

# BiCrossSNPCaduceus

---

## 1. 输入表示

设输入的两条链为A链 (A链为 $(X^a)$ ) 和B链 (B链为 $(X^b)$ ) , 长度均为 ( $L=512$ ):

$$[ X^a = [x_1^a, x_2^a, \dots, x_L^a], \quad X^b = [x_1^b, x_2^b, \dots, x_L^b] ]$$

其中, 每个位置的核苷酸:  $[x_i^a, x_i^b \in \{A,T,C,G\}]$

---

## 2. Mask 策略

随机选择 15% 的位置索引集合 ( $\mathcal{M}$ ):

$$[ \mathcal{M} = \{i_1, i_2, \dots, i_{|\mathcal{M}|}\}, \quad |\mathcal{M}| = 0.15 \times L ]$$

A,B链同时在这些位置进行Mask (替换为特殊token ( $\texttt{[MASK]}$ )) :

$$[ X_{\text{masked}}^a[i] = \begin{cases} \texttt{[MASK]}, & i \in \mathcal{M} \\ x_i^a, & i \notin \mathcal{M} \end{cases}, \quad X_{\text{masked}}^b[i] = \begin{cases} \texttt{[MASK]}, & i \in \mathcal{M} \\ x_i^b, & i \notin \mathcal{M} \end{cases} ]$$


---

## 3. 模型特征提取

对于AB链分别建立初始embedding (词表映射, 和PlantCaduceus一致) :

$$[ H_0^a = \text{Embed}(X_{\text{masked}}^a), \quad H_0^b = \text{Embed}(X_{\text{masked}}^b) ]$$

独立特征提取 (Conv & SSM) :

- **卷积特征:**  $[ H_{\text{conv}}^a = \text{Conv}(H_0^a), \quad H_{\text{conv}}^b = \text{Conv}(H_0^b) ]$
  - **状态空间模型 (SSM):**  $[ H_{\text{ssm}}^a = \text{SSM}(H_{\text{conv}}^a), \quad H_{\text{ssm}}^b = \text{SSM}(H_{\text{conv}}^b) ]$
- 

## 4. 双链交叉注意力层

分别用A链特征对B链特征进行交叉注意力, 反之亦然:

- **A链特征的Cross Attention:**

$$[ H_{\text{cross}}^a = \text{CrossAttn}(Q=H_{\text{ssm}}^a, K=H_{\text{ssm}}^b, V=H_{\text{ssm}}^b) ]$$

- **B链特征的Cross Attention:**

$$[ H_{\text{cross}}^b = \text{CrossAttn}(Q=H_{\text{ssm}}^b, K=H_{\text{ssm}}^a, V=H_{\text{ssm}}^a) ]$$

交叉注意力 (Cross Attention) 具体定义为:

$$[\text{CrossAttn}(Q, K, V) = \text{softmax}(\frac{QK^\top}{\sqrt{d_k}})V]$$

其中( $d_k$ )是attention的维度。

---

## 5. 特征融合与预测

将原特征与交叉特征Concat并经过非线性变换：

- A链融合：  $[H_{\text{fuse}}^a = \text{GELU}(H_{\text{ssm}}^a; H_{\text{cross}}^a)W^a a b^a]$
- B链融合：  $[H_{\text{fuse}}^b = \text{GELU}(H_{\text{ssm}}^b; H_{\text{cross}}^b)W^b a b^b]$

经过各自的预测头 (LM Head)：

- A链：  $[P^a(x_i) = \text{softmax}(H_{\text{fuse}}^a i) W_{\text{lm}}^a a b_{\text{lm}}^a]$
- B链：  $[P^b(x_i) = \text{softmax}(H_{\text{fuse}}^b i) W_{\text{lm}}^b a b_{\text{lm}}^b]$

其中：

- $(P^a(x_i), P^b(x_i))$  分别表示第*i*位置（被Mask位置）A链和B链的碱基概率分布，维度均为4 (A,T,C,G)。
  - $(W_{\text{lm}}^a, W_{\text{lm}}^b, b_{\text{lm}}^a, b_{\text{lm}}^b)$  是可学习的权重和偏置。
- 

## 6. 损失函数

损失函数采用**Masked Language Modeling**的交叉熵损失，分别对AB链预测概率求和：

- A链损失：  $[\mathcal{L}^a = -\frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \log P^a(x_i)]$
- B链损失：  $[\mathcal{L}^b = -\frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} \log P^b(x_i)]$

**总损失**为两者的加权和（此处可选等权重）：

$$[\mathcal{L}_{\text{total}} = \mathcal{L}^a + \mathcal{L}^b]$$


---

## 7. 模型训练目标

优化目标为最小化总损失：

$$[\theta^* = \arg \min_{\theta} \mathcal{L}_{\text{total}}(\theta)]$$

其中，( $\theta$ )代表模型中所有可学习的参数。

---

## 模型特点

- 给定双链输入，Mask相同位置；
- 独立提取A, B链特征；
- 通过交叉注意力实现链间信息交互；
- 通过LM Head分别预测AB链Mask位置碱基概率；

- 使用Masked位置标签的交叉熵损失优化整个模型。