$

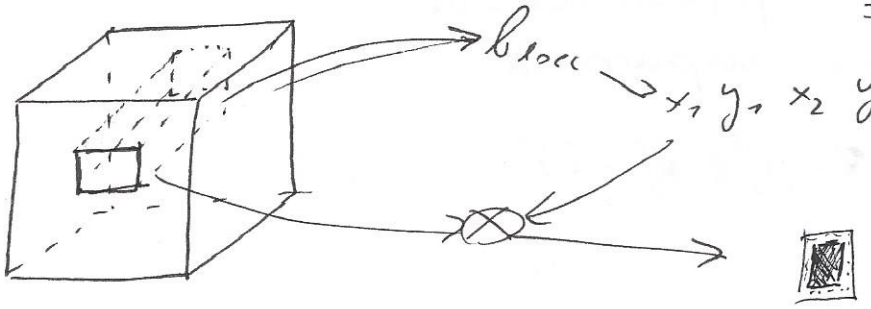
$2i$ пробегаем по эмбеддингу $\quad W@F\ (bs, q, f)$

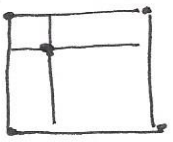
позиции слов в эмбеддингах

Self-attention $\quad f(x) = X$ $\quad W\ (bs, q, obj)$

$y_n = \sum_{i=1}^{L} \alpha_{ni} x_i$ $\quad F\ (bs, obj, f)$

Transformer $\quad (bs, nq, fd) \times (fd, hd) = (bs, nq, hd)$





$\times (hd) = (bs, nq)$

$v_{x,y} = \sum_{i=1}^{W} \sum_{j=1}^{H} u_{ij}^{u_{i+1,j+1}} \cdot \max(0, 1-|x-i|) \cdot \max(0, 1-|y-j|)$

IN $H \times W$ $\quad H \times W$

OUT $H \times W$ $\quad (H-3) \times (W-3)$ key = features

$p_3[i] = p_1[p_2[i]]$

$a = \sum_{i=1}^{L} v_i w_i$, $w_i = \frac{\exp(score(q, k_i))}{\sum_{j=1}^{L} \exp(score(q, k_j))}$

w (batch_size, queries)

features (bs, obj, f) $\quad >$ (bs, q, f)

w (bs, q)

~~w (bs, obj)~~