

# BERT 尺寸计算

前提条件

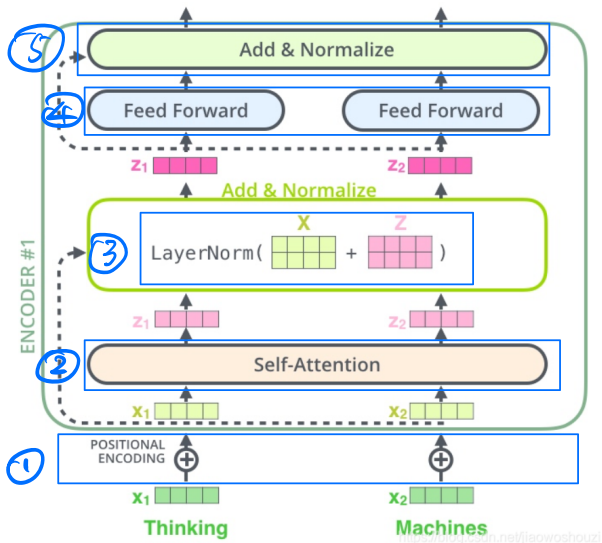
① 假设 embedding 为 768 维

② 假设. 同语句内包含 512 个 token.

$$512 \times 768$$

$$N = 512 \quad M = 768$$

$$H = 12$$



## ① Positional Encoding

⊕ 加减法移位?

sin/cos cordic? / Taylor?

加法

embedding 相加

PE - sin/cos

PE: 除法

当 10 万

$$2M \times N + N \times \text{三角函数} + 2 \times \text{除法} \times N$$

$$\approx 801792$$

乘法

PE

$$N \times (1 + 2 \times \text{除法} + \text{乘法} + \text{sin/cos}) \approx 26112$$

## ② Self-Attention.

加法

kqv

Q · K<sup>T</sup>

当 5

multihead

$$3 \times (M-1) \times N \times N + (M-1) \times N^2 + N^2 \cdot \text{Softmax} + (N+1)M \cdot N + (N+1)H \cdot M \cdot N \approx 3.42 \times 10^9$$

乘法

kqv

Q · K<sup>T</sup>

当 5

multihead

$$3 \times M \times N \times N + M \cdot N^2 + N^2 \cdot \text{Softmax} + N \cdot M \cdot N + N \cdot H \cdot M \cdot N \approx 3.56 \times 10^9$$

### ③ Layer Form

加法

$x \times 2$

$$M \cdot N + 2 \cdot MN \approx 1.18 \times 10^6$$

乘法

$x \times 2$

$$2 \cdot M \cdot N \approx 7.86 \times 10^5$$

### ④ Feed Forward

加法

$x3 \cdot x4$

$$2MN + (N-1)M \cdot N \cdot 2 + 2 \cdot MN \stackrel{\text{GELU}}{\approx} 4.034 \times 10^8$$

乘法

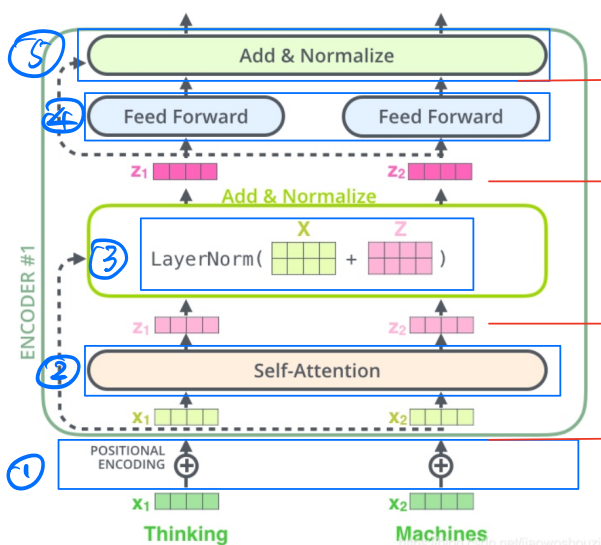
$x3 \cdot x4$

$$N \cdot MN \cdot 2 + 8 \cdot MN + MN \cdot (\oplus + \tanh + \sqrt{\phantom{x}})$$

$$\approx 4.14 \times 10^8$$

### ⑤ Normalized = ③

Result



加法	乘法
$3.42 \times 10^9$	$3.56 \times 10^9$
$4.034 \times 10^8$	$4.14 \times 10^8$
$1.18 \times 10^6$	$7.86 \times 10^5$
$3.42 \times 10^9$	$3.56 \times 10^9$
80 1792	26112