IOI2026 中国国家集训队集中培训 第一试 B. 字符串问题

IOI2026 中国国家集训队工作组

2025 年 12 月 2 日

问题（字符串问题（string））

- 给定一个长为 n 的字符串 s 和一个权值序列 $f(1),\dots,f(n)$。
- 对子串 $s[l,r]$，若其最小整周期为 d，其价值为 $f((r-l+1)/d)$。
- 对每个 i，求以 i 为右端点的子串的价值之和。

数据范围：$n \le 10^6$。

$n \leq 5000$

$n \leq 5000$

- 枚举左右端点 l, r，记 $len = (r - l + 1)$，用 KMP 顺便算出 $s[l, r]$ 的最长 border。
- 若最长 border 长度为 len'，则其最小周期为 $d' = len - len'$。

$n \le 5000$

- 枚举左右端点 l, r，记 $len = (r - l + 1)$，用 KMP 顺便算出 $s[l, r]$ 的最长 border。
- 若最长 border 长度为 len'，则其最小周期为 $d' = len - len'$。
- 为了求**最小整周期** d，需要先知道它和**最小周期** d' 的关系。注意到：
  - 对长为 len 的字符串 s，其最小整周期 d 要么为 len，要么不超过 $len/2$。
  - 若是不超过 $len/2$，则最小整周期 d 就是最小周期 d'。
  - 反证：若不然，则根据周期引理，$\gcd(d', d)$ 是一个更小的整周期，矛盾。
  - 因此，只需要判断 d' 是否整除 len，是则 $d = d'$，否则 $d = len$。

$n \le 5000$

- 枚举左右端点 l, r，记 $len = (r - l + 1)$，用 KMP 顺便算出 $s[l, r]$ 的最长 border。
- 若最长 border 长度为 len'，则其最小周期为 $d' = len - len'$。
- 为了求**最小整周期** d，需要先知道它和**最小周期** d' 的关系。注意到：
  - 对长为 len 的字符串 s，其最小整周期 d 要么为 len，要么不超过 $len/2$。
  - 若是不超过 $len/2$，则最小整周期 d 就是最小周期 d'。
  - 反证：若不然，则根据周期引理，$\gcd(d', d)$ 是一个更小的整周期，矛盾。
  - 因此，只需要判断 d' 是否整除 len，是则 $d = d'$，否则 $d = len$。
- 时间复杂度 $O(n^2)$，期望得分：25。
- 若是用更暴力的方法（例如：$O(n^3)$ 或 $O(n^2 \log n)$ 地求最小整周期），则只能得 15 分。

$n \leq 10^6$

$n \leq 10^6$

- 和前面同理可得，若 d 是 $s[l,r]$ 的最小整周期，则 $s[l,r]$ 的所有整周期均为 d 的倍数。

$n \le 10^6$

- 和前面同理可得，若 d 是 $s[l,r]$ 的最小整周期，则 $s[l,r]$ 的所有整周期均为 d 的倍数。
- 于是可以考虑莫比乌斯反演。
- 对一个串 $s[l,r]$ 和它的一个整周期 kd，我们让 $(s[l,r],kd)$ 对 r 产生 $g\left(\frac{r-l+1}{kd}\right)$ 的贡献。
- 于是，这样的贡献之和就是 $\sum_{k|\frac{r-l+1}{d}} g\left(\frac{r-l+1}{kd}\right) = (g*1)\left(\frac{r-l+1}{d}\right)$。
- 令 $g=f*\mu$，则 $g*1=f$，上式为 $f\left(\frac{r-l+1}{d}\right)$，即我们想要的贡献。

- 现在问题转化为，$s[l,r]$ 的每个整周期 d，都会对 r 产生 $g\left(\frac{r-l+1}{d}\right)$ 的贡献。
- $g(1)$ 相关的贡献是平凡的，以下我们只考虑 $g(2), g(3), \ldots$ 的贡献，也就是只考虑平方串、立方串等子串的贡献。

- 现在问题转化为，$s[l,r]$ 的每个整周期 d，都会对 r 产生 $g\left(\frac{r-l+1}{d}\right)$ 的贡献。
- $g(1)$ 相关的贡献是平凡的，以下我们只考虑 $g(2), g(3), \ldots$ 的贡献，也就是只考虑平方串、立方串等子串的贡献。
- 先枚举一个 d，考虑计算所有整周期为 d 的子串的贡献。
- 考虑一个这样的序列 $a^{(d)}$：若 $s_i = s_{i+d}$，则 $a_{i+d}^{(d)} = 1$，否则 $a_{i+d}^{(d)} = 0$。
- 这样，对于 $a^{(d)}$ 中每一个长为 kd 的全 1 子段，它就对应一个长为 $(k+1)d$ 的，以 d 为整周期的子串。

- 现在问题转化为，$s[l,r]$ 的每个整周期 d，都会对 r 产生 $g\left(\frac{r-l+1}{d}\right)$ 的贡献。
- $g(1)$ 相关的贡献是平凡的，以下我们只考虑 $g(2), g(3), \ldots$ 的贡献，也就是只考虑平方串、立方串等子串的贡献。
- 先枚举一个 d，考虑计算所有整周期为 d 的子串的贡献。
- 考虑一个这样的序列 $a^{(d)}$：若 $s_i = s_{i+d}$，则 $a_{i+d}^{(d)} = 1$，否则 $a_{i+d}^{(d)} = 0$。
- 这样，对于 $a^{(d)}$ 中每一个长为 kd 的全 1 子段，它就对应一个长为 $(k+1)d$ 的，以 d 为整周期的子串。
- 由于不考虑 $g(1)$ 的贡献，我们只需要考虑长度至少为 $2d$ 的子串。
- 对应地，我们只需先找出所有 $a^{(d)}$ 中，长度至少为 d 的极长全 1 子段，再对每个子段分别计算贡献即可。

- 具体计算贡献的方法如下：

- 具体计算贡献的方法如下:
- 对于一个 $a^{(d)}$ 中的极长全 1 子段 $[l,r]$。
- 它的任何一个长为 kd 的子区间，都应该对其右端点产生 g_k 的贡献。
- 因此，所有长为 $2d$ 的子区间对 $[l+2d-1,r]$ 产生 g_2 的贡献；所有长为 $3d$ 的子区间对 $[l+3d-1,r]$ 产生 g_3 的贡献；依此类推。
- 直接枚举 k，再用前缀和/差分的方法 $O(1)$ 地把贡献加到对应位置即可。这一步复杂度 $O((r-l+1)/d)$。

- 具体计算贡献的方法如下：
- 对于一个 $a^{(d)}$ 中的极长全 1 子段 $[l, r]$。
- 它的任何一个长为 kd 的子区间，都应该对其右端点产生 g_k 的贡献。
- 因此，所有长为 $2d$ 的子区间对 $[l+2d-1, r]$ 产生 g_2 的贡献；所有长为 $3d$ 的子区间对 $[l+3d-1, r]$ 产生 g_3 的贡献；依此类推。
- 直接枚举 k，再用前缀和/差分的方法 $O(1)$ 地把贡献加到对应位置即可。这一步复杂度 $O((r-l+1)/d)$。
- 时间复杂度分析：对于每个 d，其长度至少为 d 的极长全 1 子段，个数不超过 n/d，总长度不超过 n。
- 因此，若已经找到所有极长全 1 子段，则对一个单独的 d，计算贡献的复杂度为 $O(n/d)$。
- 总复杂度 $O(n\log n)$，足以通过本题。

- 具体寻找 $a^{(d)}$ 中极长全 1 子段有两种方法：

- 具体寻找 $a^{(d)}$ 中极长全 1 子段有两种方法：
  1. 使用类似《优秀的拆分》一题中的方法，把 $d, 2d, 3d, \ldots$ 这些位置设置成关键点，再对两个相邻的关键点算 LCP。
  2. 利用 runs 理论，求出所有 runs (l, r, p) 后，枚举 p 的倍数作为 d，若 $d \leq (r - l + 1)/2$，则 (l, r) 就是一个 $a^{(d)}$ 中的极长全 1 子段。
- 若使用后缀数组求 runs，则不难看出这和上面的方法是等价的。

- 具体寻找 $a^{(d)}$ 中极长全 1 子段有两种方法：
  1. 使用类似《优秀的拆分》一题中的方法，把 $d, 2d, 3d, \ldots$ 这些位置设置成关键点，再对两个相邻的关键点算 LCP。
  2. 利用 runs 理论，求出所有 runs (l, r, p) 后，枚举 p 的倍数作为 d，若 $d \leq (r-l+1)/2$，则 (l, r) 就是一个 $a^{(d)}$ 中的极长全 1 子段。
- 若使用后缀数组求 runs，则不难看出这和上面的方法是等价的。
- 综上所述，总时间复杂度为 $O(n \log n)$，期望得分：100。

Bonus： $n \leq 10^7$

Bonus： $n \le 10^7$

- 根据 runs 的理论，所有 runs 的指数之和是 $O(n)$ 级别的。
- 若线性求 runs，事实上可以把统计贡献的部分优化至 $O(n)$。
- 由于计算贡献时实际会使用 $g * 1$，因此莫比乌斯反演也是不必要的。
- 总时间复杂度为 $O(n)$。

特殊性质 AB

特殊性质 AB

- 沿用莫比乌斯反演的思路。
- 特殊性质 A 事实上意味着 $g_2 = 1$，其他 $g_i = 0$。也就是说，这就是算平方串个数。
- 用《优秀的拆分》一题中的方法可以直接解决这个情况。
- 特殊性质 B 类似，不过是改为算 k 次方串的个数。

特殊性质 AB

- 沿用莫比乌斯反演的思路。
- 特殊性质 A 事实上意味着 $g_2 = 1$，其他 $g_i = 0$。也就是说，这就是算平方串个数。
- 用《优秀的拆分》一题中的方法可以直接解决这个情况。
- 特殊性质 B 类似，不过是改为算 k 次方串的个数。
- 时间复杂度 $O(n \log n)$，期望得分：$25 + 10 = 35$。

$n \leq 2 \times 10^5$

$n \le 2 \times 10^5$

- 用于放过各种 $O(n \log^2 n)$ 做法，或是常数特别特别大的 $O(n \log n)$ 做法。
- 例：使用二分 $+$ 哈希求 LCP，或是在计算贡献时使用不恰当的计算方式/数据结构。

$n \le 2 \times 10^5$

- 用于放过各种 $O(n\log^2 n)$ 做法，或是常数特别特别大的 $O(n\log n)$ 做法。
- 例：使用二分 $+$ 哈希求 LCP，或是在计算贡献时使用不恰当的计算方式/数据结构。
- 期望得分：$10+15+25+10+20=80$。