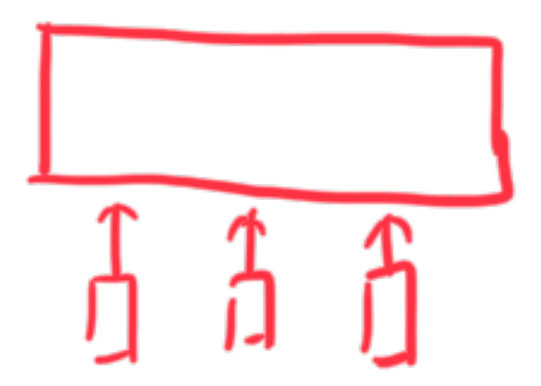

$x_1 \quad x_2 \quad \cdots \quad x_n$

bag of words

RNN

Position Encoding

Absolute Position

$(e_i+p_i)^T W_Q^T W_k (e_j+p_j)$

$$\begin{cases} P_{(k,2i)} = \sin(k/10000^{2i/d}) \\ P_{(k,2i+1)} = \cos(k/10000^{2i/d}) \end{cases}$$

$e_i^T W_Q^T W_k e_j + p_i^T W_Q^T W_k e_j + e_i^T W_Q^T W_k p_j + p_i^T W_Q^T W_k p_j$

Absolute Position → trainable embeddings

$p_i^T W_Q^T W_k p_j = g(i-j)$

$p_i^T p_j = g(i-j)$

$\cos\theta \cos\varphi + \sin\theta \sin\varphi$

$= \cos(\theta - \varphi)$

$p_i = \begin{pmatrix} \cos\theta_i \\ \sin\theta_i \end{pmatrix} \quad p_j = \begin{pmatrix} \cos\theta_j \\ \sin\theta_j \end{pmatrix}$

$\cos[(i-j)\theta_0]$

$\theta_k = k\theta_0$

$\cos(i\theta_0 - j\theta_0)$

$\cos(\theta_i - \theta_j)$

Relative Position

$(e_i+u)^T W_Q^T W_k (e_j + r_{i-j})$

- XLNet. $① + u^T W_Q^T W_k e_j + e_i^T W_Q^T W_k r_{i,j} + v^T W_Q^T W_k r_{i,j}$
- T5 $① + p_i^T W_Q^T W_k p_j \Rightarrow \beta_{i,j}$
- DeBERTa $① + r_{j,i}^T W_Q^T W_k e_j + e_i^T W_Q^T W_k r_{i,j}$

$(E+P)^T W_Q^T W_k (E+P)$

Rotate Position

Rotate Position → $f(\vec{q},m)^T f(\vec{k},n) = g(\vec{q},\vec{k},m-n)$

↓

ViT

$x_i + p_i$? $\qquad f(\cdot)$

$$f(\vec{q},m)^T f(\vec{k},n) = g(\vec{q},\vec{k},m-n)$$

$$f(\vec{q},m) = \begin{bmatrix} \cos m\theta & -\sin m\theta \\ \sin m\theta & \cos m\theta \end{bmatrix} \vec{q}$$

$$f(\vec{q},m)^T f(\vec{k},n) = \vec{q}^T \begin{bmatrix} \cos m\theta & -\sin m\theta \\ \sin m\theta & \cos m\theta \end{bmatrix}^T \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix} \vec{k}$$

$$= \vec{q}^T \begin{bmatrix} \cos(m-n)\theta & \sin(m-n)\theta \\ -\sin(m-n)\theta & \cos(m-n)\theta \end{bmatrix} \vec{k}$$

$$= g(\vec{q},\vec{k},m-n)$$

position

	0	1	…	511
$\theta_0 = \frac{1}{10000^{\frac{0}{768}}}$	$\cos(0 \times \frac{1}{10000^{\frac{0}{768}}})$	$\cos(1 \times \frac{1}{10000^{\frac{0}{768}}})$	…	$\cos(511 \times \frac{1}{10000^{\frac{0}{768}}})$
	$\sin(0 \times \frac{1}{10000^{\frac{0}{768}}})$	$\sin(1 \times \frac{1}{10000^{\frac{0}{768}}})$	…	$\sin(511 \times \frac{1}{10000^{\frac{0}{768}}})$
$\theta_0 = \frac{1}{10000^{\frac{2}{768}}}$	$\cos(0 \times \frac{1}{10000^{\frac{2}{768}}})$	$\cos(1 \times \frac{1}{10000^{\frac{2}{768}}})$	…	$\cos(511 \times \frac{1}{10000^{\frac{2}{768}}})$
	$\sin(0 \times \frac{1}{10000^{\frac{2}{768}}})$	$\sin(1 \times \frac{1}{10000^{\frac{2}{768}}})$	…	$\sin(511 \times \frac{1}{10000^{\frac{2}{768}}})$
$\theta_0 = \frac{1}{10000^{\frac{4}{768}}}$	$\cos(0 \times \frac{1}{10000^{\frac{4}{768}}})$	$\cos(1 \times \frac{1}{10000^{\frac{4}{768}}})$	…	$\cos(511 \times \frac{1}{10000^{\frac{4}{768}}})$
	$\sin(0 \times \frac{1}{10000^{\frac{4}{768}}})$	$\sin(1 \times \frac{1}{10000^{\frac{4}{768}}})$	…	$\sin(511 \times \frac{1}{10000^{\frac{4}{768}}})$
⋮	⋮	⋮	…	⋮
$\theta_0 = \frac{1}{10000^{\frac{768}{768}}}$	$\cos(0 \times \frac{1}{10000^{\frac{768}{768}}})$	$\cos(1 \times \frac{1}{10000^{\frac{768}{768}}})$	…	$\cos(511 \times \frac{1}{10000^{\frac{768}{768}}})$
	$\sin(0 \times \frac{1}{10000^{\frac{768}{768}}})$	$\sin(1 \times \frac{1}{10000^{\frac{768}{768}}})$	…	$\sin(511 \times \frac{1}{10000^{\frac{768}{768}}})$