

RoPE

AGI Group

2025-11

RoPE (Rotary Position Embedding) 的相对位置

RoPE (Rotary Position Embedding) 是一种用于自然语言处理模型中的位置编码方法。与传统的绝对位置编码不同，RoPE 通过旋转嵌入向量来表示相对位置，从而更好地捕捉序列中词语之间的相对关系。

假设输入向量 x_m^d 和 x_n^d 分别表示序列中位置 m 和 n 的词嵌入向量，其中 d 表示嵌入向量的维度。RoPE 通过对这些向量进行旋转来引入相对位置信息。

下面以 $d = 2$ 为例，介绍其数学原理。假设 k 表示在输入序列的位置，其旋转矩阵是：

$$R_k = \begin{bmatrix} \cos k\theta & -\sin k\theta \\ \sin k\theta & \cos k\theta \end{bmatrix}$$

基本数学推导

由于 $\sin(a+b) = \sin a \cos b + \cos a \sin b$, $\cos(a+b) = \cos a \cos b - \sin a \sin b$, 我们可以证明旋转矩阵具有以下性质：

$$R_m R_n = \begin{bmatrix} \cos m\theta & -\sin m\theta \\ \sin m\theta & \cos m\theta \end{bmatrix} \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix} = \begin{bmatrix} \cos(m+n)\theta & -\sin(m+n)\theta \\ \sin(m+n)\theta & \cos(m+n)\theta \end{bmatrix} = R_{m+n}$$

而由于 R_{-k} 是 R_k 的转置矩阵，我们有：

$$R_m^T R_n = R_{-m} R_n = R_{n-m}$$

在 Attention 中的应用

考虑向量 x_m 和向量 x_n ，假设 x_m 对应的查询向量是 Q_m ， x_n 对应的键向量是 K_n 。如果应用 RoPE，我们将查询向量和键向量分别旋转（考虑位置信息），即

- $Q'_m = R_m Q_m$
- $K'_n = R_n K_n$

因此，

$$Q_m'^T K'_n = (R_m Q_m)^T (R_n K_n) = Q_m^T R_m^T R_n K_n = Q_m^T R_{n-m} K_n$$

从上面的结果可知，只要 $n-m$ 相同，无论 m 和 n 具体是多少，旋转后的查询向量和键向量的点积结果是相同的。这表明 RoPE 成功地将相对位置信息引入到了注意力机制中。

扩展到更高维度

考虑通用的 d 维情况，我们可以将 d 维向量分成 $d/2$ 个二维子空间，每个子空间应用相同的旋转矩阵。

$$R_{\Theta, m}^d = \begin{pmatrix} \cos m\theta_1 & -\sin m\theta_1 & 0 & 0 & \cdots & 0 & 0 \\ \sin m\theta_1 & \cos m\theta_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cos m\theta_2 & -\sin m\theta_2 & \cdots & 0 & 0 \\ 0 & 0 & \sin m\theta_2 & \cos m\theta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cos m\theta_{d/2} & -\sin m\theta_{d/2} \\ 0 & 0 & 0 & 0 & \cdots & \sin m\theta_{d/2} & \cos m\theta_{d/2} \end{pmatrix}$$

其中, $\theta_i = 100000^{-2(i-1)/d}, i \in [1, 2, \dots, d/2]$ 。