

CBAL 解题报告

谷晟

1 题目大意

若一个字符串中的字符可以分成两个相同的多重集合，则称这个字符串是平衡的。

给出一个由 N 个小写字母组成的字符串 P 和 Q 个询问，每个询问由三个数 $L, R, Type$ 组成，要求输出

$$\sum_{\substack{L \leq s < e \leq R \\ T_{s..e} \text{ is balanced}}} |T_{s..e}|^{Type}$$

，即字符串 P 的第 L 至 R 位构成的子串 $P[L..R]$ 中，所有的平衡子串的长度为 $Type$ 次方之和。

输入数据加密，需要在线回答询问。

多组数据， $1 \leq \sum N, \sum Q \leq 100000, Type \in \{0, 1, 2\}$

时限 2 秒。

2 算法分析

2.1 基本性质

首先简化题意：一个字符串中的字符可以分成两个相同的多重集合，等价于字符串中每种字符出现次数均为偶数。

设 $A_x[i]$ 表示字符串 P 的前缀 $P[1..i]$ 中字符 x 的个数（特别的， $A_x[0] = 0$ ），则 $P[L..R]$ 中字符 x 出现次数为偶数，等价于

$$A_x[R] - A_x[L - 1] \equiv 0 \pmod{2}$$

，因此我们并不关心每种字符的实际个数，只要知道奇偶性就行了。

设 $A[i]$ 表示前缀 $P[1..i]$ 中所有字符的奇偶性状态。字符串由小写字母组成，总共有 26 种字符，奇偶性总共有 2^{26} 种情况，但是字符串的前缀只有 N 个，所以数组 A 中有效值最多只有 N 种，可以通过“离散化”将范围降到 100000 以内。

这样问题就变成了，给出包含 N 个数的数组 A 以及若干个询问 $(L, R, Type)$ ，求所有满足

$$L \leq s < e \leq R$$

$$A[e] = A[s - 1]$$

的数对 (s, e) ， $(e - s + 1)^{Type}$ 之和。

2.2 线性算法

转换了题意后，很容易得出一种 $O(N)$ 回答一个询问的算法。

先考虑 $Type = 0$ 的情况，这种情况下只要考虑满足条件的数对（子串）个数。只要开一个数组记录每种数字当前出现次数，在从左到右扫描的过程中更新答案就行了。

```
1: for  $i \leftarrow L..R$  do
2:    $ans \leftarrow ans + cnt[A[i]]$ 
3:    $cnt[A[i]] \leftarrow cnt[A[i]] + 1$ 
4: end for
```

$Type = 1$ 时，每次要累加 $\sum(i - j)$ ，可以变形为

$$Cnt \times i - \sum j$$

，此时除了记录出现次数，还要另外再开一个数组记录下标之和。

```
1: for  $i \leftarrow L..R$  do
2:    $ans \leftarrow ans + cnt[A[i]] \times i - sum[A[i]]$ 
3:    $cnt[A[i]] \leftarrow cnt[A[i]] + 1$ 
4:    $sum[A[i]] \leftarrow sum[A[i]] + i$ 
5: end for
```

$Type = 2$ 时，与 $Type = 1$ 时相似，每次要累加 $\sum(i - j)^2$ ，变形为

$$Cnt \times i^2 + \sum j^2 - 2Cnt \times i \times \sum j$$

，还需再开一个数组记录下标的平方之和。

总复杂度 $O(NQ)$ ，可以通过 20-40 分的数据。

2.3 分块算法

考虑到此题要维护的数据比较复杂，难以用常见的高级数据结构来维护。但是分块这种简单策略在本题上有良好的表现。

首先将数组 A 分块，块大小为 B 。预处理出以下信息：

- $Pre[i, j] ::= L = i \times B, R = j$ 的答案
- $Suf[i, j] ::= L = i, R = j \times B$ 的答案

设询问 L, R 时的答案为 $Ans[L, R]$ ，根据预处理得到的信息，已经知道当 L 或 R 在某个块的分界点时 $Ans[L, R]$ 的值。如果 L, R 在一个块内，可以线性算法暴力计算，时间复杂度 $O(B)$ 。

如果不在一个块内，设 L 所在的块的下一个块起点为 s ， R 所在块的上一个块终点为 e ，则

$$Ans[L, R] = Ans[L, e] + Ans[s, R] - Ans[s, e] + G[L..s - 1, e + 1..R]$$

，其中 $G[L..s - 1, e + 1..R]$ 表示左端点在 $L..s - 1$ 内，右端点在 $e + 1..R$ 内的答案。 $Ans[L, e], Ans[s, R], Ans[s, e]$ 都在预处理的过程中已算出。由于 $L..s - 1$ 和 $e + 1..R$ 长度都不超过 B ， $G[L..s - 1, e + 1..R]$ 可以暴力计算。

预处理时间复杂度 $O(\frac{N^2}{B})$ ，回答一个询问时间复杂度 $O(B)$ 。空间复杂度 $O(\frac{N^2}{B})$ 。本题极限数据 N 和 Q 同阶，取 $B = \sqrt{N}$ ，则总时间复杂度和空间复杂度均为 $O(N^{1.5})$ 。可以通过所有数据。

注意本题多组数据需要清空数组，中途记录数量和下标和的数组也要经常清空，慎用 `memset` 以防超时。