

生物进阶

首先，我们可以把 A,C,G,T 四个字母分开考虑。我们用 $matchA[i]$ 表示 $S[i-m+1] \sim S[i]$ 这段有多少字符 A 能和 T 串粗略匹配。 $matchC$, $matchG$, $matchT$ 这三个数组意义类似。将这四个数组计算出来之后，我们只需要统计

$$ans = \sum_{i=m-1}^{n-1} [matchA_i + matchC_i + matchG_i + matchT_i = m]$$

至于如何统计，这里以 A 为例，我们先将 S 串表示成一个 01 串，1 表示附近 k 位中存在 A，0 表示附近 k 位中不存在 A。例如样例中的 AGCAATTCAT，我们可以表示为 1111110111。然后再将 T 表示成一个 01 串，字符为 A 的位置为 1，否则为 0，例如 ACAT 可以表示为 1010。这样表示以后，我们发现

$$matchA_i = \sum_{j=i-m+1}^i S_j T_{j-(i-m+1)}$$

然后我们把 T 串倒一下，上面的式子就变成了：

$$matchA_i = \sum_{j=i-m+1}^i S_j T_{(m-1)-j+(i-m+1)} = \sum_{j=i-m+1}^i S_j T_{i-j} = \sum_{j=0}^i S_j T_{i-j}$$

然后 FFT 就可以了。

时间复杂度 $O(4n \log n)$