

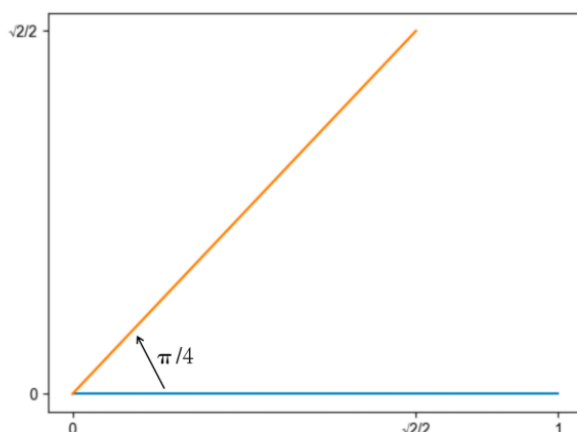
旋转位置编码：

公式推导：<https://cloud.tencent.com/developer/article/2314990>

<https://hub.baai.ac.cn/view/29979>

https://blog.csdn.net/qg_27590277/article/details/134589368

$$\begin{pmatrix} \cos \frac{\pi}{4} & -\sin \frac{\pi}{4} \\ \sin \frac{\pi}{4} & \cos \frac{\pi}{4} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \frac{\pi}{4} \\ \sin \frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}$$



上式中的旋转矩阵十分稀疏，为了节省算力，可以以下面的方式等效实现：

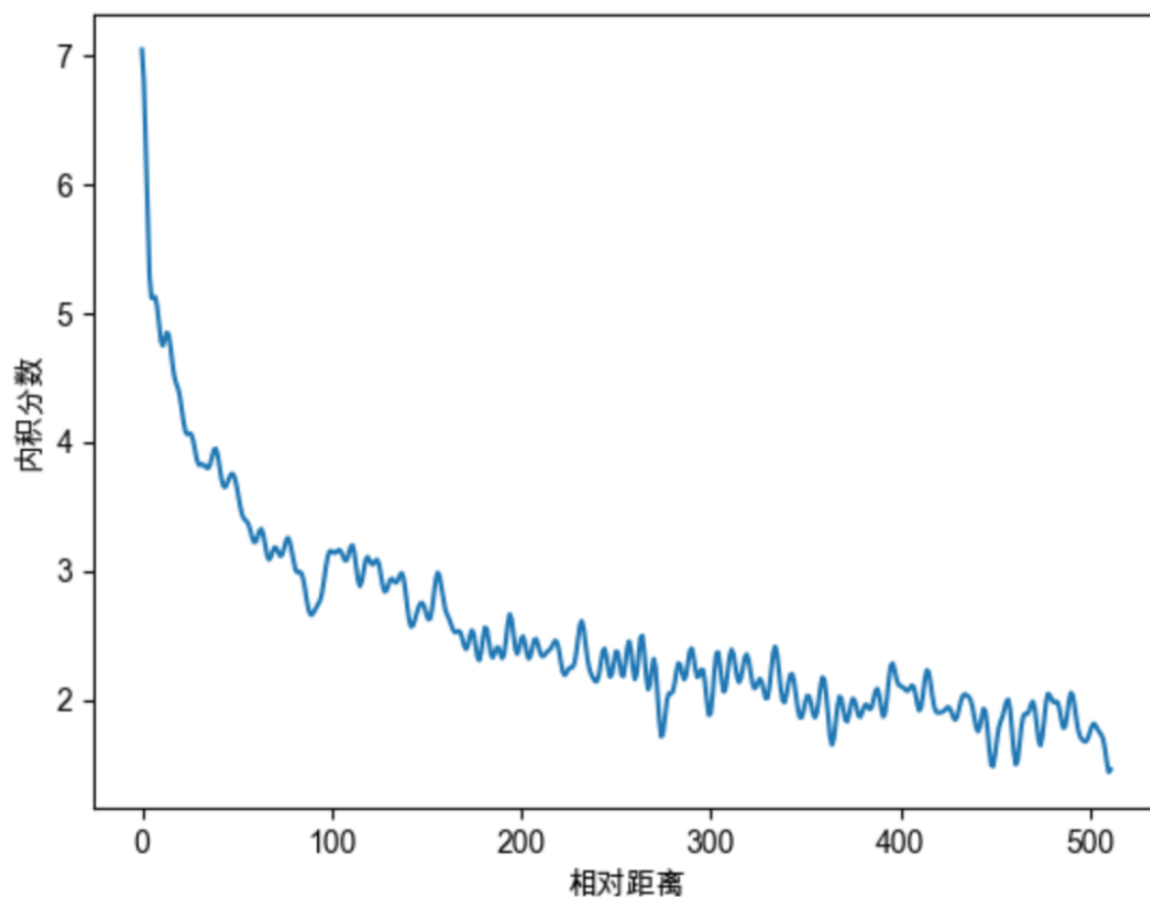
$$\begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ \vdots \\ q_{d-2} \\ q_{d-1} \end{pmatrix} \otimes \begin{pmatrix} \cos m\theta_0 & \\ & \cos m\theta_0 \\ & & \cos m\theta_1 \\ & & & \cos m\theta_1 \\ & & & & \vdots \\ & & & & & \cos m\theta_{d/2-1} \\ & & & & & & \cos m\theta_{d/2-1} \end{pmatrix} + \begin{pmatrix} -q_1 \\ q_0 \\ -q_3 \\ q_2 \\ \vdots \\ -q_{d-1} \\ q_{d-2} \end{pmatrix} \otimes \begin{pmatrix} \sin m\theta_0 & \\ & \sin m\theta_0 \\ & & \sin m\theta_1 \\ & & & \sin m\theta_1 \\ & & & & \vdots \\ & & & & & \sin m\theta_{d/2-1} \\ & & & & & & \sin m\theta_{d/2-1} \end{pmatrix}$$

$$\begin{pmatrix} \cos(m\theta) & -\sin(m\theta) \\ \sin(m\theta) & \cos(m\theta) \end{pmatrix} \begin{pmatrix} q_m^{(1)} \\ q_m^{(2)} \end{pmatrix}$$

$$p_{i,2t} = \sin\left(\frac{i}{10000^{\frac{2t}{d}}}\right)$$

$$p_{i,2t+1} = \cos\left(\frac{i}{10000^{\frac{2t}{d}}}\right)$$

意义：



1. 良好的外推性：RoPE旋转位置编码使得模型在处理超出训练长度的文本时，仍能保持较好的性能。这一特性使得RoPE在处理长文本任务中具有显著优势。
2. 远程衰减：随着文本位置的增加，RoPE旋转位置编码的影响力逐渐减弱。这种特性使得模型在关注重要信息时，能够减少冗余信息的干扰。
3. 可用于线性Attention：传统的Transformer模型在处理长文本时，由于Attention机制的计算复杂度较高，导致模型性能受限。而RoPE旋转位置编码作为一种相对位置编码方式，可以应用于线性Attention中，有效降低计算复杂度，提高模型性能。