

IOI2026 中国国家集训队集中培训 第一试 A. 异形工厂

IOI2026 中国国家集训队工作组

2025 年 12 月 2 日

问题 (异形工厂 (shapez))

- 给定两个 01 字符串 s, t , 每次可以选中 s 中一个长为 3 的子串进行左或右循环移位, 问将 s 变成 t 的最小操作次数。多组区间询问。

数据范围: $n, q \leq 5 \times 10^5$ 。

$$n, q \leq 2000$$

$n, q \leq 2000$

■ 我们首先容易观察到如下性质：

- 题目的操作等价于交换 s_i 和 s_{i+1} ，或者交换 s_i 和 s_{i+2} 。
- 令 $c_i = \sum_{j=1}^i (s_j - t_j)$ 。那么一次操作等价于给任意 c_i 加一或减一，或者同时给 c_i 和 c_{i+1} 加一或减一。我们的目标是让所有 c_i 归零。

$n, q \leq 2000$

- 我们首先容易观察到如下性质：
 - 题目的操作等价于交换 s_i 和 s_{i+1} ，或者交换 s_i 和 s_{i+2} 。
 - 令 $c_i = \sum_{j=1}^i (s_i - t_i)$ 。那么一次操作等价于给任意 c_i 加一或减一，或者同时给 c_i 和 c_{i+1} 加一或减一。我们的目标是让所有 c_i 归零。
- 假设所有 c_i 为正，那么根据霍尔定理，最小操作数等价于将数组看做一条链， c_i 为链上每个点的权值的最大独立集。对每次询问暴力求解可以通过前三个子任务。不做这一步转换而使用贪心也可以暴力求解通过前三个子任务。
- 时间复杂度 $O(nq)$ ，期望得分：45。

$$n, q \leq 2 \times 10^5$$

$$n, q \leq 2 \times 10^5$$

- 对于区间询问，我们实际上要处理给区间 $[l, r]$ 加上 c_{l-1} 之后的绝对值的最大独立集。首先考虑全正的情况，有如下性质：当我们给序列每个元素加 k 后，独立集中元素个数随 k 增加而增加。也就是说， $MSI-k$ 这一函数具有凸性。

$$n, q \leq 2 \times 10^5$$

- 对于区间询问，我们实际上要处理给区间 $[l, r]$ 加上 c_{l-1} 之后的绝对值的最大独立集。首先考虑全正的情况，有如下性质：当我们给序列每个元素加 k 后，独立集中元素个数随 k 增加而增加。也就是说， $MSI-k$ 这一函数具有凸性。
- 对于每个询问，将正负的贡献拆开来算。用线段树维护凸包并暴力合并，主定理告诉我们这部分预处理的复杂度是 $O(n \log n)$ 。
- 如果在线回答询问，则时间复杂度 $O(q \log^2 n)$ ，期望得分：65。精细实现的离线做法也可能做到 $O(q \log n)$ 。

特殊性质 A: 字符 1 只会出现在 s, t 的偶数下标位置。

特殊性质 A: 字符 1 只会出现在 s, t 的偶数下标位置。

- 该特殊性质保证对所有偶数的 i , 一定有 $c_i = c_{i+1}$ 。因此最大独立集就是绝对值之和的一半。

特殊性质 A: 字符 1 只会出现在 s, t 的偶数下标位置。

- 该特殊性质保证对所有偶数的 i , 一定有 $c_i = c_{i+1}$ 。因此最大独立集就是绝对值之和的一半。
- 同样将正负拆开计算贡献, 离线后用线段树统计答案。
- 时间复杂度 $O(n \log n)$, 期望得分: 20。

$$n, q \leq 5 \times 10^5$$

$$n, q \leq 5 \times 10^5$$

- 我们需要更精细地分析独立集的凸性：可以将独立集表示成若干“段”，每一“段”都有 01 交错的形式“01010”（其中 0 表示该位置不在独立集中，1 表示在独立集中）。段与段之间有一些全 0 的空，表示这些位置的值还是负数。当我们对序列整体加一时，有些段会变成“10101”，从而导致相邻的两段合并成一个大的段。也会有一些原来的负数变成正数，出现新的段。

$$n, q \leq 5 \times 10^5$$

- 我们需要更精细地分析独立集的凸性：可以将独立集表示成若干“段”，每一“段”都有 01 交错的形式“01010”（其中 0 表示该位置不在独立集中，1 表示在独立集中）。段与段之间有一些全 0 的空，表示这些位置的值还是负数。当我们对序列整体加一时，有些段会变成“10101”，从而导致相邻的两段合并成一个大的段。也会有一些原来的负数变成正数，出现新的段。
- 我们用并查集维护每个段的信息，包括当前的最大独立集，当前段的长度，上次更新的时间戳，以及当前段首是否为 1。通过这些信息我们可以判断一个“段”在什么时候可能会更新。当两个段合并时，或者一个新元素（从负数变正）加入段中时，我们更新段的信息。显然总共的更新次数只有 $O(n)$ 次。由于只有 $c_{l-1} = c_r$ 时有解，我们的询问都是“整段”的询问。利用简单数据结构如 BIT 就可以解决区间查询的问题。
- 时间复杂度 $O(n \log n)$ ，期望得分：100。

题外话

- 事实上，由于本题是“静态”的，只要做到第一步题意转化（也可能有其他的转化形式），就可能通过各种数据结构方法通过。除去前面提到的解法二和解法四，还可能有一些基于分治的做法。如果感兴趣大家可以交流讨论。